On the Degree of Boolean Functions as Real Polynomials

Noam Nisan *   Mario Szegedy †

December 1, 1994

Abstract

Every boolean function may be represented as a real polynomial. In this paper we characterize the degree of this polynomial in terms of certain combinatorial properties of the boolean function.

Our first result is a tight lower bound of $\Omega(\log n)$ on the degree needed to represent any boolean function that depends on n variables.

Our second result states that for every boolean function f the following measures are all polynomially related:

- The decision tree complexity of f.
- The degree of the polynomial representing f.
- The smallest degree of a polynomial *approximating* f in the L_{max} norm.

1 Introduction

1.1 Boolean Functions as Real Polynomials

Boolean functions may be represented in various forms. Some of simplest and most natural of these forms are representations as polynomials over various fields, and in particular over the real numbers. Let $f : \{F,T\}^n \rightarrow \{F,T\}$ be a boolean function. If we encode "true" as the real number 1, and "false" as the real number 0, then f becomes a function from a subset of R^n to R. We say a real multivariate polynomial $p : R^n \rightarrow R$ represents f if for every $x \in \{0,1\}^n$, $f(x) = p(x)$.

It is well known, and not difficult to see that every boolean function can be represented as a polynomial. Moreover, since for all $x \in \{0,1\}$ and integer $k \geq 1$ we have that $x^k = x$, there is no need to ever raise a variable x_i to a power greater than 1, and thus we can limit ourselves to polynomials p which are *multilinear* (i.e., each variable x_i appears with degree at most 1). In fact, it turns out that there is a unique multi-linear polynomial representing any given boolean function.

The choice of representing "true" as 1 and "false" as 0 is of course somewhat arbitrary, and is a matter of convenience. Another convenient choice is to represent "true" as -1 and "false" as 1. The representation of a function as a polynomial under these conventions is sometimes called the Fourier transform of the function (see e.g. [7, 8]). The parameters we study here remain invariant under any choice of two different real numbers to represent "true" and "false".

*The Hebrew University, Jerusalem 91904, Israel. Supported by BSF 89-00126.

†AT&T Bell Laboratories.

1.2 Previous Work

In their book "Perceptrons" [10], Minski and Papert initiated the study of the computational properties of boolean functions using their representation by polynomials. Recently there have been many more works that use these representations (or approximations) in order to study various complexity measures of the boolean functions.

The Fourier transform of boolean functions (i.e., the representation as a real polynomial with the convention $true = -1$ and $false = 1$) was used in [7] to study the *influence* of variables on boolean functions. In [8] it was used to study AC^0 functions (functions computed by constant depth, polynomial size circuits). In [8, 9] the Fourier transform was used to devise learning algorithms. In [3] it was used to characterize "polynomial threshold" functions.

In [5] a tight lower bounds for the time required to compute a boolean function on a CREW PRAM is given in terms of the degree of the function as a real polynomial. In [2, 1] lower bounds for constant depth circuits are obtained using approximation by real polynomials. Earlier, [12, 14] obtained similiar lower bounds using polynomials over finite fields.

1.3 New Results

In this paper we study the most basic parameter of the representation of a boolean function as a real polynomial, its *degree*.

Definition 1 *For a boolean function f, the degree of f, denoted by $deg(f)$, is the degree of the unique multilinear real polynomial that represents f (exactly).*

1.3.1 Minimum Possible Degree

Our first theorem answers the question of what is the smallest degree of a boolean function that depends on n variables.

Theorem 1 *Let f be a boolean function that depends on n variables. Then $deg(f) \geq \log_2 n - O(\log \log n)$.*

The proof of this theorem makes use of the relation between "Influences" and the Fourier transform due to [7].

This result is tight up to the $O(\log \log n)$ term, as can be seen by the "address" function.

1.3.2 Degree and Decision Trees

We next relate the degree of a boolean function to several combinatorial and complexity measures of the function.

The boolean decision model is perhaps the simplest computational model for boolean functions. In this model the algorithm repeatedly queries input variables until it can determine the value of the function. The algorithm is adaptive, choosing which variable to query next based on the answers to the previous queries. The only cost in this model is the number of variables queried, and the cost of an algorithm is the number of queries made for the worst case input. The decision tree complexity of f, $D(f)$, is defined to be the cost of the best algorithm for f.

The decison tree complexity is well studied in the literature in many contexts. In particular it is known that it is closely related to several other combinatorial and complexity measures. The decision tree complexity is known to be polynomially related to the *certificate* complexity (see e.g. [16]) and to the *block sensitivity* [11]. Furthermore, $\log D(f)$ is equal, up to a constant factor, to the time needed to compute f on a CREW PRAM [11]. We show that the degree of f is also polynomially related to all these measures.

Theorem 2 *For every boolean function we have*

$$deg(f) \leq D(f) \leq 16 deg(f)^8$$

The proof of this result requires results from real approximation theory [4] [6] [13].

We strongly suspect that the exponent 8 is not optimal. The strongest separation we can obtain is a function for which $D(f) = deg(f)^{1.58\ldots}$.

1.3.3 Approximation in L_{max} norm

Our techniques are strong enough to allow us to give strong bounds on the degree needed even to *approximate* boolean functions in the L_{max} norm.

Definition 2 *Let f be a boolean function, and let p be a real polynomial. We say that p approximates f if for every $x \in \{0,1\}^n$ we have that $|p(x) - f(x)| \leq 1/3$. The approximate degree of f, $\widetilde{deg}(f)$, is defined to be the minimum, over all polynomials p that approximate f, of the degree of p.*

Note: the constant $1/3$ is arbitrary and can be replace by any constant $0 < \epsilon < 1/2$, without affecting our results.

Perhaps surprisingly, we show that approximation is not much easier than exact representation.

Theorem 3 *There exists a constant c such that for every boolean function we have*

$$\widetilde{deg}(f) \leq deg(f) \leq D(f) \leq c\widetilde{deg}(f)^8$$

This theorem has been recently used in [5] to show that randomization does not give extra power to CREW PRAMs.

The best separation that we know between $deg(f)$ and $\widetilde{deg}(f)$ is quadratic, and we conjecture that this is indeed the worst case.

2 Minimum Possible Degree

In this section we prove the following theorem

Theorem 1 *Every boolean function f that depends on n variables has degree $deg(f) \geq \log_2 n - O(\log \log n)$.*

For the proof of this theorem it will be convenient to use the Fourier transform representation, i.e., -1 for true and 1 for false. (This is used in this section only.) Thus a boolean function will be viewed as a real function $f : \{-1,1\}^n \rightarrow \{-1,1\}$. For a subset $S \subseteq \{1,\ldots,n\}$ we will denote $X_S = \prod_{i \in S} x_i$.

The next two subsections provide some necessary lemmas, and the proof of the theorem appears in the third subsection.

2.1 Degree and Influences

We will require the following fact, definition, and lemma from [7].

Lemma 1 (Parseval's equality) *If we represent a boolean funcion f as $f = \sum_S \alpha_S X_S$, then*

$$\sum_S \alpha_S^2 = 1.$$

Definition 3 *For a boolean function f and a variable x_i, the influence of x_i on f, $Inf_i(f)$ is defined to be*

$$Inf_i(f) = Pr[f(x_1, \ldots, x_{i-1}, true, x_{i+1}, \ldots, x_n) \neq f(x_1, \ldots, x_{i-1}, false, x_{i+1}, \ldots, x_n)]$$

where $x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n$ are chosen at random in $\{false, true\}$.

Lemma 2 *[7] For any boolean function f on n variables, if we represent $f = \sum_S \alpha_S X_S$, then*

$$\sum_{i=1}^{n} Inf_i(f) = \sum_S |S| \alpha_S^2$$

From these lemmas we easily deduce:

Corollary 1 *For any boolean function f,*

$$\sum_{i=1}^{n} Inf_i(f) \leq deg(f)$$

2.2 Zeroes of a Multilinear Polynomial

The following simple lemma gives an upper bound for the number of zeroes of any multilinear polynomial over $\{-1, 1\}^n$. It is known as the lemma of Schwartz [15], but we prove it below for completeness.

Lemma 3 (Schwartz) *Let $p(x_1, ..., x_n)$ be a non-zero multilinear polynomial of degree d. If we choose $x_1, \ldots, x_n$ independently at random in $\{-1, 1\}$ then*

$$Pr[p(x_1, ..., x_n) \neq 0] \geq 2^{-d}$$

Proof: The proof is by induction on n. For $n = 1$, we just have a linear function of one variable which may have only one zero.

Induction step: Write

$$p(x_1, \ldots, x_n) = x_n g(x_1, \ldots, x_{n-1}) + h(x_1, \ldots, x_{n-1})$$

We can see that the non-zeroes of p over $\{-1, 1\}^n$ yield non-zeroes of $h + g$ or of $h - g$ over $\{-1, 1\}^{n-1}$: if $p(x_1, \ldots, x_{n-1}, 1) \neq 0$ then $h(x_1, \ldots, x_{n-1}) + g(x_1, \ldots, x_{n-1}) \neq 0$, and if $p(x_1, \ldots, x_{n-1}, -1) \neq 0$ then $h(x_1, \ldots, x_{n-1}) - g(x_1, \ldots, x_{n-1}) \neq 0$. We now distinguish between three cases. CASE I: $h + g$ is identically equal to zero. In this case $p = (x_n - 1)g$, where $deg(g) = d-1$ and we use the induction hypothesis on g for the x's satisfying $x_n = -1$. CASE II: $h - g$ is identically equal to zero. In this case $f = (1 + x_n)g$, where $deg(g) = d - 1$, and again we use the induction hypothesis on g for the x's satisfying $x_n = 1$. CASE III: Both $h + g$ and $h - g$ are not identically equal to zero. The degrees of $h + g$ and of $h - g$ are both bounded by d and thus we use the induction hypothesis on $h + g$ for the x's satisfying $x_n = 1$ and on $h - g$ for the x's satisfying $x_n = -1$. $\square$

2.3 Proof of Theorem

We have now assembled all that we need in order to prove the theorem.

Proof: (of Theorem 1) For each i define a function f^i on $n - 1$ variables as follows:

$$f^i(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n) =$$

$$f(x_1, \ldots, x_{i-1}, -1, x_{i+1}, \ldots, x_n) - f(x_1, \ldots, x_{i-1}, 1, x_{i+1}, \ldots, x_n)$$

Under this notation it is clear that

$$Inf_i(f) = Pr[f^i(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n) \neq 0]$$

where $x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n$ are chosen at random in $\{-1, 1\}$.

Since f depends on all of its variables, we have that for every i, f^i is not identically zero, and thus we can apply lemma 3 and conclude that for all i, $Inf_i(f) \geq 2^{-d}$.

On the other hand, from corollary 1 it follows that $\sum_i Inf_i(f) \leq d$. Combining these two bounds we get:

$$n/2^d \leq \sum_i Inf_i(f) \leq d$$

Thus $d2^d \geq n$, and the theorem follows. □

3 Degree and Decision Trees

We remind the reader that at this point we return to the representation of $true = 1$ and $false = 0$.

3.1 The Method of Symmetrization

We will use the method of symmetrization, first used by Minski and Papert [10].

Definition 4 *Let $p : R^n \rightarrow R$ be a multivariate polynomial, then the symmetrization of p is*

$$p^{sym}(x_1, \ldots, x_n) = \frac{\sum_{\pi \in S_n} p(x_{\pi(1)}, \ldots, x_{\pi(n)})}{n!}$$

The important point is that if we are only interested in inputs $x \in \{0, 1\}^n$ then p^{sym} turns out to depend only upon $x_1 + \cdots + x_n$. We can thus represent it as a univariate polynomial of $x_1 + \cdots + x_n$:

Lemma 4 *[10] If $p : R^n \rightarrow R$ is a multivariate polynomial, then there exists a unique univariate polynomial $\tilde{p} : R \rightarrow R$ of degree at most n such that for all $x_1, \ldots, x_n \in \{0, 1\}^n$ we have*

$$p^{sym}(x_1, \ldots, x_n) = \tilde{p}(x_1 + \cdots + x_n)$$

Moreover, $deg(\tilde{p}) \leq deg(p)$.

3.2 A Theorem from Approximation Theory

We will need the following result of H. Ehlich and K. Zeller [6] and T. J. Rivlin and E. W. Cheney [13]:

Theorem 2 (Echlich, Zeller; Rivlin, Cheney) *Let p be a polynomial with the following properties:*

1. *For any integer $0 \leq i \leq n$ we have $b_1 \leq p(i) \leq b_2$.*
2. *For some real $0 \leq x \leq n$ the derivative of p satifies $|p'(x)| \geq c$.*

Then $deg(p) \geq \sqrt{cn/(c + b_2 - b_1)}$.

Again, for completeness we prove this theorem. The proof is based on the following well known theorem of Markov [4]):

Theorem 3 (Markov) *Let $p : R \rightarrow R$ be a univariate polynomial of degree d such that any real number $a_1 \leq x \leq a_2$ satisfies $b_1 \leq p(x) \leq b_2$. Then for all $a_1 \leq x \leq a_2$, the derivative of p satisfies $|p'(x)| \leq \frac{d^2(b_2-b_1)}{a_2-a_1}$.*

The two theorems are similar, but in the former one we do not have the information on the value of $p(n)$ for all real x in the range but rather only for integer x. Thus the theorem can be perceived as a generalization of that of Markov. There is a surpisingly simple proof for it, however by Markov's theorem:

Proof: [Echlich, Zeller; Rivlin, Cheney] Let $c' = max_{0 \leq x \leq n}|p'(x)| \geq c$. It is clear that for all real $0 \leq x \leq n$:

$$b_1 - c'/2 \leq p(x) \leq b_2 + c'/2$$

Using the Markov inequalty we have

$$c' \leq \frac{deg(p)^2(c' + b_2 - b_1)}{n}$$

Thus,

$$deg(p)^2 \geq \frac{c'n}{c' + b_2 - b_1} \geq \frac{cn}{c + b_2 - b_1}$$

□

3.3 Main Lemma

Lemma 5 *Let f be a boolean function such that $f(000\ldots0) = 0$ and for every boolean vector $\vec{x}$ of Hamming weight 1, $f(\vec{x}) = 1$. Then*

$$deg(f) \geq \sqrt{n/2}$$

and

$$\widetilde{deg}(f) \geq \sqrt{n/6}$$

Proof: We will prove the bound for $\widetilde{deg}(f)$. The sharper bound for $deg(f)$ follows exactly the same lines. Let p be a polynomial approximating f, and consider $\tilde{p}$ the univariate polynomial giving its symmetrization. $\tilde{p}$ satisfies the following properties:

1. $deg(\tilde{p}) \leq deg(p)$. (By lemma 4.)
2. For every integer $0 \leq i \leq n$, $-1/3 \leq \tilde{p}(i) \leq 4/3$. (Since for every boolean vector $\vec{x}$, $p(\vec{x})$ is within $1/3$ of a boolean value.)
3. $\tilde{p}(0) \leq 1/3$. (Since $f(000\ldots0) = 0$.)
4. $\tilde{p}(1) \geq 2/3$. (Since for all boolean vectors $\vec{x}$ of hamming weight 1, $f(\vec{x}) = 1$.)

Properties (3) and (4) together imply that for some real $0 \leq z \leq 1$, the derivative $\tilde{p}'(z) \geq 1/3$. We can now apply lemma 5 to obtain the lower bound for $deg(\tilde{p})$, and thus also for $deg(p)$. We remark that the bound for $deg f$ is proven exactly the same way, except that the inequalities that correspond to 2.-4. contain different constants. □

The examples given below in section 3.5 show that the bound for $\widetilde{deg}(f)$ is tight (up to a constant factor). We do not know whether the bound for $deg(f)$ is tight.

3.4 General Boolean Functions

Although the main lemma concerns very special types of boolean functions, it turns out that it is enough to give good bounds for *all* boolean function. This is done by relating the degree to other combinatorial properties of boolean functions.

Notation: for a string $x \in \{0,1\}^n$ and a set $S \subseteq \{1, \ldots, n\}$, we define $x^{(S)}$ to be the boolean string which differs from x on exactly the bits in S.

Definition 5 *[11] For a boolean function f the block sensitivity of f, $bs(f)$, is defined to be the maximum number t such that there exists an input $x \in \{0,1\}^n$ and t disjoint subsets $S_1, \ldots, S_t \subset \{1, \ldots, n\}$, such that for all $1 \le i \le t$, $f(x) \ne f(x^{(S_i)})$.*

The block sensitivity of a function is known to be related to its decision tree complexity, $D(f)$.

Lemma 6 *[11] For every boolean function f, $bs(f) \le D(f) \le bs^4(f)$.*

We can easily get lower bounds for the degree in terms of the block sensitivity.

Lemma 7 *For every boolean function f,*

$$deg(f) \ge \sqrt{bs(f)/2}$$

and

$$\widetilde{deg}(f) \ge \sqrt{bs(f)/6}$$

Proof: Let f be a boolean function, and let $\vec{x}$ and $S_1, \ldots, S_t$ be the input and sets achieving the block sensitivity. Let us assume without loss of generality that $f(\vec{x}) = 0$. We define a function $f'(y_1, \ldots, y_t)$ as follows:

$$f'(y_1, \ldots, y_t) = f(\vec{x} \oplus y_1 S_1 \oplus \cdots \oplus y_t S_t)$$

i.e. the j'th bit fed to f is $x_j \oplus y_i$ if $j \in S_i$, and is x_j if j is not in any of the S_i's (the $\oplus$ operator adds bits or vectors of bits modulo 2). The following facts can easily be verified:

1. $deg(f') \le deg(f)$. (The bits x_j are constants in the definition of f'.)
2. f' satisfies the conditions of lemma 6.

Our lemma thus follows from lemma 6. □

Combining lemmas 7 and 8 we get

Theorem 4 *For every boolean function f we have*

$$deg(f) \le D(f) \le 16 deg(f)^8$$

and

$$\widetilde{deg}(f) \le deg(f) \le D(f) \le 1296 \widetilde{deg}(f)^8$$

3.5 Separations

The best separation results we know between $D(f)$, $deg(f)$ and $\widetilde{deg}(f)$ are given by the following examples.

Let $E_{12}(x_1, x_2, x_3)$ be the symmetric function taking value true if exactly one or two of the input bits are true. It is not difficult to see that $deg(E_{12}) = 2$ (we can write $E_{12} = x_1 + x_2 + x_3 - x_1x_2 - x_2x_3 - x_1x_3$). On the other hand $D(E_{12}) = 3$. For every integer k we define the function E_{12}^k on 3^k variables as a composition of E_{12} on three disjoint copies (on separate inputs) of E_{12}^{k-1}. We now have

Example 1 *The function* E_{12}^k *on* $n = 3^k$ *variables satisfies:*

$$D(E_{12}^k) = 3^k = n$$

and

$$deg(E_{12}^k) = 2^k = n^{\log_3 2} = n^{0.63\ldots}$$

Proof: The value of the degree simply follows by induction from the definition of E_{12}^k. The lower bound on the decision tree complexity follows from the fact that on input $000\ldots0$, the decision tree must look at every bit since if even one bit is changed to 1, the value of E_{12}^k changes from false to true. □

Consider the function OR_n on n variable returning true if at least one of the inputs is true. Using Chebyshev polynomials we can approximate OR_n by a rather low degree polynomial.

Example 2

$$deg(OR_n) = n$$

and

$$\widetilde{deg}(OR_n) = O(\sqrt{n})$$

Proof: We will use Chebyshev polynomials. The k'th Chebyshev polynomial, $T_k(x)$, is a real polynomial of degree k having the following properties (see [4]):

1. For every $-1 \leq x \leq 1$, $|T_k(x)| \leq 1$.
2. For all $x \geq 1$, the derivative satisfies $T'_k(x) \geq k^2$.

Now choose $k = 2\sqrt{n}$ and $c = 1/T_k(\frac{n}{n-1})$, and define the polynomial $p(x) = 1 - cT_k(x/(n-1))$. Property (2) insures that $c \leq 1/4$. By property (1) we thus have that $|p(x) - 1| \leq 1/3$ for all $0 \leq x \leq n-1$, and $p(n) = 0$. It follows that $p(x_1 + \cdots + x_n)$ approximates OR_n. □

4 Open problems

Besides the intriguing questions that remain open about the exact relation between $\widetilde{deg}(f)$, $deg f$ and $D(f)$ we would like to mention three related open problems.

The first question is known as the *question of "sensitivity versus block sensitivity."* The sensitivity of a boolean function f, $S(f)$, is the maximum of $S_x(f)$ over all inputs x, as defined below: Let x^i be the input that we obtain from x by negating its i^{th} bit but leaving all the other bits intact. $S_x(f)$ is the number of i's such that $f(x) \neq f(x^i)$. E.g. the sensitivity of the "OR" function is n, because for its "most sensitive input," the $000\ldots0$, the value of OR changes from 0 to 1 if we exchange any of the input zeros by 1. Clearly $bs(f) \geq S(f)$. Is it true that there is a polynomaial relation between the sensitivity and the block sensitivity (say, $S^2(f) \geq bs(f)$)?

The second question was asked recently by Lance Fortnow: what if we express f as a rational function $(p(x)/q(x))$ rather than as a polynomial (p and q are multivariate polynomials and we can take them to be multilinear). Can $\max(degp, degq)$ be much less than $degf$? Again, does a polynomial relation hold between $degf$ and $\max(degp, degq)$? An answer would have interesting applications in structural complexity theory.

The third question is related to the degree of *symmetric* boolean functions. Suppose that f is a symmetric boolean function, but not identically zero or one. Give a lower bound on

$deg f$ in terms of the number of variables, n. It is easy to see that $n/2$ is always a lower bound, but apparently better bounds can be obtained. Recently J. von zur Gathen showed that if the number of variables, n, is prime minus 1 then $deg f = n$. It is easy to construct symmetric functions for any odd n such, that $deg f = n-1$. Very recently J. Roche managed to find infinitely many symmetric nontrivial f's such that $deg f = n-3$, and he conjectures that $n-3$ is a general lower bound.

Acknowledgement: The authors wish to thank Ilan Newman for his insightful comments.

References

[1] J. Aspens, R. Beigel, M. Furst, and S. Rudich, *On the expressive power of voting polynomials,* STOC 1991, pages 402-409.

[2] R. Beigel, N. Reingold, and D. Spielman, *The perceptron striles back,* Technical report YALEU/DCS/TR-813, Yale University, 1990.

[3] J. Brook and R. Smolenski, *Polynomial threshold functions, AC^0 functions and spectral norms,* FOCS 1990, pages 632-642.

[4] E. W. Cheney, Introduction to approximation theory, McGraw-Hill Book Co., 1966.

[5] M. Dietzfelbinger, M. Kutylowski, R. Reischuk, *Exact Time Bounds for Computing Boolean Functions on PRAMs Without Simultaneous Writes,* SPAA 1990.

[6] H. Ehlich and K. Zeller, *Schwankung von Polynomen zwischen Gitterpunkten,* Mathematische Zeitschrift, volume 86, pages 41-44, 1964.

[7] J. Kahn, G. Kalai, and N. Linial, *The influence of variables on boolean functions,* FOCS 1988, pages 68-80.

[8] N. Linial, Y. Mansour, N. Nisan, *Constant depth circuits, Fourier transform and Learnability,* FOCS 1989, pages 574-579.

[9] E. Kushilevitz and Y. Mansour, *Learning decision trees using the Fourier transform,* STOC 1991, pages 455-464.

[10] M. Minsky and S. Papert, *Perceptrons,* MIT press, Cambridge, 1988 Expanded edition. First edition appeared in 1968.

[11] N. Nisan, *CREW PRAM's and decision trees,* STOC 1989, pages 327–335.

[12] A.A. Razborov, *Lower bounds for the size of circuits of bounded depth with basis (and,xor),* Math. Notes of the Academy of sciences of the USSR, 41(4), 1987, pages 333-338.

[13] T. J. Rivlin and E. W. Cheney, *A comparison of Uniform Approximations on an interval and a finite subset thereof,* SIAM Numer. Anal., volume 3, number 2, pages 311-320, 1966

[14] R. Smolenski, *Algebraic methods in the theory of lower bounds for Boolean circuit complexity,* STOC 1987, pages 77-82.

[15] J. T. Schwartz, *Fast probabilistic Algorithms for verification of Polynomial identities,* J. Assoc. Computing Machinery, volume 27, pages 701-717, 1980.

[16] I. Wegener, *The complexity of Boolean functions,* Wiley-Teubner Series in Comp. Sci., New York – Stuttgart, 1987.