ALTERNATING CIRCUITS OF LOW DEPTH

HAMED HATAMI

1. Introduction To Circuits

In 1949 Shannon proposed the size of Boolean circuits as a measure of computation difficulty of a function. Circuits are closely related in computational power to Turing machines, and thus they provide a nice framework for understanding the time complexity. On the other hand their especially simple definition makes them amenable to various combinatorial, algebraic, and analytic methods.

A burst of activity in circuit complexity exploded about 30 years ago with first exponential lower bounds for some circuit models, like bounded depth circuits, monotone circuits, restricted branching programs, etc. There has been quick progress made for about two decades, but soon various barriers are discovered.

A Boolean circuit is a directed acyclic graph. The vertices of indegree 0 are called inputs, and are labeled with a variable $x_i$ or with a constant 0 or 1. The vertices of indgree $k > 0$ are called gates and are labeled with a Boolean function on $k$ inputs. The indgree of a vertex is called its fanin and its outdegree is called its fanout. The most standard circuits are restricted to have gates $\land, \lor, \neg$. One of the nodes is designated the output node, and then the circuit represents a Boolean function in a natural way. The size of a circuit is its number of gates.

A simple counting argument establishes the following strong lower-bound. Roughly speaking, there are too many Boolean functions $f : \{0,1\}^n \rightarrow \{0,1\}$ compared to the number of small circuits.

**Theorem 1.1** (Muller 1956). Almost every Boolean function $f : \{0,1\}^n \rightarrow \{0,1\}$ requires fanin 2 circuits of size $\Omega(2^n/n)$. On the other hand every function $f : \{0,1\}^n \rightarrow \{0,1\}$ can be computed by a fanin 2 circuit of size $O(2^n/n)$

Theorem 1.1 has a major shortcoming. It does not provide any explicit example of a function which requires a large circuit. Also unfortunately it does not provide any example of a function in NP that requires circuits of superpolynomial size. Despite the importance of lower bounds on the circuit complexity, the best explicit known construction due to Blum 1984 provides a function which requires finin 2 circuits of size $3n - o(n)$.

2. Bounded depth circuits

Considering our inability in proving lower bounds on the circuit complexity of explicit Boolean functions, we need to impose strong restrictions on the circuits in order to be able to prove meaningful lower bounds. We will restrict to bounded depth circuits. The first strong lower bounds for bounded depth circuits were given by Ajtai (1983) and Furst, Saxe, Sipser (1984). They established a superpolynomial lower bound for constant depth circuits computing the parity function. Later Yao gave a sharper exponential lower bound. In 1987 Hastad further strengthened and simplified this argument, and obtained near optimal bounds.
Let us start by defining our constant depth circuits. As we mentioned earlier we are interested in the model where we are restricted to gates $\land$, $\lor$, $\neg$. Note that by De Morgan’s laws
\[\neg(p_1 \lor \ldots \lor p_k) = (\neg p_1) \land \ldots \land (\neg p_k),\]
and
\[\neg(p_1 \land \ldots \land p_k) = (\neg p_1) \lor \ldots \lor (\neg p_k),\]
we can assume that
- there are no $\neg$ gates in the circuit, and instead the inputs are either of the form $x_i$ or $\neg x_i$ for variables $x_i$, or constants 0 and 1.
- We shall consider circuits whose depths are much smaller than $n$, the number of inputs. Hence we need to allow arbitrary fanin so that the circuit may access the entire input.
- We will assume that the circuits are of the special form where all $\land$ and $\lor$ gates are organized into alternating levels with edges only between adjacent levels. Note that any circuit can be converted into this form without increasing the depth and by at most squaring the size (why?).

These circuits are called alternating circuits. The depth of an alternating circuit is defined as the distance from the output node to the input nodes. Let $\text{AC}[d]$ denote the set of all alternating circuits of depth at most $d$.

The alternating circuits of depth 2 are particularly important. Note that because of the “alternation” condition, there are two different types of depth 2 alternating circuits. They correspond to conjunctive normal form and disjunctive normal form formulas.

**Definition 2.1 (Conjunctive Normal Form, $\land$ of $\lor$).** A formula is in conjunctive normal form, abbreviated to CNF, if it is a conjunction (i.e. $\lor$) of clauses, where a clause is a disjunction (i.e $\lor$) of literals (i.e. $x_i$ or $\neg x_i$), where a literal and its negation cannot appear in the same clause.

For example $(x_1 \lor x_2) \land (\neg x_1 \lor x_2 \lor x_3)$ is a formula in conjunctive normal form. It corresponds to an alternating circuit of depth 2 with 3 gates.

**Definition 2.2 (Disjunctive Normal Form, $\lor$ of $\land$).** A formula is in disjunctive normal form, abbreviated to DNF, if it is a disjunction (i.e. $\lor$) of conjunctive clauses (i.e $\lor$ of literals).

Consider a fixed point $y = (y_1, \ldots, y_n) \in \{0, 1\}^n$, and $T = \{i : y_i = 1\}$. Note that the only assignment that satisfies the clause
\[
\left( \bigwedge_{i \in T} x_i \right) \land \left( \bigwedge_{i \not\in T} \neg x_i \right)
\]
is the assignment $x := y$. Hence given a Boolean function $f : \{0, 1\}^n \to \{0, 1\}$, for every point $y$ with $f(y) = 1$ we can create a clause which is satisfied only if $x = y$. By taking the $\lor$ of these clauses we create a DNF formula that represents the function $f$.

**Example 2.3.** Consider the function $f : \{0, 1\}^2 \to \{0, 1\}$ such that $f(0, 0) = f(0, 1) = f(1, 1) = 1$ and $f(1, 0) = 0$. Then the DNF
\[
(\neg x_1 \land \neg x_2) \lor (\neg x_1 \land x_2) \lor (x_1 \land x_2)
\]
represents $f$. 

By changing the role of 0’s and 1’s and $\land$ and $\lor$, we can represent $f$ in CNF. We conclude the following observation which says that the depth 2 alternating circuits are powerful enough to compute any Boolean function.
Observation 2.4. Every function \( f : \{0, 1\}^n \to \{0, 1\} \) can be represented in both DNF and CNF formulas using at most \( 2^n \) clauses.

Another key notion that we will need for this lecture is the decision tree complexity of a Boolean function.

Definition 2.5. A decision tree over variables \( x_1, \ldots, x_n \) is a binary tree where the children of each internal node have two children, left and right. Moreover each internal node is labeled with a variable, and each leaf is labeled with 0 or 1. To evaluate a decision tree at a point \( x = (x_1, \ldots, x_n) \), we start from the root, and at each internal node with label \( x_i \) we go left if \( x_i = 0 \) and right if \( x_i = 1 \) until we reach a leaf. The value of the leaf is the output of the decision tree for \( x \).

For a Boolean function \( f : \{0, 1\}^n \to \{0, 1\} \) let \( dt(f) \) be the depth of the smallest decision tree computing \( f \).

Obviously \( dt(f) \leq n \). It is an easy exercise to see that every Boolean function \( f : \{0, 1\}^n \to \{0, 1\} \) expressed as a DNF formula that satisfies the following two properties:

- Each clause has size at most \( dt(f) \);
- The clauses all accept disjoint sets of points. I.e. there is no \( x \in \{0, 1\}^n \) that satisfies more than one clause.

Similarly \( f \) can be expressed as a CNF that satisfies the following two properties:

- Each clause has size at most \( dt(f) \);
- The clauses all reject disjoint sets of points. I.e. there is no \( x \in \{0, 1\}^n \) that falsifies more than one clause.

3. Hastad’s switching lemma

The basic idea of Ajtai (1983) and Furst, Saxe, Sipser (1984) for proving lower-bounds on bounded depth AC circuits was to assign random values to a random subset of variables. This will simplify a small size AC\([d]\) circuit greatly. Consider a gate at level 1 (that is a gate directly connected to inputs \( x_i \) and \( \neg x_i \)’s). Noting that the gate is either \( \land \) or \( \lor \), if it has a large fanin, then there is a high chance that a random assignment of values to a random subset of variables will determine the value of the gate. Indeed an \( \land \) gate only needs one 0 input to be set to 0, and an \( \lor \) gate only needs one 1 on its inputs to be set to 1.

As we mentioned earlier Hastad further explored and these ideas, and obtained near optimal bounds. The core of his proof is an important lemma known as switching lemma. It is a key tool for proving lower bounds on the size of constant-depth Boolean circuits.

Definition 3.1. Let \( X = \{x_1, \ldots, x_n\} \) be the input variables to a circuit \( C \) computing a function \( f \). A restriction \( \rho \) is an element in \( \{0, 1, *\}^X \).

A restriction \( \rho \) is interpreted as setting the variables assigned 0 or 1 and leaving variables those assigned star. Under \( \rho \) we may simplify \( C \) by eliminating gates whose values become determined. Call this the induced circuit \( C_\rho \) computing the induced function \( f_\rho \).

Lemma 3.2 (Hastad’s switching lemma). Let \( f \) be given by a CNF formula where each clause has size at most \( t \). Choose a random restriction \( \rho \) by setting every variable independently to \( * \) with probability \( p \), and to 0 and 1 each with probability \( \frac{1-p}{2} \). Then

\[
\Pr[dt(f_\rho) > s] \leq (5pt)^s.
\]

We are going to prove this lemma by induction. But for the induction to work one needs to strengthen the statement:
Lemma 3.3 (Hastad’s switching lemma, stronger version). Let $f$ be given by a CNF formula where each clause has size at most $t$. Choose a random restriction $\rho$ by setting every variable independently to $\ast$ with probability $p$, and to 0 and 1 each with probability $\frac{1-p}{2}$. For every function $F : \{0,1\}^n \to \{0,1\}$, we have
\[
\Pr[dt(f_\rho) > s | F_\rho \equiv 1] \leq (5pt)^s.
\]

Proof. Set $\alpha := 5pt$, and suppose that $f = \land_{i=1}^{m} C_i$ where $C_i$’s are clauses of size at most $t$. We prove this statement by induction on $m$ the number of clauses in $f$. If $m = 0$, then $f \equiv 1$ and the lemma is obvious. For the induction step let us study what happens to $F$ in the circuit. First note that without loss of generality we can assume that there are no negated literals in $C_1$ and hence
\[
C_1 = \bigvee_{i \in T} x_i,
\]
for a subset $T \subseteq \{1,\ldots,n\}$. First note that
\[
\Pr[dt(f_\rho) > s | F_\rho \equiv 1] = \Pr[dt(f_\rho) > s, \rho T \not\subseteq \{0, \ast\}^T | F_\rho \equiv 1] + \Pr[dt(f_\rho) > s, \rho T \in \{0, \ast\}^T | F_\rho \equiv 1].
\]
Hence in order to prove (1) it suffices to prove both
\[
\Pr[dt(f_\rho) > s | F_\rho \equiv 1, \rho T \not\subseteq \{0, \ast\}^T] \leq \alpha^s,
\]
and
\[
\Pr[dt(f_\rho) > s | F_\rho \equiv 1, \rho T \in \{0, \ast\}^T] \leq \alpha^s.
\]
To prove (2) note that for $g = \land_{i=1}^{m} C_i$ (which has only $m-1$ clauses),
\[
\text{L.H.S of (2)} = \Pr[dt(g_\rho) > s | F_\rho \equiv 1, \rho T \not\subseteq \{0, \ast\}^T] = \Pr[dt(g_\rho) > s | (F \land C_1)_\rho \equiv 1] \leq \alpha^s,
\]
where in the last inequality we used the induction hypothesis applied to $g$ and $F \land C_1$. It remains to prove (3). We break (3) into $2^{|T|}$ terms based on which coordinates in $T$ are $\ast$’s and which ones are 0’s:
\[
\text{L.H.S of (3)} = \sum_{Y \subseteq T} \Pr[dt(f_\rho) > s, \rho Y = \ast, \rho T-Y = 0 | F_\rho \equiv 1, \rho T \in \{0, \ast\}^T] \]
\[
\leq \sum_{Y \subseteq T} \Pr[\rho Y = \ast, \rho T-Y = 0 | F_\rho \equiv 1, \rho T \in \{0, \ast\}^T] \times \Pr[dt(f_\rho) > s | F_\rho \equiv 1, \rho Y = \ast, \rho T-Y = 0].
\]
First note that if $Y = 0$, then $\rho T = 0$, and thus $f_\rho \equiv 0$, and consequently $dt(f_\rho) = 0$. Hence we can remove the corresponding term from the above calculation and obtain:
\[
\text{L.H.S of (3)} \leq \sum_{Y \subseteq T, Y \neq 0} \Pr[\rho Y = \ast | F_\rho \equiv 1, \rho T \in \{0, \ast\}^T] \times \Pr[dt(f_\rho) > s | F_\rho \equiv 1, \rho Y = \ast, \rho T-Y = 0].
\]
We bound the two terms in the product separately.

First observation (bounding $\Pr[\rho Y = \ast | F_\rho \equiv 1, \rho T \in \{0, \ast\}^T]$): Since setting variables in $Y$ to $\ast$ cannot increase the probability that $F_\rho \equiv 1$, we have
\[
\Pr[F_\rho \equiv 1 | \rho Y = \ast, \rho T \in \{0, \ast\}^T] \leq \Pr[F_\rho \equiv 1 | \rho T \in \{0, \ast\}^T],
\]
Hence using \( \Pr[A|B] \Pr[B] = \Pr[A \land B] \) we have
\[
\Pr[\rho_Y = \bar{s} \mid F_\rho = 1, \rho_T \in \{0,*\}^T] = \frac{\Pr[F_\rho = 1 \mid \rho_Y = \bar{s}, \rho_T \in \{0,*\}^T]}{\Pr[F_\rho = 1 \mid \rho_T \in \{0,*\}^T]} \Pr[\rho_Y = \bar{s} \mid \rho_T \in \{0,*\}^T] 
\leq \Pr[\rho_Y = \bar{s} \mid \rho_T \in \{0,*\}^T] = \left(\frac{2p}{1 + p}\right)^{|Y|} \leq (2p)^{|Y|}
\]

**Second observation:** (bounding \( \Pr[dt(f_\rho) > s \mid F_\rho = 1, \rho_Y = \bar{s}, \rho_{T-Y} = \bar{0}] \)) Note that the variables in \( Y \) can contribute by at most \(|Y|\) to the decision tree depth, or more precisely if for every \( \sigma \in \{0,1\}^{|Y|} \), we have \( dt(f_{\sigma \rho}) \leq s - |Y| \), then \( dt(f_\rho) \leq s \). Indeed to verify this, note that we can always build a decision tree of depth at most \(|Y| + \max_\sigma dt(f_{\sigma \rho}) \) as follows: In the first \(|Y|\) levels, we query all the variables \( x_i \) for \( i \in Y \) to obtain a \( \sigma \in \{0,1\}^Y \), and then we follow a decision tree of depth \( dt(f_{\sigma \rho}) \) afterwards. Hence for \( Y \neq \emptyset \), recalling that \( g = \land_{i=2}^{m} C_i \), we have
\[
\Pr[dt(f_\rho) > s \mid F_\rho = 1, \rho_Y = \bar{s}, \rho_{T-Y} = \bar{0}] \leq \Pr[\exists \sigma \in \{0,1\}^{|Y|}, \ dt(f_{\sigma \rho}) > s - |Y| \mid F_\rho = 1, \rho_{T-Y} = \bar{0}] 
\leq \sum_{\sigma \in \{0,1\}^{|Y|}} \Pr[dt(f_{\sigma \rho}) > s - |Y| \mid F_\rho = 1, \rho_{T-Y} = \bar{0}] 
\leq \sum_{\sigma \in \{0,1\}^{|Y|}} \Pr[dt(f_{\sigma \rho}) > s - |Y| \mid (F \land \land_{i \in T \backslash Y} x_i)_{\rho} \equiv 1] 
= \sum_{\sigma \in \{0,1\}^{|Y|}} \Pr[dt(g_{\sigma \rho}) > s - |Y| \mid (F \land \land_{i \in T \backslash Y} x_i)_{\rho} \equiv 1] 
\leq \sum_{\sigma \in \{0,1\}^{|Y|}} \alpha^{s-|Y|} \leq 2^{|Y|} \alpha^{s-|Y|},
\]
where we applied the induction hypothesis.

Combining the two observations with (4), we finish the proof:
\[
\text{L.H.S of (3)} \leq \sum_{\substack{Y \subseteq T \setminus Y \neq \emptyset}} 2^{|Y|} \alpha^{s-|Y|} (2p)^{|Y|} = \alpha^s \sum_{\substack{Y \subseteq T \setminus Y \neq \emptyset}} \left(\frac{4p}{\alpha}\right)^{|Y|} = \alpha^s \left(1 + \frac{4p}{\alpha}\right)^{|T|} - 1 
= \alpha^s \left(1 + \frac{4}{5t}\right)^t - 1 \leq \alpha^s.
\]

**Remark 3.4.** Since the negation of a CNF is a DNF and vice versa, the switching lemma can be used to convert a DNF formula with clauses of size at most \( t \) to a CNF with clauses of size at most \( s \) in the same way as Lemma 3.3. However the statement that “the (conjunctive) clauses in the obtained DNF accept different points” now becomes that “the (disjunctive) clauses in the obtained CNF reject different points”.

**Corollary 3.5.** Let \( f \) be a Boolean function computed by an AC circuit of size \( M \) and depth \( d \), whose output gate is \( \land \). Choose a random restriction \( p \) by setting every variable independently to \( * \) with probability \( p = \frac{1}{10^{d^2 - r}} \), and to 0 and 1 each with probability \( \frac{1-p}{2} \). Then
\[
\Pr[dt(f_\rho) > s] \leq M2^{-s}.
\]
Proof. We view the restriction $\rho$ as obtained by first having a random restriction $\rho_0$ with $\Pr[*] = 1/10$, and then $d - 1$ consecutive restrictions $\rho_1, \ldots, \rho_{d-1}$ each with $\Pr[*] = 1/10^2$. With high probability, after the restriction $\rho_0$, at the bottom level of the circuit all fanins are at most $s$. To see this, consider two cases for each gate at the bottom level of the original circuit:

(1) The original fanin is at least $2s$. In this case, the probability that the gate was not eliminated by $\rho_0$, that is, that no input to this gate got assigned a 1 (assuming without loss of generality that the bottom level is an $\lor$ level) is at most $(0.55)^{2s} < 2^{-s}$.

(2) The original fanin is at most $2s$. In this case, the probability that at least $s$ inputs got assigned a * by $\rho_0$ is at most $\left(\frac{2s}{s}\right)(1/10)^s \leq 2^{-s}$.

Thus, the probability of failure after the first restriction is at most $m_1 2^{-s}$, where $m_1$ is the number of gates at the bottom level.

We now apply the next $d - 2$ restrictions, each with $\Pr[*] = 1/10^2$. After each of these, we use H˚astad’s switching lemma (see Remark 3.4) to convert the lower two levels from CNF to DNF (or vice versa), and collapse the second and third levels (from the bottom) to one level, reducing the depth by one. For each gate of distance two from the inputs, the probability that it corresponds to a function $g$ with $dt(g_\rho) > s$, is bounded by $(5/10^2)^s \leq 2^{-s}$. The probability that a particular gate fails to satisfy the desired property is no more than $2^{-s}$. Since the top gate is $\land$, after these $d - 2$ stages we are left with a CNF formula of bottom fanin at most $s$. We now apply the last restriction and by switching lemma we get a function $f_\rho$ with $dt(f_\rho) \geq s$. The probability of failure at this stage is at most $2^{-s}$. To compute the total probability of failure, we observe that each gate of the original circuit contributed $2^{-s}$ probability of failure exactly once. \hfill \square

Note that if in the above proof we stop before applying the last restriction $\rho_{d-1}$, then we obtain the following corollary which uses a larger value for $p$.

**Corollary 3.6.** Let $f$ be a Boolean function computed by an AC circuit of size $M$ and depth $d \geq 2$ whose output gate is $\land$. Choose a random restriction $\rho$ by setting every variable independently to * with probability $p = \frac{1}{10^{d-1}s^{d-2}},$ and to 0 and 1 each with probability $\frac{1-p}{2}$. Then

$$\Pr[f_\rho \text{ does not have a CNF with fanin } \leq s] \leq M2^{-s}.$$  

Similarly if the output gate of the original circuit is $\lor$, then the probability that $f_\rho$ does not have a DNF with fanin $\leq s$ is bounded by $M2^{-s}$.

4. **Influences in bounded depth circuits**

Let us now introduce an important notion in the study of Boolean functions $f : \{0,1\}^n \rightarrow \{0,1\}$.

**Definition 4.1 (Influence).** Let $f : \{0,1\}^n \rightarrow \{0,1\}$. The influence of the $i$th variable on $f$ is the probability that changing the $i$th coordinate changes the value of $f$. That is,

$$I_i(f) = \Pr[f(x) \neq f(x \oplus e_i)],$$

where $x \in \{0,1\}^n$ is sampled uniformly and $e_i$ is the $i$th standard vector. The total influence of $f$ is defined as

$$I_f = \sum_{i=1}^n I_i(f).$$

Note that always $0 \leq I_i(f) \leq 1$ and $0 \leq I_f \leq n$. The parity function $f : x \mapsto x_1 + \ldots + x_n (\text{mod } 2)$ has total influence $n$, and a constant function has total influence 0.
The sensitivity of a point \( x \) with respect to \( f \), denoted by \( s_f(x) \), is the number of coordinates \( i \) for which \( f(x) \neq f(x \oplus e_i) \). Since
\[
I_f = \mathbb{E}[s_f(x)],
\]
sometimes \( I_f \) is called the average sensitivity of \( f \).

Our next goal is to show that the total influence of small circuits of small depth cannot be large. First we consider the CNF and the DNF circuits with small clauses.

**Lemma 4.2.** Let \( f \) be a CNF or a DNF formula where all the clauses are of size at most \( s \). Then
\[
I_f \leq 2s.
\]

**Proof.** We prove the lemma for the DNF case, and the CNF case follows by replacing \( f \) with \( 1 - f \).

For \( x \in \mathbb{Z}_2^n \), let \( s_f^1(x) \) denote the number of \( i \in [n] \) such that \( f(x) = 1 \) and \( f(x + e_i) = 0 \). Trivially if \( f(x) = 0 \), then \( s_f^1(x) \) since \( f \) is represented by a DNF with clauses of size at most \( s \), for every \( x \), we have \( s_f^1(x) \leq s \). Hence
\[
I_f = \sum_{i=1}^n \Pr[f(x) \neq f(x + e_i)] = 2 \sum_{i=1}^n \Pr[(f(x) = 1) \land (f(x + e_i) = 0)] = 2\mathbb{E}[s_{1 \rightarrow 0}(x)] \leq 2s.
\]

**Theorem 4.3 (Boppana [1]).** Let \( f \) be a Boolean function computed by an AC circuit of depth \( d \) and size \( M \), then
\[
I_f \leq 2(20 \log M)^d.
\]

**Proof.** Applying Corollary 3.6 with \( s = 2 \log M \) and \( p = \frac{1}{10^{10s^d-1}} \), and combining it with the easy fact that \( I_f \leq dt(f) \), we conclude that
\[
\Pr[I_{f_\rho} \geq s] \leq M^{2^{-s}} \leq \frac{1}{M} \leq \frac{1}{n}.
\]

Here we are assuming \( n \leq M \) by counting the input gates in the size of the circuit. Hence
\[
\mathbb{E}_\rho[I_{f_\rho}] \leq \Pr[I_{f_\rho} > s]n + s \leq \frac{1}{n}n + s \leq s + 1 \leq 2s.
\]

On the other hand trivially
\[
\mathbb{E}_\rho[I_{f_\rho}] = \mathbb{E}_{\rho, x}[s_{f_\rho}(x)] = p\mathbb{E}[s_f(x)] = pI_f.
\]

Hence
\[
I_f \leq \frac{2s}{p} \leq 2(10s)^{d-1} = 2(20 \log M)^d.
\]

One can improve the bound slightly by using Corollary 3.6 and Lemma 4.2 instead of the switching lemma.

**Theorem 4.4 (Boppana [1]).** Let \( f \) be a Boolean function computed by an AC circuit of depth \( d \) and size \( M \), then
\[
I_f \leq 2(20 \log M)^{d-1}.
\]

The function PARITY is defined as \( x \mapsto x_1 + \ldots + x_n \pmod{2} \), and the majority function MAJ is defined as \( \text{MAJ}(x) = 1 \) if and only if \( \sum x_i \geq n/2 \). As we saw above the total influence of PARITY is \( n \). It is also straightforward to see that the total influence of MAJ is \( \Theta(\sqrt{n}) \). We conclude

**Corollary 4.5.** The functions PARITY and MAJ are not in \( \text{AC}^0 \).
5. Optional: The Fourier Tail of Functions with Small Bounded Depth Circuits

For this section, it is more convenient to change the domain from \( \{0, 1\}^n \) to \( \{-1, 1\}^n \). Every function \( f : \{-1, 1\}^n \to \mathbb{R} \) has a unique representation as a polynomial in variables \( x_1, \ldots, x_n \). That is

\[
f(x) = \sum_{S \subseteq [n]} \hat{f}(S) \chi_S(x),
\]

where \( \chi_S(x) = \prod_{i \in S} x_i \) (we define \( \chi_\emptyset(x) \equiv 1 \)), and \( \hat{f}(S) \) are real numbers. This representation is sometimes referred to as the Fourier expansion of \( f \) (related to the Fourier analysis of the group \( \mathbb{Z}_2^n \)). The uniqueness of this representation follows from the fact that the monomials are orthonormal with respect to the following inner product: For \( f, g : \{-1, 1\}^n \to \mathbb{R} \),

\[
\langle g, h \rangle := \mathbb{E}[g(x)h(x)].
\]

For a positive integer \( k \), and a function \( f : \mathbb{Z}_2^n \to \mathbb{C} \), we define

\[
f =^k := \sum_{S : |S| \leq k} \hat{f}(S) \chi_S,
\]

and \( f^{\leq k}, f^{\geq k}, f^{< k} \) and \( f^{> k} \) are defined similarly. Note that by the Parseval identity,

\[
\|f\|_2^2 = \sum_{k=0}^{n} \|f =^k\|_2^2.
\]

We leave the proof of the following easy lemma as an exercise to the reader (See Exercise ??).

**Lemma 5.1.** Let \( f : \{0, 1\}^n \to \{0, 1\} \) be computed by a \( \land \)-clause of size \( s \) without repeated variables. Then \( \deg(f) = s \).

**Remark 5.2.** Identifying \( \{0, 1\}^n \) with the \( n \)-dimensional hypercube, note that \( f \) in Lemma 5.1 is the indicator function of a sub-cube of codimension \( s \). So Lemma 5.1 is equivalent to the fact that the degree of the indicator function of a subcube of the hypercube \( \{0, 1\}^n \) is its codimension.

**Corollary 5.3.** Let \( f \) be a Boolean function computed by an AC circuit of size \( M \) and depth \( d \). Choose a random restriction \( \rho \) by setting every variable independently to \( \ast \) with probability \( p = \frac{1}{10^d s^{-2}} \), and to 0 and 1 each with probability \( \frac{1}{2} - p \). Then

\[
\Pr[\deg(f_\rho) > s] \leq M2^{-s}.
\]

**Proof.** Since \( \deg(g) = \deg(1 - g) \) for every function \( g \), we can assume that the output gate of the circuit computing \( f \) is \( \lor \) (otherwise we replace \( f \) with \( 1 - f \) and negate the circuit). Now by Corollary 3.5 with probability at least \( 1 - M2^{-s} \), we have \( f_\rho = \lor_{i=1}^{m} C_i \) for \( \land \) clauses \( C_1, \ldots, C_m \), each of size at most \( s \), such that the clauses all accept disjoint sets of points (i.e. no \( x \in \{-1, 1\}^n \) satisfies more than one clause). By the latter property we can write \( f_\rho = \lor_{i=1}^{m} C_i \), where here we are identifying clauses with the functions represented by them. By Lemma 5.1, we know \( \deg(C_i) \leq s \) for all \( 1 \leq i \leq m \). Hence the degree of their sum is also at most \( s \). We conclude

\[
\Pr[\deg(f_\rho) > s] \leq M2^{-s}.
\]

Now we are at the point to prove the main theorem of this section.
Theorem 5.4 (Linial, Mansour, Nisan [2]). Let $f$ be a Boolean function computed by an AC circuit of depth $d$ and size $M$, and let $t$ be any integer. Then
\[ \sum_{|S|>t} |\hat{f}(S)|^2 \leq 2M2^{-t^{1/d}/20}. \]

Proof. Consider a random restriction $\rho \in \{-1, 1, \ast\}^n$ with $\Pr[\ast] = p \leq \frac{1}{10^d k^{45}}$ for values of $k$ and $p$ to be determined later. We sample $\rho$ in two steps. First we pick $T \subseteq [n]$ corresponding to the positions that are not assigned a $\ast$. Then we pick $x_T \in \{-1, 1\}^T$ uniformly at random, and $\rho$ is defined as $\rho := (x_T, \ast)$. Set $f_{x_T} := f_\rho = f(x_T, \cdot)$. Since $\chi_S(x) = \prod_{i \in S} x_i$, we can decompose it as
\[ \chi_S(x) = \chi_{S \cap T}(x_T)\chi_{S \setminus T}(x_T). \]
Note that $f_{x_T} : \{-1, 1\}^T \to \{0, 1\}$ and since
\[ f(x) = \sum_{S \subseteq [n]} \hat{f}(S)\chi_S(x) = \sum_{S \subseteq [n]} \hat{f}(S)\chi_{S \cap T}(x_T)\chi_{S \setminus T}(x_T) = \sum_{A \subseteq T} \left( \sum_{B \subseteq T} \hat{f}(A \cup B)\chi_B(x_T) \right) \chi_A(x_T), \]
we have
\[ \hat{f}_{x_T}(A) = \sum_{B \subseteq T} \hat{f}(A \cup B)\chi_B(x_T), \]
for every $A \subseteq T$. Hence by the Parseval identity
\[ \mathbb{E}_{x_T} |\hat{f}_{x_T}(A)|^2 = \sum_{B \subseteq T} |\hat{f}(A \cup B)|^2, \]
which shows that
\[ \mathbb{E}_{x_T} \left\| f_{x_T}^{>k} \right\|_2^2 = \mathbb{E}_{x_T} \sum_{A \subseteq T, \, |A|>k} |\hat{f}_{x_T}(A)|^2 = \sum_{A \subseteq T} \sum_{B \subseteq T, \, |A|>k} |\hat{f}(A \cup B)|^2 = \sum_{S \subseteq [n], |S \cap T|>k} |\hat{f}(S)|^2. \]
Now we use the randomness in $T$. Since $f_{x_T}^{>k} = 0$ if $\deg(f_\rho) \leq k$, and that always $\|f_{x_T}^{>k}\|_2^2 \leq \|f_{x_T}\|_2^2 \leq 1$, we have
\[ (6) \quad \mathbb{E}_T \left[ \sum_{S \subseteq [n], |S \cap T|>k} |\hat{f}(S)|^2 \right] = \mathbb{E}_T \mathbb{E}_{x_T} \left\| f_{x_T}^{>k} \right\|_2^2 = \mathbb{E}_\rho \left\| f_\rho^{>k} \right\|_2^2 \leq \Pr[\deg(f_\rho) > k] \leq M2^{-k}, \]
where the last inequality follows from Corollary 5.3 as we chose $\Pr[\ast] = p \leq \frac{1}{10^d k^{45}}$. Also we can bound the left-hand side of (6) from below:
\[ \text{L.H.S. of (6)} = \sum_{S \subseteq [n]} \Pr[|S \cap T| > k]|\hat{f}(S)|^2 \geq \sum_{|S|>t} \Pr[|S \cap T| > k]|\hat{f}(S)|^2. \]
Taking $p = \frac{1}{10^d k^{45}}$ and $k = t^{1/d}/20$, satisfies $p \leq \frac{1}{10^d k^{45}}$, and by the Chernoff bound for $|S| > t$, the probability of $|S \cap T| > k = pt/2$ is at least $1 - 2e^{-\frac{pt}{2t}} \geq \frac{1}{2}$. Hence by (6), we have
\[ \sum_{S \subseteq [n], |S|>t} \frac{1}{2} |\hat{f}(S)|^2 \leq M2^{-t^{1/d}/20}. \]
6. The Razborov-Smolensky Theorem

Taking \( g = f^\leq t \), Theorem 5.4 shows that \( \|f - g\|_2^2 \leq 2M^2^{-t^d/20} \). In other words circuits of low depth and small size can be approximated by functions of low Fourier degree in the \( L_2 \) norm. The next theorem shows a different type of approximating such functions with low degree functions.

**Theorem 6.1** ([3], [4]). Let \( f : \{0, 1\}^n \to \{0, 1\} \) be computed by a circuit of depth \( d \) and size \( M \). For any \( s \), there is a polynomial \( g \) in variables \( x_1, \ldots, x_n \) with degree \( r \leq (s \log M)^d \) such that

\[
\Pr[f(x) \neq g(x)] \leq \left(1 - \frac{1}{2e}\right)^s M.
\]

**Proof.** The function \( g \) is constructed in an inductive way. We will show how to make a step with an \( \land \) gate. Since the whole construction is symmetric with respect to 0 and 1, the step also holds for an \( \lor \) gate. Let

\[
f = \land_{i=1}^k f_i
\]

where \( k < M \). For convenience, let us assume that \( k = 2^\ell \) is a power of 2. For every \( p = 2^{-1}, 2^{-2}, \ldots, 2^{-\ell} = 1/k \) we pick \( s \) random subsets of \( \{1, \ldots, k\} \) by including every element in the subset independently with probability \( p \). We obtain a collection of sets \( S_1, \ldots, S_t \) with \( t = s\ell \leq s \log M \). Let \( g_1, \ldots, g_k \) be the approximating functions for \( f_1, \ldots, f_k \) provided by the previous inductive step. We set

\[
g := \prod_{i=1}^t (1 - |S_i|) + \sum_{j \in S_i} g_j.
\]

By the induction assumption, the degrees of \( g_j \) are \( \leq (s \log m)^d - 1 \), hence the degree of \( f \) is bounded by \( t(s \log m)^d - 1 \leq (s \log m)^d \). Next we bound the probability of \( f(x) \neq g(x) \) conditioned on the event that all of the inputs \( f_1, \ldots, f_k \) are calculated correctly. We have

\[
\Pr[f(x) \neq g(x)|g_j = f_j \text{ for all } j] = \Pr \left[ \prod_{i=1}^t \left(1 - |S_i| + \sum_{j \in S_i} f_j\right) \neq \prod_{j=1}^k f_j \right].
\]

To bound this we fix a vector of specific values \( f_1(x), \ldots, f_k(x) \) and calculate the probability of an error over the possible choices of the random sets \( S_i \). Note that if all the \( f_j(x) \)'s are 1 then the value of \( f(x) = 1 \) is calculated correctly with probability 1. Suppose that \( f(x) = 0 \) (and thus at least one of the \( f_j \)'s is 0). Let \( 1 \leq z \leq k \) be the number of zeros among \( f_1(x), \ldots, f_k(x) \), and \( \alpha \in \mathbb{Z} \) be such that \( 2^\alpha \leq z < 2^{\alpha+1} \). Let \( S \) be a random set with parameter \( p = 2^{-\alpha-1} \). Our approximation will be correct if \( S \) hits exactly one 0 among the \( z \) zeros of \( f_1(x), \ldots, f_k(x) \). The probability of this event is exactly

\[
z p(1 - p)^{z-1} \geq \frac{1}{2} (1 - p)^{1/p - 1} > \frac{1}{2e}.
\]

Hence the probability of being wrong after \( s \) such sets are being chosen is bounded by \( (1 - \frac{1}{2e})^s \) and

\[
\Pr \left[ \prod_{i=1}^t \left(1 - |S_i| + \sum_{j \in S_i} f_j\right) \neq \prod_{j=1}^k f_j \right] < \left(1 - \frac{1}{2e}\right)^s.
\]

By making the same probabilistic argument at every node, by the union bound we conclude that the probability that an error happens is at most \( M \left(1 - \frac{1}{2e}\right)^s \). \( \square \)
References


School of Computer Science, McGill University, Montréal, Canada

Email address: hatami@cs.mcgill.ca