COMP760, LECTURES 6-7: FOURIER SPECTRUM OF FUNCTIONS WITH BOUNDED DEPTH CIRCUITS

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 indegree $k > 0$ are called *gates* and are labeled with a Boolean function on k inputs. The indegree of a vertex is called its *fanin* and its outdegree is called its *fanout*. The most standard circuits are restricted to have gates $\wedge$, $\vee$, $\neg$. 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 $f : \{0,1\}^n \to \{0,1\}$ (there are 2^{2^n} of those functions) compared to the number of small circuits.

Theorem 1.1 (Muller 1956)**.** *Almost every Boolean function $f : \{0,1\}^n \to \{0,1\}$ requires fanin 2 circuits of size $\Omega(2^n/n)$. On the other hand every function $f : \{0,1\}^n \to \{0,1\}$ 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 $\wedge$, $\vee$, $\neg$. Note that by De Morgan's laws

$$\neg(p_1 \vee \ldots \vee p_k) = (\neg p_1) \wedge \ldots \wedge (\neg p_k),$$

and

$$\neg(p_1 \wedge \ldots \wedge p_k) = (\neg p_1) \vee \ldots \vee (\neg p_k),$$

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 $\wedge$ and $\vee$ 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 AC$[d]$ denote the set of a 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, $\wedge$ of $\vee$)**.** *A formula is in conjunctive normal form, abbreviated to CNF, if it is a conjunction (i.e. $\vee$) of clauses, where a clause is a disjunction (i.e $\vee$) of literals (i.e. x_i or $\neg x_i$), where a literal and its negation cannot appear in the same clause*

For example $(x_1 \vee x_2) \wedge (\neg x_1 \vee x_2 \vee x_3)$ 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, $\vee$ of $\wedge$)**.** *A formula is in disjunctive normal form, abbreviated to DNF, if it is a disjunction (i.e. $\vee$) of conjunctive clauses (i.e $\vee$ 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

$$\left(\bigwedge_{i \in T} x_i\right) \wedge \left(\bigwedge_{i \notin T} \neg x_i\right)$$

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 $\vee$ 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

$$(\neg x_1 \wedge \neg x_2) \vee (\neg x_1 \wedge x_2) \vee (x_1 \wedge x_2)$$

represents f. ■

By changing the role of 0's and 1's and $\wedge$ and $\vee$, 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.*

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 $\wedge$ or $\vee$, 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 $\wedge$ gate only needs one 0 input to be set to 0, and an $\vee$ 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 *ρ is an element in $\{0, 1, *\}^X$.*

A restriction ρ is interpreted as setting the variables assigned 0 or 1 and leaving variables those assigned star. Under ρ we may simplify C by eliminating gates whose values become determined. Call this the *induced circuit* C_ρ computing the *induced function* f_ρ.

For a Boolean function $f : \{0,1\}^n \to \{0,1\}$ let $\mathcal{D}(f)$ denote the smallest $s \geq 0$ such that f can be expressed as a DNF formula that satisfies the following two properties:

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

Note that the construction following Definition 2.2 shows that always $\mathcal{D}(f) \leq n$.

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 ρ by setting every variable independently to $*$ with probability p, and to 0 and 1 each with probability $\frac{1-p}{2}$. Then*

$$\Pr[\mathcal{D}(f_\rho) > s] \leq (5pt)^s.$$

Lecture 7

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 ρ by setting every variable independently to $*$ 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*

$$\Pr[\mathcal{D}(f_\rho) > s | F_\rho \equiv 1] \leq (5pt)^s. \tag{1}$$

Proof. Set $\alpha := 5pt$, and suppose that $f = \wedge_{i=1}^{m} C_i$ 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 C_1, the first clause 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 to prove (1) it suffices to prove both

$$\Pr[\mathcal{D}(f_\rho) > s | F_\rho \equiv 1,\ \rho_T \in \overline{\{0,*\}^T}] \leq \alpha^s, \tag{2}$$

and

$$\Pr[\mathcal{D}(f_\rho) > s | F_\rho \equiv 1,\ \rho_T \in \{0,*\}^T] \leq \alpha^s. \tag{3}$$

To prove (2) note that

$$\text{L.H.S of (2)} \quad = \quad \Pr[\mathcal{D}(f_\rho) > s \mid (F \wedge C_1)_\rho \equiv 1] = \Pr[\mathcal{D}((\wedge_{i=2}^{m} C_i)_\rho) > s \mid (F \wedge C_1)_\rho \equiv 1] \leq \alpha^s,$$

where in the last inequality we used the induction hypothesis. It remains to prove (3). Note that if $\rho_T = \vec{0}$, then $f_\rho \equiv 0$ and thus $\mathcal{D}(f_\rho) = 0$. Hence

$$\begin{aligned}
\text{L.H.S of (3)} \quad &= \sum_{\substack{Y \subseteq T \\ Y \neq \emptyset}} \Pr[\mathcal{D}(f_\rho) > s,\ \rho_Y = \vec{*},\ \rho_{T-Y} = \vec{0} \mid F_\rho \equiv 1,\ \rho_T \in \{0,*\}^T] \\
&\leq \sum_{\substack{Y \subseteq T \\ Y \neq \emptyset}} \Pr[\rho_Y = \vec{*}, \rho_{T-Y} = \vec{0} \mid F_\rho \equiv 1,\ \rho_T \in \{0,*\}^T] \times \\
&\qquad \Pr[\mathcal{D}(f_\rho) > s \mid F_\rho \equiv 1,\ \rho_Y = \vec{*},\ \rho_{T-Y} = \vec{0},\ \rho_T \in \{0,*\}^T] \\
&\leq \sum_{\substack{Y \subseteq T \\ Y \neq \emptyset}} \Pr\left[\rho_Y = \vec{*} \middle| F_\rho \equiv 1, \rho_T \in \{0,*\}^T\right] \times \Pr\left[\mathcal{D}(f_\rho) > s \middle| F_\rho \equiv 1, \rho_Y = \vec{*}, \rho_{T-Y} = \vec{0}\right]. \qquad (4)
\end{aligned}$$

Observation 1: Since setting variables in Y to $*$ cannot increase the probability that $F_\rho \equiv 1$, we have

$$\Pr[F_\rho \equiv 1 \mid \rho_Y = \vec{*},\ \rho_T \in \{0,*\}^T] \leq \Pr[F_\rho \equiv 1 \mid \rho_T \in \{0,*\}^T],$$

which using the formula $\Pr[A|B]\Pr[B] = \Pr[A \wedge B]$ implies that

$$\Pr[\rho_Y = \vec{*} \mid F_\rho \equiv 1,\ \rho_T \in \{0,*\}^T] \leq \Pr[\rho_Y = \vec{*} \mid \rho_T \in \{0,*\}^T] = \left(\frac{2p}{1+p}\right)^{|Y|} \leq (2p)^{|Y|}.$$

Observation 2: Also defining $G : \{0,1\} \to \{0,1\}$ as

$$G : x \mapsto \begin{cases} 0 & x_{T \setminus Y} \neq \vec{0} \\ F(x) & x_{T \setminus Y} = \vec{0}, \end{cases}$$

note that by induction hypothesis,

$$\begin{aligned}
&\Pr[\mathcal{D}(f_\rho) > s \mid F_\rho \equiv 1, \rho_Y = \vec{*}, \rho_{T-Y} = \vec{0}] \\
\leq\ & \Pr\left[\forall \sigma \in \{0,1\}^{|Y|},\ \mathcal{D}(f_{\sigma\rho_{\overline{Y}}}) > s - |Y| \,\middle|\, F_\rho \equiv 1, \rho_{T-Y} = \vec{0}\right] \\
\leq\ & \sum_{\sigma \in \{0,1\}^{|Y|}} \Pr[\mathcal{D}(f_{\sigma\rho_{\overline{Y}}}) > s - |Y| \mid G_\rho \equiv 1] \\
\leq\ & \sum_{\sigma \in \{0,1\}^{|Y|}} \alpha^{s-|Y|} \leq 2^{|Y|}\alpha^{s-|Y|}.
\end{aligned}$$

Combining the two observations with (4), we finish the proof:

$$\text{L.H.S of (3)} \quad \leq \quad \sum_{\substack{Y \subseteq T \\ Y \neq \emptyset}} 2^{|Y|}\alpha^{s-|Y|}(2p)^{|Y|} = \alpha^s \sum_{\substack{Y \subseteq T \\ Y \neq \emptyset}} \left(\frac{4p}{\alpha}\right)^{|Y|} \leq \alpha^s.$$

□

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 (Linial, Mansour, Nisan 1993). *Let f be a Boolean function computed by an* AC *circuit of size M and depth d whose output gate is $\wedge$. Choose a random restriction ρ by setting every variable independently to $*$ with probability $p = \frac{1}{10^d s^{d-1}}$, and to* 0 *and* 1 *each with probability $\frac{1-p}{2}$. Then*

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

Proof. We view the restriction ρ as obtained by first having a random restriction ρ_0 with $\Pr[*] = 1/10$, and then $d-1$ consecutive restrictions $\rho_1, \ldots, \rho_{d-1}$ each with $Pr[*] = \frac{1}{10s}$. With high probability, after the restriction ρ_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 ρ_0, that is, that no input to this gate got assigned a 1 (assuming without loss of generality that the bottom level is an $\vee$ 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 ρ_0 is at most $\binom{2s}{s}(1/10)^s \leq 2^{-s}$.

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 $Pr[*] = \frac{1}{10s}$. After each of these, we use Hastad'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 $\mathcal{D}(g_{\rho_i}) > s$, is bounded by $(5\frac{1}{10s}s)^s \leq 2^{-s}$. The probability that some gate fails to satisfy the desired property is no more than $m_i 2^{-s}$, where m_i is the number of gates at level i. Since the top gate is $\wedge$, 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_ρ with

$\mathcal{D}(f_\rho) \geq s$. 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. □

Lecture 8

Recall that the characters of $\mathbb{Z}_2^n$ are $\chi_S : x \mapsto (-1)^{\sum_{i\in S} x_i}$ for $S \subseteq [n]$. So in this notation, the Fourier expansion of $f : \mathbb{Z}_2^n \to \mathbb{C}$ is $f = \sum_{S\subseteq[n]} \widehat{f}(S)\chi_S$. We think of $|S|$ as the "frequency" of the character χ_S. This corresponds to the fact that when $|S|$ is small, χ_S is more stable under local changes (e.g. change of one random coordinate).

Remark 3.6. Note that χ_S has two important properties:

- $\chi_S(x)$ depends only on the coordinates in S. That is $\chi_S(x) = \chi_S(x_S)$.
- For every $i \in [n]$ and $x_{[n]\setminus\{i\}} \in \mathbb{Z}_2^{[n]\setminus\{i\}}$, we have

$$\mathbb{E}_{x_i}\left[\chi_S(x)\right] = \left\{ \begin{array}{ll} 0 & \quad i \in S \\ 1 & \quad y \notin S \end{array} \right.$$

■

The Fourier degree of a function $f : \mathbb{Z}_2^n \to \mathbb{C}$, denoted by $\deg(f)$, is the size of the largest S such that $\widehat{f}(S) \neq 0$. For a positive integer k, and a function $f : \mathbb{Z}_2^n \to \mathbb{C}$, we introduce the following notations:

$$f^{=k} = \sum_{S:|S|\leq k} \widehat{f}(S)\chi_S.$$

Also $f^{\leq k}$, $f^{\geq k}$, $f^{<k}$ and $f^{>k}$ are defined similarly. Note that by Parseval identity we can prove

$$\|f\|_2^2 = \sum_{k=0}^{n} \|f^{=k}\|_2^2,$$

and similar identities.

Lemma 3.7. *Let f be the indicator function of a restriction $\rho \in \{0,1,*\}^n$. Then* $\deg(f)$ *is at most the number of non-$*$ coordinates of ρ.*

Remark 3.8. Equivalently:

- the degree of the indicator function of a subcube of the hypercube $\{0,1\}^n$ is at most its codimension.
- if f can be expressed as an $\wedge$ clause if size at most k, then $\deg(f) \leq k$.

■

Proof of Lemma 3.7. Let T denote the set of fixed (non-$*$) coordinates by ρ. Note that f depends on the coordinates in T, that is $f(x) = f(x_T)$. Hence if there exists $i \in S \setminus T$, then by Remark 3.6 we have

$$\widehat{f}(S) = \mathbb{E}f(x)\chi_S(x) = \mathbb{E}_{x_{[n]\setminus\{i\}}}\left[\mathbb{E}_{x_i}\left[f(x_T)\chi_S(x)\right]\right] = \mathbb{E}_{x_{[n]\setminus\{i\}}}\left[f(x_T)\left[\mathbb{E}_{x_i}\chi_S(x)\right]\right] = 0.$$

We conclude that $\widehat{f}(S) = 0$ if $S \not\subseteq T$. Hence $\deg(f) \leq |T|$. □

Corollary 3.9. *Let f be a Boolean function computed by an* AC *circuit of size M and depth d. Choose a random restriction ρ by setting every variable independently to $*$ with probability $p = \frac{1}{10^d s^{d-1}}$, and to 0 and 1 each with probability $\frac{1-p}{2}$. 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 $\vee$ (otherwise we replace f with $1-f$ and negate the circuit). Now by switching lemma with probability at least $1 - M2^{-s}$, we have $f_\rho = \vee_{i=1}^m C_i$ for $\wedge$ 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 $x \in \{0,1\}^n$ satisfies more than one clause). By the latter property we can write $f_\rho = \sum_{i=1}^m C_i$ where here we are identifying clauses with the functions represented by them. By Lemma 3.7, 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 3.10 (Linial, Mansour, Nisan)**.** *Let f be a Boolean function computed by an AC circuit of depth d and size M, and let t be any integer. Then*

$$\sum_{|S|>t} |\widehat{f}(S)|^2 \leq 2M2^{t^{-1/d}/20}.$$

Proof. Consider a random restriction $\rho \in \{0,1,*\}^n$ with $\Pr[*] = p \leq \frac{1}{10^d k^{d-1}}$ for a value of k and s to be determined later. We sample ρ in two steps. First we pick $T \subseteq [n]$ corresponding to the positions that are not assigned a $*$. Then we pick $x_T \in \{0,1\}^T$ uniformly at random, and ρ is defined as $\rho := (x_T, \vec{*})$. Set $f_{x_T} := f_\rho = f(x_T, \cdot)$. Since $\chi_S(x) = \prod_{i \in S}(-1)^{x_i}$, we can decompose it as

$$\chi_S(x) = \chi_{S \cap T}(x_T)\chi_{S \setminus T}(x_{\overline{T}}).$$

Now since

$$f(x) = \sum_{S \subseteq [n]} \widehat{f}(S)\chi_S(x) = \sum_{S \subseteq [n]} \widehat{f}(S)\chi_{S\cap T}(x_T)\chi_{S\setminus T}(x_{\overline{T}}) = \sum_{A \subseteq \overline{T}} \left(\sum_{B \subseteq T} \widehat{f}(A \cup B)\chi_B(x_T) \right) \chi_A(x_{\overline{T}}),$$

we have

$$\widehat{f_{x_T}}(A) = \sum_{B \subseteq T} \widehat{f}(A \cup B)\chi_B(x_T),$$

for every $A \subseteq \overline{T}$. Hence by Parseval identity

$$\mathbb{E}_{x_T} \left| \widehat{f_{x_T}}(A) \right|^2 = \sum_{B \subseteq T} |\widehat{f}(A \cup B)|^2,$$

which shows that

$$\mathbb{E}_{x_T} \left\| f_{x_T}^{>k} \right\|_2^2 = \mathbb{E}_{x_T} \sum_{\substack{A \subseteq \overline{T} \\ |A|>k}} \left| \widehat{f_{x_T}}(A) \right|^2 = \sum_{\substack{A \subseteq \overline{T} \\ |A|>k}} \sum_{B \subseteq T} |\widehat{f}(A \cup B)|^2 = \sum_{S:|S \cap \overline{T}|>k} |\widehat{f}(S)|^2$$

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

$$\mathbb{E}_T\left[\sum_{S:|S\cap\overline{T}|>k} |\widehat{f}(S)|^2\right] = \mathbb{E}_T\mathbb{E}_{x_T}\left\|f_{x_T}^{>k}\right\|_2^2 = \mathbb{E}_\rho\left\|f_\rho^{>k}\right\|_2^2 \leq \Pr[\deg(f_\rho) > k] \leq M2^{-k}, \tag{5}$$

where the last inequality follows from the switching lemma (recall that we chose $\Pr[*] = p \leq \frac{1}{10^d k^{d-1}}$). Also we can bound the left-hand of (5) from below:

$$\text{L.H.S. of (5)} \geq \sum_{|S|>t} \Pr[|S\cap\overline{T}| > k]|\widehat{f}(S)|^2$$

Taking $p = \frac{1}{10t^{(d-1)/d}}$ and $k = t^{1/d}/20$, we satisfy $p \leq \frac{1}{10^d k^{d-1}}$, and by Chernoff bound for $|S| > t$, the probability of $|S\cap\overline{T}| > k = pt/2$ is at least $1 - 2e^{\frac{-pt}{12}} \geq \frac{1}{2}$. Hence by (5), we have

$$\sum_{S:|S|>t} \frac{1}{2}|\widehat{f}(S)|^2 \leq M2^{-t^{1/d}/20}.$$

□

Lecture 9

Taking $g = f^{\leq t}$, Theorem 3.10 shows that $\|f-g\|_2^2 \leq 2M2^{t^{-1/d}/20}$. In other words circuits of low depth and small size can be approximated by functions of low degree in the L_2 norm. The next theorem shows a different type of approximating such functions with a low degree function.

Theorem 3.11 (Razborov 87, Smolensky 87)**.** *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 degree $r \leq (s\log M)d$ function g 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 $\wedge$ gate. Since the whole construction is symmetric with respect to 0 and 1, the step also holds for an $\vee$ gate. Let

$$f = \wedge_{i=1}^k f_i$$

where $k < M$. For convenience, let us assume that $k = 2^\ell$ is a power of 2. We take a collection of $t := s\log M$ random Poisson subsets of $\{1,\ldots,k\}$: at least s of each of the $p = 2^{-1}, 2^{-2}, \ldots, 2^{-\ell} = 1/k$. Denote the sets by $S_1,\ldots,S_t$. In addition, we make sure to include $\{1,\ldots,k\}$ as one of the sets. Let $g_1,\ldots,g_k$ be the approximating functions for $f_1,\ldots,f_k$ provided by the previous inductive step. We set

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

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

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

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 α be such that $2^\alpha \leq z < 2^{\alpha+1}$. Let S be a random Poisson set with $p = 2^{-\alpha-1}$. 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

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

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

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

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

School of Computer Science, McGill University, Montréal, Canada
E-mail address: hatami@cs.mcgill.ca