COMS 6998-2 Advanced Topics in Computational Learning Theory — Spring 2005

Lecture 4 : February 9, 2005

Lecturer: Rocco A. Servedio — *Scribe: David Hessing & Homin Lee*

1 Preliminaries

Admin: Project page is up.

Last Time:

- Discussed learning LTFs and PTFs
- We showed we can learn d-degree PTFs in $n^{O(d)}$ time.
- We want PTF degree bounds

Today:

- Upper and Lower bounds on PTF degree for PTFs of Parity and DNF

Reading:

- Beigel paper.
- Minsky/Papert book.
- Klivans/Servedio *Learning DNF in Time* $2^{O(n^{1/3})}$

2 Degree of PTF for Parity

The parity function is defined as:

$$PAR(x_1, x_2, \ldots, x_n) = -1^{\sum_{i=0}^{n} x_i} \tag{1}$$

Theorem 1 *Any PTF for PAR must have degree greater than n.*

Proof: Let $p(x_1, x_2, \ldots, x_n)$ be degree-d PTF for PAR.
Consider the following function Q:

$$Q(x_1, x_2, \ldots, x_n) = \sum_{\forall \pi} p(x_{\pi(1)}, x_{\pi(2)}, \ldots, x_{\pi(n)}) \tag{2}$$

where the summ is over every permutation $\pi : \{1, \ldots, n\} \rightarrow \{1, \ldots, n\}$.

Regarding equation (2) we can say the following:

- $deg(Q) < d$
- Q is symmetric, multilinear, and a PTF for parity

Recall that a symmetric multilinear polynomial has the property that all coefficients of the monomials of equal degrees are equal, e.g. if the coefficient of $x_3x_4x_5$ is 7, the coefficients of all monomials of degree 3 are also 7. Formally, we express this property as follows:

$$Q(x_1, x_2, \ldots, x_n) = \sum_{i=0}^{d} \alpha_i \sum_{s \subseteq [n] \text{ s.t. } |s|=i} x_s \tag{3}$$

where α is the coefficient and $\sum_{s \subseteq [n] \text{ s.t. } |s|=i} x_s$ means that for $s \subseteq [n]$, the variable x_s is $\prod_{j \in s} x_j$ e.g. if $s = \{1, 3, 4\}$ then $x_s = x_1x_3x_4$.

Fact 1 *If Q is a symmetric multilinear polynomial of degree d then there is a 1-variable polynomial $R(t)$ of degree d such that for every $x \in \{0, 1\}^n$ we have that $Q(x_1, x_2, \ldots, x_n) = R(x_1 + x_2 + \cdots + x_n)$*

Proof: The value of any symmetric polynomial function is dependent only on the sum of the inputs.

The proof is a straightforward induction and is not enumerated here. ■

Thus, $R(x_1 + x_2 + \cdots + x_n) > 0$ iff $\sum_{i=1}^{n} x_i$ is even and therefore $R(t) \geq 0$ for $t = 0, 1, 2, \ldots, n$ iff t is even.

Thus, if we plot $R(t)$ it must look like the graph of Figure 1.

Therefore, $deg(R) \geq n$ because of the number of sign changes.

■

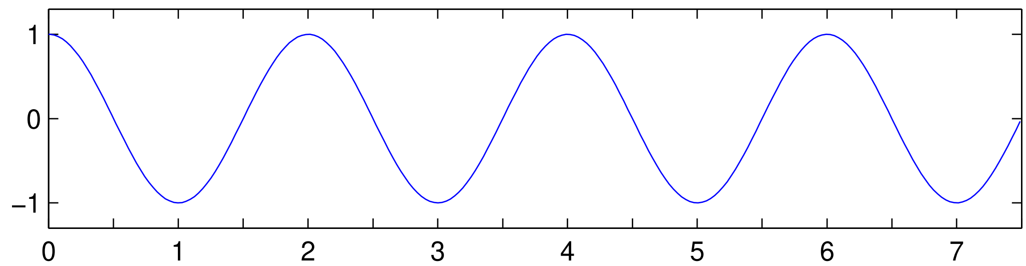

Figure 1: Plot of the function $R(t)$

3 Degree of PTF for Decision Trees

Given what we just proved about the degree of PTFs for PAR, can we learn anything about PTFs for Decision Trees?

Fact 2 *There exist DTs of size s which have PTF degree greater than* $\log(s)$.

Proof: Given what we proved in the previous section, this follows directly from a fact learned in the first lecture, namely that $PAR(x_1, x_2, \ldots x_n)$ has a DT of size 2^n constructed as in the Figure 2.

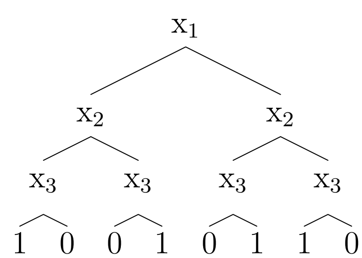

Figure 2: A DT that computes $PAR(x_1, x_2, x_3)$

■

Fact 3 *Any size-s DT has a PTF of at most degree* $\log s$.

Proof: Any DT has a ($\log s$)-DL. Construct the DL by explicitly enumerating all paths down the tree, making a conjunction C_i for each path. Each path is of size $\log s$. String together the conjunctions, each with the correct output, to make the DL.

Given any d-DL, we learned earlier how to convert the DL to a PTF of degree d. As a simple example consider a 3-DL where C_1 yielded 1 and equals $x_1\overline{x_2}x_3$ and C_2 yielded -1 and equals $\overline{x_1}x_2x_3$ etc. Then the corresponding PTF would be $2^n[x_1(1-x_2)x_3] - 2^{n-1}[(1-x_n)x_4x_5] \pm \cdots \geq 0$ ■

4 Degree of PTF for DNFs

The following theorem is due to Minsky and Papert.

Theorem 2 *Let $n = 4m^3$. Construct the function f as follows:*

$$f(x_1, x_2, \ldots x_n) = (x_1x_2 \ldots x_{4m^2}) \vee (x_{4m^2+1} \ldots x_{8m^2}) \vee \cdots \vee (x_{n-4m^2+1} \ldots x_n) \quad (4)$$

where f has m terms, each with $4m^2$ variables, and each variable is in exactly one term. Any PTF for f must have degree $\geq m$. In terms of n, size s for f is $n^{1/3}$ and the PTF degree is $n^{1/3}$

Proof: Let $P(x_1, x_2, \ldots x_n)$ be a degree-d PTF for f (we'll show that $d \geq m$). Group the inputs of f into sets A_1, A_2, etc. where $A_1 = \{x_1, \ldots, x_{4m^2}\}$, $A_2 = \{x_{4m^2+1}, \ldots x_{8m^2}\}$ and so on. For any permutation $\pi_1, \ldots \pi_m$ on the set $\{1, 2, \ldots 4m^2\}$ we have that $P(\pi_1(A_1), \pi_2(A_2), \ldots, \pi_m(A_m))$ is a PTF for f. Define the function Q as follows:

$$Q(x_1, \ldots, x_n) = \sum_{\pi_1, \ldots, \pi_m} P(\pi_1(A_1), \pi_2(A_2), \ldots, \pi_m(A_m)) \quad (5)$$

Regarding equation (5) we can say the following:

- $deg(Q) \leq d$
- Q is multilinear, is a PTF for f and is symmetric in the elements of the sets A_1, A_2, etc.

Fact 4 *Given these properties of Q, there is a polynomial $R(v_1, \ldots v_n)$ of degree-(Q) such that for all $x \in \{0,1\}^n$*
$Q(x_1, x_2, \ldots, x_n) = R(x_1 + \cdots + x_{4m^2}, (x_{4m^2+1} + \cdots + x_{8m^2}), \ldots, (x_{n-4m^2+1} + \cdots + x_n))$

Thus, for any m integers $a_1, a_2, \ldots, a_m$ with $0 \leq a_i \leq 4m^2$ we have that $R(a_1, a_2, \ldots, a_m) > 0$ iff some $a_i = 4m^2$. Now, let $S(t) = R(a_i(t), \ldots, a_m(t))$ where $a_i(t) = 4m^2 - (2i - 1 - t)^2$ for $i = 1 \ldots m$. So, for example:

- $a_1(t) = 4m^2 - (1 - t)^2$
- $a_2(t) = 4m^2 - (3 - t)^2$

Note that if t is even then none of $a_i(t)$ equal $4m^2$. If t is odd then some $a_i(t)$ equals $4m^2$. Note also that $deg(S) \leq 2d$. Looking at S we see that for any $t = 0, 1, 2, \ldots, 2m^2$ and for any $i = 1, 2, 3, \ldots, m$ we have that $a_i(t) \in \{0, 2, \ldots, 4m^2\}$. Thus, $S(0) < 0$, $S(1) > 0$, $S(2) < 0$, ..., $S(2m) < 0$. Thus, there are $2m$ sign changes and $deg(S) \geq 2m$. Since we constructed S such that $deg(S) \leq 2d$, we conclude that $d \geq m$ ■

Theorem 3 *Any size-s DNF over $\{0,1\}^n$ has a PTF of degree $O(n^{1/3} \log s)$*

Proof: The essence of the proof is a transformation from a DNF to a PTF in stages. We have covered almost all of the required transformations in previous lectures, but we need the following lemma. After we have proved the following lemma, all the ingredients will be in place, and we give the outline of the steps needed to complete the transformation.

In slightly more detail, what we plan to do is take the following two facts and combine them.

1. Recall the Bshouty theorem, which states that any s-term DNF has a $O(\sqrt{n \log n \log s})$-DL. This tells us that any s-term DNF has an $O(\sqrt{n \log n \log s})$-degree PTF
2. The lemma we are about to introduce and prove tell us that any s-term t-DNF has a $O(\sqrt{t} \log s)$ degree PTF. Since in our case we always have $t \leq n$, this lemma gives us a $O(\sqrt{n} \log s)$ PTF bound for any s-term DNF.

Lemma 1 *Any s-term t-DNF (recall, s terms each of length t) has a $O(\sqrt{t} \log s)$ degree PTF*

Proof: First note that a simple bound of just t is trivial to prove. Simply convert the DNF in the standard way we've already done several times; e.g. $x_1 x_2 \ldots x_t \vee \overline{x_1 x_2} \ldots \overline{x_t}$ translates to the t-degree PTF $x_1 x_2 \ldots x_t + (1 - x_1)(1 - x_2) \ldots (1 - x_t) \geq \frac{1}{2}$

Roughly, the lemma lets us simulate certain degree-t monomials with degree-($\sqrt{t}\log s$) monomials. We now consider how to achieve this substitution.

Consider the graph of the monomial $x_1x_2\ldots x_t$, shown in Figure 3. Note that the graph is a simple point; $x_1x_2\ldots x_t = 1$ only when all inputs are 1, and is 0 otherwise.

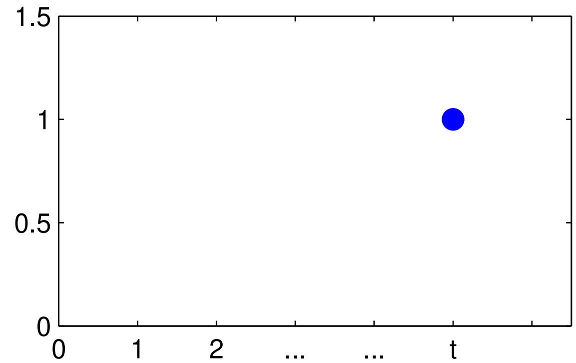

Figure 3: Plot of $x_1x_2\ldots x_t$

Now consider the graph of $(\frac{x_1+x_2+\cdots+x_t}{t})^t$ shown in Figure 4. Note that at $t-1$ we have $(\frac{t-1}{t})^t = (1-\frac{1}{t})^t \approx \frac{1}{e} \approx 0.37$. In other words, for all points below t the function is well below 1, but then shoots up to 1 at t and increases rapidly thereafter.

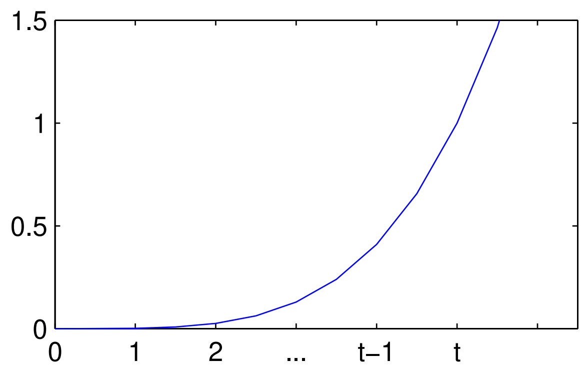

Figure 4: Plot of $(\frac{x_1+x_2+\cdots+x_t}{t})^t$

We would like to get a similar looking polynomial of degree $\approx \sqrt{t}$. Is this possible?

Math Question: Framing this question as a pure math question, we ask if we fix an integer $d \geq 1$, what is the polynomial $p(x)$ of degree d that has the following properties:

1. $|p(x)| \leq 1$ for $x \in [-1, 1]$
2. $p(1) = 1$
3. $p'(1)$ is a as big as possible

Math Answer: The Chebyshev Polynomial of the First Kind, which we will call $C_d(x)$, satisfies these properties.

For more information on the Chebyshev Polynomial of the First Kind, see `http://mathworld.wolfram.com/ChebyshevPolynomialoftheFirstKind.html`. For our purposes, the following definition of the function suffices:

$$C_d(x) = \begin{cases} \cos(d \cos^{-1} x) & |x| \leq 1 \\ \cosh(d \cosh^{-1} x) & |x| > 1 \end{cases} \tag{6}$$

Fact 5 *$C_d(x)$ satisfies the following properties:*

1. *$C_d(x) \leq 1$ for $|x| \leq 1$*
2. *$C_d(1) = 1$ and $C'_d(1) = d^2$*
3. *$C'_d(x) \geq d^2$ for $|x| \geq 1$*
4. *At $x = 1 + \frac{1}{d^2}$, $C_d(x) \geq 2$*

Thus, over $[0, 1 + \frac{1}{d^2}]$ we see that $C_d(x)$ mimics x^{d^2}. Now that we have defined $C_d(x)$ we can prove the lemma. Let $f = T_1 \vee T_2 \vee \cdots \vee T_s$ be a t-DNF. As we've seen several times, given a literal l_i, the arithmetization of l_i is x_j if $l_i = x_j$ and is $1 - x_j$ if $l_i = \overline{x_j}$. Given T_i let S_i equal the sum of the arithmetization of the literals in T_i. e.g. if $T_i = \overline{x_1} x_2 \overline{x_3}$ then $S_i = 1 - x_1 + x_2 + 1 - x_3$. Let t_i equal the number of literals in T_i. Now define the function Q as follows:

$$Q_i(x_1 \ldots x_n) = C_{\sqrt{t}}\left(\frac{S_i}{t_i}\left(1 + \frac{1}{t}\right)\right) \tag{7}$$

The properties of the Chebyshev polynomial imply that if $T_i(x) = 1$ then $S_i = t_i$ and $Q_i(x) = C_{\sqrt{t}}(1 + \frac{1}{t}) \geq 2$. Otherwise, if $T_i(x) = 0$ then $S_i \leq t_i - 1$ so $\frac{S_i}{t_i} \leq 1 - \frac{1}{t_i} \leq 1 - \frac{1}{t}$

which in turn implies $\frac{S_i}{t_i}(1+\frac{1}{t}) \leq 1-\frac{1}{t^2} < 1$ and $|Q_i(x)| \leq 1$. Now, define the following polynomial, $P(x)$:

$$P(x) = \sum_{i=1}^{s} Q_i(x)^{\log 2s} \tag{8}$$

Consider $P(x) \geq s + \frac{1}{2}$. (Recall lowercase s is the number of terms in f, not the arithmetization of a particular term.) $P(x) \geq s + \frac{1}{2}$ is a PTF for f. Take any $x \in \{0,1\}^n$:

- If $f(x) = 1$ some $T_i(x) = 1$ so $Q_i(x)^{\log 2s} \geq 2s$. If every other $T_i(x) = 0$ then all $Q_j(x)^{\log 2s} \geq -1$ so $P(x) = 2s - (s-1) = s + 1$.
- If $f(x) = 0$ all $T_i(x) = 0$, so every $|Q_i(x)| \leq 1$, so every $Q_j(x)^{\log 2s} \leq 1$ so $P(x) \leq s$.

This completes the proof of the lemma. ■

We now can prove Theorem 3. The transformation from DNF to PTF involves the following steps:

1. Convert DNF to DT with t-DNFs as the leaves using modified Bshouty theorem.
2. Convert the t-DNF leaves to $(\sqrt{t}\log s)$-degree PTFs using the lemma we just proved.
3. Convert the entire DT to a DL using the Avrim Blum transformation.
4. Convert the DL to the final PTF.

We explore each step in more detail.

Step 1: DNF to DT with t-DNF leaves

The Bshouty s-length n-DL proof specialized to s-term DNFs gives the following lemma:

Lemma 2 *Let $h = O(\sqrt{n \log n \log s})$. Any s-term DNF over $x_1, x_2, \ldots, x_n$ is equivalent to a DT of size 2^h with s-term h-DNFs at the leaves.*

Proof: The proof is exactly the same as before, just using DNFs. The following are also true of the Bshouty theorem:

1. For any $t = 1, 2, \ldots, n$ the constructed DT will need to be of size $\approx 2^{n/t}$ in order to get t-DNFs at the leaves.

2. The Bshouty proof can be done with rank instead of size.

■

Thus the Bshouty theorem allows us to state the following lemma:

Lemma 3 *Let $f : \{0,1\}^n \to \{-1,1\}$ be an s-term DNF. Then for any $t = 1, 2, \ldots, n$ f is equivalent to a decision tree T with $rank(T) \leq \frac{2n}{t} \ln s + 1$ where each leaf of T is an s-term t-DNF.*

Step 2: t-DNF leaves to $(\sqrt{t}\log s)$-degree PTFs

Using Lemma 1 we apply the transformation to each leaf, and state the following lemma:

Lemma 4 *Let $f : \{0,1\}^n \to \{-1,1\}$ be an s-term DNF. For $1 \leq t \leq n$, f is equivalent to a decision tree T with $rank(T) \leq \frac{2n}{t} \ln s + 1$ where each leaf of T is a $O(\sqrt{t}\log s)$-degree PTF.*

Step 3: DT to DL using Blum Trick

Using the Avrim Blum transformation learned earlier, we can state the following lemma:

Lemma 5 *Let $f : \{0,1\}^n \to \{-1,1\}$ be an s-term DNF. f can be expressed as a R-DL with $O(\sqrt{t}\log s)$-degree PTFs as the output "bits". $R = (\frac{2n}{t} \ln s + 1)$ (from Step 2).*

Step 4: DL to PTF

Figure 5 depicts L, the DL obtained after performing steps $1 - 3$ on f. (L is depicted as a degenerate DT.) Each leaf, $P_i(x)$ is a $(\sqrt{t}\log s)$-PTF.

Claim 1 *Consider some $P_i(x)$. Without loss of generality, we have that for all $x \in \{0,1\}^n$, $|P_i(x)| \geq 1$. This is trivially true due to the properties of polynomials.*

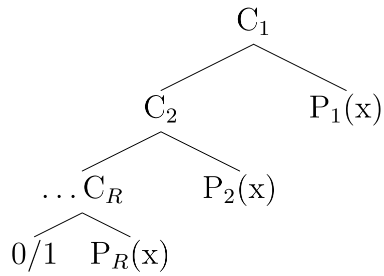

Figure 5: L, the DL obtained after steps $1-3$

Now, consider each conjunction C_i. For each one, let $\widetilde{C_i}$ be the corresponding polynomial of degree $\leq \frac{2n}{t}\ln s+1$ (since C_i is a conjunction of at most this many variables). $\widetilde{C_i}$ outputs 0/1 on all $x \in \{0,1\}^n$ that exactly agrees with C_i. So, for example if $C_i(x) = x_3\overline{x_4 x_5}x_8$ then $\widetilde{C_i}(x) = x_3(1-x_4)(1-x_5)x_8$.

Using this conversion, the final PTF for L will be:

$$Z(x) = A_1\widetilde{C_1}(x)P_1(x) + A_2\widetilde{C_2}(x)P_2(x) + \cdots + A_R\widetilde{C_R}(x)P_R(x) \geq 0 \tag{9}$$

where $A_R = 1$, $A_{R-1} >> A_R$, $A_{R-2} >> A_{R-1}$, etc.

By considering any $x \in \{0,1\}^n$ once can easily convince oneself that equation (9) is a PTF that agrees with L on all $x \in \{0,1\}^n$.

Finally, since each $\widetilde{C_i}$ is of degree $R = (\frac{2n}{t}\ln s + 1)$, and each P_i is of degree $\sqrt{t}\log s$ we conclude that:

$$deg(Z) \leq \frac{2n}{t}\ln s + 1 + \sqrt{t}\log s \tag{10}$$

To minimize this we take $t = n^{2/3}$ we obtain the desired bound of $O(n^{1/3}\log s)$

■

5 Some Closing Points

We conclude with two open questions.

First, regarding the degree of PTFs for s-term DNFs over n variables, we've seen both lower bounds and upper bounds. We saw lower bounds of $\log s$ and $n^{1/3}$. We just proved the upper bound of $n^{1/3} \log s$. Is there an exact answer and if so what is it?

Second, consider bounds on the degree of PTSs for depth-3 boolean circuits of s-gates. The best bounds are pretty poor, with the best lower bound being $n^{1/3} \log^{1/3} s$ and $\log^2 s$ and the best upper bound being $n-1$. Are there better bounds, and if so what are they?