CSG399: Gems of Theoretical Computer Science. Lectures 4,5. Jan. 20,23, 2009.
Instructor: Emanuele Viola Scribe: Aldo Cassola

Parity Requires Large Constant-Depth Circuits (I)

In the following lectures we will examine how the PARITY function correlates with circuits.

1 Parity Requires Large Circuits

Theorem 1. *Consider the* PARITY *function on n bits. There is an absolute constant $\epsilon > 0$ such that for every circuit $C : \{0,1\}^n \to \{0,1\}$ of depth d and size $w := 2^{n^{\epsilon/d}}$ we have*

$$\mathrm{COR}(\mathrm{PARITY}, C) \leq 1/w.$$

We will prove this theorem in 2 stages. First, we will establish that the correlation is less than 1. That is, PARITY cannot be computed exactly by such circuits. Then, we will prove that said correlation is at most $1/w$. But before we start with the proof, let's do some warmup that will be useful for it.

1.1 Warmup

Consider the circuits of type $\wedge$.

Claim 1. *Such circuits cannot compute* PARITY *regardless of size.*

Proof. Take one such circuit C, and say it is fed input x as a series of bits. Consider the input $x = 1000\ldots0$ (a one followed only by zeros). Then $\mathrm{PARITY}(x) = 1$, $C(x) = 0$ □

The second round of warmup considers circuits of DNF form, i.e. $(x_1 \wedge x_2) \vee (x_3 \wedge \neg x_4 \wedge \neg x_1) \vee \ldots$.

Claim 2. *Any DNF for* PARITY *on n bits requires size $w \geq 2^{n-1}$.*

Recall (for comparison) the fact that any function can be computed by circuits of size $O(n \cdot 2^n)$

Proof. Suppose some $\wedge$ gate has fan-in less than n. Therefore, it doesn't depend on some x_i. Fix an input that makes $\wedge = 1$. This will make the whole circuit output 1. If we flip one of the x_i on which $\wedge$ does not depend, the value of the circuit will stay the same, but the parity will flip.

This means that each $\wedge$ gate has a fan-in of exactly n, which in turn implies there is only one input that will make the gate output 1 (namely, the input on which every one of the in-wires carries 1.) On the other hand, there exist 2^{n-1} inputs that give a parity of 1. Therefore, we need 2^{n-1} $\wedge$ gates. □

This concludes the warmup. We have seen a function (parity) that requires size 2^{n-1} $\wedge\vee$ circuits. Challenge: Prove bounds of the form $2^{0.001n}$ for $\wedge\vee\wedge$ circuits. Just one more layer!

It seems hard to extend previous techniques to higher depth. We now use different techniques to show a bound of the form 2^{n^ϵ}. Moreover our techniques will also give correlation bounds. The argument has two stages. Stage 1 shows that small constant-depth circuits can be "well-approximated" by low-degree polynomials, Stage 2 shows that parity cannot. The combination of these two facts establishes the lower bound.

1.2 Stage 1: Parity Cannot be computed exactly by small circuits

1.2.1 Low Degree Polynomials

A polynomial P of degree d is a function from $\{0,1\}^n \to \mathbb{R}$:

$$\sum_{M\subseteq[n],|M|\leq d} C_M \prod_{i\in M} x_i$$

with the sum and product over the reals.

For instance, the polynomial $P = x_1x_2 + x_3 + x_3x_1$ has $n = 3, \text{degree}(P) = 2$.

Theorem 2 (Small circuits are well approximated by low-degree polynomials)**.** $\forall\epsilon$ *For any circuit* $C : \{0,1\}^n \to \{0,1\}$ *of size* $w \geq n$ *and depth* d*, there exists a polynomial* $P : \{0,1\}^n \to \mathbb{R}$ *such that*

$$\Pr_{x\in\{0,1\}^n}[C(x) = P(x)] \geq 1 - \epsilon,$$

and $\text{degree}(P) = \log^{O(d)}(w/\epsilon)$

You can think of $w = \text{poly}(n), \epsilon = 0.001, d = 10, \text{degree}(P) = \text{polylog}(n)$.

The plan to prove the above theorem is to replace each gate by a polynomial. The final polynomial will be the composition of the ones coming before it. The degree at each gate will be $O(\log(w/\epsilon))$, and composing will raise this to a power. The final degree bound will be (the degree at each gate)^d. This construction will be probabilistic.

First, we write the circuit with the basis $\vee, \neg$ (using DeMorgan's law $\bigwedge_i x_i = \neg\bigvee_i \neg x_i$). We show how to build a polynomial from every kind of gate. We start with the $\neg$ gate.

Lemma 3. *There is a polynomial* P *of degree 1 such that* $P(b) = \neg b$ *for all* $b \in \{0,1\}$*.*

Proof. $P(b) := 1 - b$. □

The next more challenging lemma deals with $\vee$ gates. Note that the lemma crucially holds for any distribution. This is because we need to apply the lemma to internal gates of the circuit that can be fed inputs with distributions we do not have a grasp of. In the next lemma we use n for the input length of the $\vee$ gate, which should not be confused with the input length of the parity function we will ultimately be proving a lower bound for. We use n here for convenience; later we will bound from above this parameter by w, the size of the circuit.

Lemma 4. *For all distributions D on $\{0,1\}^n$, for all ϵ, there exists a polynomial $P : \{0,1\}^n \to \mathbb{R}$ such that*

$$\Pr_{x\sim D}[\vee(x) = P(x)] \geq 1 - \epsilon,$$

and $\text{degree}(P) = O(\log n \cdot \log(1/\epsilon))$.

Proof. Let $S_0 := \{1, \ldots, n\}$ be the input variables to $\vee$. Build the set $S_i = S_{i-1}$ by keeping each element in the set independently with probability $1/2$ for $i = 1, \ldots, \log n + 1$. (i.e. before adding each element to S_i, toss a coin to determine if that element will go into the set or not.)

Let

$$P_i(x) := \sum_{i \in S_i} x_i$$

for $i = 0, \ldots, \log n + 1$.

The proof of the following claim is obvious.

Claim 3. *$x = 0^n$ implies $P_i(x) = 0$ for all i.*

Claim 4. *For all $x \in \{0,1\}^n$, $x \neq 0^n$, with probability at least $1/6$ over $P_0, \ldots, P_{\log n+1}$ $\exists i$ such that $P_i(x) = 1$.*

Proof. If x has weight 1 (i.e. $\sum x_i = 1$), then $P_0(x) = 1$ and we are done.

If $\sum x_i > 1$, we can prove the claim by showing that $\Pr[\forall i, P_i(x) \neq 1]$ has an upper bound of $5/6$. Let F_i be the event that among $(P_0(x), \ldots, P_{\log n+1}(x))$, $P_i(x)$ gives the first value ≤ 1. Then

$$\begin{aligned}
\Pr[\forall i, P_i(x) \neq 1] &\leq \sum_{j=0}^{\log n+1} \Pr[\forall i, P_i(x) \neq 1 | F_j] \cdot \Pr[F_j] + \Pr[\forall i, P_i(x) > 1] \\
&\leq \sum_{j=0}^{\log n+1} \Pr[\forall i, P_i(x) \neq 1 | F_j] \cdot \Pr[F_j] + \frac{1}{2}.
\end{aligned}$$

Consider $\Pr[\forall i, P_i(x) | F_j]$. We have that $j = 0$ gives $\Pr[\forall i, P_i(x) \neq 1 | F_0] = 0$. Also, for $j > 0$, we have

$$\Pr[P_j(x) = 1 | F_j] = \frac{P_{j-1}(x)}{P_{j-1}(x) + 1} \geq \frac{2}{3}$$

since $P_{j-1} > 1$. So,

$$\Pr[\forall i, P_i(x) \neq 1] \leq \frac{1}{3} \sum_{j=0}^{\log n+1} \Pr[F_j] + \frac{1}{2} \leq \frac{1}{3} + \frac{1}{2} = \frac{5}{6}.$$

□

Define

$$P'(x) := 1 - \prod_{i=0}^{\log n+1} (1 - P_i(x))$$

If $x = 0$, then $P'(x) = 0$. If $x \neq 0$ then $\Pr[P'(x) = 1] \geq 1/6$ by Claim 4. Also degree$(P') = O(\log n + 1)$

We can reduce the error in P' by using the same trick. Let

$$P_\epsilon(x) := 1 - \prod_{i=0}^{O(\log 1/\epsilon)} (1 - P'_i(x))$$

where $P'_i(x)$ are independent copies of P' above. If $x = 0$ then $P_\epsilon(x) = 0$. If $x \neq 0$,

$$\begin{aligned}
\Pr[P_\epsilon(x) = 1] &\geq \Pr[\exists i : P'_i(x) = 1] \\
&= 1 - \Pr[\forall i, P'_i(x) \neq 1] \\
&= 1 - \Pr[P'(x) \neq 1]^{O(\log 1/\epsilon)} \\
&\geq 1 - (5/6)^{O(\log 1/\epsilon)} = 1 - \epsilon.
\end{aligned}$$

Also, degree$(P_\epsilon) = O(\log n \cdot \log 1/\epsilon)$.

We have constructed P_ϵ such that

$$\forall x, \Pr_{P_\epsilon}[P_\epsilon(x) \neq \vee(x)] \leq \epsilon.$$

In particular,

$$\Pr_{x \sim D, P_\epsilon}[P_\epsilon(x) \neq \vee(x)] \leq \epsilon.$$

This implies that $\exists \hat{P}$ such that

$$\Pr_{x \sim D}\left[\hat{P}(x) \neq \vee(x)\right] \leq \epsilon.$$

Noting that degree$\left(\hat{P}\right) = O(\log n \cdot \log 1/\epsilon)$ concludes the proof. □

We are now ready to prove Theorem 2. We will use the above lemmas on every gate.

Proof of Theorem 2. Given ϵ, set $\epsilon_{lemma} = \epsilon/w$. Invoke lemmas for every gate g with respect to the distribution D_g induced by a uniform input at that gate. This gives a polynomial P_g of degree $O(\log w \cdot \log(w/\epsilon))$ that approximates g with respect to D_g, and has error ϵ/w.

Now we compose the polynomials for every gate. Let P be the composition of P_g that corresponds to the circuit.

$$\begin{aligned}
\Pr_{x \in \{0,1\}^n}[P(x) \neq C(x)] &\leq \sum_g \Pr_{y \sim D_g \text{on input } x}[P_g(y) = g(y)] \\
&\leq w \cdot \epsilon/w = \epsilon.
\end{aligned}$$

The degree of P is $\log^{O(d)}(w/\epsilon)$. □

1.3 Computing parity

Claim 5. *Suppose $P : \{0,1\}^n \to \mathbb{R}$ is a polynomial of degree d such that*

$$\Pr_{x\in\{0,1\}^n}[P(x) = \text{PARITY}(x)] = \delta$$

Then there exists another polynomial $P' : \{-1,1\}^n \to \mathbb{R}$ of degree d such that

$$\Pr_{x\in\{-1,1\}^n}\left[P'(x) = \prod_{i=1}^{n} x_i\right] = \delta \tag{1}$$

Proof. Observe the map $b \in \{-1,1\} \to \frac{b+1}{2} \in \{0,1\}$. This is linear and invertible, so $P'(x_1,\ldots,x_n) = 2P\left(\frac{x_1+1}{2},\ldots,\frac{x_n+1}{2}\right) - 1$. Since the map is linear, then both the degrees of P and P' are d. □

We will show that achieving Equation 1 requires large degree by showing how we can build another polynomial $\bar{P}$ that "weakly computes" parity, and that the latter requires maximum degree n. Recall the sign function that outputs $+1$ if the input is positive, and -1 otherwise (it is irrelevant how we define the output when the input is 0).

Definition 5. *A polynomial $\bar{P}$* weakly computes *parity when*

1. *$\bar{P}$ is not the 0 polynomial, and*
2. *If $\bar{P}(x) \neq 0$, then $\text{sign}(\bar{P}(x)) = \prod_{i=1}^{n} x_i$, with $x \in \{-1,1\}^n$.*

Claim 6. *To weakly compute parity on n bits, degree n is required.*

Proof. Let $p : \{-1,1\}^n \to \mathbb{R}$ be a polynomial of degree $n-1$:

$$p(x) := \sum_{M\subseteq[n],|M|\leq n-1} c_M \prod_{i\in M} x_i.$$

On the one hand, $E_{x\in\{-1,1\}^n}[p(x)\cdot\prod_{i=1}^{n} x_i] > 0$, since the polynomial is not identically 0 and when it is not zero its sign is $\prod_{i=1}^{n} x_i$.

On the other hand, by linearity of expectation,

$$E_{x\in\{-1,1\}^n}\left[p(x)\cdot\prod_{i=1}^{n} x_i\right] = \sum_{M\subseteq[n],|M|\leq n-1} c_M E_{x\in\{-1,1\}^n}\left[\prod_{i\in M} x_i \cdot \prod_{i=1}^{n} x_i\right]$$

$$= \sum_{M\subseteq[n],|M|\leq n-1} c_M E_{x\in\{-1,1\}^n}\left[\prod_{i\notin M} x_i\right] = 0,$$

where the last equality holds because $|M| < n$, and so the product $\prod_{i\notin M} x_i$ contains at least one variable, and thus over a uniform choice of x this product will be $+1$ with probability $1/2$ and -1 also with probability $1/2$, giving 0 expectation. □