Smolensky's Lower Bound

Yuval Filmus

September 2010

1 Introduction

In this writeup we're going to discuss Smolensky's algebraic method according to his 1993 paper *On Representations by Low-Degree Polynomials*. Lecture notes usually only cover the case of parity, but Smolensky's method is more general. In particular, it directly gives lower bounds for majority.

Smolensky's circuit lower bound approach consists of two parts (the first of which is due to Razborov):

1. Show that a small, bounded depth circuit can be approximated by a low degree polynomial.
2. Show that a specific function (for example, parity or majority) cannot be approximated by a low degree polynomial.

The first step is described in many lecture notes. For the second step, only an argument for parity is given usually. We reproduce this argument below for completeness in section 5. However, Smolensky's 1993 paper takes a slightly different approach, which generalizes to majority as well as to other functions.

2 Framework

We will be interested in Boolean functions over the Boolean cube $\{0,1\}^n$ and their polynomial approximations. Razborov's method for approximating an $\mathsf{AC}_0[p]$ circuit produces a polynomial over $\mathbb{Z}_p$ (in fact, this generalizes to prime powers); Smolensky's lower bounds work for any field $\mathbb{F}$. *For the rest of the writeup, we will assume all polynomials are over some fixed field $\mathbb{F}$, unless explicitly noted.* Razborov's polynomial takes the values $0, 1 \in \mathbb{Z}_p$; Smolensky's method works for the following weaker concept.

Definition 2.1. *A polynomial $P(x_1, \ldots, x_n)$ over $\mathbb{F}$* represents *a Boolean function P_B with domain $\{0,1\}^n$ defined by*

$$P_B(x_1, \ldots, x_n) = \begin{cases} 1 & \text{if } P(x_1, \ldots, x_n) = 0, \\ 0 & \text{if } P(x_1, \ldots, x_n) \neq 0. \end{cases}$$

Smolensky's method uses the concept of the (affine) Hilbert function, which is a measure of complexity of a subset of the Boolean cube; we will apply it to zero-sets of Boolean functions.

Definition 2.2. *Let $S \subseteq \{0,1\}^n$. For each polynomial P over $\mathbb{F}$, its image $P|_S$ on S can be considered as a vector of length $|S|$. The* (affine) Hilbert function *$h_k(S)$ is the dimension of the vector space spanned by $P|_S$ for all polynomials P of degree at most k.*

Clearly $h_k(S) \leq |S|$ and $h_k(S) \leq \sum_{t \leq k} \binom{n}{t}$ (since $x^2 = x$ for Boolean x). Below we will see several example where the latter inequality is tight.

It turns out that the Hilbert function of low degree polynomials is small, in the sense that $h_k(S) \leq |S|/2$. This prompts the following definition.

Definition 2.3. *Let $S \subseteq \{0,1\}^n$. The* Hilbert excess function *$\alpha_k(S)$ is*

$$\alpha_k(S) = 2h_k(S) - |S|.$$

We will be comparing the value of the Hilbert function on the zero-set of a difficult function to that of the non-zero-set of a low-degree polynomial (this is the zero-set of the Boolean function the low-degree polynomial repersents). We will show that the Hilbert functions are far apart, and we'd like to conclude that the sets themselves are quite different. Is is natural therefore to consider what happens to the Hilbert function when a point is added or removed from S. It is easy to see that adding a point to S can at most increase the Hilbert function by 1. In particular, we obtain the following:

Lemma 2.4. *For any k and two subsets S, T of the Boolean cube we have*

$$|\alpha_k(S) - \alpha_k(T)| \leq |S \triangle T|.$$

Proof. Clearly, it is enough to prove the lemma for the case $T = S \cup \{x\}$ for some $x \notin S$. It is easy to see that

$$h_k(S) \leq h_k(T) \leq h_k(S) + 1,$$

from which we easily deduce

$$\alpha_k(S) - 1 \leq \alpha_k(T) \leq \alpha_k(S) + 1. \qquad \square$$

We can now outline Smolensky's method:

1. Show that the non-zero-set of a low-degree polynomial has small excess, for an appropriate value of k.
2. Show that the zero-set of some hard function (e.g. parity or majority) has high excess, by explicitly calculating its Hilbert function.
3. Conclude that the corresponding zero-sets are far apart, i.e. the low degree polynomial correlates weakly with the hard function.

In the next section, we will prove the first step, and conclude Smolensky's end result. The following section will prove the second step for parity and majority. We will also mention how the usual proof for parity appears in this framework.

3 Low-degree polynomials

In this section we will show that the non-zero-set of a low-degree polynomial has small excess α_k for k slightly smaller than $n/2$.

Theorem 3.1. *Suppose* $\deg P = d$*, and let* $k < (n-d)/2$*. Put* $S = \{x : P(x) \neq 0\}$*. Then* $\alpha_k(S) \leq 0$*.*

Proof. The proof is by contradiction: we assume that $h_k(S) > |S|/2$, and reach a contradiction.

Recall that $h_k(S)$ is the dimension of the linear subspace $\{Q|_S : \deg Q \leq k\}$. This subspace is spanned by all monomials of degree at most k. We can construct a matrix whose rows correspond to monomials of degree at most k and whose columns correspond to points of S, such that $h_k(S)$ is the rank of this matrix. In particular, there is a set $U \subseteq S$ of $h_k(S)$ linearly independent columns. The set U satisfies

$$h_k(U) = h_k(S) = |U|.$$

By assumption $2h_k(S) > |S|$, and so $h_k(S) > |S| - |U| = |S \setminus U|$. Therefore we can find a non-zero k-degree polynomial A which is equal to zero on all of $S \setminus U$. We will transform A into a delta-function for the entire Boolean cube. The following notation will be useful: $\delta(T, x)$ is the collection of functions which are non-zero on x and zero on the rest of T.

The first step is transforming A to a delta-function on S. Since $A(S \setminus U) = 0$, all we need to do is multiply A by an appropriate delta-function on U. Notice that since $h_k(U) = |U|$, we can realize any function on U by a k-degree polynomial. In particular, picking any $x \in U$ such that $A(x) \neq 0$, there is a k-degree polynomial $B \in \delta(U, x)$. Multiplying, we find that $AB \in \delta(S, x)$.

In order to transform AB to a delta-function on the entire Boolean cube, all we need to do is multiply AB by the original polynomial P: indeed, $ABP \in \delta(\{0,1\}^n, x)$. Notice that $\deg ABP \leq 2k + d < n$. By translation, we can assume that x is the all-1 point. We get a contradiction since every monomial of degree less than n is zero on x. □

Smolensky's method is an easy corollary of the preceding theorem.

Theorem 3.2. *Let* f *be any Boolean function,* P *any* d*-degree polynomial, and* P_B *the Boolean function represented by* P*. Then*

$$\Pr[f \neq P_B] \geq 2^{-n} \alpha_{(n-d-1)/2}(f^{-1}(0)).$$

Proof. Let $k = (n - d - 1)/2$. Theorem 3.1 implies that

$$\alpha_k(f^{-1}(0)) - \alpha_k(P_B^{-1}(0)) \geq \alpha_k(f^{-1}(0)).$$

The theorem now follows from lemma 2.4. □

4 Hilbert function of parity and majority

In this section we will show that for $k < n/2$, the Hilbert function of the zero-sets of both the parity function and majority is maximal. In particular, we get the same correlation bounds for both.

Corollary 4.1. *Let f be a Boolean function such that for some $k < n/2$,*

$$h_k(f^{-1}(0)) = \sum_{t \le k} \binom{n}{t}.$$

Then for any polynomial P of degree smaller than $n - 2k$,

$$\Pr[f \neq P_B] \geq 2 \Pr[\mathrm{Bin}(n, \tfrac{1}{2}) \leq k] - \Pr[f = 0],$$

where P_B is the Boolean function represented by P.

Proof. Easy corollary of theorem 3.2. □

The conditions of the corollary are true for both parity and majority, as we are going to show presently. We will use the following test.

Lemma 4.2. *Let S be some subset of the Boolean cube. Then for any k, we have*

$$h_k(S) = \sum_{t \le k} \binom{n}{t}$$

if and only if the only polynomial P of degree at most k which is zero on all of S is the zero polynomial.

Proof. Recall that $h_k(S)$ is the dimension of the linear subspace $\{P|_S\}$ for all polynomials of degree at most k. This subspace is spanned by $\{M|_S\}$ for all *monomials* of degree at most k. The monomials are linearly independent iff there is no non-trivial linear combination of them which gives the zero vector. Any such linear combination corresponds to a non-zero polynomial of degree at most k which is zero on all of S. □

4.1 Parity

In order to analyze parity, we will use a transformation $\{0, 1\} \to \{1, -1\}$, which is reversible if $\mathrm{char}\, \mathbb{F} \neq 2$; naturally, if $\mathrm{char}\, \mathbb{F} = 2$ then parity is easy. We need an easy lemma.

Lemma 4.3. *Let $B = \{a, b\}^n$ be some Boolean cube ($a \neq b$). The set of square-free monomials over the cube $\{M|_B\}$ forms a basis for the linear space of polynomials over the cube $\{P|_B\}$.*

Proof. Clearly the set $\{M|_B\}$ spans the linear space $\{P|_B\}$. The degree of the latter is 2^n, which is also the number of monomials. Hence the monomials must be linearly independent. □

Theorem 4.4. *Denote by S the set of vectors with even parity. If* $\operatorname{char}\mathbb{F} \neq 2$ *then for all $k < n/2$,*

$$h_k(S) = \sum_{t \leq k} \binom{n}{t}.$$

Proof. We use the technique of lemma 4.2. Assume we are given a polynomial P of degree at most k which is zero on all of S. Since $\operatorname{char}\mathbb{F} \neq 2$, the element 2 is invertible. Substitute $x_i = (1 - y_i)/2$ to get a new polynomial $P_y(y_1, \ldots, y_n)$ of the same degree. Notice that $x_i = 0$ corresponds to $y_i = 1$, and that $x_i = 1$ corresponds to $y_i = -1$. Therefore, the original condition that P is zero on S translates to the condition that P_y is zero on

$$S_y = \{y : y_1 \cdots y_n = 1\}.$$

Now form the polynomial

$$Q_y = P_y(1 + y_1 \cdots y_n).$$

This polynomial is clearly zero on the entire Boolean cube $\{1, -1\}^n$. The polynomial Q_y breaks as the sum

$$Q_y = P_y + y_1 \cdots y_n P_y.$$

Using the identity $y_i^2 = 1$, we see that the first summand is a combination of monomials of degree less than $n/2$, and the second summand is a combination of monomials of degree more than $n/2$. Lemma 4.3 implies that $P_y = 0$ and so $P = 0$. □

The proof can be generalized for a version of the mod-q function.

Theorem 4.5. *Let q be a prime different from* $\operatorname{char}\mathbb{F}$*. Denote by S the set of vectors whose Hamming weight is* not *a multiple of q. For all $k < n/2$,*

$$h_k(S) = \sum_{t \leq k} \binom{n}{t}.$$

Proof. The proof is very similar to the proof of the previous theorem. We can assume, without loss of generality, that $\mathbb{F}$ is algebraically closed (enlarging $\mathbb{F}$ can only lower the dimension of the linear space of polynomials). The polynomial $x^q - 1$ is coprime to its derivative, and so has no repeated roots. Therefore some $\lambda \neq 1$ is a primitive q-th root of unity in $\mathbb{F}$.

In view of using lemma 4.2, consider some polynomial P of degree at most k which is zero on all of S. The substitution $x_i = (1 - y_i)/(1 - \lambda)$ translates P into a polynomial P_y over $\{1, \lambda\}^n$. The condition that P is zero on S translates to the condition that P_y is zero on

$$S_y = \{y : y_1 \cdots y_n \neq 1\}.$$

Therefore the polynomial

$$Q_y = P_y(1 - y_1 \cdots y_n) = P_y - y_1 \cdots y_n P_y$$

is zero over $\{1, \lambda\}^n$. Over this cube, we have $y_i^2 = (\lambda+1)y_i - \lambda$. Thus, for every monomial M, we have (over $\{1, \lambda\}^n$)

$$y_1 \cdots y_n M = c\overline{M} + \cdots,$$

where $\overline{M}$ is the monomial formed by complementing the set of indices, c is some non-zero constant, and the dots represent monomials of higher degree. In this way Q_y corresponds to some polynomial R_y, all of whose monomials are square-free, which is zero on $\{1, \lambda\}^n$. Lemma 4.3 implies that R_y is the zero polynomial. We wish to conclude that P_y is the zero polynomial. Otherwise, pick a monomial $M \in P_y$ with maximal degree. Notice that this condition forces $\overline{M} \in R_y$ (here we use $k < n/2$), which contradicts our assumption that $R_y = 0$. Thus $P_y = 0$, and so $P = 0$. □

4.2 Majority

The analysis for majority has a completely different flavor. The technique actually gives a lower bound for any Boolean function, but this bound is tight only for monotone functions. We will use the following correspondence between points and monomials.

Definition 4.6. *Let $z \in \{0,1\}^n$ be some point in the Boolean cube. Its corresponding monomial $\mathcal{M}(z)$ is*

$$\mathcal{M}(z) = \prod_{i=1}^{n} x_i^{z_i} = \prod_{z_i=1} x_i.$$

The reverse mapping $\mathcal{P}(M)$ takes a monomial $\prod_{i \in S} x_i$ to the point z defined by $z_i = 1$ for $i \in S$ and $z_i = 0$ for $i \notin S$.

Theorem 4.7. *Denote by S the vectors with Hamming weight smaller than $n/2$. For $k < n/2$,*

$$h_k(S) = \sum_{t \leq k} \binom{n}{t}.$$

Proof. We use lemma 4.2. Let P some polynomial of degree at most k which is zero on all of S. Suppose that P is not the zero polynomial, and let M be some monomial of minimal degree in P. Set $z = \mathcal{P}(M)$. Notice that $M(z) = 1$ whereas $N(z) = 0$ for all other monomials $N \in P$. This is contradictory since it implies $P(z) \neq 0$. We conclude that necessarily P is the zero polynomial. □

The proof technique can be generalized to arbitrary subsets of the Boolean cube. It is tight for downward-closed sets.

Definition 4.8. *A subset $S \subseteq \{0,1\}^n$ is* downward-closed *if for any two points $x, y \in \{0,1\}^n$ such that $x_i \leq y_i$, $y \in S$ implies $x \in S$.*

Theorem 4.9. *Denote by H_k the vectors with hamming weight at most k. Then for any subset S of the Boolean cube,*

$$h_k(S) \geq |S \cap H_k|.$$

Moreover, equality holds if S is downward-closed.

Proof. The first part follows by showing that the set

$$B = \{\mathcal{M}(z) : z \in S \cap H_k\}$$

of monomials of degree at most k is linearly independent over S. If it is not, then there is some non-zero polynomial P with monomials from B which is zero on all of S. If $M \in P$ is a monomial of minimal degree, then just as in the proof of the previous theorem, $P(\mathcal{P}(M)) \neq 0$. This contradiction shows that the set B is linearly independent.

Now suppose that S is downward-closed. If $z \in H_k \setminus S$ then we claim that $\mathcal{M}(z)$ is zero on S. Indeed, if $\mathcal{M}(z)(y) \neq 0$ then $y \geq z$, contradicting the fact that $z \notin S$. Thus the monomials in B span the space of all polynomials of degree at most k. □

5 Another proof for parity

The usual circuit lower bound for parity is proved by a variant of the method explained in the previous section. In particular, we will argue directly with the Hilbert function.

Lemma 5.1. *Let $S \subseteq T \subseteq \{0,1\}^n$ be two subsets of the Boolean cube. Then for every d,*

$$h_d(S) \leq h_d(T) \leq h_d(S) + |T \setminus S|.$$

Proof. The lemma follows from the fact that adding a point to S can increase the dimension of the linear space of polynomials by at most 1. □

Theorem 5.2. *Let P by any polynomial of degree d. Denote by S the set of points on which P agrees with parity. Then*

$$h_n(S) \leq \sum_{t \leq \frac{n+d}{2}} \binom{n}{t}.$$

Proof. We show that any monomial M can be represented by some monomial of degree at most $(n+d)/2$ on S. If $\deg M \leq (n+d)/2$ then M represents itself. Otherwise, let $\tilde{M}$ denote the monomial such that $x_i \in M$ iff $x_i \notin \tilde{M}$, i.e. the corresponding points $\mathcal{P}(M)$ and $\mathcal{P}(\tilde{M})$ are complementary. Since on S we have $M = \tilde{M}P$, the monomial M is represented by the monomial $\tilde{M}P$ of degree at most $(n-d)/2 + d = (n+d)/2$. □

Corollary 5.3. *Let P by any polynomial of degree d. Denote by π the parity function. We have*

$$\Pr[P = \pi] \leq \Pr\left[\mathrm{Bin}(n, \tfrac{1}{2}) \leq \frac{n+d}{2}\right].$$

Proof. Denote by S the set of points on which $P = \pi$. Since $h_n(\{0,1\}^n) = 2^n$, lemma 5.1 shows that $h_n(S) \geq |\{z : P(z) = \pi(z)\}|$. The corollary now follows from the preceding theorem. □

6 Circuit lower bounds

In order to obtain circuit lower bounds, we use the following well-known approximation obtained by Smolensky.

Theorem 6.1. *Let C be an $\mathsf{AC}_0[p]$ circuit of size S and depth h. For every l there is a polynomial P over $\mathbb{Z}_p$ of degree $((p-1)l)^h$ such that*

$$\Pr[C \neq P] \leq \frac{S}{2^l}.$$

Proof. See any lecture notes on the subject. □

Using this, we can upper bound the correlation between an arbitrary AC_0 circuit and either parity or majority.

Theorem 6.2. *Let C be an $\mathsf{AC}_0[p]$ circuit of size S and depth h. Let f be either parity or majority. Then*

$$Pr[C = f] \leq \frac{1}{2} + \frac{O(\log(nS))^h}{\sqrt{n}} + \frac{1}{n}.$$

Proof. Choose $l = \log(nS)$ in the preceding theorem. We get a polynomial P over $\mathbb{Z}_p$ of degree $d = ((p-1)l)^h$ which differs from C with probability at most $1/n$. Combining corollary 4.1 with either theorem 4.4 or theorem 4.7, we get

$$\Pr[C \neq f] \leq 2\Pr\left[\mathrm{Bin}(n, \tfrac{1}{2}) \leq \frac{n}{2} - t\right] - \frac{1}{2} - \frac{1}{n},$$

where the parameter t is given by

$$t = \frac{n}{2} - \frac{n-d-1}{2} = \frac{d+1}{2} = O(\log(nS))^h.$$

In order to estimate the binomial tail, we use

$$\Pr\left[\mathrm{Bin}(n, \tfrac{1}{2}) \leq \frac{n}{2} - t\right] = \frac{1}{2} - 2^{-n}\sum_{k=0}^{t-1}\binom{n}{k} = \frac{1}{2} - O\left(\frac{t}{\sqrt{n}}\right),$$

where we estimated each binomial coefficient with the central one. □