Week 6

Polynomials, Part 2: Dual polynomials

Last week, we saw that a quantum algorithm computing a Boolean function gives rise to an approximating polynomial for that function, and that the degree of this polynomial is (up to constant factors) at most the number of queries used by the algorithm. We also saw that the degree of a polynomial approximating a Boolean function can be lower bounded using *symmetrization*: this is a technique for converting a polynomial in many variables into a polynomial in one (or few) variables with the same degree, which computes something related to the original problem. We can then use some tools from approximation theory to put constraints on what bounded single-variate polynomials of degree d can do.

We will now look more generally at the task of lower bounding $\widetilde{\deg}(f)$, the approximate degree of a Boolean function f.

6.1 Duality for polynomials

We have defined $\widetilde{\deg}_\epsilon(f)$ as the minimum degree of a polynomial p which approximates f to error ϵ. Consider the flip side of this question: the task of minimizing the error ϵ to which a polynomial p of fixed degree d can approximate f. This flipped question is closely related to the original one; if we knew, for each d, the smallest error ϵ_d to which polynomials of degree at most d may approximate f, we could simply define $\widetilde{\deg}_\epsilon(f) = \min\{d \in \{0, 1, \ldots, n\} : \epsilon_d \leq \epsilon\}$.

It turns out that this flipped question can be viewed as a linear program. This is perhaps most cleanly expressed when f (and the polynomial approximating it) have $\{+1, -1\}$ outputs, in which case we want to approximate f to error 2ϵ. Indeed, it is the linear program

$$\begin{array}{lll} \min & \epsilon & \\ \text{s.t.} & f(x)\cdot\sum_{m\in M_d} m(x)c_m \geq 1-2\epsilon & \forall x \in \mathrm{Dom}(f) \\ & \sum_{m\in M_d} m(x)c_m \leq 1 & \forall x \in \{+1,-1\}^n \\ & \sum_{m\in M_d} m(x)c_m \geq -1 & \forall x \in \{+1,-1\}^n. \end{array}$$

In the above program, the variables are ϵ and c_m for all $m \in M_d$, where M_d is the set of all monomials which have degree at most d. The variable c_m represent the coefficients of a polynomial of degree d, so that the polynomial itself is $p(x) = \sum_{m\in M_d} c_m m(x)$. For the purpose of the linear program, we treat the terms $m(x)$ as constants and the terms c_m as variables, where recall that we are working in the $\{+1, -1\}$-basis so that each $m(x)$ has value either $+1$ or -1 (depending on m and on x). The first constraint then says that we must have $f(x)p(x) \geq 1 - 2\epsilon$ for all $x \in \mathrm{Dom}(f)$, which means that $p(x)$ must have the same sign as $f(x)$ and must have absolute value at least $1-2\epsilon$. The second and third constraints together say that $|p(x)| \leq 1$ for all $x \in \{+1, -1\}^n$ (this can be

directly generalized to non-Boolean alphabets, as long as we are careful to define the monomials $m \in M_d$ appropriately).

Now that we have written down a linear program for (the flipped version of) approximate degree, let's take the dual and see what we get. In the dual, we have one variable for each constraint; this gives us a variable μ_x for each $x \in \mathrm{Dom}(f)$, as well as variables ν_x^+ and ν_x^- for each $x \in \{+1,-1\}^n$. The dual becomes

$$\begin{array}{lrll} \max & (1/2)\sum_x \mu_x - \sum_x \nu_x^+ - \sum_x \nu_x^- & & \\ \text{s.t.} & \sum_x m(x)(f(x)\mu_x/2 - \nu_x^+ + \nu_x^-) & = 0 & \forall m \in M_d \\ & \sum_x \mu_x & = 1 & \\ & \mu_x & \geq 0 & \forall x \in \mathrm{Dom}(f) \\ & \nu_x^+, \nu_x^- & \geq 0 & \forall x \in \{+1,-1\}^n. \end{array}$$

We can simplify this. First, we can substitute $\sum_x \mu_x = 1$ in the objective function, since this is a constraint. Next, observe that if for any $x \in \{+1,-1\}^n$ it holds that $\nu_x^+ > 0$ and $\nu_x^- > 0$, we can decrease both by the same amount, until one of them becomes 0; this only increases the objective value, and does not break any constraints. Hence we can assume that for each x, either $\nu_x^+ = 0$ or $\nu_x^- = 0$. We can set $\nu_x = \nu_x^+$ if $\nu_x^+ \geq 0$ and $\nu_x = -\nu_x^-$ if $\nu_x^- \geq 0$. This way, we will get

$$\begin{array}{lrll} \max & 1/2 - \|\nu\|_1 & & \\ \text{s.t.} & \sum_x m(x)(f(x)\mu_x/2 - \nu_x) & = 0 & \forall m \in M_d \\ & \sum_x \mu_x & = 1 & \\ & \mu_x & \geq 0 & \forall x \in \mathrm{Dom}(f). \end{array}$$

Note that we can always achieve objective value at least 0 by taking $\nu_x = f(x)\mu_x/2$. Now, note that for any x, if $f(x)\mu_x/2$ and ν_x are both positive in the optimal solution, we can decrease them both by the same amount (say $\delta/2$), preserving the first constraint; we can then scale μ and ν up by the same factor of $1/(1-\delta)$, which restores the second constraint. This new solution has objective value $1/2 - (\|\nu\|_1 - \delta/2)/(1-\delta)$, which is at least $1/2 - \|\nu\|_1$ when the latter is at least 0. Hence we can assume that either $f(x)\mu_x/2$ or ν_x is not positive in the optimal solution. Similarly, if they are both negative, we can add $\delta/2$ weight and the same calculation goes through; hence they are not both negative. Finally, if one is negative and the other is positive, we can add $\delta/2$ to one and subtract $\delta/2$ from the other, and once again scale ν and μ by $1/(1-\delta)$. This lets us conclude that either $\mu_x = 0$ or $\nu_x = 0$ for all $x \in \{+1,-1\}^n$.

Let $v_x = f(x)\mu_x/2 - \nu_x$. Then we now know that $\|v\|_1 = \|\mu/2\|_1 + \|\nu\|_1 = 1/2 + \|\nu\|_1$. We also have $\sum_x v_x f(x) = \sum_x f(x)^2\mu_x/2 - \sum_x f(x)\nu_x = 1/2 - \sum_x f(x)\nu_x \geq 1/2 - \|\nu\|_1$, and for each $m \in M_d$, $\sum_x m(x)v_x = 0$. It is also not hard to get back μ and ν if someone gives us v satisfying these conditions. Letting $\gamma = \|\nu\|_1$, the program can be written

$$\begin{array}{lrll} \max & 1/2 - \gamma & & \\ \text{s.t.} & \sum_x m(x)v_x & = 0 & \forall m \in M_d \\ & \sum_x f(x)v_x & \geq 1/2 - \gamma & \\ & \|v\|_1 & = 1/2 + \gamma. & \end{array}$$

Finally, let $\delta = 1/2 - \gamma$, and let $u = v/(1/2+\gamma)$. Then $\|u\|_1 = 1$, and $\sum_x f(x)u_x \geq \delta/(1-\delta)$. The program is then

$$\begin{array}{lrll} \max & \delta & & \\ \text{s.t.} & \sum_x m(x)u_x & = 0 & \forall m \in M_d \\ & \sum_x f(x)u_x & \geq \delta/(1-\delta) & \\ & \|u\|_1 & = 1. & \end{array}$$

How shall we interpret this? Well, since $\|u\|_1 = 1$, we can split it up into a probability distribution μ with $\mu_x = |u_x|$ and a function $f' \colon \{+1,-1\}^n \to \{+1,-1\}$ specifying the sign, so that $u_x = f'(x)\mu_x$ (note that we reused the variable μ_x, which has a different meaning now than in the first linear program). In terms of μ and f', the constraints are

$$\mathop{\mathbb{E}}_{x\sim\mu}[m(x)f'(x)] = 0 \quad \forall m \in M_d$$

$$\mathop{\mathbb{E}}_{x\sim\mu}[f(x)f'(x)] \geq \delta/(1-\delta).$$

In other words, we have the following conclusion.

Theorem 6.1. *To show that there is no bounded polynomial of degree d which approximates f to error ϵ, it suffices to show that there is a total function f' and a distribution μ over the hypercube such that*

1. *f' has correlation greater than $\epsilon/(1-\epsilon)$ with f when measured against distribution μ (i.e. $\mathbb{E}_{x\sim\mu}[f(x)f'(x)] > \epsilon/(1-\epsilon)$ in the $\{+1,-1\}$ basis, where we set $f(x) = 0$ if $x \notin \mathrm{Dom}(f)$),*
2. *f' has zero correlation with all monomials m of degree at most d when measured against μ (i.e. $\mathbb{E}_{x\sim\mu}[f'(x)m(x)] = 0$ for all such monomials m in the $\{+1,-1\}$ basis).*

Moreover, this technique is tight: if there is no bounded polynomial of degree d approximating f to error ϵ, then such f' and μ must exist.

6.2 Sign degree and discrepancy

Before we continue with the study of polynomial lower bounds, let's pause to say a few things about smaller query measures. One important measure is the *threshold degree* or *sign degree* of a Boolean function.

Definition 6.2. *The* threshold degree *or* sign degree *of a (possibly partial) Boolean function f, denoted* $\deg_\pm(f)$*, is the minimum degree of a polynomial p which approximates f to some error less than $1/2$; in other words, in the $\{+1,-1\}$ basis, it is the minimum degree of a polynomial p such that $p(x)f(x) > 0$ for all $x \in \mathrm{Dom}(f)$.*

Note that unlike in the definition of $\widetilde{\deg}(f)$, we did not require p to be bounded. It turns out that a polynomial which only approximates f in sign (i.e. to arbitrarily small bias) can always be assumed to be bounded: we can scale down p by the maximum value of $|p(x)|$ over $x \in \{+1,-1\}^n$, ensuring that $|p(x)| \leq 1$ for all x after the scaling, and this does not affect the property that $p(x)f(x) > 0$ (though it may make the bias even smaller, meaning it may make the error even closer to $1/2$).

The sign degree of a function is similar to the complexity class PP, the class corresponding to a model of computation where you are allowed to use randomness and yet are only required to succeed with any non-zero bias (rather than with bounded error, as in BPP). For reasons we'll see later, the sign degree is usually considered to correspond to the class UPP, because a different query measure more closely corresponds to PP.

Although we defined the sign degree of a Boolean function in terms of polynomials, it turns out that one gets an equivalent definition if one tries to make an analogous definition for randomized or quantum algorithms.

Lemma 6.3. *Let f be a (possibly partial) Boolean function. Let $\mathrm{R}_{\pm}(f)$ denote the minimum number of worst-case queries that one needs to make in order to compute f to an error less than $1/2$; that is, $\mathrm{R}_{\pm}(f) \coloneqq \inf\{\mathrm{R}_{\epsilon}(f) : \epsilon < 1/2\}$. Similarly, define $\mathrm{Q}_{\pm}(f) \coloneqq \inf\{\mathrm{Q}_{\epsilon}(f) : \epsilon < 1/2\}$. Then*

$$\mathrm{R}_{\pm}(f) = \deg_{\pm}(f)$$

and

$$\frac{1}{2}\deg_{\pm}(f) \leq \mathrm{Q}_{\pm}(f) \leq \deg_{\pm}(f).$$

Proof. Note that $\mathrm{R}_{\pm}(f) \geq \mathrm{Q}_{\pm}(f) \geq (1/2)\deg_{\pm}(f)$; the first inequality follows from the fact that quantum algorithms can simulate classical ones, and the second from the fact that a quantum algorithm gives rise to an approximating polynomial. To complete the proof, all we need to show is that $\deg_{\pm}(f) \geq \mathrm{R}_{\pm}(f)$. To this end, consider a polynomial p of degree $\deg_{\pm}(f)$ satisfying $p(x)f(x) > 0$ for all $x \in \mathrm{Dom}(f)$. Let $\{c_m\}_m$ be the set of coefficients of the monomials of p. Divide p by $\sum_m |c_m|$, so that the sum of absolute values of the coefficients of p becomes 1; this does not affect the property that $p(x)f(x) > 0$, nor does this change the degree of p.

Define a randomized algorithm R which picks a monomial m of degree d with probability $|c_m|$ (recall that the absolute coefficients $|c_m|$ now sum to 1, making them a probability distribution over the monomials). The algorithm R then queries all the indices x_i of the input that occur in the monomial m it sampled; this is at most d queries. Next, R computes $m(x)$ using those queries, and outputs the result (i.e. it outputs 1 if $m(x) = 1$ and -1 if $m(x) = -1$). Note that since $p(x)f(x) > 0$, more than half of the monomials m of p agree with f in sign, when weighted by $|c_m|$. This means that the probability that $R(x) = f(x)$ is more than $1/2$, so this algorithm makes error strictly less than $1/2$. We conclude that $\mathrm{R}_{\pm}(f) = \deg_{\pm}(f)$, as desired. □

This lemma shows that all the different query versions of UPP collapse together; they are equal up to constant factors. However, as mentioned previously, there is a different notion of PP in query complexity, which is not equal to sign degree. To distinguish between these, the sign degree is usually referred to as UPP rather than PP.

The difference has to do with just how small the bias of a PP algorithm should be allowed to be. In the definitions we've seen so far, the bias was allowed to be arbitrarily small (in other words, the error was allowed to be arbitrarily close to $1/2$). However, intuitively, if an algorithm uses only T amount of resources (in our case, queries), one may wish to also only let it flip $O(T)$ coins, or maybe $O(\mathrm{poly}(T))$ coins, when making its random choices. Putting such a restriction on the randomness would mean that the probabilities of the algorithm are discretized at around size $2^{-O(T)}$, and in particular, if the algorithm achieves non-zero bias, then it will achieve bias at least $2^{-O(T)}$. However, sign degree has no such guarantees.

There are several different query versions of PP which add some restriction on the number of coins, or on the bias achieved. These measures are not all exactly identical to one another, but they generally differ only by a factor of $O(\log n)$, where n is the input size. This is in contrast to the sign degree measure (which we will consider from now on to be a query version of UPP), which can be exponentially smaller than the query versions of PP.

We will define one particular query version of PP, which has reasonably nice properties.

Definition 6.4. *Let f be a (possibly partial) Boolean function. For any algorithm R which computes f to some non-zero bias, let $|R|$ denote the worst-case number of queries R makes, and let $\mathrm{bias}(R)$ denote the minimum value of $\Pr[R(x) = f(x)] - \Pr[R(x) \neq f(x)]$ over $x \in \mathrm{Dom}(f)$. Then define*

$$\mathrm{R}_{\mathrm{PP}}(f) \coloneqq \min\{T : \exists R \text{ s.t. } |R| \leq T \text{ and } \mathrm{bias}(R) \geq n^{-T}\}.$$

Define $\mathrm{Q}_{\mathrm{PP}}(f)$ *and* $\deg_{\mathrm{PP}}(f)$ *similarly (for polynomials, the bias is the minimum value of* $p(x)f(x)$ *over* $x \in \mathrm{Dom}(f)$ *in the* $\{+1,-1\}$ *basis, and we minimize only over polynomials* p *that are bounded on the Boolean hypercube).*

It turns out that these PP measures can be exponentially larger than the sign degree. An example of a function which gives an exponential separation is a function called Odd-max-bit. This function is defined on n bits, and on input x, it looks at the largest $i \in [n]$ such that $x_i = 1$ (set $i = 0$ if there is no such $i \in [n]$). The function then returns 1 if this largest i is odd, and 0 if the largest i is even.

Let us construct a $\deg_{\pm}$ polynomial for Odd-max-bit. This is easiest to do when the input alphabet is $\{0,1\}$ but the outputs of the function are $\{+1,-1\}$. In this case, we simply set $p(x) = 1 + \sum_{i=1}^{n}(-2)^i x_i$. This is a polynomial of degree 1. However, note that the absolute coefficients of this polynomials are $1, 2, 4, 8, \ldots$, and that they alternate in sign; this means that the sign of $p(x)$ is the sign of the largest $(-2)^i$ when x_i is non-zero, which is the parity of the largest i such that $x_i = 1$. Hence p agrees with Odd-max-bit in sign. If we scale p to be bounded in $[-1,1]$ (by dividing it by something like 2^{n+1}), the bias achieved by p would only be $2^{-O(n)}$, even though the degree of p is 1; hence p does not give us an upper bound on $\deg_{\mathrm{PP}}(\text{Odd-max-bit})$ that is better than $n/\log n$.

It turns out that $\deg_{\mathrm{PP}}(\text{Odd-max-bit}) = \widetilde{\Theta}(n^{1/3})$, though this is somewhat tricky to prove. The first lower bound was provided by [Bei94]. Hence Odd-max-bit shows that $\deg_{\mathrm{PP}}$ and $\deg_{\pm}$ can be vastly different, or in other words, that PP and UPP are not the same in query complexity.

Next, we will show that all the PP measures are approximately equal to each other.

Lemma 6.5. *Let* f *be a (possibly partial) Boolean function on* n *bits. Then*

$$\mathrm{R}_{\mathrm{PP}}(f) \geq \mathrm{Q}_{\mathrm{PP}}(f) \geq (1/2)\deg_{\mathrm{PP}}(f) \geq (1/3)\mathrm{R}_{\mathrm{PP}}(f).$$

Proof. Since quantum algorithms can simulate quantum ones and since they give rise to polynomials, we get $\mathrm{R}_{\mathrm{PP}}(f) \geq \mathrm{Q}_{\mathrm{PP}}(f) \geq (1/2)\deg_{\mathrm{PP}}(f)$. We now show how to convert a PP polynomial p into a randomized algorithm. Let $T = \deg_{\mathrm{PP}}(f)$, and suppose p is a multilinear bounded polynomial with $\deg(p) \leq T$ and $p(x)f(x) \geq n^{-T}$ for all $x \in \mathrm{Dom}(f)$.

The polynomial p is bounded, meaning it lies in $[-1,1]$ on $\{+1,-1\}^n$. In Lemma 6.6, we will see that if p is bounded in this way, then the sum of absolute values of its coefficients is at most $n^{\deg(p)/2}$. Let $p' := p/n^{\deg(p)/2}$. Then we have $\deg(p') = \deg(p) \leq T$, $\mathrm{bias}(p') = \mathrm{bias}(p)/n^{\deg(p)/2} \geq n^{-3T/2}$, and the sum of absolute values of coefficients of p' is at most 1. We can now construct a randomized algorithm R for f which samples a monomial m of p' in proportion to the probabilities $|c_m|$, queries the positions x_i of the input that occur in the monomial m, and then outputs $m(x)$. The expected output of this algorithm, when run on x, is $p'(x)$, so the algorithm achieves bias at least $\mathrm{bias}(p')$ for computing f. It also makes at most $\deg(p')$ queries. Hence $|R| \leq T$ and $\mathrm{bias}(R) \geq n^{-3T/2}$, so $\mathrm{R}_{\mathrm{PP}}(f) \leq (3/2)\deg_{\mathrm{PP}}(f)$. □

The proof of this lemma relied on the following bound for the sum of absolute values of coefficients of a bounded multilinear polynomial.

Lemma 6.6. *Let* p *be a bounded multilinear polynomial in* n *variables in the* $\{+1,-1\}$ *basis. Then the sum of absolute coefficients of* p *is at most*

$$\sqrt{\sum_{i=0}^{\deg(p)} \binom{n}{i}},$$

and also at most $n^{\deg(p)/2}$.

Proof. For a multilinear polynomial q in n variables, let $\mathbb{E}[q]$ denote the expectation of $q(x)$ over the uniform distribution over $x \in \{+1,-1\}^n$. Note that $\mathbb{E}[m] = 0$ for any non-constant monomial m, because each non-constant monomial is a parity function on some subset of the bits, and parity is equally likely to be $+1$ and -1 over the Boolean hypercube. Now, consider $\mathbb{E}[p^2]$. Note that we know $p^2(x) \leq 1$ for all $x \in \{+1,-1\}^n$, since $p(x) \in [-1,1]$ on the Boolean hypercube; hence $\mathbb{E}[p^2] \leq 1$. On the other hand, if we expand out $p(x)^2 = (\sum_m c_m m(x))^2$, we get a sum of $c_m c_{m'} m(x)m'(x)$. The product $m(x)m'(x)$ may simplify (since recall that $x_i^2 = 1$ over $x \in \{+1,-1\}^n$), but it will be non-constant unless $m = m'$. Now, $\mathbb{E}[p^2]$ is the sum of $c_m c_{m'} \mathbb{E}[mm']$, and we have $\mathbb{E}[mm'] = 0$ unless $m = m'$, in which case $\mathbb{E}[mm'] = \mathbb{E}[1] = 1$. We conclude that $\mathbb{E}[p^2] = \sum_m c_m^2$, and hence $\sum_m c_m^2 \leq 1$.

Finally, we use Cauchy-Schwartz. We have

$$\sum_m |c_m| = \sum_m |c_m| \cdot 1 \leq \sqrt{\sum_m |c_m|^2 \sum_m 1} \leq \sqrt{\sum_m 1}.$$

Hence we've bounded the sum of absolute coefficients of p by the square root of the number of monomials of p. Now, the total number of multilinear monomials on n variables that exist of degree i is $\binom{n}{i}$. Since p has degree $\deg(p)$, the number of monomials in p is at most $\sum_{i=0}^{\deg(p)} \binom{n}{i}$, from which the desired result follows.

As for the upper bound of $n^{\deg(p)}$, it is not hard to show that $\sum_{i=0}^{k} \binom{n}{i} \leq n^k$ unless $k = 1$, in which case $\sum_{i=0}^{1} \binom{n}{i} = n + 1$. We can deal with degree-1 polynomials separately: they have the form $p(x) = c_0 + c_1 x_1 + c_2 x_2 + \ldots c_n x_n$. By plugging in the correct $x \in \{+1,-1\}^n$, we can ensure the signs of all the terms $c_0, c_1 x_1, c_2 x_2, \ldots, c_n x_n$ match; since $|p(x)| \leq 1$ for such x, we therefore have $|c_0| + |c_1| + \cdots + |c_n| \leq 1 \leq n^{\deg(p)/2}$. □

So far, we have seen that the three measures $\deg_\pm$, $\mathrm{R}_\pm$, and $\mathrm{Q}_\pm$ are equivalent up to constant factors; these are usually referred to by the name "sign degree" or "threshold degree', or sometimes by the notation UPP (the "U" stands for "unbounded", so this is unbounded PP).

We also saw that the three measures $\deg_{\mathrm{PP}}$, R_{PP}, and Q_{PP} are equivalent up to constant factors. These measures (or variants of them) are often called "discrepancy", and they correspond more closely to the class PP. We also noted that discrepancy can be much larger than sign degree for some functions.

6.2.1 Duality for the discrepancy and sign degree

We have introduced two types of small-bias measures: $\deg_\pm$ and $\deg_{\mathrm{PP}}$. It will be useful to look at the dual objects of these measures, which end up being simpler than the dual of approximate degree in Theorem 6.1. First, we have the following for $\deg_\pm$.

Lemma 6.7. *Let f be a (possibly partial) Boolean function. To show that $\deg_\pm(f) > T$, it suffices to show that there exists a distribution μ on $\mathrm{Dom}(f)$ such that for every multilinear monomial m of degree at most T, we have $\mathbb{E}_{x\sim\mu}[m(x)f(x)] = 0$.*

Moreover, this technique is tight: if $\deg_\pm(f) > T$, such a distribution μ always exists.

Note that this lemma is simpler than Theorem 6.1, since it makes no reference to a modified function f'; to prove a sign degree lower bound, all you need to provide is a hard distribution for the original function f.

Proof. First, if such a distribution μ exists, then $\mathbb{E}_{x\sim\mu}[m(x)f(x)] = 0$ for all m of degree at most T; since each polynomial of degree at most T is a linear combination of such monomials, we also

have $\mathbb{E}_{x\sim\mu}[p(x)f(x)] = 0$ for all polynomials of degree at most T. In this case, there clearly cannot be a polynomial of degree T with $p(x)f(x) > 0$ for all x, so $\deg_\pm(f) > T$.

To show tightness, we can directly use Theorem 6.1 in the limiting case of $\epsilon \to 1/2$. If $\deg_\pm(f) > T$, then for all $\epsilon < 1/2$, there is no polynomial of degree T approximating f to error ϵ; by Theorem 6.1, we get dual objects f'_ϵ, μ_ϵ for all such $\epsilon < 1/2$. Let's take a sequence of these objects $(f'_1, \mu_1), (f'_2, \mu_2), \ldots$ which corresponds to a sequence of errors ϵ that approaches $1/2$. Some Boolean function f' must occur infinitely often in this sequence; restrict to the subsequence where all the f'_i are the function f'. The sequence of corresponding μ_i are vectors which lie in a closed and bounded subset of $\mathbb{R}^n$, which means they have a limit point: some subsequence of the μ_i must converge to a distribution μ. It is not hard to check that the correlation between f and f' against this distribution μ must be at least $1-\delta$ for all δ, so it must equal 1. This means we can take $f' = f$, since f' agrees with f on the support of μ. It is also not hard to verify that $\mathbb{E}_{x\sim\mu}[f(x)m(x)] = 0$ must hold for all monomials m of degree at most T, as desired. □

The measure $\deg_{\text{PP}}$ also has a dual object which is simply a probability distribution, with no need for a modified function f'.

Lemma 6.8. *Let f be a (possibly partial) Boolean function. To show that $\deg_{\text{PP}}(f) > T$, it suffices to show that there exists a distribution μ on $\text{Dom}(f)$ such that every multilinear monomial m of degree at most T satisfies $|\mathbb{E}_{x\sim\mu}[m(x)f(x)]| < n^{-3T/2}$.*

Moreover, this technique is tight up to constant factors: if $\deg_{\text{PP}}(f) > T$, then a distribution certifying a lower bound of $> T/4$ always exists.

Proof. Let p be a bounded multilinear polynomial of degree at most T. Recall that the sum of absolute coefficients of p is at most $n^{T/2}$. Now,

$$\mathop{\mathbb{E}}_{x\sim\mu}[p(x)f(x)] = \sum_m c_m \mathop{\mathbb{E}}_{x\sim\mu}[m(x)f(x)] < n^{T/2}n^{-3T/2} = n^{-T}.$$

This means there must be some x for which $p(x)f(x) < n^{-T}$, so $\text{bias}(p) < n^{-T}$ and hence $\deg(p) + \log(1/\text{bias}(p))/\log n > T$. Since this applies to all polynomials p of degree at least T, we conclude that $\deg_{\text{PP}}(f) > T$.

In the other direction, we use Theorem 6.1. If $\deg_{\text{PP}}(f) > T$, we know that all polynomials of degree at most $T/4$ must have bias less than $n^{-3T/4}$, which means they must make error at least $(1 - n^{-3T/4})/2$; Theorem 6.1 then gives us a modified function f' and a distribution μ such that f and f' have correlation greater than $(1 - n^{-3T/4})/(1 + n^{-3T/4}) \geq 1 - 2n^{-3T/4}$ against μ, and such that $\mathbb{E}_{x\sim\mu}[f'(x)m(x)] = 0$ for all monomials m of degree at most $T/4$. Now, we know that μ places probability mass less than $2n^{-3T/4}$ on $x \in \{+1, -1\}^n$ where $f(x) \neq f'(x)$ (where we define $f(x) = 0$ if $x \notin \text{Dom}(f)$). Then for any monomial m of degree at most $T/3$, we have

$$|\mathop{\mathbb{E}}_{x\sim\mu}[f(x)m(x)]| = |\mathop{\mathbb{E}}_{x\sim\mu}[(f(x) - f'(x))m(x)] + \mathop{\mathbb{E}}_{x\sim\mu}[f'(x)m(x)]| < 2n^{-3T/4}.$$

Now, if $T < 4$, then the only monomial of degree at most $T/4$ is the constant monomial, and then if the function is not constant we can put equal weight on $+1$ and -1 inputs to get a correlation of 0 with this monomial; hence we can restrict to the case $T \geq 4$. In this case, we also have $n \geq 4$, so $2n^{-3T/4} \leq n^{1/2-3T/4} \leq n^{T/8-3T/4} = n^{-5T/8} \leq n^{-(3/2)T/4}$. Hence μ certifies that $\deg_{\text{PP}}(f) > T/4$. □

Discrepancy lower bounds sometimes prove an even stronger statement: they show that for some distribution μ on $\text{Dom}(f)$, it holds that $|\mathbb{E}_{x\sim\mu}[f(x)m(x)]| \leq n^{-\Omega(T)}$ for all monomials m (not

merely for the monomials of degree at most T). This is of course sufficient to prove an $\Omega(T)$ lower bound, but it is not necessary.

The technique of showing a distribution against which f poorly correlates with all monomials is what is traditionally referred to as "discrepancy". Note that if you manage to provide such a distribution, it gives not only a lower bound on $\mathrm{Q}(f)$, but even on the small-bias measure $\mathrm{Q}_{\mathrm{PP}}(f)$.

6.3 Back to bounded-error polynomials

We can relax the conditions in Theorem 6.1 slightly, making it a little easier to show polynomial lower bounds using this method.

Theorem 6.9. *Let f be a (possibly partial) Boolean function on n bits. Suppose we had a total Boolean function f' and a distribution μ on the Boolean hypercube satisfying the following properties:*

1. *f' has correlation at least $2/3$ with f against μ*
2. *For any monomial m of degree at most d, the correlation of f' with m against μ is at most n^{-d}.*

Then $\widetilde{\deg}(f) = \Omega(d)$.

Note that this theorem is very similar to Theorem 6.1; the main difference is that we are now allowing f' to have a small n^{-d} correlation with the degree-d monomials rather than having zero correlation with them.

The constant $2/3$ above can be replaced by any other constant in $(1/2, 1)$, and the statement of the theorem will still hold.

Proof. Consider any bounded polynomial p of degree at most d. Then for any $x \in \mathrm{Dom}(f)$, we have $p(x)m(x) \geq \mathrm{bias}(p)$, as $\mathrm{bias}(p)$ is defined as the minimum of $p(x)m(x)$ over $x \in \mathrm{Dom}(f)$.

Consider where μ places its weight. Suppose it puts weight a on $x \notin \mathrm{Dom}(f)$ and weight b on $x \in \mathrm{Dom}(f)$ with $f(x) \neq f'(x)$, with weight $1 - a - b$ on $x \in \mathrm{Dom}(f)$ with $f(x) = f'(x)$. Then $\mathbb{E}_{x\sim\mu}[f(x)f'(x)] = 1 - a - 2b \leq 2/3$, so $a + 2b \geq 1/3$.

Since μ places weight $1 - a$ on $\mathrm{Dom}(f)$, we have $\mathbb{E}_{x\sim\mu}[f(x)p(x)] \geq (1 - a)\,\mathrm{bias}(p)$. On the other hand,

$$\mathop{\mathbb{E}}_{x\sim\mu}[f(x)p(x)] = \mathop{\mathbb{E}}_{x\sim\mu}[(f(x) - f'(x))p(x)] + \mathop{\mathbb{E}}_{x\sim\mu}[f'(x)p(x)]$$

$$\leq a + 2b + \sum_m c_m \mathop{\mathbb{E}}_{x\sim\mu}[f'(x)m(x)] \leq a + 2b + n^{-d}\sum_m |c_m| \leq a + 2b + n^{-d/2}.$$

Hence we have $a + 2b + n^{-d/2} \geq (1 - a)\,\mathrm{bias}(p)$, or $\mathrm{bias}(p) \leq (a + 2b + n^{-d/2})/(1 - a)$. Using $a \geq 1/3 - 2b$, we get $\mathrm{bias}(p) \leq (1/3 + n^{-d/2})/(2b + 2/3) \leq (1 + 3n^{-d/2})/2$. Hence p must make error at least $(1 - \mathrm{bias}(p))/2 \geq (1 - 3n^{-d/2})/4$ when estimating f. Note that if d is constant, an $\Omega(d)$ lower bound on the degree always holds (since the degree is at least 1 for all non-constant functions). When d is larger than some constant, say 10, we must also have $n \geq 10$, and then the error of p when computing f becomes strictly greater than $1/5$. Since p was arbitrary, we get a lower bound on $\widetilde{\deg}_{1/5}(f)$.

However, it turns out that approximate degree can be amplified: if there is a polynomial approximating f to error $1/3$, we can turn it into a polynomial estimating f to error $1/5$ while increasing its degree by at most a constant factor. Combined with amplification, we get an $\Omega(d)$ lower bound for $\widetilde{\deg}(f)$ with any constant error level. □

The lower bound technique in Theorem 6.9 is called *generalized discrepancy*. The idea in discrepancy is to show that f has low correlation with every monomial when measured against some hard distribution μ. In generalized discrepancy, we show that a modified function f' has low correlation with every monomial against μ, and also that f correlates reasonably well with f' when measured against the same distribution μ. In other words, we are trying to show that f is not too far from a very hard function f', but the distance between f and f' must be measured according to the hard distribution for f'.

6.4 Applications

This lower bound technique (called either dual polynomials or generalized discrepancy) is not easy to use in practice, but we review some applications.

6.4.1 Composition theorems

Recall that if f and g are (possibly partial) Boolean functions defined on n and m bits respectively, then $f \circ g$ is a (possibly partial) Boolean function on nm bits. We can get a polynomial for $f \circ g$ by composing a polynomial p_f for f with n copies of a polynomial p_g for g. However, there is a problem if p_g only approximates g: in this case, there is no guarantee that $p_f(p_g, p_g, \ldots, p_g)$ approximates $f \circ g$, because p_f is only known to approximates f on the Boolean hypercube (not when fed in noisy input bits). It is possible to overcome this issue by amplifying the polynomial p_g before composing, increasing its degree by a factor of $O(\log \widetilde{\deg}(f))$ and decreasing its approximation error to $1/\operatorname{poly}(\widetilde{\deg}(f))$. This gives a polynomial approximating $f \circ g$ of degree $O(\widetilde{\deg}(f)\,\widetilde{\deg}(g)\log\widetilde{\deg}(f))$.

It turns out that the extra log factor is not necessary when composing polynomials, just as it was not necessary for quantum algorithms. This was shown by Sherstov [She12b]. This means that $\widetilde{\deg}(f \circ g) = O(\widetilde{\deg}(f)\,\widetilde{\deg}(g))$. On the other hand, the other direction is still open.

Conjecture 6.10. *For all Boolean functions f and g, $\widetilde{\deg}(f \circ g) = \Omega(\widetilde{\deg}(f)\,\widetilde{\deg}(g))$.*

While the conjecture is still open, some partial results are known. Sherstov [She12a] showed the following using the dual characterization of approximate degree.

Theorem 6.11. *Let f act on n bits and let g act on m bits. Then*

$$\widetilde{\deg}(f \circ g) = \Omega\left(\frac{\widetilde{\deg}(f)^2\,\widetilde{\deg}(g)}{n}\right).$$

In particular, this theorem gives a tight composition theorem when $\widetilde{\deg}(f) = \Omega(n)$, but it gets weaker the smaller $\widetilde{\deg}(f)$ gets.

Other partial progress towards a composition theorem includes work on composition with AND and OR.

6.4.2 Explicit lower bounds

Apart from composition theorems, generalized discrepancy or dual polynomials can be explicitly constructed to give lower bounds on approximate degree. One example of a paper doing so is [BKT18], in which dual polynomials are combined with the symmetrization technique for functions

with large input alphabets to give several new lower bounds on quantum query complexity, including for functions like k-DISTINCTNESS and IMAGE-SIZE-TESTING.

References

[Bei94] Richard Beigel. "Perceptrons, PP, and the polynomial hierarchy". In: *Computational complexity* 4.4 (1994) (p. 5).

[BKT18] Mark Bun, Robin Kothari, and Justin Thaler. "The polynomial method strikes back: Tight quantum query bounds via dual polynomials". In: *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing*. 2018 (p. 9).

[She12a] Alexander A Sherstov. "Strong direct product theorems for quantum communication and query complexity". In: *SIAM Journal on Computing* 41.5 (2012) (p. 9).

[She12b] Alexander A. Sherstov. "Making Polynomials Robust to Noise". In: *Proceedings of the 44th Symposium on Theory of Computing*. 2012. DOI: 10.1145/2213977.2214044 (p. 9).