THE GAUSSIAN SURFACE AREA AND NOISE SENSITIVITY OF DEGREE-D POLYNOMIAL THRESHOLD FUNCTIONS

Daniel M. Kane

Abstract. We prove asymptotically optimal bounds on the Gaussian noise sensitivity and Gaussian surface area of degree-d polynomial threshold functions. In particular we show that for f a degree-d polynomial threshold function that $\mathrm{GNS}_\epsilon(f) \leq \frac{d \arcsin(\sqrt{2\epsilon-\epsilon^2})}{\pi}$. This bound translates into an optimal bound on the Gaussian surface area of such functions, namely that the Gaussian surface area is at most $\frac{d}{\sqrt{2\pi}}$. Finally we note that the later result implies bounds on the runtime of agnostic learning algorithms for polynomial threshold functions.

Keywords. Gaussian Noise, Polynomial Threshold Functions, Machine Learning

Subject classification. 68T05, 60E10

1. Introduction

1.1. Background. We call a function f a polynomial threshold function if it is of the form $f(x) = \mathrm{sgn}(p(x))$ for some polynomial p. We say that f is a degree-d polynomial threshold function if the degree of p is at most d. Polynomial threshold functions are a natural generalization of linear threshold functions (the $d = 1$ case). Polynomial threshold functions find application in fields such as circuit complexity (Beigel 1993), communication complexity (Sherstov 2009), and learning theory (Klivans & Servedio 2004).

One property of interest for a Boolean function is that of sensitivity. That is some measure of how stable the value of the function is under small changes to its input. Various notions of sensitivity have found applications in a number of areas including hardness of approximation (Subhash Khot & ODonnell 2007), hardness amplification (ODonnell 2004), quantum complexity (Shi 2000), and learning theory (Adam R. Klivans & Servedio 2008). We will discuss the last of these in more detail shortly.

1.2. Measures of Sensitivity. We will concern ourselves with four different measures of sensitivity of a Boolean function; two for Bernoulli inputs (x chosen uniformly from $\{-1,1\}^n$) and two for Gaussian inputs. Our results shall focus only on the latter two measures, but the first two are useful for motivational purposes.

Perhaps the most natural notion of sensitivity is that of average sensitivity. This measures the probability that flipping a randomly chosen bit of a random Bernoulli input changes the value of the function. In particular we define:

$$\mathrm{AS}(f) := \mathrm{E}_x[\text{Number of bits of } x \text{ that when flipped would change the value of } f].$$

A slightly less precise notion is that of noise sensitivity. This is a measure of the probability that randomly changing some fraction of the bits will change the value of f. In particular we define the noise sensitivity with noise rate ϵ to be

$$\mathrm{NS}_\epsilon(f) := \mathrm{Pr}_{x,z}(f(x) \neq f(z))$$

where x is Bernoulli and z is obtained from x by flipping each bit independently with probability ϵ.

The concept of noise sensitivity translates naturally into the Gaussian setting. In particular we define the Gaussian noise sensitivity with noise rate ϵ to be

$$\mathrm{GNS}_\epsilon(f) := \mathrm{Pr}(f(X) \neq f(Z))$$

where X is an n-dimensional Gaussian random variable, and $Z = (1-\epsilon)X + \sqrt{2\epsilon - \epsilon^2}Y$ for Y an independent n-dimensional Gaussian. Notice that Z is chosen here to be a standard Gaussian whose covariance with X is $1-\epsilon$ in each coordinate, agreeing with the noise sensitivity in the Bernoulli case. In fact we can think of the Gaussian noise sensitivity as a special case of the Bernoulli noise sensitivity by approximating each Gaussian random variable as a properly scaled average of a large number of Bernoulli random variables.

There is a closely related concept called the Gaussian surface area, which, instead of directly measuring the sensitivity of a function, measures the boundary between its $+1$ and -1 regions. We define the Gaussian surface area of a set A to be

$$\Gamma(A) := \liminf_{\delta \to 0} \frac{\mathrm{GaussianVolume}(A_\delta \backslash A)}{\delta}.$$

Where the Gaussian volume of a region R is $\mathrm{Pr}(X \in R)$ for X a Gaussian random variable, and where A_δ is the set of points x so that $d(x, A) \leq \delta$ (under the Euclidean metric). We note that if A is a sufficiently nice region with a smooth boundary, then its Gaussian surface area should be equal to

$$\int_{\partial A} \phi(x) d\sigma.$$

Here $\phi(x)$ is the Gaussian density, and $d\sigma$ is the surface measure on ∂A. The basic idea is that the integral above is a natural surface integral with a correction factor of ϕ to account for the Gaussian measure. Furthermore if A is such a region, then its Gaussian surface area should be equal to

$$\lim_{\delta \to 0} \frac{\text{GaussianVolume}((\partial A)_\delta)}{2\delta}.$$

The factor of 2 above comes from the fact that $(\partial A)_\delta$ has volume both inside and outside of A. These equalities are proven for the case of A the set of positive values of a square-free polynomial threshold function in Proposition A.3.

For f a Boolean function, we define

$$\Gamma(f) := \Gamma(f^{-1}(1)).$$

The concepts of Gaussian noise sensitivity and surface area are related to each other by noting that the noise sensitivity is roughly the probability that X is close enough to the boundary that wiggling it will push it over the boundary. In particular it is proved in Ledoux (1994) that for any Boolean function, f,

$$\frac{\text{GNS}_\epsilon(f)}{2} \leq \frac{\arccos(1-\epsilon)}{\sqrt{2\pi}} \Gamma(f).$$

We essentially prove an asymptotic converse to this statement for f a polynomial threshold function in Section 3.

1.3. Previous Work. In this paper we study the sensitivity of polynomial threshold functions. Before stating our major results, we will provide a brief overview of the previous work in this area.

The most ambitious conjecture in this field is the Gotsman-Linial conjecture (Gotsman & Linial 1994), which states that the largest average sensitivity of a degree-d polynomial threshold function is obtained by taking the product of linear threshold functions slicing through the middle d layers of hypercube. This would imply that

$$\text{AS}(f) \leq d\sqrt{\frac{2n}{\pi}}.$$

This would in turn imply optimal or nearly optimal bounds for our other measures. In particular, by I. Diakonikolas & Tan (2009) Theorem 7.1 it would imply that

$$\mathrm{NS}_\epsilon(f) \leq d\sqrt{\frac{2\epsilon}{\pi}}.$$

This bound will be asymptotically correct for small ϵ and properly chosen f.

Treating Gaussian random variables as averages of Bernoulli random variables we get a bound on the Gaussian noise sensitivity. In particular I. Diakonikolas & Tan (2009) Proposition 9.2, this would imply

$$\mathrm{GNS}_\epsilon(f) \leq d\sqrt{\frac{2\epsilon}{\pi}}.$$

As we shall see, this bigger than the correct asymptotic by a factor of $\sqrt{\pi}$. I believe that this discrepancy comes from the fact that in the Bernoulli case things can be arranged so that the boundary makes particular angles with edges of the hypercube, while the Gaussian case is necessarily isotropic.

As we shall wee in section 3, this bound would imply a bound on the Gaussian surface area of

$$\Gamma(f) \leq \frac{d}{\sqrt{2}},$$

which is again off from the optimal by a factor of $\sqrt{\pi}$.

Unfortunately, we are far from proving these conjectures. The Gotsman-Linial Conjecture is known for $d = 1$, but is much more difficult for larger degrees. Non-trivial bounds on the average sensitivity were proved independently by I. Diakonikolas & Tan (2009) and P. Harsha & Meka (2009). They each proved bounds of $O(n^{1-1/O(d)})$ on the average sensitivity and $O_d(\epsilon^{1-1/O(d)})$ on the Gaussian and Bernoulli noise sensitivity. These results are proved by obtaining results in the Gaussian setting in such a way that they can be carried over to the Bernoulli setting. The proofs involve a number of advanced techniques including the invariance principle, and concentration and anti-concentration results for polynomials of Gaussians.

Little at all was previously known about Gaussian surface area for $d > 1$. For $d = 1$, a simple computation yields the optimal bound of $\frac{1}{\sqrt{2\pi}}$. Adam R. Klivans & Servedio (2008) proved a bound of 1 for the Gaussian surface area of balls, and essentially nothing else was known for polynomial threshold functions.

1.4. Statement of Results. We focus on proving two main results.

Theorem 1.1. *If f is a degree d polynomial threshold function, then*

$$GNS_\epsilon(f) \leq \frac{d\arcsin(\sqrt{2\epsilon - \epsilon^2})}{\pi} \sim \frac{d\sqrt{2\epsilon}}{\pi} = O(d\sqrt{\epsilon}).$$

Furthermore this bound is asymptotically tight as $\epsilon \to 0$ for f the threshold function of any square-free product of homogeneous linear functions.

Theorem 1.2. *If f is a degree-d polynomial threshold function then $\Gamma(f) \leq \frac{d}{\sqrt{2\pi}}$. This bound is again optimal for f the threshold function of any square-free product of homogeneous linear functions.*

1.5. Application to Agnostic Learning. We briefly describe one of the applications of our work to agnostic learning algorithms for polynomial threshold functions.

1.5.1. Overview of Agnostic Learning. Agnostic learning is a model for machine learning. Suppose that you have an unknown distribution D on $X \times \{-1, 1\}$ (for some set X) with known marginal distribution D_X on X. Think of X as the observable data about some object and the element of $\{-1, 1\}$ as a bit we are trying to predict. What we would like is an algorithm that given a number of samples from D outputs a function $h : X \to \{-1, 1\}$ so that for $(x, y) \sim D$, $h(x) = y$ with high probability. For arbitrary D we have little hope of success as we have no way to predict what value y should have on values of x that we have never seen in our training data. To fix this, the agnostic learning model merely requires that your function be competitive against some class of predictors. We define a concept class to be some set C of functions $X \to \{-1, 1\}$. We let $opt = \inf_{f \in C} \Pr_{(x,y)\sim D}(f(x) \neq y)$. In the agnostic learning model, our objective is now to find an h so that $\Pr_{(x,y)\sim D}(h(x) \neq y) \leq opt + \epsilon$.

1.5.2. Connection to Gaussian Surface Area. The connection with Gaussian surface was discovered by Adam R. Klivans & Servedio (2008). In particular they prove

Theorem (Adam R. Klivans & Servedio 2008 Theorem 25). *Let C be a collection of Borel sets in $\mathbb{R}^n$ each of which has Gaussian surface area at most s. Suppose furthermore that C is invariant under affine transformations. Then C is agnostically learnable with respect to any Gaussian distribution in time $n^{O(s^2/\epsilon^4)}$.*

Hence our results imply that

Corollary 1.3. *The concept class of degree-d polynomial threshold functions is agnostically learnable with respect to any Gaussian distribution in time $n^{O(d^2/\epsilon^4)}$.*

1.6. Outline. Section 2 will be devoted to the proof of Theorem 1.1, Section 3 to the proof of Theorem 1.2, and Section 4 will provide some closing notes. The proof of Theorem 1.2 requires several technical results whose proofs are not very enlightening and are put in the Appendix so as not to interfere with the flow of the paper.

2. Proof of the Noise Sensitivity Bound

Proof (of Theorem 1.1). We begin by letting $\theta = \arcsin(\sqrt{2\epsilon - \epsilon^2})$. We need to bound

$$p := \mathrm{GNS}_\epsilon(f) = \Pr(f(X) \neq f(\cos(\theta)X + \sin(\theta)Y)). \tag{2.1}$$

In general we let

$$X_\phi = \cos(\phi)X + \sin(\phi)Y.$$

Hence, $\mathrm{GNS}_\epsilon(f) = \Pr(f(X_0) \neq f(X_\theta))$.

For purposes of intuition we have found it useful to consider the XY-plane, whose points consist of random variables which are linear combinations of X and Y. Both X and Z are unit vectors in this plane separated by an angle of θ. The X_ϕ trace out the unit circle in this plane. Once we fix the values of X and Y, f is now a degree-d polynomial threshold function on the XY-plane, and we ask the probability that $f(X_0) \neq f(X_\theta)$. This question is difficult because we need to average over the choice of X and Y, and hence the plane in $\mathbb{R}^n$ cut out by the XY-plane. We solve this by symmetrizing first over rotations of the XY-plane. This reduces the problem to that of counting the number of sign changes of f on the unit circle, which is at most $2d$ for any choice of X and Y.

To manage this symmetrization, we note that the value of p given in Equation (2.1) remains the same if X and Y are replaced by any X' and Y' that are i.i.d. standard Gaussian distributions. The key observation is that X_ϕ and $X_{\phi+\pi/2}$ are such i.i.d. Gaussians. Additionally we note that for any ϕ

$$\begin{aligned}\cos(\theta)X_\phi + \sin(\theta)X_{\phi+\pi/2} =& \cos(\theta)\cos(\phi)X - \sin(\theta)\sin(\phi)X + \\ & \cos(\theta)\sin(\phi)Y + \sin(\theta)\cos(\phi)Y \\ =& \cos(\theta+\phi)X + \sin(\theta+\phi)Y \\ =& X_{\theta+\phi}.\end{aligned}$$

Therefore, we have that for any ϕ,

$$\mathrm{GNS}_\epsilon(f) = \Pr(f(X_\phi) \neq f(X_{\phi+\theta})).$$

Averaging over ϕ, we find that

$$\mathrm{GNS}_\epsilon(f) = \frac{1}{2\pi}\int_0^{2\pi} \Pr(f(X_\phi) \neq f(X_{\phi+\theta}))d\phi. \tag{2.2}$$

We define the random function $F : \mathbb{R} \to \{-1, 1\}$ by

$$F(\phi) = f(X_\phi).$$

(F depends on X and Y as well as ϕ). Hence

$$\mathrm{GNS}_\epsilon(f) = \frac{1}{2\pi}\int_0^{2\pi} \Pr(F(\phi) \neq F(\phi+\theta))d\phi = \frac{1}{2\pi}\mathrm{E}_{X,Y}\left[\int_0^{2\pi} \mathbf{1}_{F(\phi)\neq F(\phi+\theta)}d\phi\right].$$

Here $\mathbf{1}_{F(\phi)\neq F(\phi+\theta)}$ is the indicator function of the event that $F(\phi) \neq F(\phi+\theta)$. We may now consider the value of the integral above for fixed X and Y and then take the expectation. In such a case, F is some Boolean function. We note that $F(\phi)$ can only be different from $F(\phi+\theta)$ if F has a sign change in $[\phi, \phi+\theta]$ (with these values taken modulo 2π). This means that $F(\phi) \neq F(\phi+\theta)$ only on a union of intervals of length θ preceding each sign change of F. Therefore, the value of the above integral is at most θ times the number of sign changes of F on $[0, 2\pi)$. Hence,

$$\mathrm{GNS}_\epsilon(f) \leq \frac{\theta \mathrm{E}_{X,Y}[\text{number of sign changes of } F \text{ on } [0, 2\pi)]}{2\pi}. \tag{2.3}$$

We now make use of the fact that f is a degree d polynomial threshold function. In particular we will show that for any X and Y, F changes signs at most $2d$ times on $[0, 2\pi)$. We let $f = \mathrm{sgn}(g)$ for some degree d polynomial g. We note that the number of sign changes of F is at most the number of zeroes of the function $g(\cos(\phi)X + \sin(\phi)Y)$ (unless this function is identically 0, in which case there are no sign changes). It should be noted though that $g(\cos(\phi)X + \sin(\phi)Y) = 0$ if and only if $z = e^{i\phi}$ is a root of the degree-$2d$ polynomial

$$z^d g\left(\left(\frac{z + z^{-1}}{2}\right)X + \left(\frac{z - z^{-1}}{2i}\right)Y\right).$$

Therefore the expectation in Equation (2.3) is at most $2d$. Alternatively, we could say that these zeroes correspond the number of joint solutions of $g(aX +$

$bY) = 0$ and $a^2 + b^2 = 1$, which is at most $2d$ by Bezout's Theorem. Therefore we have

$$\mathrm{GNS}_\epsilon(f) \leq \frac{2d\theta}{2\pi} = \frac{d\theta}{\pi}$$

as desired.

We also note the ways in which the above bound can fail to be tight. Firstly, there may be some probability that F changes signs less than $2d$ times on a full circle. Secondly, the integral of $\mathbf{1}_{F(\phi)\neq F(\phi+\theta)}$ may be less than θ times number of times F changes signs if two of these sign changes are within θ of each other. On the other hand it should be noted that if f is the threshold function for a product of d homogeneous linear functions, none of which are scalar multiples of each other, the first case happens with probability 0, and the probability of the second case occurring will necessarily go to 0 as ϵ does. Therefore for such functions our bound is asymptotically correct as $\epsilon \to 0$. □

3. Proof of the Gaussian Surface Area Bounds

The basic idea for our proof of Theorem 1.2 is to relate the Gaussian surface area of f to its noise sensitivity. In particular we claim that:

LEMMA 3.1. *If f is a polynomial threshold function, and if X and Y are independent Gaussians then,*

$$\lim_{\epsilon\to 0} \frac{Pr(f(X) = -1 \text{ and } f(X+\epsilon Y) = 1)}{\epsilon} = \frac{\Gamma(f)}{\sqrt{2\pi}}. \tag{3.2}$$

The basic idea here is that the surface area determines how likely it is that X will lie within some distance of the boundary. In particular we expect that the probability that X is distance t from the boundary (and on the appropriate side) to be roughly $\Gamma(f)dt$. Suppose that we have fixed X. We may assume that X is close to the boundary (since otherwise the probability that $X+\epsilon Y$ lies on the other side is negligible). Since the boundary is smooth (at most points anyway) it is approximately linear at small scales. Therefore the probability that $X + \epsilon Y$ is on the opposite side of the boundary is roughly the probability that the projection of ϵY onto the direction perpendicular to the boundary is more than the distance from X to the boundary. Thus the probability on the left is roughly

$$\int_{\epsilon y > t > 0} \phi(y)\Gamma(f) dt dy = \int_0^\infty \epsilon\Gamma(f) y\phi(y) dy = \frac{\Gamma(f)\epsilon}{\sqrt{2\pi}}.$$

The actual proof is technical and not terribly enlightening and so is put off until the Appendix.

We note that the $\Pr(f(X) = -1$ and $f(X + \epsilon Y) = 1)$ term in Lemma 3.1 is very nearly a noise sensitivity. Unfortunately it fails to be for two reasons. Firstly, we require that $f(X) = -1$ and $f(X + \epsilon Y) = 1$ rather than merely asking that they are unequal. Secondly, this probability fails to be a noise sensitivity since X and $X + \epsilon Y$ are normalized differently. We solve the first of these problems by noting that the asymptotic probability that $f(X) = -1$ and $f(X + \epsilon Y) = 1$ should equal the asymptotic probability that $f(X) = 1$ and $f(X + \epsilon Y) = -1$ (which follows from Lemma 3.1 and Lemma 3.3 below). The second difficulty is overcome by showing that $f(X + \epsilon Y)$ is very likely equal to $f\left(\frac{X+\epsilon Y}{\sqrt{1+\epsilon^2}}\right)$, which is properly normalized. This follows from Lemma 3.4 below.

Lemma 3.3. *For f a polynomial threshold function, $\Gamma(f) = \Gamma(-f)$.*

The idea is that they both measure the surface area of the same boundary between $f^{-1}(1)$ and $f^{-1}(-1)$. We put off the proof until the Appendix.

Lemma 3.4. *If f is a degree d polynomial threshold function in n dimensions, $\epsilon > 0$ and X a random Gaussian variable, then*

$$Pr(f(X) \neq f(X(1 + \epsilon))) \leq d\epsilon\sqrt{\frac{n}{4\pi}}.$$

Proof. First note that by conditioning on the line through the origin that X lies on, we may reduce this problem to the case of a one dimensional distribution. Note that f changes sign at most d times along this line. We need to bound the probability that at least one of these sign changes is in between X and $(1 + \epsilon)X$. It therefore suffices to prove that for any one of these sign changes, that it lies between X and $(1 + \epsilon)X$ with probability at most $\epsilon\sqrt{\frac{n}{4\pi}}$. Note that the probability that X is on the same side of the origin as this sign change is $\frac{1}{2}$. Beyond that, $|X|^2$ satisfies the χ^2 distribution with n degrees of freedom, namely $\frac{1}{2^{n/2}\Gamma(n/2)} x^{n/2-1}e^{-x/2}dx$. Letting $y = \ln(x) = 2\ln(|X|)$ we find that that y has distribution

$$\frac{1}{2^{n/2}\Gamma(n/2)} e^{ny/2}e^{-e^y/2}dy.$$

We want the probability that y is within a particular interval of width $2\ln(1+\epsilon)$. This is at most 2ϵ times the maximum value of the density function. The maximum is achieved when $ny - e^y$ is maximal, or when $y = \ln(n)$. Then the density is

$$\frac{1}{2^{n/2}\Gamma(n/2)} n^{n/2} e^{-n/2} = \sqrt{\frac{n}{4\pi}} \frac{(n/2)^{n/2} e^{-n/2} \sqrt{2\pi(n/2)}}{(n/2)!} \leq \sqrt{\frac{n}{4\pi}}.$$

Multiplying this by d, 2ϵ and $\frac{1}{2}$ (the probability that X is on the same side of 0 as the sign change), we get our bound. □

Notice that this bound should be nearly sharp if the polynomial giving f is a product of terms of the form $|X|^2 - r_i$ for r_i approximately n and spaced apart by factors of $(1+\epsilon)^2$.

We can now prove a bound on a quantity more relevant to Gaussian surface area:

Corollary 3.5. *If f is an n dimensional, degree d polynomial threshold function, $\epsilon > 0$ and X and Y independent Gaussians, then*

$$Pr(f(X) \neq f(X+\epsilon Y)) \leq \frac{d\epsilon}{\pi} + \frac{d\epsilon^2}{4}\sqrt{\frac{n}{\pi}}.$$

Proof. We let $r = \sqrt{1+\epsilon^2}$, $\theta = \arctan(\epsilon)$, and let $Z = \cos(\theta)X + \sin(\theta)Y$ be a normal random variable. Note that $X + \epsilon Y = rZ$. We then have that

$$\Pr(f(X) \neq f(X+\epsilon Y)) \leq \Pr(f(X) \neq f(Z)) + \Pr(f(Z) \neq f(rZ)).$$

By Theorem 1.1 and Lemma 3.4 this is at most

$$\frac{d\theta}{\pi} + d(r-1)\sqrt{\frac{n}{4\pi}} \leq \frac{d\epsilon}{\pi} + \frac{d\epsilon^2}{4}\sqrt{\frac{n}{\pi}}.$$

(since $\theta \leq \tan(\theta) = \epsilon$ and $\sqrt{1+\epsilon^2} \leq 1 + \epsilon^2/2$) □

We are now ready to prove Theorem 1.2

Proof (of Theorem 1.2). The result follows immediately from Lemma 3.1 (for both f and $-f$), Lemma 3.3, and Corollary 3.5 after noting that

$$\Pr(f(X) = -1, f(X+\epsilon Y) = 1) \sim \frac{\epsilon\Gamma(f)}{\sqrt{2\pi}} = \frac{\epsilon\Gamma(-f)}{\sqrt{2\pi}} \sim \Pr(f(X) = 1, f(X+\epsilon Y) = -1)$$

and thus

$$\Gamma(f) = \sqrt{\frac{\pi}{2}} \lim_{\epsilon \to 0} \frac{\Pr(f(X) \neq f(X + \epsilon Y))}{\epsilon} \leq \frac{d}{\sqrt{2\pi}}.$$

□

4. Conclusion

We have shown nearly tight bounds on the Gaussian surface area and noise sensitivity of polynomial threshold functions. One might hope to generalize these results to work for other distributions, such as the uniform distribution on vertices of the hypercube. Unfortunately, several aspects of this proof are difficult to generalize. Perhaps most significantly, we lose the symmetry that allowed us to prove our original result on noise sensitivity. Another difficulty would be in the relation between noise sensitivity and surface area. In our case the two are essentially equivalent quantities of study. On the other hand Adam R. Klivans & Servedio (2008) defined a notion of surface area for the hypercube distribution and proved that for even linear threshold functions there could be a gap between noise sensitivity and surface area of as much as $\Theta(\sqrt{\log(n)})$.

On the other hand, our results still provide hope for a proof of the full Gotsman-Linial Conjecture. For one, we have proved what can be considered to be an important special case of this conjecture. Secondly, many of the results for average sensitivity and Bernoulli noise sensitivity come from generalizing results in the Gaussian setting, and hence there is hope that our results might be a starting point for proofs in these more difficult contexts.

References

Ryan O'Donnell Adam R. Klivans & Rocco A. Servedio (2008). Learning geometric concepts via Gaussian surface area. *Proceedings of the 2008 49th Annual IEEE Symposium on Foundations of Computer Science* 541–550.

Richard Beigel (1993). The polynomial method in circuit complexity. *Proc. of 8th Annual Structure in Complexity Theory Conference* 82–95.

Ryan O'Donnell Elchanan Mossel & Krzysztof Oleszkiewicz (2005). Noise stability of functions with low influences: invariance and optimality. Short version in Proceedings of 46th Annual IEEE Symposium on the Foundations of Computer Science (FOCS). Full manuscript available at http://arxiv.org/abs/math/0412377.

C. Gotsman & N. Linial (1994). Spectral properties of threshold functions. *Combinatorica* **14**(1), 35–50.

R. Servedio I. Diakonikolas, P. Raghavendra & L.-Y. Tan (2009). Average sensitivity and noise sensitivity of polynomial threshold functions. Unpublished. Manuscript available at http://arxiv.org/abs/0909.5011.

Adam R. Klivans & Rocco A. Servedio (2004). Learning DNF in time $2^{O(n^{1/3})}$. *J. Computer and System Sciences* **68**, 303–318.

M. Ledoux (1994). Semigroup proofs of the isoperimetric inequality in Euclidean and Gauss space. *Bull. Sci. Math.* **118**, 485–510.

Ryan ODonnell (2004). Hardness amplification within NP. *J. Computer and System Sciences* **69**, 68–94.

A. Klivans P. Harsha & R. Meka (2009). Bounding the sensitivity of polynomial threshold functions. Unpublished. Manuscript available at http://arxiv.org/abs/0909.5175.

Alexander A. Sherstov (2009). Separating AC0 from depth-2 majority circuits. *SIAM J. Computing* **38**, 2113–2129.

Yaoyun Shi (2000). Lower bounds of quantum black-box complexity and degree of approximating polynomials by influence of Boolean variables. *Inf. Process. Lett.* **75**, 79–83.

Elchanan Mossel Subhash Khot, Guy Kindler & Ryan ODonnell (2007). Optimal inapproxima-bility results for MAX-CUT and other 2-variable CSPs? *SIAM J. Computing* **37**, 319–357.

A. Proofs of Technical Statements about Surface Area

Here we prove some technical results about the Gaussian surface area for polynomial threshold function. In the below, μ denotes the Gaussian measure. Throughout we consider p a non-zero degree d polynomial, with $f = \mathrm{sgn}(p)$.

Lemma A.1. *Let p be a non-constant, degree d polynomial. Then $\mu(p^{-1}([-\epsilon, \epsilon])) = O(\epsilon^{1/d})$, where the implied constant depends on p.*

It should be noted that much more precise versions of this are true. See for example Elchanan Mossel & Oleszkiewicz (2005) Theorem 3.22. We include a simple proof of what we need for completeness.

PROOF. The Lemma follows easily for a monic polynomial of one-variable. This is because p will only be small if x is near at least one of the roots of p. Our idea will be to make a change of variables and consider our polynomial as a function of just one variable to reduce to this case.

By making a generic orthogonal change of variables, we may assume that the leading coefficient of p in terms of x_1 is a constant. By rescaling, we may assume that p is a degree d monic polynomial in $x = x_1$ whose coefficients are polynomials in $y = (x_2, \ldots, x_n)$ (rescaling changes ϵ by a factor that depends on p but nothing else). We have that $p(x) = (x-r_1(y))(x-r_2(y))\cdots(x-r_d(y))$ (here $r_i(y)$ are the roots of $p(x, y)$ thought of as a polynomial in x for fixed y). Therefore if $|p(x, y)| \leq \epsilon$, then x must be within $\epsilon^{1/d}$ of at least one of the $r_i(y)$. Hence after fixing y,

$$\mu\left(\{x : |p(x, y)| \leq \epsilon\}\right) = O(\epsilon^{1/d}).$$

Integrating with respect to y proves the Lemma. $\square$

LEMMA A.2. *Let p be a non-zero degree d polynomial in n variables such that p is not divisible by the square of any polynomial. Let $S = \{x : p(x) = 0\}$. For $\epsilon > \delta > 0$, let $S^\epsilon = \{x \in S : |\nabla p(x)| < \epsilon\}$. Then*

$$\mu((S^\epsilon)_\delta) = O(\epsilon^{1/d^2}\delta).$$

Where $(S^\epsilon)_\delta$ is the set of points within δ of some point of S^ϵ. (again the asymptotic constant here and those throughout the proof depend on p)

PROOF. This proof is considerably more difficult than the previous one. The basic idea is as follows. First, ignoring the condition on ϵ we would like to show that $\mu(S_\delta) = O(\delta)$. Suppose we look at points x' within δ of some point x with $p(x) = 0$ and $\frac{\partial p(x)}{\partial x_1}$ reasonably large. Then it would follow that $\left|\frac{\frac{\partial p(x')}{\partial x_1}}{p(x')}\right| = \Omega(\delta^{-1})$. Considering p as a function of x_1 (with coefficients polynomials in the other x_i), it is not hard to see that the probability of this happening is $O(\delta)$ (since if the derivative is constant sized, you leave the region where p is small after $O(\delta)$ distance). If $\frac{\partial p}{\partial x_1}$ is small, but $\frac{\partial^2 p}{\partial^2 x_1}$ is reasonably large we can play the same trick with $\frac{\partial p}{\partial x_1}$ instead of p. In general we deal with this by splitting S into subsets $S_{k,i}$ where k indexes the first degree of derivative for which p is not small at x, and i indexes a direction at which the k^{th} derivative is reasonably large.

We deal with ϵ by noting that if both p and p' are small then their resolvent polynomial is also small. We combine the above ideas with Lemma A.1 to get the ϵ dependence.

We begin by partitioning S^ϵ into parts. First we pick a large constant C (we will specify how large later). For $k \leq d$ and $x \in \mathbb{R}^n$ we let $|p^{(k)}(x)|$ be the sum of the absolute values of the k^{th} order mixed partial derivatives of p at x. We let S^ϵ_k be the set of $x \in S^\epsilon$ so that $(C\delta)^k|p^{(k)}(x)| \geq (C\delta)^i|p^{(i)}(x)|$ for all $1 \leq i \leq d$.

Next we pick some finite number of orthonormal bases, $B_i = (x_{1,i}, \ldots, x_{n,i})$ so that:

1. When written in basis B_i, p has non-zero x_1^d coefficient.
2. For each k there exists a $C_k > 0$ so that for any x, $|p^{(k)}(x)| \leq C_k \max_i \left|\frac{\partial^k p(x)}{\partial x_{1,i}^k}\right|$.

The first condition holds as long as the B_i are chosen generically. The second condition holds as long as the B_i are chosen generically and there are enough of them. For the second condition it suffices to check that $|q^{(k)}(0)| \leq C_k \max_i \left|\frac{\partial^k q(0)}{\partial x_{1,i}^k}\right|$ for q an arbitrary normalized (for example by making the sum of squares of coefficients equal to 1) homogeneous degree k polynomial (use q equal to the normalized version of the degree k part of the Taylor expansion of p about x, recentered around 0). If we picked sufficiently many generic B_i it will be the case that for any such q, the $\frac{\partial^k q(0)}{\partial x_{1,i}^k}$ will not be 0 for all i. Then there will exist a C_k by compactness. Note that C_k will likely depend on n.

Having picked the B_i, we ensure that $C > 3C_k$ for each k. We then further partition S^ϵ by letting $S^\epsilon_{k,i}$ be the subset of S^ϵ_k so that $|p^{(k)}(x)| \leq C_k \left|\frac{\partial^k p(x)}{\partial x_{1,i}^k}\right|$. We are now prepared to show that

$$\mu\left(\left(S^\epsilon_{k,i}\right)_\delta\right) = O(\delta\epsilon^{1/d^2}).$$

First off we use the coordinates $x = x_{1,i}$ and $y = (x_{2,i}, \ldots, x_{n,i})$. We rescale p so that it is a monic, degree d polynomial in x with coefficients that are polynomials in y. Now consider $(x, y) \in S^\epsilon_{k,i}$. We have that

$$C_k \left|\frac{\partial^k p(x,y)}{\partial x^k}\right| \geq |p^{(k)}(x,y)| \geq (C\delta)^{i-k}|p^{(i)}(x,y)|.$$

(the inequality on the left because $(x, y) \in S^\epsilon_{k,i}$ and the one on the right holds because $(x, y) \in S^\epsilon_k$.) In particular, if (x', y') is within δ of (x, y) then by Taylor's

Theorem:

$$\begin{aligned}\left|\frac{\partial^k p(x',y')}{\partial x^k} - \frac{\partial^k p(x,y)}{\partial x^k}\right| &\leq \sum_{j=1}^{d-k} \frac{\delta^j}{j!}|p^{(k+j)}(x,y)| \\ &\leq \left(C_k \sum_{j=1}^{d-k} \frac{C^{-j}}{j!}\right)\left|\frac{\partial^k p(x,y)}{\partial x^k}\right| \\ &\leq \frac{1}{2}\left|\frac{\partial^k p(x,y)}{\partial x^k}\right|.\end{aligned}$$

Hence $\left|\frac{\partial^k p(x',y')}{\partial x^k}\right| \geq \frac{1}{2}\left|\frac{\partial^k p(x,y)}{\partial x^k}\right|$.

Similarly we have that

$$\begin{aligned}\left|\frac{\partial^{k-1} p(x',y')}{\partial x^{k-1}}\right| &\leq \sum_{j=0}^{d-k+1} \frac{\delta^j}{j!}|p^{(k-1+j)}(x,y)| \\ &\leq C\delta \sum_{j=0}^{d-k+1} \frac{C^{-j}}{j!}\left|\frac{\partial^k p(x,y)}{\partial x^k}\right| \\ &\leq 2C\delta\left|\frac{\partial^k p(x,y)}{\partial x^k}\right|.\end{aligned}$$

Therefore for $(x,y) \in (S^\epsilon_{k,i})_\delta$ we have that

$$\left|\frac{\left(\frac{\partial^k p(x,y)}{\partial x^k}\right)}{\left(\frac{\partial^{k-1} p(x,y)}{\partial x^{k-1}}\right)}\right| \geq \frac{1}{4C\delta}.$$

Letting $g(x,y) = \frac{\partial^{k-1}p(x,y)}{\partial x^{k-1}}$, we have that $\left|\frac{g'(x)}{g(x)}\right| \geq \frac{1}{4C\delta}$. On the other hand if we let $g(x,y) = a_k(x - r_1(y)) \cdots (x - r_{d-k+1}(y))$ (here $r_i(y)$ are the roots of $g(x,y)$ thought of as a polynomial in x for fixed y) we have that $\frac{g'(x)}{g(x)} = \frac{1}{x-r_1(y)} + \cdots + \frac{1}{x-r_{d-k+1}(y)}$. Hence for the above to hold, x must be within $4Cd\delta$ of at least one of the $r_j(y)$. Therefore the measure of the intersection of $(S^\epsilon_{k,i})_\delta$ with the line defined by fixing the value of y is at most $4Cd^2\delta$.

Note also that if $(x,y) \in S^\epsilon$ and if (x',y') is within δ of (x,y) that $|p(x',y')| \leq \epsilon\delta + O(r^d\delta^2)$, and $|p'(x',y')| \leq \epsilon + O(\delta r^d)$, where $r = |(x,y)| + 1$. Considering $p(x,y) = (x - r_1(y)) \cdots (x - r_d(y))$ again as a function of x consider its discriminant $R(y)$. $R(y)$ is a non-zero polynomial of degree at most $d(d-1)$

in y. Furthermore, $R(y)$ can be written as $pq_1 + p'q_2$ for some polynomials q_1 and q_2 of degrees $d(d-2)$ and $d(d-2)+1$. Therefore if $(x', y') \in (S^\epsilon)_\delta$, then $|R(y)| \leq \epsilon r^{d^2}$.

Now the measure of the set of points with $r > \epsilon^{1/d^3}$ is small enough to ignore. Throwing away this region, we find that if we project $\left(S^\epsilon_{k,i}\right)_\delta$ onto its y-coordinate, it covers only points where $|R(y)| \leq \epsilon^{1+1/d}$, which by Lemma A.1 has measure at most $O(\epsilon^{(d+1)/(d^2(d-1))}) = O(\epsilon^{1/d^2})$. Furthermore by the above, once we fix such a y the measure over x of this set is $O(\delta)$. Therefore for each k, i, $\mu((S^\epsilon_{k,i})_\delta) = O(\epsilon^{1/d^2}\delta)$. Hence, summing over all k and i,

$$\mu((S^\epsilon)_\delta) = O(\epsilon^{1/d^2}\delta).$$

□

Proposition A.3. *Let p be a non-zero polynomial. Let $A = \{x : p(x) < 0\}$. Let $S = \partial \bar{A}$. Then*

$$\Gamma(A) = \lim_{\delta \to 0} \frac{\mu(A_\delta \backslash A)}{\delta} = \lim_{\delta \to 0} \frac{\mu(S_\delta)}{2\delta} = \int_S \phi(x) d\sigma(x).$$

Proof. The idea here is that we can use Lemma A.2 to bound the contribution to the volume from x near singular points of S. Away from these points, we can parameterize $A_\delta \backslash A$ by $S \times [0, \delta]$ using $x \to$ (nearest point of S to $x, d(x, S)$). Avoiding the singular set, it is easy to see that this converges to our desired integral.

First we note that we can remove any square factors from p (as removing them only changes $A_\delta - A$ by a set of measure 0 and does not change S), and thus assume that p is a square-free polynomial. By Lemma A.2, when calculating $\mu(A_\delta \backslash A)$ in the computation of $\Gamma(A)$, we may ignore points in $(S^{\delta^{1/6}})_\delta$. We may also ignore points with absolute value more than $r = \log(\delta^{-1})$ since the total measure of the set of such points goes rapidly to 0 as $\delta \to 0$. Let $S_0 = S \backslash S^{\delta^{1/6}}$.

We begin with the following claim. Suppose that $|a| \leq r$ and that b is the closest point in S to a (such a point exists since S is closed and non-empty). Suppose also that $d(a, b) < \delta$ and that $b \in S_0$. Then for δ sufficiently small b is the only point of S within δ of a for which $d(a, b')$ obtains a local minimum.

To prove this claim we may begin by assuming that $b = 0$ and p a polynomial of degree d with coefficients of size at most $O(r^d)$. Let us pick a coordinate system where we describe points as (x, y) where x is the distance from the origin in the direction of $p'(0)$, and y is a coordinate in the orthogonal plane. Hence $p'(0) = (t, 0)$ for some $t > \delta^{1/6}$. Furthermore since b attains a local minimum on S of distance to a, it must be the case that $a = (s, 0)$ for some $\delta > s > 0$.

Now we have that for $c = (x, y)$ within δ of a that $p(c) = tx + O(r^d|c|^2)$ and that $p'(c) = (t, 0) + O(r^d|c|)$. If such a point were to be a local minimum of distance to a on S we would need that $p(c) = 0$ and $p'(c)$ is parallel to $a - c$. The first implies that $x = O(r^d\delta^{-1/6}y^2)$. The latter implies that

$$\frac{O(r^d|y|)}{|y|} = \frac{t + O(r^d|y|)}{s}.$$

But this is impossible since $t \gg sr^d + |y|r^d$.

Now for $x \in S_0$ let $n(x) = \frac{p'(x)}{|p'(x)|}$ be the outward pointing normal vector to x. The above claim along with Lemma A.2 says that $\mu(A_\delta \backslash A)$ is $o(\delta)$ plus the volume of the region parameterized by $x + tn(x)$ for $x \in S_0, t \in [0, \delta]$. Hence we have that

$$\begin{aligned}
&\mu(A_\delta \backslash A) \\
&\quad = \int_{S_0} \int_0^\delta \phi(x + tn(x)) |\det(J(x, t))| dt d\sigma(x) + o(\delta)
\end{aligned}$$

where $J(x, t)$ is the Jacobian of our parametrization $(x, t) \to x + tn(x)$. We first claim that $J(x, t)$ is $O(r^{3d}\delta^{1/2})$ plus the matrix sending a tangent vector to S to itself and sending dt to $n(x)$. This is because the t derivative is exactly $n(x)$, and the derivative with respect to x is $I + t\frac{\partial n}{\partial x}$. But $|t| \leq \delta$ and $\frac{\partial n}{\partial x}$ is a degree $3d$ polynomial divided by $|p'|^3$, and therefore is $O(r^{3d}\delta^{-1/2})$, hence multiplying by t we get an error of size $O(r^{3d}\delta^{1/2})$. Hence $|\det(J(x, t))| = 1 + O(r^{3d}\delta^{1/2})$. Furthermore, $\phi(x + tn) = \phi(x)(1 + O(r\delta))$. Hence we have that:

$$\begin{aligned}
\mu(A_\delta \backslash A) &= \int_{S_0} \int_0^\delta \phi(x)(1 + O(r^{3d}\delta^{1/2})) dt d\sigma(x) + o(\delta) \\
&= \int_{S_0} \int_0^\delta \phi(x) dt d\sigma(x)(1 + O(r^{3d}\delta^{1/2})) + o(\delta) \\
&= \delta \int_{S_0} \phi(x) d\sigma(x) + o(\delta).
\end{aligned}$$

Finally noting that

$$\lim_{\delta \to 0} \int_{S_0} \phi(x) d\sigma(x) = \int_S \phi(x) d\sigma(x)$$

proves the first two equalities.

The last equality comes from this after noting that if $B = \{x : p(x) > 0\}$, then

$$S_\delta = (A_\delta \backslash \bar{A}) \cup (B_\delta \backslash \bar{B}).$$

□

Lemma 3.3 now follows immediately

PROOF (of Lemma 3.3). Let $f = \text{sgn}(p(x))$. Let $A = \{x : p(x) > 0\}, A' = \{x : p(x) < 0\}$ and $S = \partial\bar{A} = \bar{A} \cap \bar{A}' = \partial\bar{A}'$. Then by Proposition A.3 we have that

$$\begin{aligned}\Gamma(f) &= \lim_{\delta\to 0} \frac{\mu(A_\delta \backslash A)}{\delta} \\ &= \int_S \phi(x) d\sigma(x) \\ &= \lim_{\delta\to 0} \frac{\mu(A'_\delta \backslash A')}{\delta} \\ &= \Gamma(-f).\end{aligned}$$

□

PROOF (of Lemma 3.1). Again we use Lemma A.2 to show we can afford to ignore the case where X is near singular points of p. We use a parametrization of X near the zero set of p similar to that in the proof of Proposition A.3, and derive a rigorous version of the intuition we had before.

First recall that if $A = f^{-1}(1)$, then

$$\Gamma(f) = \lim_{\epsilon\to 0} \frac{\mu(A_\epsilon \backslash A)}{\epsilon}.$$

Next note that since the probability that $|\epsilon Y| > \epsilon^{1-1/12d^2}$ goes rapidly to 0 as $\epsilon \to 0$, we can throw away all cases where X is not within $\epsilon^{1-1/12d^2}$ of $S = \partial A$ from the left hand side of Equation (3.2).

Define S_0 as in the proof of Proposition A.3. By Lemma A.2 we have that X is within $\epsilon^{1-1/12d^2}$ of $(S\backslash S_0)$ with probability at most $O(\epsilon^{1+1/12d^2})$ and so we may throw these events away. Hence we may restrict ourselves to considering only X whose nearest neighbor in S lies in S_0 and is at distance at most $\epsilon^{1-1/12d^2}$. We may also assume that $|X| < \log(\epsilon^{-1})$.

Letting n be the derivative of p at the point of S nearest to X, we have that $p(X + \epsilon Y) = p(X) + \epsilon n \cdot Y + O(r^d\epsilon^2 Y^2)$. For Y in the range we are considering, $O(r^d\epsilon^2 Y^2) = O(r^d\epsilon^{3/2})$. Hence $p(X) = d(X, S)|n| + O(\epsilon^{3/2})$. Hence

$p(X+\epsilon Y) = |n| \left(d(X,S) + \epsilon \left(\frac{n}{|n|}\right) \cdot Y\right) + O(\epsilon^{3/2})$. Note that up to $O(\epsilon^{3/2})$, the sign of this only depends on the dot product of Y with a unit vector and on $d(X,S)$. On the other hand since $|n|\epsilon \gg \epsilon^{3/2}$, the probability that $p(X+\epsilon Y)$ is negative is within $\epsilon^{1/3}$ of the probability that a random one dimensional normal is more than $\frac{d(X,S)}{\epsilon}$.

Hence

$$\begin{aligned}
&\Pr(f(X) = -1 \text{ and } f(X+\epsilon Y) = 1) \\
=&o(\epsilon) + O(\epsilon^{1/3})\Pr(d(X,S) \leq \epsilon^{1-1/12d^2}) \\
&+ \Pr(f(X) = -1 \text{ and } d(X,S) < \epsilon N) \\
=&o(\epsilon) + \Pr(f(X) = -1 \text{ and } d(X,S) < \epsilon N).
\end{aligned}$$

where N is a random normal variable.

After fixing the value of N this probability is just $\mu(A_{\epsilon N} \backslash A)$. Hence integrating we get

$$\begin{aligned}
&\int_0^\infty \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \mu(A_{\epsilon N} \backslash A) dx \\
&= \int_0^\infty \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \Gamma(f) \epsilon x (1 + o(1)) dx \\
&= \frac{\Gamma(f)\epsilon}{\sqrt{2\pi}} + o(\epsilon).
\end{aligned}$$

This completes our proof. □

Acknowledgements

The author is currently supported by an NDSEG fellowship. The author would also like to thank Ilias Diakonikolas for suggesting this problem to him.

Manuscript received 5/3/2010

Daniel M. Kane