Influences in low-degree polynomials

Artūrs Bačkurs

December 12, 2012

1 Introduction

In [3] it is conjectured that every bounded real polynomial has a "highly influential" variable. The conjecture is known as Aaronson-Ambainis conjecture. In this survey we describe recent results that resolves some special cases of the conjecture. We also describe some consequences in quantum computing assuming the conjecture.

2 Aaronson-Ambainis Conjecture

Let $p : \mathbb{R}^n \to \mathbb{R}$.

Definition 1

$$\mathrm{Var}(p) := \mathrm{E}_{x,y \in \{-1,1\}^n} \left[(p(x) - p(y))^2 \right].$$

Definition 2 *The influence of the* i^{th} *variable is*

$$\mathrm{Inf}_i(p) := \mathrm{E}_{x \in \{-1,1\}^n} \left[(p(x) - p(x^i))^2 \right],$$

where x^i *is* x *with the* i^{th} *bit flipped.*

Now we state Aaronson-Ambainis conjecture:

Conjecture 1 *Suppose that* p *is degree-*d *polynomial and* $|p(x)| \leq 1$ *for all* $x \in \{-1,1\}^n$. *Then there exists an* $i \in [n]$ *such that* $\mathrm{Inf}_i(p) \geq (\mathrm{Var}(p)/d)^{O(1)}$.

3 The Boolean case

In [4] the Boolean case $p : \{-1,1\}^n \to \{-1,1\}$ of the conjecture is solved. First we define decision tree as in [5]:

Definition 3 *A decision tree is a binary tree* T. *Each internal node of* T *is labeled with a variable* x_i *and each leaf is labeled with a value* -1 *or* 1. *Given an input* $x \in \{-1,1\}^n$, *the tree is evaluated as follows. Start at the root; if this is a leaf then stop. Otherwise, query the value of the variable* x_i. *If* $x_i = -1$ *then recursively evaluate the left subtree, if* $x_i = 1$ *then recursively evaluate the right subtree. The output of the tree is the value (*-1 *or* 1*) of the leaf that is reached eventually. Note that an input* x *deterministically determines the leaf, and thus the output, that the procedure ends up in.*

We say a decision tree computes p *if its output equals* $p(x)$, *for all* $x \in \{-1,1\}^n$. *We define* $D(p)$, *the decision tree complexity of* p, *as the depth of an optimal (= minimal-depth) decision tree that computes* p.

Theorem 1 *[4] Let T be a decision tree computing a function $p : \{-1,1\}^n \to \{-1,1\}$. Then*

$$\mathrm{Var}(f) \leq \sum_{i=1}^{n} \delta_i(T)\,\mathrm{Inf}_i(f),$$

where

$$\delta_i(T) = \Pr_{x \in \{-1,1\}^n} [T \text{ queries } x_i] .$$

Corollary 1 *[4] For every $p : \{-1,1\}^n \to \{-1,1\}$ we have*

$$\mathrm{Inf}_{\max}(p) \geq \frac{\mathrm{Var}(p)}{D(p)}.$$

Proof: Let T be the optimal decision tree computing p. From Theorem 1,

$$\mathrm{Var}(f) \leq \sum_{i=1}^{n} \delta_i(T)\,\mathrm{Inf}_i(f)$$

$$\leq \mathrm{Inf}_{\max}(p) \sum_{i=1}^{n} \delta_i(T) \leq \mathrm{Inf}_{\max}(p) D(p),$$

where the last inequality follows from

$$\sum_{i=1}^{n} \delta_i(T) = \mathrm{E}_{x \in \{-1,1\}^n} [\text{number of coordinates } T \text{ queries on } x] .$$

■

Since $D(p) \leq O(\deg(p)^4)$ (by a result of Nisan and Smolensky [5]), Corollary 1 implies that the maximum influence satisfies $\mathrm{Inf}_{\max}(p) \geq \Omega(\mathrm{Var}[p]/\deg(p)^4)$.

4 The case of symmetric function

Let $p : R^n \to R$ be a symmetric polynomial ($p(x)$ depends only on the Hamming weight of x) of degree d and $|p(x)| \leq 1$ for all $x \in \{-1,1\}^n$. Let $\mathrm{Var}(p) = \epsilon$, where $\epsilon = o(1)$ (proofs are similar in the case $\epsilon = \Omega(1)$).

We write $p(x) = \sum_{S \subseteq [n]} \widehat{p}(S) \chi_S(x)$, where $\chi_S(x) = \prod_{i \in S} x_i$. This is called the Fourier expansion of p. We define $\|p\|_2 := \sqrt{\sum_{S \subseteq [n]} \widehat{p}(S)^2}$. It can be verified that $\|p\|_2^2 = \mathrm{E}_x[p^2(x)]$ and $\widehat{p}(S_1) = \widehat{p}(S_2)$ if $|S_1| = |S_2|$ for symmetric p.

Theorem 2 *For all i,*

$$\mathrm{Inf}_i(p) = \Omega\left(\frac{\epsilon^3}{\deg^4(p)\ln(1/\epsilon)}\right).$$

Proof: It can be shown that $\mathrm{Inf}_i(p) = 4\sum_{S : i \in S} \widehat{p}^2(S)$. Thus

$$\sum_i \mathrm{Inf}_i(p) = 4 \sum_{S \subseteq [n]} |S| \widehat{p}^2(S) \geq 4 \sum_{S : S \neq \emptyset} \widehat{p}^2(S) = 4\left(\|p\|_2^2 - \widehat{p}^2(\emptyset)\right) = 2\,\mathrm{Var}(p) = 2\epsilon,$$

where the second to last equality follows from $2\left(\|p\|_2^2 - \widehat{p}^2(\emptyset)\right) = \mathrm{Var}(p)$. Now it suffices to prove $\frac{1}{n} = \Omega\left(\frac{\epsilon^2}{\deg^4(p)\ln(1/\epsilon)}\right)$.

From inequality $2\,\mathrm{E}_{x,y}[|p(x)-p(y)|] \geq \mathrm{Var}(p)$ we get $\mathrm{E}_{x,y}[|p(x)-p(y)|] \geq \epsilon/2$.

We choose A_n such that

$$\Pr_x[||x|-n/2| > A_n] = \epsilon/16,$$

which gives

$$\mathrm{E}_{\substack{x:||x|-n/2|\leq A_n \\ y:||y|-n/2|\leq A_n}}[|p(x)-p(y)|] = \Omega(\epsilon).$$

By Chernoff bound, $A_n = O\left(\sqrt{\ln(1/\epsilon)n}\right)$.

Let $g(k) = \mathrm{E}_{x:|x|=k}[p(x)]$. Now

$$\max_x[|g'(x)|] = \Omega\left(\frac{\epsilon}{\sqrt{\ln(1/\epsilon)n}}\right).$$

By Markov brothers' inequality,

$$\deg(p) \geq \sqrt{\frac{n\max_{x\in[0,n]}[|g'(x)|]}{2\max_{x\in[0,n]}[|g(x)|]}}$$

$$= \Omega\left(\min\left(\sqrt{n}, \sqrt{\epsilon}\left(\frac{n}{\ln(1/\epsilon)}\right)^{1/4}\right)\right) = \Omega\left(\sqrt{\epsilon}\left(\frac{n}{\ln(1/\epsilon)}\right)^{1/4}\right).$$

■

5 Exponential lower bound

First we define junta of boolean function:

Definition 4 *A function $p : \{-1,1\}^n \to \mathbb{R}$ is called an (δ, j)-junta if there exists a function $g : \{-1,1\}^n \to \mathbb{R}$ depending on at most j coordinates such that $\|p-g\|_2^2 \leq \delta$.*

In [6] the following result is given:

Theorem 3 *[6] Let $p : \{-1,1\}^n \to [-1,1]$, $k \geq 1$, and $\delta > 0$. Suppose*

$$\sum_{|S|>k} \widehat{p}(S)^2 \leq \exp(-O(k^2\log k)/\delta).$$

Then p is an $(\delta, 2^{O(k)}/\delta^2)$-junta.

Now we prove exponential version of Aaronson-Ambainis conjecture:

Theorem 4 *Suppose that p is degree-d polynomial and $|p(x)| \leq 1$ for all $x \in \{-1,1\}^n$. Then there exists an $i \in [n]$ such that $\mathrm{Inf}_i(p) \geq (\mathrm{Var}(p)/2^d)^{O(1)}$.*

Proof: In the proof of Theorem 3, the junta of p is constructed as

$$g = \sum_{|S|\leq k, S\subseteq J} \widehat{p}(S)\chi_S,$$

where J is a set of variables that g depends on. Because the influence of a variable is a sum of squares of Fourier coefficients, it suffices to prove a lower bound on $\mathrm{Inf}_{\max}(g)$. From [4] (Corollary 3.4) we get that $\mathrm{Inf}_{\max}(g) \geq \frac{\mathrm{Var}(g)}{|J|^2}$. We apply Theorem 3 with $k = \deg p$. Because $\sum_{|S|>\deg p}\widehat{p}(S)^2 = 0$, we can choose δ arbitrarily small. Choosing $\delta = c\,\mathrm{Var}(p)^2$ for sufficiently small c gives the result. ■

6 Proof sketch for Theorem 3

We give a sketch of the proof from [6].

In the proof of Theorem 3 the following two lemmas are used:

Lemma 1 *[2] There is a universal constant K such that the following holds: Let $l(x) = \sum_{i=1}^{n} a_i x_i$, where the a_i's satisfy $\sum a_i^2 = 1$ and the x_i's are independent and uniform random variables taking value from $\{-1, 1\}$. Let $t \geq 1$ and suppose that $|a_i| < \frac{1}{Kt}$ for all i. Then*

$$\Pr[|l(x)| > t] \geq \exp(-Kt^2).$$

Lemma 2 *[6] There is a universal constant K' such that the following holds: Suppose $p : \{-1, 1\}^n \to \mathbb{R}$ has degree at most k and $\sum_{i=1}^{n} \widehat{p}(i)^2 \geq 1$. Let $t \geq 1$ and suppose that $|\widehat{p}(i)| < \frac{1}{K'tk}$ for all i. Then*

$$\Pr[|p| \geq t] \geq \exp(-K't^2k^2).$$

Lemma 1 shows a supergaussian estimate for linear functions whose coefficients are all small. Lemma 2 shows that the supergaussian behaviour is maintained even if terms of degree at most k are added to the function.

The proof of Lemma 2 first considers the linear part of p. According to Lemma 1, when evaluated on a random point x_0, the linear part has a non-negligible probability of obtaining a large value. It may be that $p(x_0)$ is still small because of cancellations contributed by the non-linear part. To evade this cancellation random noise with rate ρ is introduced and the value of $T_\rho p(x_0) = \sum_{S \subseteq [n]} \rho^{|S|} \widehat{p}(S) \chi_S(x_0)$ is considered. Some extremal properties of the Chebyshev polynomials are used to show the existence of successfull noise rate.

In the proof of Theorem 3 slightly modified version of Lemma 2 is used.

Lemma 3 *[6] There is a universal constant K such that the following holds: Suppose $p : \{-1, 1\}^n \to \mathbb{R}$ has degree at most k. Let $T \subseteq [n]$, and let $t \geq 1$. Suppose that $\sum_{i \in T} \widehat{p}(i)^2 \geq 1$ and that $|\widehat{p}(i)| < \frac{1}{Ktk}$ for all $i \in T$. Then*

$$\Pr[|p| \geq t] \geq \exp(-Kt^2k^2).$$

The proof of Lemma 3 is almost the same as of Lemma 2.

Now Lemma 3 is used to prove the following Theorem.

Theorem 5 *[6] There is a universal constant C such that the following holds: Suppose $p : \{-1, 1\}^n \to \mathbb{R}$ has degree at most k and assume $\sum_{S \neq \emptyset} \widehat{p}(S)^2 = 1$. Let $t \geq 1$ and suppose that $\mathrm{Inf}_i(p) \leq t^{-2}C^{-k}$ for all i. Then*

$$\Pr[|p| \geq t] \geq \exp(-Ct^2k^2 \log k).$$

In the proof of Theorem 5 s is chosen such that the weight of the Fourier transform of p on levels between 2^{s-1} and 2^s is at least $1/\log k$. By performing an appropriate random restriction which fixes many of the coordinates of p, a function is obtained where much of this weight is brought down to the first level (with non-negligible probability). Since p has very small influences, the restricted function has to have small first level Fourier coefficients (hypercontractive of [1] is used to control the amount by which the coefficients deviate from their expectation). When it happens, Lemma 3 is applied to the restricted function.

In the proof of Theorem 3 the following modified version of Theorem 5 is used.

Theorem 6 *[6] There is a universal constant C such that the following holds: Suppose $p : \{-1, 1\}^n \to \mathbb{R}$ has degree at most k, $j \subseteq [n]$, and assume $\sum_{S \setminus J \neq \emptyset} \widehat{p}(S)^2 \geq \epsilon$. Let $t \geq \sqrt{\epsilon}$ and suppose that $\mathrm{Inf}_i(p) \leq \epsilon^2 t^{-2} C^{-k}$ for all $i \notin J$. Then*

$$\Pr[|p| \geq t] \geq \exp(-(Ct^2k^2 \log k)/\epsilon).$$

In the proof of 3 function p is approximated by truncating its Fourier expansion, thus getting a low degree polynomial. Now there are two possibilities: either this polynomial depends on few coordinates or it has a significant component contributed by all coordinates with small influence. In the later case we use 6 and show that the initial polynomial is not bounded by -1 and 1.

7 Consequences of Aaronson-Ambainis conjecture

We will use the fact that the acceptance probability of the quantum query algorithm is $2T$-degree multilinear polynomial in input variable, where T is the number of queries made by quantum algorithm.

Theorem 7 *Suppose Conjecture 1 holds. Let Q be a quantum algorithm that makes T queries to a Boolean input $X = (x_1, ..., x_N)$, and let $\epsilon > 0$. Then there exists a deterministic classical algorithm C that makes $\mathrm{poly}(T, 1/\epsilon, 1/\delta)$ queries to the x_i's, and that approximates Q's acceptance probability to within an additive arror ϵ on a $1 - \delta$ fraction of inputs.*

Classical algorithm C is as follows. We query additional variable with highest influence (that exists assuming Conjecture 1) as long as the polynomial of the acceptance probability has large variance. After querying each additional variable we update polynomial with the polynomial induced by the answer.

Let $\mathrm{D}_\epsilon(f)$ be the minimum number of queries made by a deterministic algorithm that evaluates f on at least $1 - \epsilon$ fraction of inputs. Similarly define $\mathrm{Q}_\epsilon(f)$.

Theorem 8 *Suppose Conjecture 1 holds. Then $\mathrm{D}_{\epsilon+\delta}(f) \leq (\mathrm{Q}_\epsilon(f)/\delta)^{O(1)}$ for all Boolean functions f and all $\epsilon, \delta > 0$.*

We run C from Theorem 7 on the polynomial p (which is the acceptance probability of Q_ϵ) until we obtain estimate $\tilde{p}$ of p such that

$$\Pr_{x \in \{-1,1\}^n} \left[|\tilde{p}(x) - p(x)| \leq \frac{1}{10} \right] \geq 1 - \delta.$$

Output $f(x) = 1$ if $\tilde{p}(x) \geq 1/2$ and $f(x) = -1$ otherwise. By Theorem 7, this requires $\mathrm{poly}(T, 1/\delta)$ queries to x, and by the union bound it successfully computes f on at least $1 - \epsilon - \delta$ fraction of inputs.

AvgP is the class of languages for which there exists a polynomial-time algorithm that solves a $1 - o(1)$ fraction of instances.

Theorem 9 *Suppose Conjecture 1 holds. Then $\mathrm{P} = \mathrm{P}^{\#\mathrm{P}}$ implies $\mathrm{BQP}^A \subset \mathrm{AvgP}^A$ with probability 1 for a random oracle A.*

It follows that separating BQP from AvgP relative to a random oracle would as hard as separating complexity classes in the unrelativized world. In the proof of Theorem 9 it is shown that there exists a deterministic polynomial-time classical algorithm such that for all quantum algorithms and $x \in \{-1, 1\}^n$,

$$\Pr_A \left[|\tilde{p}_x(A) - p_x(A)| > \frac{1}{10} \right] < \frac{1}{n^3},$$

where $\tilde{p}_x(A)$ is the output of the classical algorithm given input x and oracle A. $p_x(A)$ is the probability of acceptance of x of the quantum algorithm given oracle A. Suppose we had an extremely strong variant of Aaronson-Ambainis conjecture, one that implied something like

$$\Pr_A \left[|\tilde{p}_x(A) - p_x(A)| > \frac{1}{10} \right] < \frac{1}{\exp(n)}.$$

Then we could eliminate the neet for AvgP in Theorem 9, and show that $\mathrm{P} = \mathrm{P}^{\#\mathrm{P}}$ implies $\mathrm{P}^A = \mathrm{BQP}^A$ with probability 1 for a random oracle A.

Theorem 3 have the following corollaries.

Corollary 2 *Suppose a quantum algorithm makes T queries to a Boolean input $x \in \{0,1\}^n$. Then for all $\alpha, \delta > 0$, we can approximate the acceptance probability to within an additive constant α, on a $1-\delta$ fraction of inputs, by making $\frac{2^{O(T)}}{\alpha^4 \delta^4}$ deterministic classical queries.*

Corollary 3 $\mathrm{D}_{\epsilon+\delta}(f) \leq 2^{O(Q_\epsilon(f))}/\delta^4$ *for all Boolean functions f and all $\epsilon, \delta > 0$.*

Acknowledgments

The author would like to thank Scott Aaronson for helpful discussions and teaching a great class.

References

[1] W. Beckner. Inequalities in Fourier analysis. *Annals of Mathematics*, 102:159-182, 1975.

[2] Michel Ledoux and Michel Talagrand. *Probability in Banach Spaces.* Springer, 1991.

[3] Scott Aaronson and Andris Ambainis. The need for structure in quantum speedups. In *Proceedings of the 2nd Annual Symposium on Innovations in Computer Science*, 2011.

[4] Ryan ODonnell, Michael Saks, Oded Schramm, and Rocco Servedio. Every decision tree has an inuential variable. In *Proceedings of the 46th Annual IEEE Symposium on Foundations of Computer Science*, pages 3139, 2005.

[5] H. Buhrman and R. de Wolf. Complexity measures and decision tree complexity: a survey. *Theoretical Computer Science*, 288(1):2143, 2002.

[6] Irit Dinur, Ehud Friedgut, Guy Kindler, and Ryan O'Donnell. On the Fourier tails of bounded functions over the discrete cube. *Israel Journal of Mathematics*, 160(1):389412, 2007.