On Boolean functions with low sensitivity

Hing Yin Tsang

September 5, 2014

Abstract

We review several complexity measures for Boolean functions which are related to the study of the decision tree model. Among those measures, sensitivity is the simplest to define, but the most difficult to prove upper bounds in terms of it. We discuss the known upper bounds on other measures in terms of sensitivity and the proof techniques. By a simple decision tree construction, we suggest two conjectures that, if both are true, would imply a polynomial upper bound on decision tree complexity in terms of sensitivity.

1 Introduction

The complexity of a Boolean function can be studied in many ways by considering different resources as the measures of complexity. The decision tree model appears naturally when we are interested in the number of bits we need to look at x in order to compute $f(x)$. The decision tree complexity $D(f)$ is the depth of the shortest tree that computes f. Although in principle we can determine it by enumerating all possible decision trees that computes f, it is more desirable to study the relation between $D(f)$ and other complexity measure so that we can have a better understanding on the power and limitation of decision tree model.

A number of complexity measures are known to be closely related to the decision tree complexity. Some of them are motivated by the study of other computation models. Sensitivity was introduced by Cook and Dwork [CD82], and Reischuk [Rei82] to give a lower bound on CREW PRAM. Nisan introduced a variant of sensitivity known as the block sensitivity in [Nis91] and showed that it captures the complexity of CREW PRAM. Now we have a fairly complete understanding on the decision tree complexity. We know that block sensitivity, certificate complexity, Fourier degree and approximate degree are polynomially related to decision tree complexity. And allowing the decision tree to flip a coin or to use quantum mechanics does not increase its power by much. However, the picture is still incomplete and the only missing puzzle is the sensitivity. Among the measures we just mentioned, sensitivity is arguably one of the oldest and also one of the simplest to define. But three decades after it is introduced, we still know very little about it. By the current understanding, it is completely possible that sensitivity is logarithmic in block sensitivity.

The *sensitivity vs. block sensitivity* problem of Nisan and Szegedy [NS94] asks what is the correct relation between them. In particular, the *Sensitivity Conjecture* says that sensitivity and block sensitivity are actually polynomially related. The progress on the Sensitivity Conjecture has been very limited. The first general upper bound on any measures we mentioned in terms of sensitivity was observed by Simon [Sim83]. His exponentially upper bound has been the best known upper bound for over twenty years until Kenyon and Kutin [KK04] improved the exponent by using different argument. The next improvement is due to Ambainis et al. [ABG$^+$14] using a combinatorial argument similar to Simon's proof. They show $bs(f) = O(2^{s(f)}s(f))$, and it is the currently best known upper bound on block sensitivity in terms of sensitivity.

For the separation between sensitivity and block sensitivity, No super-quadratic gap between them is known. The first quadratic gap is obtained by Rubinstein [Rub95] and the currently best separation is due to Ambainis and Sun [AS11], which is also quadratic.

1.1 Organization

In Section 2, we introduce the notation used in this paper and define the complexity measures we will discuss in the remaining sections. In Section 3, we discuss in detail the known upper bounds on block sensitivity and certificate complexity in terms of sensitivity. Finally, in Section 4, we give a simple decision tree construction that suggests two combinatorial conjectures about the structure of low-sensitivity functions as the first step to attack the Sensitivity Conjecture.

2 Complexity measures

Notation and defintions. Given $x \in \{0,1\}^n$ and $B \subseteq [n]$, we denote x^B the point that differs from x at the i-th position for each $i \in B$, i.e. $(x^B)_j = x_j$ if $j \notin B$ and $(x^B)_j = 1 - x_j$ if $j \in B$. When $B = \{i\}$ is a singleton, we write x^i instead of $x^{\{i\}}$. The *n-dimensional Boolean cube* is the graph $G = (V, E)$ with $V = \{0,1\}^n$ and $E = \{\{x, y\} \subseteq V \times V \mid d(x, y) = 1\}$, where $d(x, y) = |\{i \in [n] \mid x_i \neq y_i\}|$ is the Hamming distance between x and y. A subgraph of the n-dimensional Boolean cube is called a *subcube* if it is a subgraph induced by the vertex set $Q = \{x \in \{0,1\}^n \mid x_{i_1} = b_i, \ldots, x_{i_k} = b_k\}$ for some $i_j \in [n]$ and $b_j \in \{0,1\}$, where k is called the *codimension* of the subcube. We will abuse the notation to denote both the subcube and its vertex set by Q.

Let $f : \{0,1\}^n \to \{0,1\}$ be a Boolean function. We now define the complexity measures used in the later sections.

Definition 2.1 (**sensitivity**)**.** We say that f is *sensitive to the i-th variable at x* if $f(x) \neq f(x^i)$. The *sensitivity $s(f, x)$ of f at x* is the number of variables that f is sensitive to each of them at x. The *sensitivity of f* is the maximum sensitivity of f over all x, i.e. $s(f) = \max_{x \in \{0,1\}^n} s(f, x) = \max_{x \in \{0,1\}^n} |\{i \in [n] \mid f(x) \neq f(x^i)\}|$. For $b \in \{0,1\}$, the *b-sensitivity $s^b(f)$ of f* is $\max_{x \in f^{-1}(b)} s(f, x)$.

Definition 2.2 (**influence**)**.** The *influence $\mathbf{Inf}_i[f]$ of the i-th variable* is the probability that f is sensitive to the i-th variable at uniformly random x, i.e. $\mathbf{Inf}_i[f] = \Pr_x[f(x) \neq f(x^i)]$. The *total influence (or average sensitivity) of f* is $\mathbf{Inf}[f] = \mathbb{E}_x[s(f, x)] = \sum_{i=1}^n \mathbf{Inf}_i[f]$.

Definition 2.3 (**block sensitivity**)**.** We say that f is *sensitive to the block $B \subseteq [n]$ at x* if $f(x) \neq f(x^B)$. The *block sensitivity $bs(f, x)$ of f at x* is the largest integer r such that there exists r disjoint blocks $B_1, \ldots, B_r$ of $[n]$ such that f is sensitive to each of them at x. The *block sensitivity of f* is $bs(f) = \max_{x \in \{0,1\}^n} bs(f, x)$. For $b \in \{0,1\}$, the *b-block sensitivity $bs^b(f)$ of f* is $\max_{x \in f^{-1}(b)} bs(f, x)$.

One useful observation about block sensitivity is that for any x, we can always assume each block in the family of disjoint sensitive blocks $\mathcal{B}$ at x have size at most $s(f)$. Indeed, for a sensitive block B at x, since $f(x) \neq f(x^B)$, we can take the minimal size $B' \subseteq B$ such that $f(x) \neq f(x^{B'})$. By minimality of B', $f((x^{B'})^i) \neq f(x^{B'})$ for each $i \in B'$. It follows that f is sensitive to each variable in B' at $x^{B'}$. So, the size of a minimal sensitive block is at most $s(f)$. This fact would be useful in the proofs in later sections.

For any f, $\mathbf{Inf}[f] \leq s(f) \leq bs(f)$. And the gap between $\mathbf{Inf}[f]$ and $s(f)$ can be arbitrarily large. For instance, let $f = \text{AND}_n$. Then $\mathbf{Inf}[f] = \frac{n}{2^{n-1}}$ but $s(f) = n$. For $s(f)$ and $bs(f)$, the largest known gap is quadratic, i.e. there exists f such that $bs(f) = \Omega(s(f)^2)$

Definition 2.4 (**certificate complexity**)**.** The *certificate complexity $C(f, x)$ of f at x* is the smallest integer m such that there exists a subcube Q of codimension m such that $x \in Q$ and f is a constant on Q. The *certificate complexity $C(f)$ of f* is $\max_{x \in \{0,1\}^n} C(f, x)$. The *minimum certificate $C_{\min}$ of f* is $\min_{x \in \{0,1\}^n} C(f, x)$. For $b \in \{0,1\}$, the *b-certificate complexity $C^b(f)$ of f* is $\max_{x \in f^{-1}(b)} C(f, x)$.

In the above definition, we consider subcube on which f is a constant. Any subcube with this property is called the *monochromatic subcube of f*.

Note that $C_{\min}(f) \leq C(f)$ by definition. And the gap can be arbitrarily large. For instance, let $f = \text{AND}_n$, then $C_{\min}(f) = 1$ but $C(f) = n$.

For any real-valued function f defined on $\{0,1\}^n$, by interpolation, one can always construct a multilinear polynomial p over real such that $p(x) = f(x)$ for all $x \in \{0,1\}^n$. If we further restrict f to be a Boolean function with range $\{0,1\}$, then we can actually construct in a similar way a multilinear polynomial over *any* field $\mathbb{F}$ such that $p(x) = f(x)$ for all $x \in \{0,1\}^n$ (we consider 0 and 1 as elements in $\mathbb{F}$ when evaluating p). We say p is the *$\mathbb{F}$-polynomial representation (or $\mathbb{F}$-representation) of f*. It is not difficult to see that such (multilinear) polynomial is unique and hence the following definitions are well-defined.

Definition 2.5 (**degree**)**.** The *Fourier degree* $\deg(f)$ *of* f is the degree of the real polynomial representation of f. The $\mathbb{F}_2$*-degree* $\deg_{\mathbb{F}_2}(f)$ *of* f is the degree of the $\mathbb{F}_2$-polynomial representation of f.

It is not difficult to see that $\deg_{\mathbb{F}_2}(f) \leq \deg(f)$ for any Boolean function f. Indeed, one can obtain the $\mathbb{F}_2$-polynomial representation of f by first finding the $\mathbb{R}$-polynomial representation of f and then taking the coefficients module 2. Note that the gap between them can be arbitrarily large. For instance, let $f = \text{PARITY}_n$. Then $\deg_{\mathbb{F}_2}(f) = 1$ but $\deg(f) = n$.

A *decision tree* is a rooted binary tree $\mathcal{T}$ such that each internal node is labelled by a variable x_i for some $i \in [n]$ and each leaf is labelled by 0 or 1. Given $x \in \{0,1\}^n$, $\mathcal{T}$ computes $\mathcal{T}(x)$ as follows. Start at the root. If it is a leaf then output the value of this leaf, otherwise query the variable labeling the current node. Recursively compute the left and right subtree if the value of the query is 0 and 1 respectively. We say that $\mathcal{T}$ *computes* f if $\mathcal{T}(x) = f(x)$ for all $x \in \{0,1\}^n$. The *depth* $\text{depth}(\mathcal{T})$ *of* $\mathcal{T}$ is the number of internal nodes in the longest root-to-leaf path in $\mathcal{T}$.

Definition 2.6 (**deterministic decision tree complexity**)**.** The *(deterministic) decision tree complexity* $D(f)$ *of* f is the smallest integer d such that there exists a depth-d decision tree that computes f.

Relations among them. It is not difficult to show that both sensitivity, block sensitivity, certificate complexity and Fourier degree are lower bounds on decision tree complexity. It turns out that they are also upper bounds of decision tree complexity up to a polynomial. To be more specific, we say that two complexity measures M_1 and M_2 are *polynomially related* if there exists $0 < a \leq b$ such that $M_1(f)^a \leq M_2(f) \leq M_1(f)^b$ for all Boolean f. We have the following theorem:

Theorem 2.7. *[BI87], [Nis91], [NS94] Block sensitivity, certificate complexity, Fourier degree and decision tree complexity are polynomially related to each other.*

The proof of it is contained in several papers from several different authors. We refer to the excellent survey of Buhrman and de Wolf [BdW02], which contains the proofs of the polynomial relations between all those measures with a more complete references, and also many other results related to decision trees. Thanks to Theorem 2.7, we know that for all complexity measures we defined earlier in this section, either they can be arbitrarily smaller then $D(f)$ (including $\mathbf{Inf}[f]$, $C_{\min}(f)$ and $\deg_{\mathbb{F}_2}(f)$), or they are in fact polynomially related to $D(f)$, *except* sensitivity. It is natural to ask whether we can also put sensitivity into the list of measures polynomially related to $D(f)$. This question was asked by Nisan and Szegedy in [NS94] and it turns out to be an extremely difficult problem. By Theorem 2.7, it suffices to ask whether sensitivity and block sensitivity are polynomially related. In particular, the following conjecture is known as the *Sensitivity Conjecture*:

Conjecture 1 (**Sensitivity Conjecture**)**.** *There exists $c > 0$ such that for all $f : \{0,1\}^n \to \{0,1\}$, $bs(f) \leq s(f)^c$.*

Two decades after this question was asked, we are still far from resolving it. This question can be reformulated as questions in extremal graph theory, communication complexity or Fourier analysis. See the survey of Hatami et al. [HKP11] for related or equivalent conjectures. This conjecture is difficult even for special classes of functions. Other than monotone functions and symmetric functions, the conjecture is known to be true for graph properties [Tur84] and minterm transitive functions [Cha05]. Both Turan's and Chakraborty's results follow from a $n^{\Omega(1)}$ lower bound on sensitivity. For general Boolean functions, the best known upper bound on block sensitivity in terms of sensitivity is exponential, and it would be our main topic in the next section. For the separation, the largest known gap is quadratic.

Theorem 2.8. *There exists Boolean function f such that $bs(f) = \Omega(s(f)^2)$.*

The first quadratic separation is due to Rubinstein [Rub95], who achieved $bs(f) = \frac{1}{2}s(f)^2$. The currently best separation is $bs(f) = \frac{2}{3}s(f)^2 - \frac{1}{3}s(f)$ due to Ambainis and Sun [AS11].

3 Upper bounds in terms of sensitivity

Although it is easy to see sensitivity is a lower bound on many other complexity measures, it turns out that obtaining any upper bound in terms of sensitivity is much more difficult. It is not obvious at all that complexity measures like block sensitivity is in fact upper bounded by some function of sensitivity. The first upper bound on those measures in terms of sensitivity was observed by Simon [Sim83]. Simon's upper bound is exponential in sensitivity. However, three decades after Simon's result, the currently best known upper bound on block sensitivity is still exponential in sensitivity, but with a smaller exponent. There are very few upper bounds on block sensitivity in terms of sensitivity known in the literature. In this section, we review the results and proof techniques by Simon [Sim83], Kenyon and Kutin [KK04] and Ambanins et al. [ABG+14]. Their arguments fall into two types. One is to use the isoperimetric inequality for the Boolean cube and the other is to find a point with high sensitivity. Both methods have different strengths and they can solve some special cases of the Sensitivity Conjecture.

3.1 Using isoperimetric inequality

One approach to establish an inequality relating sensitivity and other complexity measures is to consider the total sensitivity of all points in some subset of the Boolean cube. From which we can always upper bound the sensitivity at each point by $s(f)$ (or more precisely $s^0(f)$ or $s^1(f)$ depending on the value of f at that point) and then sum them up. The lower bound needs extra works and one way to get it is to use the *isoperimetric inequality* for the Boolean cube.

Lemma 3.1 ([Har64] **edge-isoperimetric inequality for the Boolean cube**)**.** *Let $G = (V, E)$ be a n-dimensional Boolean cube. Let $S \subseteq V$, then*

$$|E(S, \overline{S})| \geq |S|(n - \log |S|)$$

where $E(S, \overline{S}) = \{(u, v) \in S \times \overline{S} \mid \{u, v\} \in E\}$ is the number of edges between S and $\overline{S}$.

This inequality asserts that subgraphs of the Boolean cube with large *average* degree must be large (i.e. with many vertices). It can be used to get a lower bound on the size of the subgraph induced by function with low sensitivity. The following fact is used in Simon's proof and can be proved by simple induction on dimension without using Lemma 3.1 (in fact Lemma 3.1 also has a simple induction proof [CFGS88]).

Corollary 3.2 (**Simon's lemma**)**.** *Let $G = (V, E)$ be a subgraph of the n-dimensional Boolean cube with minimum degree $d_{\min}$. Then $|V| \geq 2^{d_{\min}}$.*

Proof. By Lemma 3.1, $|E(V, \overline{V})| \geq |V|(n - \log |V|)$. The claim follows by observing that $|E(V, \overline{V})| \leq (n - d_{\min})|V|$ and rearranging terms. □

We are now ready to prove Simon's result.

Theorem 3.3. *[Sim83] If $f : \{0,1\}^n \to \{0,1\}$ depends on all n variables, then $s(f) \geq \frac{1}{2}\log n - \frac{1}{2}\log\log n + \frac{1}{2}$.*

Proof. Let $s = s(f)$. We first show that $2^n\mathbf{Inf}_i[f] = |\{x \in \{0,1\}^n \mid f(x) \neq f(x^i)\}| \geq 2^{n-2s+2}$ if f depends on the i-th variable, i.e. $f(x) \neq f(x^i)$ for some x.

Let $g(x) = f(x) \oplus f(x^i)$ and $A = g^{-1}(1)$, it suffices to show $|A| \geq 2^{n-2s+2}$. Since $s(g,x) \leq \big(s(f,x) - 1\big) + \big(s(f,x^i) - 1\big) \leq 2s - 2$. It follows that the subgraph induced by the set A has minimum degree at least $n - 2s + 2$. By Corollary 3.2, we have $|A| \geq 2^{n-2s+2}$ and hence $\mathbf{Inf}_i[f] \geq 2^{-2s+2}$.

Since the total influence of f is at most s, we obtain the inequality

$$s \geq \mathbf{Inf}[f] = \sum_{i=1}^{n}\mathbf{Inf}_i[f] \geq n2^{-2s+2}.$$

From this inequality, It is not difficult to check that $s \geq \frac{1}{2}\log n - \frac{1}{2}\log\log n + \frac{1}{2}$. □

Since all the complexity measures we discussed are trivially upper bounded by the number of variables on which f depends, Simon's result implies an $O(s(f)4^{s(f)})$ upper bound on all of them. Note that this relation between sensitivity and the number of relevant variables of f is actually optimal up to an additive $O(\log\log n)$ factor. Wegner [Weg85] constructed a function depending on all variables but has sensitivity $\frac{1}{2}\log n + O(\log\log n)$.

By considering the total sensitivity of all points, we get Theorem 3.3. It is natural to ask whether we can achieve better bound by considering the total sensitivity of points in some other set. It turns out that it is indeed possible. Ambainis et al. [ABG+14] observed that it can be done by considering a maximal monochromatic subcube and some of its neighboring subcubes.

Theorem 3.4. *[ABG+14] $C^b(f) \leq 2^{s^{1-b}(f)-1}s^b(f) - \big(s^{1-b}(f) - 1\big)$. In particular, $C(f) \leq 2^{s(f)-1}s(f)$.*

Proof. We proof $C^0(f) \leq 2^{s^1(f)-1}s^0(f) - \big(s^1(f) - 1\big)$, and the other case is similar. By shifting f if necessary, we can assume $f(0^n) = 0$ and $C(f, 0^n) = C^0(f)$. Let $m = C^0(f)$ and Q_0 be the largest subcube containing 0^n on which f is 0. We may assume $Q_0 = \{x \in \{0,1\}^n \mid x_1 = x_2 = \cdots = x_m = 0\}$. Let $Q_i = Q_0 + e_i$, $i = 1, \ldots, m$, and $A_i = Q_i \cap f^{-1}(1)$. We are interested in the total sensitivity S of all points in $\bigcup_{i=1}^m Q_i$. It is clear that

$$S = \sum_{i=1}^{m}\sum_{x\in Q_i} s(f,x) \leq \sum_{i=1}^{m}|A_i|s^1(f) + \sum_{i=1}^{m}(2^{n-m} - |A_i|)s^0(f). \tag{1}$$

For the lower bound, we divide the total sensitivity into three parts and then lower bound them separately. More precisely, we let $S = S_1 + S_2 + S_3$ where S_1 is the sensitivity between Q_i and Q_0, S_2 is the sensitivity within Q_i, and S_3 is the sensitivity between Q_i and $\{0,1\}^n \setminus \bigcup_{j=0}^m Q_j$.

For S_1, every point in A_i contributes 1 to the total sensitivity, and points in $Q_i \setminus A_i$ contribute nothing. Thus we have

$$S_1 = \sum_{i=1}^{m}|A_i|.$$

For S_2, the total sensitivity within Q_i is twice the number of edges between A_i and $Q_i \setminus A_i$. Since Q_i can be viewed as a $(n-m)$-dimensional Boolean cube, we consider A_i as a subgraph of Q_i and apply Lemma 3.1 to get

$$S_2 = 2\sum_{i=1}^{m}|E(A_i, \overline{A_i})| \geq 2\sum_{i=1}^{m}|A_i|(n - m - \log|A_i|).$$

For S_3, observe that a point in Q_i is of the form $x + e_i$ for some $x \in Q_0$, and its neighbor in $\{0,1\}^n \setminus \bigcup_{j=0}^m Q_j$ is of the form $x + e_i + e_j$ for some $j \neq i$ and $j \leq m$. Thus we have

$|f(x+e_i) - f(x+e_i+e_j)| + |f(x+e_j) - f(x+e_i+e_j)| \geq |f(x+e_i) - f(x+e_j)|$. We now can lower bound S_3,

$$
\begin{aligned}
S_3 &= \sum_{x \in Q_0} \sum_{1 \leq i,j \leq n, i \neq j} |f(x+e_i) - f(x+e_i+e_j)| \\
&\geq \sum_{x \in Q_0} \sum_{1 \leq i < j \leq n} |f(x+e_i) - f(x+e_j)| \\
&= \sum_{x \in Q_0} \left(\sum_{i=1}^{m} f(x+e_i) \right) \left(m - \sum_{i=1}^{m} f(x+e_i) \right) \\
&= \sum_{x \in Q_0} s(f,x)\big(m - s(f,x)\big).
\end{aligned}
$$

To summarize, we now have the lower bound

$$
S = S_1 + S_2 + S_2 \geq \sum_{i=1}^{m} |A_i| + 2 \sum_{i=1}^{m} |A_i|(n - m - \log |A_i|) + \sum_{x \in Q_0} s(f,x)\big(m - s(f,x)\big). \tag{2}
$$

The rest of the proof is to deduce from (1) and (2) that

$$
w\left(1 + m - s^1(f) - \frac{2\big(s^1(f) - 1\big)}{2^{s^1(f)-1} - 1}\right) \leq m 2^{n-m} \left(s^0(f) - \frac{2\big(s^1(f) - 1\big)}{2^{s^1(f)-1} - 1}\right), \tag{3}
$$

where $w = \sum_{i=1}^{m} |A_i|$. Assuming (3) and by applying Corollary 3.2 to each A_i-induced subgraph in the $(n-m)$-dimensional Boolean cube Q_i, we have $w = \sum_{i=1}^{m} |A_i| \geq m 2^{n-m-s^1(f)+1}$ (since $A_i \subseteq f^{-1}(1)$, every point in A_i has at most $s^1(f) - 1$ neighbors in the $(n-m)$-dimensional subcube Q_i that is not in A_i, it follows that every vertex in this subgraph has degree at least $n - m - s^1(f) + 1$). The proof is then completed by a simple algebraic manipulation.

It remains to derive (3). Let $w = \sum_{i=1}^{m} |A_i|$ and by rearranging terms, we have

$$
w\big(1 + 2n - 2m - s^1(f) + s^0(f)\big) \leq m 2^{n-m} s^0(f) + 2 \sum_{i=1}^{m} |A_i| \log |A_i| - \sum_{x \in Q_0} s(f,x)\big(m - s(f,x)\big).
$$

We first deal with the last term on the right. For $x \in Q_0$, $s(f,x) \leq s^0(f)$ and hence

$$
\sum_{x \in Q_0} s(f,x)\big(m - s(f,x)\big) \geq \sum_{x \in Q_0} s(f,x)\big(m - s^0(f)\big) = \big(m - s^0(f)\big) \sum_{i=1}^{m} |A_i| = \big(m - s^0(f)\big) w.
$$

Plugging it back to the inequality we just get, we have

$$
w\big(1 + 2n - m - s^1(f)\big) \leq m 2^{n-m} s^0(f) + 2 \sum_{i=1}^{m} |A_i| \log |A_i|. \tag{4}
$$

We now upper bound $|A_i| \log |A_i|$ by some affine function of $|A_i|$. Observe that the function $x \log x$ is convex, and $2^{n-m-s^1(f)+1} \leq |A_i| \leq 2^{n-m}$ (the first inequality follows by Lemma 3.2). So we can upper bound $|A_i| \log |A_i|$ by

$$
\theta\big(n - m - s^1(f) + 1\big) 2^{n-m-s^1(f)+1} + (1 - \theta)(n - m) 2^{n-m},
$$

where θ is chosen such that $\theta 2^{n-m-s^1(f)+1} + (1 - \theta) 2^{n-m} = |A_i|$. By a straightforward computation, it is not difficult to show the quantity above is

$$
|A_i|(n - m) - \frac{2^{n-m} - |A_i|}{2^{s^1(f)-1} - 1} \big(s^1(f) - 1\big).
$$

Thus we have

$$\begin{aligned}\sum_{i=1}^{m}|A_i|\log|A_i| &\leq \sum_{i=1}^{m}\left(|A_i|(n-m)-\frac{2^{n-m}-|A_i|}{2^{s^1(f)-1}-1}\big(s^1(f)-1\big)\right)\\ &= w\left(n-m+\frac{s^1(f)-1}{2^{s^1(f)-1}-1}\right)-\frac{m2^{n-m}\big(s^1(f)-1\big)}{2^{s^1(f)-1}-1}.\end{aligned}$$

Plugging it into (4) and rearranging terms gives (3). ☐

By Theorem 3.4 and the fact that $bs(f) \leq C(f)$, we get the best known upper bound on block sensitivity in terms of sensitivity. In fact such bound can be improved significantly when either 0-sensitivity or 1-sensitivity is small. We need the following lemma of Kenyon and Kutin:

Lemma 3.5. *[KK04] $bs^b(f) \leq 2C^{1-b}(f)s^b(f)$.*

Corollary 3.6. *[ABG+14] $bs(f) \leq \min\{2^{s^0(f)}, 2^{s^1(f)}\}s^0(f)s^1(f)$.*

Proof. Since $bs^0(f) \leq C^0(f)$, by Theorem 3.4 we have $bs^0(f) \leq 2^{s^1(f)}s^0(f) \leq 2^{s^1(f)}s^0(f)s^1(f)$. On the other hand, by Lemma 3.9 and Theorem 3.4, $bs^0(f) \leq 2C^1(f)s^0(f) \leq 2^{s^0(f)}s^0(f)s^1(f)$. Thus we have $bs^0(f) \leq \min\{2^{s^0(f)}, 2^{s^1(f)}\}s^0(f)s^1(f)$. The case for $bs^1(f)$ is similar and the proof is completed. ☐

Finally we state without proof the following upper bound on Fourier degree. The proof is by induction and it also use Lemma 3.1 (in fact Corollary 3.2) to bound the size of certain set from which we can establish an inequality relating degree and sensitivity.

Theorem 3.7. *[ABG+14] $\deg(f)^{1-O(1/\log\log\deg(f))} \leq 2^{s(f)}$.*

3.2 Finding a highly sensitive point

The proofs of Theorem 3.3 and Theorem 3.4 can be regarded as counting arguments which show there must be a point in certain sets (the whole Boolean cube in Theorem 3.3 and $\bigcup_{i=1}^{m} Q_i$ in Theorem 3.4) with sensitivity $\frac{1}{2}\log n - O(\log\log n)$ and $\log C(f) - O(\log\log C(f))$ respectively. They tell very little about the highly sensitive point.

The following result of Kenyon and Kutin was obtained by demonstrating certain carefully chosen point has significant sensitivity.

Theorem 3.8. *[KK04] $bs(f) \leq \left(\frac{e}{\sqrt{2\pi}}\right)e^{s(f)}\sqrt{s(f)}$*

The proof of Kenyon and Kutin is very different from the proofs of Theorem 3.2 and Theorem 3.4. The main idea behind the proof is to assign weights to the sensitive blocks according to the block size. Suppose we assign weight 1 to blocks of size 1 and 0 to blocks of larger size, then for the pair $(x, \mathcal{B})$ of point and family of disjoint sensitive blocks that maximizes the total weight, x must be one of the points with highest sensitivity and $\mathcal{B}$ contains all variables (blocks of size 1) for which f is sensitive to at x. On the other hand, if we assign weight 1 to all blocks, then the pair $(x, \mathcal{B})$ achieving the maximum total weight contains x with highest block sensitivity and $\mathcal{B}$ achieves the block sensitivity of f. One may expect by choosing the weights in a specific way, we may be able to "interpolate" sensitivity and block sensitivity and hence obtaining an inequality relating them. By carefully choosing the weights, Kenyon and Kutin were able to argue that the point with maximum total weight must have sensitivity that satisfies the inequality in Theorem 3.8.

The key of their proof is the following lemma. A block $B \subseteq [n]$ is a ℓ-*block* if its size is at most ℓ, i.e. $|B| \leq \ell$. For $1 \leq \ell \leq n$, we define $bs_\ell(f)$ as the variant of block sensitivity that requires each sensitive block to be a ℓ-block.

Lemma 3.9. *Let $f : \{0,1\}^n \to \{0,1\}$ with sensitivity $s(f) = s$. Then for any $w_1 \geq w_2 \geq \cdots \geq w_\ell \geq 1$, there exists $x \in \{0,1\}^n$ and a family of disjoint sensitive ℓ-blocks $\mathcal{B}$ at x such that*

$$\sum_{i=1}^{\ell} m_i w_i (s - i) \geq \sum_{i=2}^{\ell} i m_i (w_{i-1} - w_i) \tag{5}$$

and

$$\sum_{i=1}^{\ell} m_i w_i \geq bs_\ell(f),$$

where m_i is the number of blocks of size i in $\mathcal{B}$.

Proof. For any $x \in \{0,1\}^n$ and any family $\mathcal{B} = \{B_1, \ldots, B_r\}$ of disjoint sensitive ℓ-blocks at x. We define the total weight $t(x, \mathcal{B})$ of the pair $(x, \mathcal{B})$ by $\sum_{k=1}^{r} w_{|B_k|}$. Let $(x, \mathcal{B})$ be the pair that achieves the maximum total weight. We can assume $\mathcal{B}$ only contains minimal sensitive blocks. Now we show that such x and $\mathcal{B}$ satisfy the conditions in Lemma 3.9.

First, if $(x', \mathcal{B}')$ is the pair that achieves the maximum ℓ-block sensitivity, then $\sum_{i=1}^{\ell} m'_i w_i \geq \sum_{i=1}^{\ell} m'_i = bs_\ell(f)$ since $w_i \geq 1$ for each i. Since $t(x', \mathcal{B}') = \sum_{i=1}^{\ell} m'_i w_i \geq bs_\ell(f)$, and the pair $(x, \mathcal{B})$ maximizes the total weight, we have $\sum_{i=1}^{\ell} m_i w_i = t(x, \mathcal{B}) \geq t(x', \mathcal{B}') \geq bs_\ell(f)$.

To show (5), let K be the subset of $[n]$ such that $k \in K$ if $k \in B$ for some $B \subseteq \mathcal{B}$ of size at least 2. For each $k \in K$, we define a pair $(x^k, \mathcal{A}_k)$ where the family $\mathcal{A}_k$ is defined as follows. If $B \in \mathcal{B}$ contains k, then put $B \setminus \{k\}$ into $\mathcal{A}_k$. If $B \in \mathcal{B}$ does not contain k, then put it into $\mathcal{A}_k$ if it is a sensitive block of f at x^k.

The main observation is that if f has low sensitivity, then each sensitive block $B \in \mathcal{B}$ must appear in $\mathcal{A}_k$ for many k. Specifically, for each $B \in \mathcal{B}$, $|\{k \in K \setminus B \mid B \notin \mathcal{A}_k\}| \leq s - |B|$. Indeed, if $B \notin \mathcal{A}_k$ for some $k \notin B$, then by definition B is not a sensitive block at x^k. It implies that f is sensitive to the k-th variable at x^B since $f(x^B) \neq f(x)$ and $f((x^B)^k) = f((x^k)^B) = f(x^k) = f(x)$ where the last equality follows from the fact that k is in some minimal sensitive block of size at least 2. Thus, $f(x^B) \neq f(x) = f((x^B)^k)$, which means f is sensitive to each such k at x^B. On the other hand, since B is a minimal sensitive block, f is sensitive to each variable in B at x^B. So there are at most $s - |B|$ possible $k \notin B$ such that $B \notin \mathcal{A}_k$.

Now, for $B \in \mathcal{B}$ of size at least 2 and $k \in B$. The difference between $t(x, \mathcal{B})$ and $t(x^k, \mathcal{A}_k)$ is given by

$$t(x, \mathcal{B}) - t(x^k, \mathcal{A}_k) = (w_{|B|} - w_{|B|-1}) + \sum_{C \in \mathcal{B} \setminus (\mathcal{A}_k \cup \{B\})} w_{|C|}.$$

Note that B appears in $|B|$ different k, and each $w_{|C|}$ appears in at most $s - |C|$ different k. So if we sum the differences over all $k \in K$, and since $t(x, \mathcal{B})$ is the largest possible total weight, we have

$$0 \leq \sum_{k \in K} \Big(t(x, \mathcal{B}) - t(x^k, \mathcal{A}_k) \Big) \leq \sum_{i=2}^{\ell} i m_i (w_i - w_{i-1}) + \sum_{i=1}^{\ell} (s - i) m_i w_i,$$

as desired. □

Recall that we want to find a point x together with a family of disjoint sensitive blocks $\mathcal{B}$ at x so that many blocks have small size (ideally, we want to have many blocks of size 1, so that we can conclude $s(f, x)$ is large). If we rearrange the terms in (5), we have

$$w_1 m_1 + \sum_{i=2}^{\ell} i m_i w_{i-1} \leq s \sum_{i=1}^{\ell} m_i w_i.$$

So if w_{i-1} is much larger than w_i, we expect that the pair $(x, \mathcal{B})$ in Lemma 3.9 should not have too much large blocks. We now demonstrate how it gives us a pair $(x, \mathcal{B})$ such that many blocks have size less than ℓ.

Corollary 3.10. $bs_{\ell-1}(f) \geq \frac{\ell}{4s(f)} bs_\ell(f)$.

Proof. Let $w_i = w > 1$ for $i < \ell$, $w_\ell = 1$ and $w_i = 0$ for $i > \ell$. By Lemma 3.9, there exists $x \in \{0,1\}^n$ and a family $\mathcal{B}$ of disjoint sensitive blocks such that $\sum_{i=1}^{\ell} m_i w_i = w(\sum_{i=1}^{\ell-1} m_i) + m_\ell \geq bs_\ell(f)$ and

$$\sum_{i=1}^{\ell-1} m_i w(s-i) + m_\ell(s-\ell) \geq \ell m_\ell(w-1).$$

Upper bounding the summation by $sw\sum_{i=1}^{\ell-1} m_i$. By rearranging terms and assuming $\ell w - s > 0$, we have

$$\frac{sw}{\ell w - s}\sum_{i=1}^{\ell-1} m_i \geq m_\ell.$$

Finally, adding $w\sum_{i=1}^{\ell-1} m_i$ to both sides and letting $w = 2s/\ell$ (which satisfies $\ell w - s > 0$), we have

$$\frac{4s}{\ell}\sum_{i=1}^{\ell-1} m_i \geq w\Big(\sum_{i=1}^{\ell-1} m_i\Big) + m_\ell \geq bs_\ell(f).$$

Since $bs_{\ell-1}(f,x) = \sum_{i=1}^{\ell} m_i$, the proof is completed. ☐

With a more careful choice of the weights w_i, we can establish Theorem 3.8.

Proof of Theorem 3.8. To see what w_i we should choose, we rearrange (5) as

$$m_1 w_1(s-1) \geq \sum_{i=2}^{\ell} m_i(iw_{i-1} - sw_i).$$

Now if we choose w_i such that $iw_{i-1} - sw_i = yw_i$ for $2 \leq i \leq \ell$ and add ym_1w_1 to both sides, we have

$$m_1 w_1(s-1+y) \geq y\sum_{i=1}^{\ell} m_i w_i \geq y bs_\ell(f).$$

By our choice of w_i, we see that $w_1 = (s+y)^{\ell-1}/\ell!$. We simply upper bound the left hand side by $m_1w_1(s+y)$ to simplify the expression. Now we have

$$\frac{(s+y)^\ell}{y\ell!} m_1 \geq bs_\ell(f).$$

The upper bound is minimized by taking $y = s/(\ell-1)$. Since $m_1 \leq s(f)$, a simple calculation shows that

$$bs_\ell(f) \leq \frac{e}{(\ell-1)!} s(f)^\ell.$$

Since $bs_{s(f)}(f) = bs(f)$, by Stirling's approximation, $s! \geq \sqrt{2\pi s}(s/e)^s$ and hence we have

$$bs(f) \leq \frac{es^{s+1}}{s!} \leq \frac{e}{\sqrt{2\pi}} e^{s(f)}\sqrt{s(f)}.$$

☐

Corollary 3.10 asserts an interesting relation between ℓ-block sensitivity and sensitivity. In particular, it says that for any constant ℓ, we have $bs_\ell(f) = O(s(f)^\ell)$ and hence they are polynomially related. It can be used to show if $bs(f) = \Omega(n)$, then $s(f) = n^{\Omega(1)}$. Indeed, if $bs(f) = Kn$, then the average block size at most $1/K$. So, by Markov inequality, at least half of the blocks have size at most $2/K$, which implies $bs_{2/K}(f) = \Omega(n)$ and we are done. However, it is known that Corollary 3.10 is tight for all $2 \leq \ell \leq s(f)$ [KK04].

4 Decision tree via $\mathbb{F}_2$-representation

The exponential upper bounds on block sensitivity in terms of sensitivity may indicate that our understanding on sensitivity is very limited (in case they are not actually tight). Like the polynomial relations between some measures are established via decision tree complexity. Probably we should ask what structures are implied by being low-sensitivity, and then try to use those structures to construct a short decision tree. However, most of the decision tree constructions depend on measures which are polynomially related to block sensitivity. It does not suggest any weaker questions that may help us to understand more about low-sensitivity functions. In order to obtain some possibly weaker questions about sensitivity, we would like to upper bound decision tree complexity by some measures which can be *exponentially smaller* than block sensitivity, and then ask whether those measures can be upper bounded a polynomial in sensitivity. In this section, we give a simple decision tree construction that depends on two measures which can be *arbitrarily smaller* than sensitivity.

Many results about decision tree are based on the $\mathbb{R}$-representation of f. It may be natural to ask whether representing f as a polynomial over other field can bring us more. We believe the answer is *yes*. We start with a decision tree construction that is based on $\mathbb{F}_2$-representation. We need the following variant of minimum certificate.

Definition 4.1. $C'_{\min}(f)$ is the largest minimum certificate of the restriction of f on any subcube. More precisely, let $\mathcal{Q}_n$ be the set of all subcubes of the n-dimensional Boolean cube, and $f|_Q$ is the restriction of f on the subcube Q. Then $C'_{\min}(f) = \max_{Q\in\mathcal{Q}_n}\{f|_Q\}$.

Proposition 4.2. $D(f) \le C'_{\min}(f)\deg_{\mathbb{F}_2}(f)$.

Proof. By definition $C_{\min}(f|_Q) \le C'_{\min}(f)$ for any subcube Q. Let p be the $\mathbb{F}_2$-representation of f and Q be the subcube of codimension $m = C_{\min}(f)$. Note that restricting f on Q is equivalent to fixing some variables of f. Since $f|_Q$ is a constant, it follows that the set of fixed variables corresponding to Q intersect all maxonomials (which are the monomials of degree $\deg_{\mathbb{F}_2}(2)$), since otherwise some maxonomial will be remained and the resulting function cannot be a constant. So querying all those variables (and hence fixing them to the values of the queries) will reduce the degree by at least 1. By repeating it at most $\deg_{\mathbb{F}_2}(f)$ times the function would become a constant and hence we know what $f(x)$ is. $\square$

Note that the same proof works for any field $\mathbb{F}$. One can obtain a weaker relation $D(f) \le C'_{\min}(f)\deg(f)$ by considering $\mathbb{R}$-polynomial representation [BdW02].[1]

4.1 Two combinatorial conjectures

Proposition 4.2 suggests the conjecture that asserts both $C'_{\min}(f)$ and $\deg_{\mathbb{F}_2}(f)$ are upper bounded by a polynomials in $s(f)$. In fact, since $s(f)$ is non-increasing under taking restriction. we can replace $C'_{\min}(f)$ by $C_{\min}(f)$ and ask whether $C_{\min}(f)$ is upper bound by a polynomial in $s(f)$. Specifically, we have the following conjecture:

Conjecture 2. *There exists $c > 0$ such that If $s(f) = k$, then there exists a subcube of codimension k^c on which f is a constant.*

The question about $\deg_{\mathbb{F}_2}(f)$ seems a bit unnatural as it is about $\deg_{\mathbb{F}_2}(2)$, which is an algebraic quantity. But in fact we can reformulate it as a purely combinatorial question. Observe that we can assume without loss of generality that $\deg_{\mathbb{F}_2}(f) = n$ since we can always find a subfunction on $\deg_{\mathbb{F}_2}(f)$ variables with degree $\deg_{\mathbb{F}_2}(f)$, and sensitivity of that subfunction cannot be larger than $s(f)$. Since the $\mathbb{F}_2$-representation can be obtained by taking the coefficients of the $\mathbb{R}$-representation module 2 and the coefficient of $\prod_{i=1}^n x_i$ in the $\mathbb{R}$-representation is a sum of $|f^{-1}(1)|$ many $+1$ or -1 (since in the interpolation, each point in $|f^{-1}$ contributes either $+1$ or -1 to the coefficients of $\prod_{i=1}^n x_i$). It follows that $\deg_{\mathbb{F}_2}(f) = n$ is equivalent to $|f^{-1}(1)|$ is odd. So we have the following equivalent formulation.

[1]They stated $D(f) \le C^1(f)\deg(f)$ in [BdW02], but their proof actually showed $D(f) \le C'_{\min}(f)\deg(f)$.

Conjecture 3. *There exists $c > 0$ such that if $|f^{-1}(1)|$ is odd, then $s(f) \geq n^c$.*

Clearly both Conjecture 2 and Conjecture 3 are consequences of the Sensitivity Conjecture. To the best of our knowledge, both of them are open. But there are examples that shows the constant in both conjectures cannot be too small.

It was observed by Wegener and Zádori in [WZ89] that there exists Boolean function f such that $s(f) = C_{\min}(f) \geq s(f)^{\log_2 3} \approx s(f)^{1.58}$. They gave a function on 4 variables with $C_{\min}(f) = 3$ and $s(f) = 2$, then observed $C_{\min}(f \circ g) \geq C_{\min}(g) \cdot C_{\min}(f)$ and $s(f \circ g) \leq s(f) \cdot s(g)$. We note that the SORT function of Ambainis [Amb06] defined by $f(x) = 1$ if either $x_1 \geq x_2 \geq x_3 \geq x_4$ or $x_1 \leq x_2 \leq x_3 \leq x_4$, and $f(x) = 0$ otherwise, has the required properties. For separation between $\deg_{\mathbb{F}_2}(f)$ and $s(f)$, consider the function $f = \text{AND}_{\sqrt{n}} \circ \text{OR}_{\sqrt{n}}$. It is straightforward to see that $s(f) = \sqrt{n}$ but $\deg_{\mathbb{F}_2} f = n$. We summarize such separations as follows.

Proposition 4.3. *(i). There exists Boolean function f such that $C_{\min}(f) = \Omega(s(f)^{\log_2 3})$.*
(ii). There exists Boolean function f such that $\deg_{\mathbb{F}_2}(f) = \Omega(s(f)^2)$.

Acknowledgement

We are most graceful to Laci Babai for organizing the REU program and inviting us to join it. We thank Yuan Li for many interesting discussions and Toyota Technological Institute at Chicago (TTIC) for the support during the program.

References

[ABG+14] Andris Ambainis, Mohammad Bavarian, Yihan Gao, Jieming Mao, Xiaoming Sun, and Song Zuo. Tighter relations between sensitivity and other complexity measures. In Javier Esparza, Pierre Fraigniaud, Thore Husfeldt, and Elias Koutsoupias, editors, *Automata, Languages, and Programming*, volume 8572 of *Lecture Notes in Computer Science*, pages 101–113. Springer Berlin Heidelberg, 2014. 1, 4, 5, 7

[Amb06] Andris Ambainis. Polynomial degree vs. quantum query complexity. *Journal of Computer and System Sciences*, 72(2):220 – 238, 2006. JCSS FOCS 2003 Special Issue. 11

[AS11] Andris Ambainis and Xiaoming Sun. New separation between $s(f)$ and $bs(f)$. *Electronic Colloquium on Computational Complexity (ECCC)*, 18(116), 2011. 1, 4

[BdW02] Harry Buhrman and Ronald de Wolf. Complexity measures and decision tree complexity: A survey. *Theoretical Computer Science*, 288(1):21–43, 2002. 3, 10

[BI87] Manuel Blum and Russell Impagliazzo. Generic oracles and oracle classes. In *Proceedings of the 28th Annual Symposium on Foundations of Computer Science*, SFCS '87, pages 118–126, Washington, DC, USA, 1987. IEEE Computer Society. 3

[CD82] Stephen Cook and Cynthia Dwork. Bounds on the time for parallel RAM's to compute simple functions. In *Proceedings of the Fourteenth Annual ACM Symposium on Theory of Computing*, STOC '82, pages 231–233, New York, NY, USA, 1982. ACM. 1

[CFGS88] Fan R.K. Chung, Zoltán Füredi, Ronald L. Graham, and Paul Seymour. On induced subgraphs of the cube. *Journal of Combinatorial Theory, Series A*, 49(1):180–187, 1988. 4

[Cha05] Sourav Chakraborty. On the sensitivity of cyclically-invariant boolean functions. *IEEE Conference on Computational Complexity*, pages 163–167, 2005. 4

[Har64] L. H. Harper. Optimal assignments of numbers to vertices. *Journal of the Society for Industrial and Applied Mathematics*, 12(1):pp. 131–135, 1964. 4

[HKP11] Pooya Hatami, Raghav Kulkarni, and Denis Pankratov. *Variations on the Sensitivity Conjecture*. Number 4 in Graduate Surveys. Theory of Computing Library, 2011. 4

[KK04] Claire Kenyon and Samuel Kutin. Sensitivity, block sensitivity, and ℓ-block sensitivity of boolean functions. *Information and Computation*, 189(1):43–53, 2004. 1, 4, 7, 9

[Nis91] Noam Nisan. Crew prams and decision trees. *SIAM Journal on Computing*, 20(6):999–1007, 1991. 1, 3

[NS94] Noam Nisan and Mario Szegedy. On the degree of boolean functions as real polynomials. *computational complexity*, 4(4):301–313, 1994. 1, 3

[Rei82] Rüdiger Reischuk. A lower time-bound for parallel random access machines without simultaneous writes. *Technical Report RJ3431, IBM, New York*, 1982. 1

[Rub95] David Rubinstein. Sensitivity vs. block sensitivity of boolean functions. *Combinatorica*, 15(2):297–299, 1995. 1, 4

[Sim83] Hans-Ulrich Simon. A tight $\Omega(\log \log n)$-bound on the time for parallel RAM's to compute nondegenerated boolean functions. *Foundations of Computation Theory*, 158:439–444, 1983. 1, 4, 5

[Tur84] György Turán. The critical complexity of graph properties. *Information Processing Letters*, 18(3):151–153, 1984. 4

[Weg85] Ingo Wegener. The critical complexity of all (monotone) boolean functions and monotone graph properties. *Information and Control*, 67:212–222, 1985. 5

[WZ89] Ingo Wegener and László Zádori. A note on the relations between critical and sensitive complexity. *Journal of Information Processing and Cybernetics (EIK)*, 25(8/9):417–421, 1989. 11