Electronic Colloquium on Computational Complexity, Report No. 62 (2016)

On the Sensitivity Conjecture

Avishay Tal *

April 18, 2016

Abstract

The sensitivity of a Boolean function $f : \{0,1\}^n \to \{0,1\}$ is the maximal number of neighbors a point in the Boolean hypercube has with different f-value. Roughly speaking, the block sensitivity allows to flip a set of bits (called a block) rather than just one bit, in order to change the value of f. The sensitivity conjecture, posed by Nisan and Szegedy (CC, 1994), states that the block sensitivity, $bs(f)$, is at most polynomial in the sensitivity, $s(f)$, for any Boolean function f. A positive answer to the conjecture will have many consequences, as the block sensitivity is polynomially related to many other complexity measures such as the certificate complexity, the decision tree complexity and the degree. The conjecture is far from being understood, as there is an exponential gap between the known upper and lower bounds relating $bs(f)$ and $s(f)$.

We continue a line of work started by Kenyon and Kutin (Inf. Comput., 2004), studying the ℓ-block sensitivity, $bs_\ell(f)$, where ℓ bounds the size of sensitive blocks. While for $bs_2(f)$ the picture is well understood with almost matching upper and lower bounds, for $bs_3(f)$ it is not. We show that any development in understanding $bs_3(f)$ in terms of $s(f)$ will have great implications on the original question. Namely, we show that either $bs(f)$ is at most sub-exponential in $s(f)$ (which improves the state of the art upper bounds) or that $bs_3(f) \geq s(f)^{3-\varepsilon}$ for some Boolean functions (which improves the state of the art separations).

We generalize the question of $bs(f)$ versus $s(f)$ to bounded functions $f : \{0,1\}^n \to [0,1]$ and show an analog result to that of Kenyon and Kutin: $bs_\ell(f) = O(s(f))^\ell$. Surprisingly, in this case, the bounds are close to being tight. In particular, we construct a bounded function $f : \{0,1\}^n \to [0,1]$ with $bs(f) \geq n/\log n$ and $s(f) = O(\log n)$, a clear counterexample to the sensitivity conjecture for bounded functions.

Finally, we give a new super-quadratic separation between sensitivity and decision tree complexity by constructing Boolean functions with $\mathrm{DT}(f) \geq s(f)^{2.115}$. Prior to this work, only quadratic separations, $\mathrm{DT}(f) = s(f)^2$, were known.

*Institute for Advanced Study, Princeton, NJ. Email: avishay.tal@gmail.com. Research supported by the Simons Foundation, and by the National Science Foundation grant No. CCF- 1412958. Any opinions, findings and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation.

ISSN 1433-8092

1 Introduction

A long-standing open problem in complexity and combinatorics asks what is the relationship between two complexity measures of Boolean functions: the sensitivity and block-sensitivity. We first recall the definition of the two complexity measures.

Definition 1.1. *Let $f : \{0,1\}^n \to \{0,1\}$ be a Boolean function and $x \in \{0,1\}^n$ be a point. The* sensitivity of f at x *is the number of neighbors y of x in the Hamming cube such that $f(y) \neq f(x)$, i.e., $s(f,x) \triangleq |\{i \in [n] : f(x) \neq f(x \oplus e_i)\}|$.*[1] *The (maximal)* sensitivity of f *is defined as $s(f) \triangleq \max_{x\in\{0,1\}^n} s(f,x)$.*

Definition 1.2. *Let $f : \{0,1\}^n \to \{0,1\}$ be a Boolean function and $x \in \{0,1\}^n$ be a point. For a block $B \subseteq [n]$, denote by $\mathbb{1}_B \in \{0,1\}^n$ its characteristic vector, i.e., $(\mathbb{1}_B)_i = 1$ iff $i \in B$. We say that a block B is* sensitive *for f on x if $f(x) \neq f(x \oplus \mathbb{1}_B)$. The* block-sensitivity of f at x *$x \in \{0,1\}^n$ is the maximal number of disjoint sensitive blocks for f at x, i.e.,*

$$bs(f,x) = \max\{r : \exists \text{ disjoint } B_1, B_2, \ldots, B_r \subseteq [n]\ , f(x) \neq f(x \oplus \mathbb{1}_{B_i})\}\ .$$

The (maximal) block-sensitivity of f *is defined as $bs(f) \triangleq \max_{x\in\{0,1\}^n} bs(f,x)$.*

For shorthand, we will denote $(x \oplus e_i)$ and $(x \oplus \mathbb{1}_B)$ by $(x + e_i)$ and $(x + B)$ respectively. By definition, the block-sensitivity is at least the sensitivity by considering only blocks of size 1. The sensitivity conjecture, posed by Nisan and Szegedy [NS94], asks if a relation in the other direction holds as well.

Conjecture 1.3 (The Sensitivity Conjecture)**.** *$\exists d\ \forall f :\ bs(f) \leq s(f)^d$.*

A stronger variant of the conjecture states that d can be taken to be 2. Despite much work on the problem [Nis89, NS94, Rub95, KK04, Cha11, Vir11, AS11, HKP11, Bop12, ABG+14, AP14, AV15, APV15, GKS15, Sze15, GNS+16] there is still an exponential gap between the best known separations and the best known relations connecting the two complexity measures.

Known Separations. An interesting example due to Rubinstein [Rub95] shows a quadratic separation between the two measures: $bs(f) = \frac{1}{2} \cdot s(f)^2$. This example was improved by [Vir11] and then by [AS11] to $bs(f) = \frac{2}{3} \cdot s(f)^2 \cdot (1 - o(1))$ which is current state of the art.

Known Relations. Simon [Sim83] proved (implicitly) that $bs(f)$ is at most $4^{s(f)} \cdot s(f)$. The upper bound was improved by Kenyon and Kutin [KK04] who showed that $bs(f) \leq O(e^{s(f)} \cdot \sqrt{s(f)})$. Recently, Ambainis et al. [ABG+14] improved this bound to $bs(f) \leq 2^{s(f)-1} \cdot s(f)$. Even more recently, Ambainis et al. [APV15] improved this bound slightly to $bs(f) \leq 2^{s(f)-1} \cdot (s(f) - 1/3)$.

To sum up, while the best known upper bound on the block-sensitivity in terms of sensitivity is exponential, the best known lower bound is quadratic. Indeed, we seem far from understanding the right relation between the two complexity measures.

1.1 ℓ-block sensitivity

All mentioned examples that exhibit quadratic separations between the sensitivity and block sensitivity ([Rub95, Vir11, AS11]) have the property that the maximal block sensitivity is achieved on blocks of size at most 2. For this special case, Kenyon and Kutin [KK04] showed that the block sensitivity is at most $2 \cdot s(f)^2$. Hence, these examples are essentially tight for this subcase.

[1] e_i is the vector whose i-th entry equals 1 and all other entries equal 0.

Kenyon and Kutin introduced the notion of ℓ-block sensitivity (denoted $bs_\ell(f)$): the maximal number of disjoint sensitive blocks where each block is of size at most ℓ. Note that without loss of generality we may consider only sensitive blocks that are minimal with respect to set-inclusion (since otherwise we could of picked smaller blocks that are still disjoint). A well-known fact (cf. [BdW02, Lemma 3]) asserts that any minimal sensitive block for f is of size at most $s(f)$, thus $bs(f) = bs_{s(f)}(f)$. Kenyon and Kutin proved the following inequalities relating the ℓ-block sensitivity of different ℓ-s:

$$bs_\ell(f) \le \frac{4}{\ell} \cdot s(f) \cdot bs_{\ell-1}(f) \tag{1}$$

$$bs_\ell(f) \le \frac{e}{(\ell-1)!} \cdot s(f)^\ell \tag{2}$$

for all $2 \le \ell \le s(f)$. Plugging $\ell = s(f)$ gives the aforementioned bound $bs(f) \le O(e^{s(f)} \cdot \sqrt{s(f)})$.

1.2 Our Results

1. In Section 2, we refine the argument of Kenyon and Kutin giving a better upper bound on the ℓ-block sensitivity in terms of the $(\ell-1)$-block sensitivity. We show that

$$bs_\ell(f) \le \frac{e}{\ell} \cdot s(f) \cdot bs_{\ell-1}(f) \tag{3}$$

   improving the bound in Eq. (1). On the other hand, Kenyon and Kutin gave examples with $bs_\ell(f) \ge \frac{1}{\ell} \cdot s(f) \cdot bs_{\ell-1}(f)$. Hence, Eq. (3) (and in fact, also Eq. (1)) is tight up to a constant. Interestingly, our analysis uses (a very simple) ordinary differential equation.

2. In Section 3, we put focus on understanding $bs_3(f)$ in terms of the sensitivity. We show that an upper bound of the form $bs_3(f) \le s(f)^{3-\varepsilon}$ for some constant ε implies a sub-exponential upper bound for the sensitivity conjecture: $\forall f : bs(f) \le 2^{s(f)^{1-\delta}}$, for $\delta > 0$. On the other hand, the best known separation (i.e., the aforementioned example by [AS11]) gives examples with $bs_3(f) \ge bs_2(f) \ge \Omega(s(f)^2)$. Thus, improving either the upper or lower bound for $bs_3(f)$ in terms of $s(f)$ will imply a breakthrough in our understanding of the sensitivity conjecture.

3. In Section 4, we consider an extension of the sensitivity conjecture to bounded functions $f : \{0,1\}^n \to [0,1]$. We show that while Kenyon and Kutin's approach works in this model, it is almost tight, i.e., we give functions for which $bs_\ell(f) = \Omega((s(f)/\ell)^\ell)$. In particular, we give a function with sensitivity $O(\log n)$ and block sensitivity $\Omega(n/\log n)$ – a clear counterexample for the sensitivity conjecture in this model.

4. In Section 5, we find better-than-quadratic separations between the sensitivity and the decision tree complexity. We construct functions based on minterm cyclic functions (as coined by Chakraborty [Cha11]), that were found using computer search. In particular, we give an infinite family of functions $\{f_n\}_{n\in I}$ with $\mathrm{DT}(f_n) = n$ and $s(f_n) = O(n^{0.48})$. In addition, we give an infinite family of functions $\{g_n\}_{n\in I}$ with $s(g_n) = O(\mathrm{DT}(g_n)^{0.473})$.

2 Improving The Bound on bs_ℓ

In this section, we improve the bound on $bs_\ell(f)$ as a function of $bs_{\ell-1}(f)$ and $s(f)$. We start by recalling the analysis of [KK04], and then improve it using new ideas.

2.1 Kenyon-Kutin Argument

Let $x \in \{0,1\}^n$ be a point in the Boolean hypercube and $\mathcal{B}$ a collection of disjoint minimal blocks such that $f(x) \neq f(x+B)$ for any $B \in \mathcal{B}$. We assign weights $w_1 \geq \ldots \geq w_\ell \geq 1$ to sets of size $1, 2, \ldots, \ell$ respectively, and we seek to maximize $t(x, \mathcal{B}) = \sum_{B \in \mathcal{B}} w_{|B|}$. Since all weights are at least 1, we have $t(x, \mathcal{B}) \geq |B|$. Thus, upper bounding the value of t yields an upper bound on the ℓ-block sensitivity.

We choose $w_1 = w_2 = \ldots = w_{\ell-1} = w$ and $w_\ell = 1$ for some parameter $w \geq 1$. Let $(x, \mathcal{B})$ be a point and a collection of disjoint minimal sensitive blocks maximizing $t(\cdot,\cdot)$ w.r.t. the parameter w. Let $m_1, \ldots, m_\ell$ be the number of blocks of size $1, \ldots, \ell$ respectively in $\mathcal{B}$. We have $t(x, \mathcal{B}) = w \cdot (m_1 + \ldots + m_{\ell-1}) + m_\ell$.

Lemma 2.1. *Suppose $(x, \mathcal{B})$ maximize $t(\cdot,\cdot)$ w.r.t. $w \geq 1$ and let $m_1, \ldots, m_\ell$ be the number of blocks of size $1, \ldots, \ell$ in $\mathcal{B}$ respectively. Then,*

$$m_\ell \cdot (\ell w - s(f)) \leq (m_1 + \ldots + m_{\ell-1}) \cdot w \cdot s(f) \,.$$

Proof. We would derive the above inequality by examining the value of $t(\cdot,\cdot)$ on neighbors of x, and using the fact that all of these values are smaller or equal to $t(x, \mathcal{B})$.

Let $B \in \mathcal{B}$ be a block of size ℓ. By the minimality of the block B, it means that any subset of B does not flip the value of f on x. Thus, for each $i \in B$, we have $f(x+e_i) = f(x)$. In addition, the block $B' = B \setminus \{i\}$ is a sensitive block (of size $\ell - 1$) for $x + e_i$, but is not a sensitive block for x. Consider all such $\ell \cdot m_\ell$ neighbors $y = x + e_i$ where $i \in B, B \in \mathcal{B}$ and $|B| = \ell$. Denote by $\mathcal{A}_i$ the collection of all blocks B'' in $\mathcal{B}$ such that $f(y) = f(y + B'')$ (i.e., we are only considering disjoint blocks that were sensitive on x and minimal). Looking at a specific block $B'' \in \mathcal{B}$, we count for how many y's it is not a sensitive block, i.e., $f(y) = f(y+B'')$. Since $f(x) = f(y)$ and $f(x) \neq f(x+B'')$ the block B'' is not sensitive for $y = x + e_i$ if and only if $f(x + B'') \neq f(x + B'' + e_i)$. In other words, for B'' to be non-sensitive on $y = x + e_i$, i must be a sensitive coordinate of $x + B''$. Hence, each block $B'' \in \mathcal{B}$ may appear in at most $s(f)$ of the sets $\mathcal{A}_i$.

By our design for $y = x + e_i$ the block $B' = B \setminus \{i\}$ and the blocks in $B'' \in \mathcal{B} \setminus \mathcal{A}_i$ are sensitive. In order to show that they are disjoint it is enough to show that $B \in \mathcal{A}_i$. This is indeed the case since $x + e_i + B = x + B'$ and by the minimality of B, we have $f(x+e_i+B) = f(x+B') = f(x) = f(x+e_i)$, hence B is not a sensitive block for $x + e_i$. We got that $\{B'\} \cup (\mathcal{B} \setminus \mathcal{A}_i)$ is a family of disjoint sensitive blocks for $x + e_i$.

Using the fact that $t(x, \mathcal{B})$ is maximal, and summing over all neighbors of x considered above, we get

$$\begin{aligned} \ell \cdot m_\ell \cdot t(x, \mathcal{B}) &\geq \sum_{i \in B, |B| = \ell} t\left(x + e_i,\ \{B \setminus \{i\}\} \cup (\mathcal{B} \setminus \mathcal{A}_i)\right) \\ &\geq \sum_{i \in B, |B| = \ell} \left(w_{\ell-1} + t(x, \mathcal{B}) - \sum_{B'' \in \mathcal{A}_i} w_{|B''|} \right) . \end{aligned}$$

Rearranging we get

$$\ell \cdot m_\ell \cdot w_{\ell-1} \leq \sum_{i \in B, |B| = \ell} \sum_{B'' \in \mathcal{A}_i} w_{|B''|} = \sum_{B''} w_{|B''|} \cdot |\{(i, B) : i \in B, |B| = \ell, B'' \in A_i\}| \leq \sum_{B''} w_{|B''|} \cdot s(f) \,.$$

Substituting $w_1, \ldots, w_{\ell-1}$ with w and w_ℓ with 1 and rearranging gives

$$m_\ell \cdot (\ell \cdot w - s(f)) \leq (m_1 + \ldots + m_{\ell-1}) \cdot w \cdot s(f)$$

which completes the proof. $\square$

In order to get something meaningful from Lemma 2.1 we need $\ell \cdot w - s(f)$ to be greater than 0. Writing w as $\alpha \cdot s(f)/\ell$, this means that $\alpha > 1$. So we can choose any $\alpha > 1$ and get that the optimal $(m_1, \ldots, m_\ell)$ for that α fulfills the following inequality:

$$m_\ell \le (m_1 + \ldots + m_{\ell-1}) \cdot \frac{\alpha \cdot s^2/\ell}{\alpha \cdot s - s} = (m_1 + \ldots + m_{\ell-1}) \cdot \frac{s}{\ell} \cdot \frac{\alpha}{\alpha - 1} .$$

Overall we got that the maximal value of $t(\cdot,\cdot)$ with respect to $w = \frac{\alpha}{\ell} \cdot s(f)$ is at most the value of following linear program:

$$\begin{array}{ll} \text{maximize} & \frac{\alpha \cdot s(f)}{\ell} \cdot (m_1 + \ldots + m_{\ell-1}) + m_\ell \\ \\ \text{subject to} & m_\ell \le \frac{\alpha}{\alpha-1} \cdot \frac{s(f)}{\ell} \cdot (m_1 + \ldots + m_{\ell-1}) \\ & (m_1 + \ldots + m_{\ell-1}) \le bs_{\ell-1}(f) \\ & m_i \ge 0 \qquad\qquad\qquad\qquad \text{for } i = 1, \ldots, \ell \end{array} \tag{4}$$

Substituting $x_1 \triangleq (m_1+\ldots+m_{\ell-1})/bs_{\ell-1}$ and $x_2 \triangleq m_\ell/(bs_{\ell-1} \cdot s(f)/\ell)$ gives the following equivalent linear program:

$$\begin{array}{ll} \text{maximize} & \frac{s(f)}{\ell} \cdot bs_{\ell-1}(f) \cdot (\alpha \cdot x_1 + x_2) \\ \\ \text{subject to} & x_2 \le \frac{\alpha}{\alpha-1} \cdot x_1 \\ & x_1 \le 1 \\ & x_i \ge 0 \qquad\qquad\qquad\qquad \text{for } i = 1, 2 \end{array} \tag{5}$$

The value of this linear program is $\frac{s(f)}{\ell} \cdot bs_{\ell-1}(f) \cdot (\alpha + \frac{\alpha}{\alpha-1})$ (achieved at $x_1 = 1$ and $x_2 = \frac{\alpha}{\alpha-1}$). This value attains its minimum at $\alpha = 2$, which gives a value of $\frac{s(f)}{\ell} \cdot bs_{\ell-1}(f) \cdot 4$ to the LP.

What does that mean? It means that $(m_1 + \ldots + m_{\ell-1}) \cdot s(f) \cdot 2/\ell + m_\ell \le \frac{s(f)}{\ell} \cdot bs_{\ell-1} \cdot 4$ for any $(m_1, \ldots, m_\ell)$ disjoint sensitive blocks of size $(1, \ldots, m_\ell)$ respectively. In particular, since $s(f) \cdot 2/\ell \ge 1$ (because $\ell \le s(f)$ WLOG) this inequality bounds $bs_\ell(f)$ from above by $\frac{s(f)}{\ell} \cdot bs_{\ell-1} \cdot 4$.

2.2 Improved Bounds

Kenyon-Kutin [KK04] stopped at this point, seemingly getting the best bound this analysis could offer. This is indeed true if we use only one choice of α, however, one can consider using several different α's to get a better bound, as we do next.

For starters, we show that using two different weights α_1, α_2 gives better bounds on $bs_\ell(f)$ in terms of the $bs_{\ell-1}(f)$ and $s(f)$. The idea is that the solution for the linear program for a certain α_1 implies a new equation for the feasible region of the linear program for α_2.

Recall that choosing $\alpha_1 = 2$ implies that $2 \cdot x_1 + x_2 \le 4$. We now rewrite the linear program for an arbitrary α adding this constraint.

$$\begin{array}{ll} \text{maximize} & \frac{s(f)}{\ell} \cdot bs_{\ell-1}(f) \cdot (\alpha \cdot x_1 + x_2) \\ \\ \text{subject to} & x_2 \le \frac{\alpha}{\alpha-1} \cdot x_1 \\ & 2 \cdot x_1 + x_2 \le 4 \\ & x_1 \le 1 \\ & x_i \ge 0 \qquad\qquad\qquad\qquad \text{for } i = 1, 2 \end{array} \tag{6}$$

One can check that for $\alpha_2 = \frac{4}{3}$ the optimal value for the LP is $\frac{32}{9} \cdot \frac{s(f)}{\ell} \cdot bs_{\ell-1}(f)$. One can now get a new constraint from the linear program for α_2 and continue repeating this process by choosing a sequence of α's. Instead of defining a sequence of α's we will use a continuous strategy.

Theorem 2.2. $\forall f : bs_\ell(f) \leq \frac{e}{\ell} \cdot s(f) \cdot bs_{\ell-1}(f)$.

Proof. We calculate the optimal value for α given an optimal value for $\alpha+\delta$, for an infinitely small $\delta > 0$. Let $\mathrm{OPT}(\alpha)$ be the optimal value of $t(\cdot,\cdot)$ for parameter α, and in order to avoid carrying the multiplicative factor of $bs_{\ell-1}(f) \cdot \frac{s(f)}{\ell}$ let $F(\alpha) = \frac{\mathrm{OPT}(\alpha)}{bs_{\ell-1}(f)\cdot s(f)/\ell}$. The value of the next linear program upper bounds $F(\alpha)$:

$$\begin{array}{ll} \text{maximize} & \alpha \cdot x_1 + x_2 \\ & \\ \text{subject to} & x_2 \leq \frac{\alpha}{\alpha-1} \cdot x_1 \\ & x_1 \leq 1 \\ & x_i \geq 0 \qquad\qquad \text{for } i = 1,2 \end{array} \tag{7}$$

By the definition of $F(\alpha)$ as the normalized optimal value of $t(\cdot,\cdot)$ w.r.t. α we get a new linear equation $\alpha \cdot x_1 + x_2 \leq F(\alpha)$ for all feasible (x_1, x_2). We wish to invoke the equation given by $\alpha+\delta$ on the linear program upper-bounding $F(\alpha)$, for an infinitely small $\delta > 0$.

$$F(\alpha) \leq \begin{cases} \text{maximize} & \alpha \cdot x_1 + x_2 \\ & \\ \text{subject to} & x_2 \leq \frac{\alpha}{\alpha-1} \cdot x_1 \\ & (\alpha+\delta)\cdot x_1 + x_2 \leq F(\alpha+\delta) \\ & x_1 \leq 1 \\ & x_i \geq 0 \qquad\qquad\qquad\qquad \text{for } i = 1,2 \end{cases} \tag{8}$$

Let $(x_1^{\mathrm{OPT}}, x_2^{\mathrm{OPT}})$ be the optimal point for the above LP. In the above LP, x_2 is upper bounded by two linear functions on x_1:

$$x_2 \leq \frac{\alpha}{\alpha-1} \cdot x_1 \quad \text{and} \quad x_2 \leq F(\alpha+\delta) - (\alpha+\delta)\cdot x_1 \; .$$

Since one linear function is increasing and the other is decreasing, the optimal value is achieved either at the intersection of these two lines or at $x_1 = 1$. The intersection point of the two lines, denoted by x_1^{int} is given by

$$x_1^{\mathrm{int}} = \frac{F(\alpha+\delta)}{\frac{\alpha}{\alpha-1} + \alpha + \delta} \; .$$

x_1^{int} is smaller than 1 for $\alpha > 1$ since $F(\alpha+\delta) \leq \frac{\alpha+\delta}{(\alpha+\delta)-1} + \alpha + \delta$ and $\frac{x}{x-1}$ is decreasing for $x > 1$. After the intersection, x_2 decreases faster than $\alpha \cdot x_1$ increases, hence the optimal value of the LP is achieved at the intersection, $x_1^{\mathrm{OPT}} = x_1^{\mathrm{int}}$. The optimal value of x_2 is given by $x_2^{\mathrm{OPT}} = \frac{\alpha}{\alpha-1} \cdot x_1^{\mathrm{OPT}}$, which yields

$$\begin{aligned} F(\alpha) &\leq x_1^{\mathrm{OPT}} \cdot \alpha + x_2^{\mathrm{OPT}} = x_1^{\mathrm{OPT}} \cdot \left(\frac{\alpha}{\alpha-1} + \alpha \right) \\ &= \frac{F(\alpha+\delta)}{\frac{\alpha}{\alpha-1} + \alpha + \delta} \cdot \left(\frac{\alpha}{\alpha-1} + \alpha \right) \\ &= F(\alpha+\delta) \cdot \left(1 - \frac{\delta}{\frac{\alpha}{\alpha-1} + \alpha + \delta} \right) \end{aligned}$$

Rearranging the equation gives

$$\frac{F(\alpha+\delta) - F(\alpha)}{\delta} \leq \frac{F(\alpha+\delta)}{\frac{\alpha}{\alpha-1} + \alpha + \delta} \; ,$$

and as δ tends to 0 we get $F'(\alpha) \le \frac{F(\alpha)}{\frac{\alpha}{\alpha-1}+\alpha} = F(\alpha)\cdot\frac{\alpha-1}{\alpha^2}$. The solution for this ODE is $F(\alpha) \le \alpha\cdot e^{\frac{1}{\alpha}}\cdot c$ for some constant $c > 0$. Taking an initial condition on $\alpha \gg 1$: $F(\alpha) \le \alpha + \frac{\alpha}{\alpha-1}$ gives

$$c \le \frac{F(\alpha)}{\alpha \cdot e^{\frac{1}{\alpha}}} \le \frac{\alpha\cdot(1+\frac{1}{\alpha-1})}{\alpha \cdot e^{\frac{1}{\alpha}}} \underset{\alpha\to\infty}{\to} 1 \ .$$

Hence, $F(\alpha) \le \alpha\cdot e^{\frac{1}{\alpha}}$. When α approaches 1 we get $\lim_{\alpha\to 1^+} F(\alpha) \le e$, thus $bs_\ell(f) \le \frac{e}{\ell}\cdot s(f)\cdot bs_{\ell-1}(f)$ completing the proof. □

As a special case, Theorem 2.2 implies that $bs_2(f) \le \frac{e}{2}\cdot s(f)^2$, which leads us to the following open problem.

Open Problem 1. *What is the smallest constant $c > 0$ such that $bs_2(f) \le c\cdot s(f)^2$ for all Boolean functions?*

An example with $bs_2(f) = \frac{2}{3}\cdot s(f)^2\cdot(1-o(1))$ is given in [AS11], thus $\frac{2}{3} \le c \le \frac{e}{2}$.

3 Understanding $bs_3(f)$ is Important

As the upper and lower bounds for $bs_2(f)$ are almost matching, it seems that the next challenge is understanding the asymptotic behavior of $bs_3(f)$. A more modest challenge is the following.

Open Problem 2. *Improve either the upper or lower bound on $bs_3(f)$.*

Recall that the upper bound on $bs_3(f)$ is $O(s(f)^3)$ (see Eq.(2)) and the lower bound is $(2/3)\cdot s(f)^2\cdot(1-o(1))$. It is somewhat surprising that any slight improvement on either the lower or upper bound on bs_3 would be a significant step forward in our understanding of the general question. The following claim shows that a slightly better than quadratic gap on a single example implies a better than quadratic gap on an infinite family of examples.

Claim 3.1. *If there exists a function such that $bs_3(f) > s(f)^2$ then there exists a family of functions $\{f_n\}_{n\in\mathbb{N}}$ with $bs(f_n) > s(f_n)^{2+\varepsilon}$ for some constant $\varepsilon > 0$ (dependant on f).*

This family is simply $f_1 = f$, $f_n = f \circ f_{n-1}$ where $\circ$ stands for Boolean function composition as in [Tal13]. Next, we prove a theorem exhibiting the self-reducibility nature of the problem.

Theorem 3.2. *Let $k, \ell, a \in \mathbb{N}$ such that $\ell > k$ and let $T : \mathbb{N} \to \mathbb{R}$ be a monotone function.*
If $\forall f:\ bs_\ell(f) \le T(bs_k(f))$, then $\forall f':\ bs_{\ell a}(f') \le T(bs_{ka}(f'))$.

Proof. Assume by contradiction that there exists a function f' such that $bs_{\ell a}(f') > T(bs_{ka}(f'))$. We will show that there exists a function f such that $bs_\ell(f) > T(bs_k(f))$. We shall assume WLOG that the maximal $bs_{\ell a}$ of f' is achieved on $\vec{0}$. Let $B_1, B_2, \ldots, B_m$ be a family of disjoint sensitive blocks for f at $\vec{0}$, each B_i of size at most ℓa. Split every block B_i to ℓ sets $B_{i,1}, \ldots, B_{i,\ell}$ of size at most a. The function f will have a variable $x_{i,j}$ corresponding to every set $B_{i,j}$ of size at most a. The value of $f(x_{1,1}, \ldots, x_{m,\ell})$ is defined to be the value of f' where the variable in each $B_{i,j}$ equal $x_{i,j}$, and all other variables equal 0. $bs_\ell(f, \vec{0}) \ge bs_{\ell a}(f', \vec{0})$, since for any sensitive block $B_1, \ldots, B_m$ for f', there exists a corresponding sensitive block $B'_1, \ldots, B'_m$ for f of size ℓ, where $B'_i = \{x_{i,j} : j \in [\ell]\}$.

On the other hand, any set of disjoint sensitive blocks of size at most k for f corresponds to a disjoint set of sensitive blocks of size at most ka for f'. Thus $bs_k(f) \le bs_{ka}(f')$, giving

$$T(bs_k(f)) \le T(bs_{ka}(f')) < bs_{\ell a}(f') \le bs_\ell(f) \ ,$$

where we used the monotonicity of T in the first inequality. □

Using Theorem 3.2 we get that any upper bound of the form $bs_\ell(f) \le s(f)^{\ell-\varepsilon}$ implies a sub-exponential upper bound on $bs(f)$ in terms of $s(f)$.

Theorem 3.3. *Let $k \in \mathbb{N}, \varepsilon > 0$ be constants. If for all Boolean functions $bs_k(f) \le s(f)^{k-\epsilon}$, then for the constant $\gamma = \frac{\log(k-\varepsilon)}{\log(k)} < 1$ it holds that $bs(f) \le 2^{O(s(f)^\gamma \cdot \log s(f))}$ for all f.*

For example, Theorem 3.3 shows that if $\forall f : bs_3(f) \le s(f)^2$, then $\forall f : bs(f) \le 2^{O(s^{0.631} \cdot \log(s))}$.

Proof. Using the hypothesis and Theorem 3.2 one can show by induction on t that

$$\forall f : bs_{k^t}(f) \le s(f)^{(k-\epsilon)^t} . \tag{9}$$

The base case $t = 1$ is simply the hypothesis. We assume the claim is true for $1, \ldots, t-1$, and show the claim is true for t. Using Theorem 3.2 with $T(x) = x^{k-\epsilon}$ and $a = k^{t-1}$ we get $bs_{k^t}(f) \le T(bs_{k^{t-1}}(f)) = (bs_{k^{t-1}}(f))^{k-\epsilon}$. By induction $bs_{k^{t-1}}(f) \le s(f)^{(k-\epsilon)^{t-1}}$. Hence, we get $bs_{k^t}(f) \le s(f)^{(k-\epsilon)^t}$, which finishes the induction proof.

Fix f and let $s = s(f)$. Recall that $bs(f) = bs_s(f)$ since each minimal block that flips the value of f is of size at most s. Hence,

$$\begin{aligned} bs(f) &= bs_s(f) = bs_{k^{\lceil \log_k(s) \rceil}}(f) \\ &\le s^{(k-\epsilon)^{\lceil \log_k(s) \rceil}} \le s^{(k-\epsilon)^{\log_k(s)+1}} = 2^{\log(s) \cdot s^{\log(k-\epsilon)/\log(k)} \cdot (k-\epsilon)} = 2^{O(s^\gamma \cdot \log(s))} . \end{aligned}$$

□

4 The Sensitivity Conjecture for Bounded Functions

In this section, we generalize the definitions of sensitivity and block sensitivity to bounded functions $f : \{0,1\}^n \to [0,1]$, extending the definitions for Boolean functions. We generalize the result of Kenyon and Kutin to this setting (after removing some trivial obstucles). Given that, one may hope that the sensitivity conjecture holds also for bounded functions, i.e., that the block-sensitivity is at most polynomial in the sensitivity. However, we give a counterexample to this question, by constructing functions on n variables with sensitivity $O(\log n)$ and block sensitivity $n/\log(n)$. In fact, we show that the result of Kenyon and Kutin is essentially tight by giving examples for which $bs_\ell(f) = n/\ell$ and $s(f) = O(\ell \cdot n^{1/\ell})$ for any $\ell \le \log n$.

We begin by generalizing the definitions of sensitivity and block-sensitivity. For $f : \{0,1\}^n \to [0,1]$ and $x \in \{0,1\}^n$, we denote the sensitivity of f at a point x by

$$s(f,x) = \sum_{i=1}^{n} |f(x) - f(x \oplus e_i)|. \tag{10}$$

Similarly we define the block sensitivity and ℓ-block sensitivity as

$$bs(f,x) = \max \left\{ \sum_i |f(x) - f(x+B_i)| \ : \ B_1, \ldots, B_k \subseteq [n] \text{ are disjoint} \right\} . \tag{11}$$

and

$$bs_\ell(f,x) = \max \left\{ \sum_i |f(x) - f(x+B_i)| \ : \ B_1, \ldots, B_k \subseteq [n] \text{ are disjoint and } \forall i . |B_i| \le \ell \right\} .$$

Naturally we denote by $s(f) = \max_x s(f,x)$, by $bs(f) = \max_x bs(f,x)$ and by $bs_\ell(f) = \max_x bs_\ell(f,x)$. It is easy to see that for a Boolean function these definitions match the standard definitions of sensitivity, block sensitivity and ℓ-block sensitivity.

We wish to prove an analog of Kenyon-Kutin result, showing that $bs_\ell(f) \le c_\ell \cdot s(f)^\ell$. However, stated as is the claim is false for a "silly" reason. Take any Boolean function f with a gap between the sensitivity and the ℓ-block sensitivity and take $g(x) = f(x)/s(f)$. Then, we get $s(g) = 1$ and $bs_\ell(g) = bs_\ell(f)/s(f)$. As there are examples with $bs_2(f) = n/2$ and $s(f) = \sqrt{n}$, we get that $bs_2(g) = \sqrt{n}/2$ while $s(g) = 1$, where n grows to infinity. This seems to rule out any relation between the sensitivity and block sensitivity (and even 2-block sensitivity) in the case of bounded functions. To overcome this triviality, we insist that the block sensitivity is close to n, or alternatively that changing each block dramatically changes the value of the function. Surprisingly, under this requirement we are able to retrieve known relations between sensitivity and block sensitivity that were established in the Boolean setting by Kenyon and Kutin [KK04].

Theorem 4.1. *Let $c > 0$ and $f : \{0,1\}^n \to [0,1]$. Assume that there exists a point $x_0 \in \{0,1\}^n$ and disjoint blocks $B_1, \ldots, B_k$ of size at most ℓ such that $|f(x_0) - f(x_0 + B_i)| \ge c$ for all $i \in k$. Furthermore, assume that $2 \le \ell \le \log(k)$. Then, $s(f) \ge \Omega(k^{1/\ell} \cdot c)$.*

We get the following corollary, whose proof is deferred to Appendix A.

Corollary 4.2. *Let $f : \{0,1\}^n \to [0,1]$ with $bs(f) \ge n/\ell$. Then, $s(f) \ge \Omega(n^{1/2\ell}/\ell)$.*

Unlike in the Boolean case, we are able to show that Theorem 4.1 is essentially tight! That is, for any ℓ and n we have a construction with $bs_\ell(f) \ge n/\ell$ and $s(f) = O(\ell \cdot n^{1/\ell})$. In particular, picking $\ell = \log(n)$ gives an exponential separation between block sensitivity (which is at least $n/\log n$) and sensitivity (which is $O(\log n)$).

Theorem 4.3. *Let $\ell, n \in \mathbb{N}$ with $2 \le \ell \le n$. Then, there exists a function $h : \{0,1\}^n \to [0,1]$ with $bs_\ell(h) \ge \lfloor n/\ell \rfloor$ and $s(h) \le 3 \cdot \ell \cdot n^{1/\ell}$.*

4.1 Proof of Kenyon-Kutin Result for Bounded Functions

Proof Overview. We start by giving a new proof for Kenyon-Kutin result, based on random walks on the hypercube. We assume by contradiction that $f(x_0) = 0$ and $f(x_0 + B_i) = 1$ for all $i \in [k]$ and that the sensitivity is $o(k^{1/\ell})$. Taking a random walk of length $r = n/k^{1/\ell}$ starting from x_0 will end up in point y where with high probability $f(y) = f(x_0)$. This is true since in each step with probability at least $1 - s(f)/n$ we are maintaining the value of f, hence by union bound with probability at least $1 - r \cdot s(f)/n$ we maintain the value of f in the entire walk. On the contrast, choosing a random $i \in [k]$ and starting a random walk of length $r - |B_i|$ starting from $(x_0 + B_i)$ will lead to a point y' where with high probability $f(y') = f(x_0 + B_i) = 1$. However, as we show in the proof below, the distributions of y and y' are similar (close in statistical distance). This leads to a contradiction as $f(y)$ tends to be equal to 0 and $f(y')$ tends to be equal to 1.

A simple observation, which allows us to generalize the argument above to bounded function, is that for a given point $x \in \{0,1\}^n$ and a random neighbor in the hypercube, $y \sim x$, the expected value of $f(y)$ is close to $f(x)$. This follows from Eq. (10). Thus, the only difference in the argument for bounded functions will be that $\mathbf{E}[f(y)]$ is close to 0 and $\mathbf{E}[f(y')]$ is close to 1, leading to a contradiction as well.

Proof of Theorem 4.1. First, we make a few assumptions that are without loss of generality, in order to make the argument later clearer. We assume $x_0 = 0^n$ and $f(x_0) = 0$. We assume $n = k \cdot \ell$

and that the blocks are given by $B_i = \{(i-1)\ell+1, \ldots, i\ell\}$ for $i \in [k]$. We assume that $c = 1$, since for $c < 1$ one can take $f'(x) = \min\{f(x)/c, 1\}$, and note that f' is a bounded function with $f'(x_0 + B_i) = 1$. Proving the theorem for f' gives $s(f) \geq s(f') \cdot c \geq \Omega(c \cdot k^{1/\ell})$.

Let $r = \lfloor \frac{n}{(2k)^{1/\ell}} \rfloor$, by the assumption $2 \leq \ell \leq \log(k)$ we have $\sqrt{n} \leq r \leq n/2$. Assume by contradiction that $s(f) \leq \varepsilon \cdot k^{1/\ell}$ for some sufficiently small constant $\varepsilon > 0$ to be determined later. Consider the following two random processes.

Algorithm 1 Process A

$X_0 \leftarrow 0^n$
for $t = 1, \ldots, r$ **do**
    Select a random $i \in [n]$ among the coordinates for which X_{t-1} is 0 and let $X_t \leftarrow X_{t-1} + e_i$.
end for

Algorithm 2 Process B

Select uniformly $i \in [k]$ and let $Y_0 \leftarrow B_i$
for $t = 1, \ldots, r - \ell$ **do**
    Select a random $i \in [n]$ among the coordinates for which Y_{t-1} is 0 and let $Y_t \leftarrow Y_{t-1} + e_i$.
end for

For each $t \in \{0, \ldots, r-1\}$, we claim that

$$\begin{aligned}\mathbf{E}[f(X_{t+1}) - f(X_t)] &= \mathbf{E}\left[\frac{1}{n-t} \cdot \sum_{i:(X_t)_i=0} f(X_t + e_i) - f(X_t)\right] \\ &\leq \frac{1}{n-t} \cdot \mathbf{E}[s(f(X_t))] \leq \frac{s(f)}{n-t} \, .\end{aligned}$$

By telescoping this implies that

$$\mathbf{E}[f(X_r)] = \mathbf{E}[f(X_0)] + \sum_{t=0}^{r-1} \mathbf{E}[f(X_{t+1}) - f(X_t)] \leq 0 + \frac{r \cdot s(f)}{n-r} \leq O(\varepsilon) \, .$$

In a symmetric fashion, for each $t \in \{1, \ldots, r-\ell\}$ we have $\mathbf{E}[f(Y_{t+1}) - f(Y_t)] \geq -\frac{s(f)}{n-t-\ell}$. Again, telescoping implies that

$$\mathbf{E}[f(Y_{r-\ell})] \geq \mathbf{E}[f(Y_0)] - \frac{(r-\ell) \cdot s(f)}{n-r} \geq 1 - \frac{r \cdot s(f)}{n-r} \geq 1 - O(\varepsilon) \, .$$

So it seems that the distribution of X_r and $Y_{r-\ell}$ are very different from one another. However, we shall show that conditioned on a probable event, X_r and $Y_{r-\ell}$ are identically distributed. To define the event, consider the sets

$$U_i = \{\mathbb{1}_A \mid A \subseteq [n], |A| = r, B_i \subseteq A, \forall j \neq i : B_j \nsubseteq A\}$$

for $i \in [k]$ and their union

$$U = \bigcup_{i=1}^{k} U_i = \{\mathbb{1}_A \mid A \subseteq [n], |A| = r, \exists! i \in [k] : B_i \subseteq A\} \, .$$

Let E_X be the event that $X_r \in U$, and E_Y be the event that $Y_{r-\ell} \in U$. We show that

Claim 4.4. *The following hold:*

1. $X_r|E_X$ *is identically distributed as* $Y_{r-\ell}|E_Y$.
2. $\mathbf{Pr}[E_Y] = \Omega(1)$
3. $\mathbf{Pr}[E_X] = \Omega(1)$

We defer the proof of Claim 4.4 for later. We derive a contradiction from all of the above by showing that $\mathbf{E}[f(X_r)|E_X] < \mathbf{E}[f(Y_{r-\ell})|E_Y]$ (this is indeed a contradiction because by the claim $X_r|E_X$ and $Y_{r-\ell}|E_Y$ should be identically distributed and hence the expected values of $f(\cdot)$ on each of them should be the same). To show this, we note that

$$\begin{aligned}\mathbf{E}[f(X_r)|E_X] &= \mathbf{E}[f(X_r) \cdot \mathbb{1}_{E_X}]/\,\mathbf{Pr}[E_X] \\ &\le \mathbf{E}[f(X_r)]/\,\mathbf{Pr}[E_X] = O(\mathbf{E}[f(X_r)]) = O(\varepsilon)\ .\end{aligned}$$

On the other hand

$$\begin{aligned}\mathbf{E}[f(Y_{r-\ell})|E_Y] &= 1 - \mathbf{E}[1 - f(Y_{r-\ell})|E_Y] \\ &\ge 1 - \mathbf{E}[1 - f(Y_{r-\ell})]/\,\mathbf{Pr}[E_Y] = 1 - O(\mathbf{E}[1 - f(Y_{r-\ell})]) = 1 - O(\varepsilon)\ .\end{aligned}$$

Choosing ε to be a small enough constant implies that $\mathbf{E}[f(X_r)|E_X] < \mathbf{E}[f(Y_{r-\ell})|E_Y]$, which completes the proof. $\square$

Proof of Claim 4.4. We shall use in the proof of Items 2 and 3 the fact that $1/3 \le \frac{r^\ell k}{n^\ell} \le 1/2$ which follows from the choice of $r = \lfloor \frac{n}{(2k)^{1/\ell}} \rfloor$ (for large enough n and k).

1. First note that X_r is distributed uniformly over the set of vectors in $\{0,1\}^n$ with hamming weight r. In particular, conditioning that X_r is in a set U of such vectors, makes it uniform over U. We are left to show that $Y_{r-\ell}|E_Y$ is distributed uniformly over U. Given that $Y_0 = B_i$, we have that $Y_{r-\ell}$ is the OR of $\mathbb{1}_{B_i}$ with a random vector of weight $r-\ell$ on $[n] \setminus B_i$. Conditioned on E_Y the only way to reach U_i is if $Y_0 = B_i$, hence, by the above, all points in U_i are attained with the same probability. Using symmetry, all points in $U = \bigcup_i U_i$ are attained with the same probability.

2. Let B_i be the block selected in the first step of Process B. We analyze the probability that all indices in B_j for some $j \neq i$ are chosen in the $r-\ell$ iterations of Process B.

$$\begin{aligned}\mathbf{Pr}[B_j \text{ is selected}] &= \frac{(\#\text{ of sequences where } B_j \text{ is selected})}{(\#\text{ of sequences})} \\ &= \frac{(r-\ell)^{\underline{\ell}} \cdot (n-2\ell)^{\underline{r-2\ell}}}{(n-\ell)^{\underline{r-\ell}}} = \frac{(r-\ell)!(n-2\ell)!(n-r)!}{(r-2\ell)!(n-r)!(n-\ell)!} \\ &= \frac{(r-\ell)!(n-2\ell)!}{(r-2\ell)!(n-\ell)!} = \frac{(r-\ell)\cdots(r-2\ell+1)}{(n-\ell)\cdots(n-2\ell+1)} \le \left(\frac{r}{n}\right)^\ell\end{aligned}$$

(recall that $n^{\underline{k}} \triangleq \frac{n!}{(n-k)!}$). Hence, $\mathbf{Pr}[\exists j \neq i : B_j \text{ is selected}] \le k \cdot (r/n)^\ell \le 1/2$ and we have $\mathbf{Pr}[E_Y] \ge 1/2$.

3. Let $\pi_1, \ldots, \pi_r \in [n]$ be the sequence of choices made by Process A. For $i \in [k]$, let $E_{X,i}$ be the event that $X_r \in U_i$. By the uniqueness of the block contained in X_r the events $E_{X,i}$ are disjoint, hence $\mathbf{Pr}[E_X] = \sum_{i=1}^k \mathbf{Pr}[E_{X,i}]$. By symmetry, $\mathbf{Pr}[E_X] = k \cdot \mathbf{Pr}[E_{X,1}]$. The event

$E_{X,1}$ is simply the event that there exists a set $S \subseteq [r]$ of size ℓ such that $\{\pi_j\}_{j\in S} = B_1$ and the sequence $\{\pi_j : j \in [r] \setminus S\}$ is a sequence of choices for which E_Y holds, when starting Process B from $Y_0 = B_1$. This shows that $\mathbf{Pr}[E_{X,1}] = \mathbf{Pr}[E_Y|Y_0 = B_1] \cdot \mathbf{Pr}[B_1 \subseteq \{\pi_1, \dots, \pi_r\}]$. By Symmetry, $\mathbf{Pr}[E_Y|Y_0 = B_i] = \mathbf{Pr}[E_Y] = \Omega(1)$ from the previous item. In addition,

$$\begin{aligned}\mathbf{Pr}[B_1 \subseteq \{\pi_1, \dots, \pi_r\}] &= \frac{r^{\underline{\ell}} \cdot (n-\ell)^{\underline{r-\ell}}}{n^{\underline{r}}} = \frac{r!(n-\ell)!(n-r)!}{(r-\ell)!(n-r)!n!} \\ &= \frac{r!(n-\ell)!}{(r-\ell)!n!} = \frac{r\cdots(r-\ell+1)}{n\cdots(n-\ell+1)} \geq \left(\frac{r-\ell}{n}\right)^{\ell} \\ &= \left(\frac{r}{n}\right)^{\ell} \cdot (1-\ell/r)^{\ell} = \left(\frac{r}{n}\right)^{\ell} \cdot (1-o(1))\end{aligned}$$

where $(1-\ell/r)^{\ell} = 1 - o(1)$ follows from $\ell \leq \log(k)$ and $r \geq \sqrt{n} \geq \sqrt{k}$. Thus,

$$\begin{aligned}\mathbf{Pr}[E_X] = k \cdot \mathbf{Pr}[E_{X,1}] &= k \cdot \mathbf{Pr}[B_1 \text{ is selected}] \cdot \mathbf{Pr}[E_Y|Y_0 = B_1] \\ &\geq k \cdot \left(\frac{r}{n}\right)^{\ell} \cdot (1-o(1)) \cdot \frac{1}{2} \geq \frac{1}{3} \cdot (1-o(1) \cdot \frac{1}{2} = \Omega(1) .\end{aligned}$$ □

4.2 Separating Sensitivity and Block Sensitivity of Bounded Functions

The Lattice Variant of The Sensitivity Conjecture The proof of Theorem 4.3 is more natural in the lattice-variant of the sensitivity conjecture as suggested by Aaronson (see [Bop12]). In this variant, instead of talking about functions over $\{0,1\}^n$ we are considering functions over $\{0,1,\dots,\ell\}^k$ for $\ell, k \in \mathbb{N}$. Given a function $g : \{0,1,\dots,\ell\}^k \to \mathbb{R}$ one can define a Boolean function $f : \{0,1\}^{\ell\cdot k} \to \mathbb{R}$ by the following equation:

$$f(x_{1,1}, \dots, x_{k,\ell}) = g\left(\sum_{i=1}^{\ell} x_{1,i}, \dots, \sum_{i=1}^{\ell} x_{k,i}\right) . \tag{12}$$

For a point $y \in \{0,1,\dots,\ell\}^k$ and function $g : \{0,\dots,\ell\}^k \to \mathbb{R}$ one can define the sensitivity of g at y as

$$s(g,y) = \sum_{y'\sim y} |g(y') - g(y)|$$

where $y' \sim y$ if $y' \in \{0,\dots,\ell\}^k$ is a neighbor of y in the grid $\{0,\dots,\ell\}^k$, i.e., if y and y' agree on all coordinates except for one coordinate, say $j \in [k]$, on which $|y_j - y'_j| = 1$. The following claim relates the sensitivity of f to that of g.

Claim 4.5. *Let $g : \{0,\dots,\ell\}^k \to \mathbb{R}$ and let f be the function defined by Eq. (12). Then $s(f) \leq \ell \cdot s(g)$.*

Proof. Let $x = (x_{1,1}, \dots, x_{k,\ell}) \in \{0,1\}^{kl}$ and and let $x' \in \{0,1\}^{kl}$ be a neighbor of x, obtained by flipping the (i,j)-th coordinate. Let $y = (\sum_{i=1}^{\ell} x_{1,i}, \dots, \sum_{i=1}^{\ell} x_{k,i})$ and similarly let $y' = (\sum_{i=1}^{\ell} x'_{1,i}, \dots, \sum_{i=1}^{\ell} x'_{k,i})$. Then y and y' differ only on the i-th coordinate, and on this coordinate they differ by a ± 1. If $y'_i = y_i + 1$, then the number of neighbors $x' \sim x$ that are mapped to y' by $y' = (\sum_i x'_{1,i}, \dots, \sum_i x'_{k,i})$ equals the number of zeros in the i-th block of x, i.e., it equals $\ell - y_i$. Similarly, in the case $y'_i = y_i - 1$ the number of $x' \sim x$ that are mapped to y' equals y_i. In both cases, there are between 1 to ℓ points $x' \sim x$ that are mapped to each neighbor $y' \sim y$. Thus,

$$\sum_{x'\sim x} |f(x') - f(x)| = \sum_{x'\sim x} |g(y') - g(y)| \leq \ell \cdot \sum_{y'\sim y} |g(y') - g(y)| .$$ □

Construction of a Separation. Let k,ℓ be integers. We construct $f:\{0,1,\dots,\ell\}^k \to [0,1]$ such that $f(0)=0$, $f(e_i\cdot\ell)=1$ for all $i\in[k]$ and $s(f)\le O(k^{1/\ell})$.

Define a weight function $w:\{0,1,\dots,\ell\}\to[0,1]$ as follows: $w(a)=k^{a/\ell}/k$ for $a\in\{1,\dots,\ell\}$ and $w(0)=0$. Take $g:\{0,\dots,\ell\}^k\to\mathbb{R}^+$ to be the function $g(x_1,\dots,x_n)=\sum_{i=1}^k w(x_i)$ and take $f:\{0,\dots,\ell\}^k\to[0,1]$ to be $f(x)=\min\{1,g(x)\}$. Then $f(0^k)=0$ and $f(\ell\cdot e_i)=1$ for all $i\in[k]$.

Theorem 4.6. $s(f)\le 3\cdot k^{1/\ell}$.

Proof. Let $x\in\{0,1,\dots,\ell\}^k$ be a point in the lattice. We distinguish between two cases $g(x)\ge 2$ and $g(x)<2$. In the first case, all neighbors $x'\sim x$ have $g(x')\ge 1$ since the sums $\sum_i w(x_i)$ and $\sum_i w(x'_i)$ differ by at most 1. Since both $g(x)$ and $g(x')$ are at least 1 we get that $f(x)=f(x')=1$ and the sensitivity of f at x is 0.

In the latter case, $g(x)<2$, we bound the sensitivity as well. For ease of notation we extend w to be defined over $\{-1,\dots,\ell+1\}$ by taking $w(\ell+1)=w(\ell)$ and $w(-1)=w(0)$. We extend also g to $\{-1,0,\dots,\ell+1\}\to\mathbb{R}^+$ by taking $g(x_1,\dots,x_n)=\sum_i w(x_i)$. We have

$$
\begin{aligned}
s(f,x)\le s(g,x) &= \sum_{i=1}^k |g(x+e_i)-g(x)|+|g(x)-g(x-e_i)|\\
&= \sum_{i=1}^k |w(x_i+1)-w(x_i)|+|w(x_i)-w(x_i-1)|\\
&= \sum_{i=1}^k w(x_i+1)-w(x_i-1) && (w \text{ is monotone})\\
&\le \sum_{i=1}^k w(x_i+1) && (w \text{ is non-negative})\\
&\le \sum_{i:x_i=0} w(1)+\sum_{i:x_i>0} w(x_i)\cdot k^{1/\ell}\\
&\le k\cdot\frac{k^{1/\ell}}{k}+\sum_i w(x_i)\cdot k^{1/\ell}\\
&= k^{1/\ell}+g(x)\cdot k^{1/\ell}\le 3k^{1/\ell}. \qquad\square
\end{aligned}
$$

We show that Theorem 4.3 is a corollary of Theorem 4.6.

Proof of Theorem 4.3. Let $k=n/\ell$. Let $f:\{0,1,\dots,\ell\}^k\to[0,1]$ be the function in Theorem 4.6. Take $h(x_{1,1},\dots,x_{k,\ell})=f\left(\sum_{i=1}^\ell x_{1,i},\dots,\sum_{i=1}^\ell x_{n,i}\right)$. For $x=0^n$, there are k disjoint blocks $B_1,\dots,B_k$ of size ℓ each such that $h(x+B_i)=1$. Hence, $bs_\ell(h)\ge k=n/\ell$. By Claim 4.5, the sensitivity of h is at most $s(f)\cdot\ell\le 3\cdot k^{1/\ell}\cdot\ell\le 3\cdot n^{1/\ell}\cdot\ell$ which completes the proof. $\square$

5 New Separations between Decision Tree Complexity and Sensitivity

We report a new separation between the decision tree complexity and the sensitivity of Boolean functions. We construct an infinite family of Boolean functions with

$$\mathrm{DT}(f_n)\ge s(f_n)^{1+\log_{14}(19)}\ge s(f_n)^{2.115}\,.$$

Our functions are transitive functions, and are inspired by the work of Chakraborty [Cha11].

Our construction is based on finding a "gadget" Boolean function f, defined over a constant number of variables, such that $s^0(f) = 1$, $s^1(f) = k$ and $\mathrm{DT}(f) = \ell$ for $\ell > k$ (recall that $s^0(f) = \max_{x:f(x)=0} s(f,x)$ and similarly $s^1(f) = \max_{x:f(x)=1} s(f,x)$). Given the gadget f, we construct an infinite family of functions with super-quadratic gap between the sensitivity and the decision tree complexity using compositions (which is a well-used trick in query complexity separations, cf. [Tal13]).

Lemma 5.1. *Let $f : \{0,1\}^c \to \{0,1\}$ such that $s^0(f) = 1$, $s^1(f) = k$ and $\mathrm{DT}(f) = \ell > k$. Then, there exists an infinite family of functions $\{g_i\}_{i\in\mathbb{N}}$ such that $s(g_i) = k^i$ and $\mathrm{DT}(g_i) = (k\ell)^i = s(g_i)^{1+\log(k)/\log(\ell)}$.*

Proof. Take $g = \mathrm{OR}_k \circ f$. It is easy to verify that $s(g) = k$, and that $\mathrm{DT}(g) = \mathrm{DT}(\mathrm{OR}_k)\cdot\mathrm{DT}(f) = k\ell$ (for the latter, one can use [Tal13, Lemma 3.1]). For $i \in \mathbb{N}$, we take $g_i = g^i$. It is well-known (cf. [Tal13, Lemma 3.1]) that $s(g^i) \le s(g)^i$ and that $\mathrm{DT}(g^i) = \mathrm{DT}(g)^i$, which completes the proof. ☐

5.1 Finding a Good Gadget

The gadget f will be a minterm-cyclic function. Roughly speaking, a function $f : \{0,1\}^n \to \{0,1\}$ is minterm-cyclic if there exists pattern $p \in \{0,1,*\}^n$ such that the function f simply checks if x matches one of the cyclic shifts of p. The formal definition follows

Definition 5.2. *A* pattern *$p \in \{0,1,*\}^n$ is a partial assignment to the variables $x_1,\ldots,x_n$. We say that a point $x \in \{0,1\}^n$* matches *the pattern p, denoted by $p \subseteq x$, if for all $i \in [n]$ such that $p_i \in \{0,1\}$ we have $p_i = x_i$. Given a pattern p, let $CS(p) = \{p^1,\ldots,p^n\}$ be the set of cyclic shifts of p, where the i-th cyclic shift of p is given by $p^i = (p_i, p_{i+1},\ldots,p_n,p_1,\ldots,p_{i-1})$. For a pattern $p \in \{0,1,*\}^n$ we denote by $f_p : \{0,1\}^n \to \{0,1\}$ the function defined by*

$$f_p(x) = 1 \quad\Leftrightarrow\quad \exists p^i \in CS(p) : p^i \subseteq x$$

and call f_p the minterm cyclic function defined by *p.*

For example, the pattern $p = \texttt{0011**}$ defines a function f_p that checks if there's a sequence of two zeros followed by two ones in x, when x is viewed as a cyclic string. We say that two patterns $p, q \in \{0,1,*\}^n$ disagree on a coordinate i if both p_i and q_i are in $\{0,1\}$ and $p_i \neq q_i$.

Claim 5.3. *Let $p \in \{0,1,*\}^n$ be a pattern defining $f_p : \{0,1\}^n \to \{0,1\}$. Assume that any two different cyclic-shifts of p disagree on at least 3 coordinates. Then, $s^0(f_p) = 1$.*

Proof. Let $x \in \{0,1\}^n$ with $f_p(x) = 0$ and assume by contradiction that $s(f_p,x) \ge 2$. In such a case, there are two indices i and j such that $f_p(x+e_i) = 1$ and $f_p(x+e_j) = 1$. Let q and q' be the patterns among $CS(p)$ that $x+e_i$ and $x+e_j$ satisfy respectively. If $q = q'$, then since both $x+e_i$ and $x+e_j$ satisfy q and they differ on coordinates i and j, it must be the case that $q_i = q_j = *$. However, this implies that x satisfy q as well, which is a contradiction. If $q \neq q'$, then we get that q and q' may disagree only on coordinates i and j, which is also a contradiction. ☐

The following fact is easy to verify.

Fact 5.4. *Let $p \in \{0,1,*\}^n$ be a pattern defining $f_p : \{0,1\}^n \to \{0,1\}$. Then, $s^0(f_p) \le c^0(f_p) \le |\{i \in [n] : p_i \in \{0,1\}\}|$.*

Next, we demonstrate a simple example with better-than-quadratic separation between $\mathrm{DT}(f)$ and $s(f)$. Take the pattern $p = *001011$. Denote by $p^1, \ldots, p^7$ all the cyclic shifts of p, where in p^i the i-th coordinate equals $*$. It is easy to verify that any p^i and p^j for $i \neq j$ disagree on at least 3 coordinates. Hence, $s^0(f_p) = 1$ and $s^1(f_p) \le 6$. We wish to show that any decision tree T for f_p is of depth 7. Let x_i be the first coordinate read by a decision tree T for f_p. Our adversary will answer 0, and will continue to answer as if x matches p^i. Assume the decision tree made a decision before reading the entire input. The decision tree must decide 1 since the adversary answered according to x which satisfies p^i. However, if the decision tree hasn't read the entire input, there is still an unread coordinate j, where $j \neq i$. Let $x' = x + e_j$. Then, the decision tree answers 1 on x' as well. However x' does not match pattern p^i as $(p^i)_j \in \{0,1\}$ and it must be the case that $x_j = (p^i)_j \neq x'_j$.

We also need to rule out that x' matches some other pattern. Indeed, if x' matches some other pattern p^k it means that p^k and p^i disagree only on at most one coordinate, which as discussed above cannot happen.

Using Lemma 5.1 the function f_p can be turned into an infinite family of functions g_i with $\mathrm{DT}(g_i) = (6 \cdot 7)^i$ and $s(g_i) \le 6^i$. This gives a super-quadratic separation since

$$\mathrm{DT}(g_i) \ge s(g_i)^{1+\log(7)/\log(6)} \ge s(g_i)^{2.086} .$$

In a similar fashion, one can show that for the pattern $p =$ `**0*10000*101` after reading any two input bits from the input there exists a cyclic shift p^i of the pattern from which no $\{0,1\}$ coordinate has been read yet. However, to verify that the input x matches p^i we must read all $\{0,1\}$ positions in p^i, which gives $\mathrm{DT}(f_p) \ge 9 + 2$ where 9 is the number of $\{0,1\}$-s in the pattern p.

The decision tree complexity analysis for the other patterns written below is more involved. We computed it using a computer program written to calculate the decision tree complexity in this special case. In the list below, we report several patterns yielding super-quadaratic separations. For each pattern p we report its length n, the decision tree complexity of f_p, the maximal sensitivity of f_p (which equals the number of $\{0,1\}$-s in p) and the resulting exponent one get by using Lemma 5.1 (i.e., $1 + \frac{\log \mathrm{DT}(f_p)}{\log s(f_p)}$).

```
p = *001011,                         n = 7,  DT = 7,  s = 6,  exp = 2.086
p = **0*10000*101,                   n = 13, DT = 11, s = 9,  exp = 2.091
p = ******01*1*01100000,             n = 19, DT = 14, s = 11, exp = 2.100
p = ******00*0*0010**1*00*011,       n = 25, DT = 17, s = 13, exp = 2.104
p = ******1**0**0**1**0**00*0*10*1011, n = 33, DT = 19, s = 14, exp = 2.115
```

Acknowledgements. I wish to thank my PhD advisor, Ran Raz, for lots of stimulating and helpful discussions about this problem. I wish to thank Scott Aaronson for his encouragement.

References

[ABG+14] A. Ambainis, M. Bavarian, Y. Gao, J. Mao, X. Sun, and S. Zuo. Tighter relations between sensitivity and other complexity measures. In *ICALP (1)*, pages 101–113, 2014.

[AP14] A. Ambainis and K. Prusis. A tight lower bound on certificate complexity in terms of block sensitivity and sensitivity. In *MFCS*, pages 33–44, 2014.

[APV15] A. Ambainis, K. Prusis, and J. Vihrovs. Sensitivity versus certificate complexity of boolean functions. *CoRR*, abs/1503.07691, 2015.

[AS11] A. Ambainis and X. Sun. New separation between $s(f)$ and $bs(f)$. *Electronic Colloquium on Computational Complexity (ECCC)*, 18:116, 2011.

[AV15] A. Ambainis and J. Vihrovs. Size of sets with small sensitivity: A generalization of simon's lemma. In *Theory and Applications of Models of Computation - 12th Annual Conference, TAMC 2015, Singapore, May 18-20, 2015, Proceedings*, pages 122–133, 2015.

[BdW02] H. Buhrman and R. de Wolf. Complexity measures and decision tree complexity: a survey. *Theor. Comput. Sci.*, 288(1):21–43, 2002.

[Bop12] M. Boppana. Lattice variant of the sensitivity conjecture. *Electronic Colloquium on Computational Complexity (ECCC)*, 19:89, 2012.

[Cha11] S. Chakraborty. On the sensitivity of cyclically-invariant boolean functions. *Discrete Mathematics & Theoretical Computer Science*, 13(4):51–60, 2011.

[GKS15] J. Gilmer, M. Koucký, and M. E. Saks. A new approach to the sensitivity conjecture. In *Proceedings of the 2015 Conference on Innovations in Theoretical Computer Science, ITCS 2015, Rehovot, Israel, January 11-13, 2015*, pages 247–254, 2015.

[GNS+16] P. Gopalan, N. Nisan, R. A. Servedio, K. Talwar, and A. Wigderson. Smooth boolean functions are easy: Efficient algorithms for low-sensitivity functions. In *ITCS*, pages 59–70, 2016.

[HKP11] P. Hatami, R. Kulkarni, and D. Pankratov. Variations on the sensitivity conjecture. *Theory of Computing, Graduate Surveys*, 2:1–27, 2011.

[KK04] C. Kenyon and S. Kutin. Sensitivity, block sensitivity, and l-block sensitivity of boolean functions. *Inf. Comput.*, 189(1):43–53, 2004.

[Nis89] N. Nisan. Crew prams and decision trees. In *STOC*, pages 327–335, 1989.

[NS94] N. Nisan and M. Szegedy. On the degree of Boolean functions as real polynomials. *Computational Complexity*, 4:301–313, 1994.

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

[Sim83] H. U. Simon. A tight $\Omega(\log \log n)$-bound on the time for parallel ram's to compute nondegenerated boolean functions. In *Foundations of computation theory*, pages 439–444. Springer, 1983.

[Sze15] M. Szegedy. An $O(n^{0.4732})$ upper bound on the complexity of the GKS communication game. *Electronic Colloquium on Computational Complexity (ECCC)*, 22:102, 2015.

[Tal13] A. Tal. Properties and applications of boolean function composition. In *ITCS*, pages 441–454, 2013.

[Vir11] M. Virza. Sensitivity versus block sensitivity of boolean functions. *Inf. Process. Lett.*, 111(9):433–435, 2011.

A Proof of Corollary 4.2

Proof. Let $x \in \{0,1\}^n$ and $B_1, \ldots, B_m$ be the blocks that achieve $bs(f)$. Assume without loss of generality that $B_1, \ldots, B_{m'}$ are of size at most 2ℓ and that $B_{m'+1}, \ldots, B_m$ are of size larger than 2ℓ. Then, by the disjointness of $B_{m'+1}, \ldots, B_m$ we have that $m - m' \leq \frac{n}{2\ell}$. Thus,

$$\begin{aligned} bs_\ell(f,x) &\geq \sum_{i=1}^{m'} |f(x) - f(x+B_i)| = \sum_{i=1}^{m} |f(x) - f(x+B_i)| - \sum_{i=m'+1}^{m} |f(x) - f(x+B_i)| \\ &\geq bs(f,x) - (m - m') \geq bs(f,x) - \frac{n}{2\ell} \geq \frac{n}{2\ell}\,. \end{aligned}$$

Assume without loss of generality that $B_1, \ldots, B_{m''}$ are blocks such that $|f(x) - f(x+B_i)| \geq \frac{1}{4\ell}$ and that $B_{m''+1}, \ldots, B_{m'}$ are not. Then, $\sum_{i=m''+1}^{m'} |f(x) - f(x+B_i)| \leq \frac{m''-m'}{4\ell} \leq \frac{n}{4\ell}$. This implies that $\sum_{i=1}^{m''} |f(x) - f(x+B_i)| \geq \frac{n}{4\ell}$, and in particular that $m'' \geq \frac{n}{4\ell}$. Thus, there are $m'' \geq n/4\ell$ disjoint blocks of size at most 2ℓ which change the value of f by at least $\frac{1}{4\ell}$. Theorem 4.1 gives that $s(f) \geq \Omega((m'')^{1/2\ell}/\ell) \geq \Omega(n^{1/2\ell}/\ell)$. $\square$

ECCC ISSN 1433-8092

http://eccc.hpi-web.de