Degree and Sensitivity: Tails of Two Distributions

Parikshit Gopalan[1], Rocco A. Servedio*[2], and Avi Wigderson†[3]

1 Microsoft Research, Redmond, USA
parik@microsoft.com
2 Department of Computer Science, Columbia University, New York, USA
rocco@cs.columbia.edu
3 Institute for Advanced Study, Princeton University, Princeton, USA
avi@ias.edu

Abstract

The *sensitivity* of a Boolean function f is the maximum, over all inputs x, of the number of sensitive coordinates of x (namely the number of Hamming neighbors of x with different f-value). The well-known *sensitivity conjecture* of Nisan (see also Nisan and Szegedy) states that every sensitivity-s Boolean function can be computed by a polynomial over the reals of degree $\mathrm{poly}(s)$. The best known upper bounds on degree, however, are exponential rather than polynomial in s.

Our main result is an approximate version of the conjecture: every Boolean function with sensitivity s can be ϵ-approximated (in ℓ_2) by a polynomial whose degree is $s \cdot \mathrm{polylog}(1/\epsilon)$. This is the first improvement on the folklore bound of s/ϵ. We prove this via a new "switching lemma for low-sensitivity functions" which establishes that a random restriction of a low-sensitivity function is very likely to have low decision tree depth. This is analogous to the well-known switching lemma for AC^0 circuits.

Our proof analyzes the combinatorial structure of the graph G_f of sensitive edges of a Boolean function f. Understanding the structure of this graph is of independent interest as a means of understanding Boolean functions. We propose several new complexity measures for Boolean functions based on this graph, including *tree sensitivity* and *component dimension*, which may be viewed as relaxations of worst-case sensitivity, and we introduce some new techniques, such as *proper walks* and *shifting*, to analyze these measures. We use these notions to show that the graph of a function of full degree must be sufficiently complex, and that random restrictions of low-sensitivity functions are unlikely to lead to such complex graphs.

We postulate a robust analogue of the sensitivity conjecture: if **most** inputs to a Boolean function f have low sensitivity, then **most** of the Fourier mass of f is concentrated on small subsets. We prove a lower bound on tree sensitivity in terms of decision tree depth, and show that a polynomial strengthening of this lower bound implies the robust conjecture. We feel that studying the graph G_f is interesting in its own right, and we hope that some of the notions and techniques we introduce in this work will be of use in its further study.

1998 ACM Subject Classification F.1.3 Complexity Measures and Classes

Keywords and phrases Boolean functions, sensitivity, decision trees, Fourier degree

Digital Object Identifier 10.4230/LIPIcs.CCC.2016.13

1 Introduction

The smoothness of a continuous function captures how gradually it changes locally (according to the metric of the underlying space). For Boolean functions on $\{0,1\}^n$, a natural analog is

* Partially supported by NSF grants CCF-1319788 and CCF-1420349.
† Partially supported by NSF grant CCF-1412958.

31st Conference on Computational Complexity (CCC 2016).
Editor: Ran Raz; Article No. 13; pp. 13:1–13:23
Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany
COMPUTATIONAL COMPLEXITY CONFERENCE

sensitivity, capturing how many neighbors of a point have different function values. More formally, the *sensitivity* of $f : \{0,1\}^n \to \{\pm 1\}$ at input $x \in \{0,1\}^n$, written $s(f,x)$, is the number of neighbors y of x in the Hamming cube $\{0,1\}^n$ such that $f(y) \neq f(x)$. The *max sensitivity* of f, written $s(f)$ and often referred to simply as the "sensitivity of f", is defined as $s(f) = \max_{x \in \{0,1\}^n} s(f,x)$. Hence we have $0 \leq s(f) \leq n$ for every $f : \{0,1\}^n \to \{\pm 1\}$; while not crucial, it may be helpful to consider this parameter as "low" when e.g. either $s(f) \leq (\log n)^{O(1)}$ or $s(f) \leq n^{o(1)}$ (note that both these notions of "low" are robust up to polynomial factors).

A well known conjecture, sometimes referred to as the "sensitivity conjecture," states that every smooth Boolean function is computed by a low degree real polynomial, specifically of degree polynomial in its sensitivity. This conjecture was first posed in the form of a question by Nisan [20] and Nisan and Szegedy [19] but is now (we feel) widely believed to be true:

▶ **Conjecture 1.1** ([20, 19]). *There exists a constant c such that every Boolean function f is computed by a polynomial of degree $\deg(f) \leq s(f)^c$.*

Despite significant effort ([17, 1, 2, 3, 4]) the best upper bound on degree in terms of sensitivity is exponential. Recently several consequences of Conjecture 1.1, e.g. that every f has a *formula* of depth at most $\mathrm{poly}(s(f))$, have been unconditionally established in [11]. Nisan and Szegedy proved the converse, that every Boolean function satisfies $s(f) = O(\deg(f)^2)$.

In this work, we make progress on Conjecture 1.1 by showing that functions with low max sensitivity are very well *approximated* (in ℓ_2) by low-degree polynomials. We exponentially improve the folklore $O(s/\epsilon)$ degree bound (which follows from average sensitivity and Markov's inequality) by replacing the $1/\epsilon$ error dependence with $\mathrm{poly}\log(1/\epsilon)$. The following is our main result:[1]

▶ **Theorem 1.2.** *For any Boolean function $f : \{0,1\}^n \to \{\pm 1\}$ and any $\epsilon > 0$, there exists a polynomial $p : \{0,1\}^n \to \mathbb{R}$ with $\deg(p) \leq O(s(f) \cdot (\log(1/\epsilon))^3)$ such that $\mathbb{E}_{x \in \{0,1\}^n}[|p(x) - f(x)|^2] \leq \epsilon$.*

En route to proving this result, we make two related contributions which we believe are interesting in themselves:

- Formulating a robust variant of the sensitivity conjecture (which would generalize Theorem 1.2).
- Defining and analyzing some natural graph-theoretic complexity measures, essential to our proof and which we believe may hold the key to progress on the original and robust sensitivity conjectures.

1.1 A robust variant of the sensitivity conjecture

A remarkable series of developments, starting with [20], showed that real polynomial degree is an extremely versatile complexity measure: it is polynomially related to many other complexity measures for Boolean functions, including PRAM complexity, block sensitivity, certificate complexity, deterministic/randomized/quantum decision tree depth, and approximating polynomial degree (see [6, 15] for details on many of these relationships). Arguably the one natural complexity measure that has defied inclusion in this equivalence class is

[1] In a subsequent version of this work [12] the exponent "3" in Theorem 1.2 is improved to 1, and it is shown that any further improvement to an exponent strictly less than 1 implies Conjecture 1.1.

sensitivity. Thus, there are many equivalent formulations of Conjecture 1.1; indeed, Nisan's original formulation was in terms of sensitivity versus block sensitivity [20].

Even though progress on it has been slow, over the years Conjecture 1.1 has become a well-known open question in the study of Boolean functions. It is natural to ask *why* this is an important question: will a better understanding of sensitivity lead to new insights into Boolean functions that have eluded us so far? Is sensitivity qualitatively different from the other concrete complexity measures that we already understand?

We believe that the answer is yes, and in this paper we make the case that Conjecture 1.1 is just the (extremal) tip of the iceberg: it hints at deep connections between the *combinatorial* structure of a Boolean function f, as captured by the graph G_f of its sensitive edges in the hypercube, and the *analytic* structure, as captured by its Fourier expansion. This connection is already the subject of some of the key results in the analysis of Boolean functions, such as [16, 9], as well as important open problems like the "entropy-influence" conjecture [10] and its many consequences.

Given any Boolean function f, we conjecture a connection between the distribution of the sensitivity of a random vertex in $\{0,1\}^n$ and the distribution of f's Fourier mass. This conjecture, which is an important motivation for the study in this paper, is stated informally below:

Robust Sensitivity Conjecture (Informal Statement): *If* ***most*** *inputs to a Boolean function f have low sensitivity, then* ***most*** *of the Fourier mass of f is concentrated on small subsets.*

Replacing both occurrences of **most** by *all* we recover Conjecture 1.1, and hence the statement may be viewed as a robust formulation of the sensitivity conjecture. Theorem 1.2 corresponds to replacing the first **most** by *all*. There are natural classes of functions which do not have low max sensitivity, but for which most vertices have low sensitivity; the robust sensitivity conjecture is relevant to these functions while the original sensitivity conjecture is not. (A prominent example of such a class is AC^0, for which the results of [18] establish a weak version of the assumption (that most inputs have low sensitivity) and the results of [18, 26] establish a strong version of the conclusion (Fourier concentration).)

In order to formulate a precise statement, for a given Boolean function $f : \{0,1\}^n \to \{\pm 1\}$ we consider the random experiment which samples from the following two distributions:

1. The Sensitivity distribution: sample a uniform random vertex $\boldsymbol{x} \in \{0,1\}^n$ and let $\boldsymbol{s} = s(f, \boldsymbol{x})$.
2. The Fourier distribution: sample a subset $\mathbf{T} \subset [n]$ with probability $\hat{f}(\mathbf{T})^2$ and let $\boldsymbol{d} = |\mathbf{T}|$.

We conjecture a close relation between the k^{th} moments of these random variables:

▶ **Conjecture 1.3** (Robust Sensitivity Conjecture)**.** *For all Boolean functions f and for all integers $k \geq 1$, there is a constant a_k such that $\mathbb{E}[\boldsymbol{d}^k] \leq a_k \, \mathbb{E}[\boldsymbol{s}^k]$.*

The key here is that there is no dependence on n. To see the connection with the informal statement above, if a function has low sensitivity for most $x \in \{0,1\}^n$, then it must have bounded k^{th} sensitivity moments for fairly large k; in such a case, Conjecture 1.3 implies a strong Fourier concentration bound by Markov's inequality. The classical Fourier expansion for average sensitivity tells us that when $k = 1$, $\mathbb{E}[\boldsymbol{s}] = \mathbb{E}[\boldsymbol{d}]$. It is also known that $\mathbb{E}[\boldsymbol{s}^2] = \mathbb{E}[\boldsymbol{d}^2]$ (see e.g. [7, Lemma 3.5]), but equality does not hold for $k \geq 3$. Conjecture 1.3 states that if we allow constant factors depending on k, then one direction still holds.

It is clear that Conjecture 1.3 (with a_k a not-too-rapidly-growing function of k) is a strengthening of our Theorem 1.2. To see its relation to Conjecture 1.1 observe that

Conjecture 1.1 implies that for $k \to \infty$, $\mathbb{E}[\boldsymbol{d}^k] \leq a^k(\mathbb{E}[\boldsymbol{s}^k])^b$ for constants a, b. On the other hand, via Markov's inequality, Conjecture 1.3 only guarantees Fourier concentration rather than small degree for functions with small sensitivity. Thus the robust version Conjecture 1.3 seems incomparable to Conjecture 1.1.

It is possible that the reverse direction of the robust conjecture also holds: for every k there exists a'_k such that $\mathbb{E}[\boldsymbol{s}^k] \leq a'_k \, \mathbb{E}[\boldsymbol{d}^k]$; settling this is an intriguing open question. We note that the Nisan-Szegedy result that $s(f) \leq O(\deg(f)^2)$ implies that as $k \to \infty$ we have $\mathbb{E}[\boldsymbol{s}^k] \leq C^k \, \mathbb{E}[\boldsymbol{d}^k]^2$ for some constant C.

Both our proof of Theorem 1.2, and our attempts at Conjecture 1.3, follow the same general path. We apply random restrictions, which reduces these statements to analyzing some natural new graph-theoretic complexity measures of Boolean functions. These measures are relaxations of sensitivity: they look for occurrences of various subgraphs in the sensitivity graph, rather than just high degree vertices. We establish (and conjecture) connections between different graph-theoretic measures and decision tree depth (see Theorem 5.4, which relates decision tree depth and the length of "proper walks", and Conjecture 4.10, which conjectures a relation between "tree sensitivity" and decision tree depth). These connections respectively enable the proof of Theorem 1.2 and provide a simple sufficient condition implying Conjecture 1.3, which suffices to prove the conjecture for $k = 3$ and 4. We elaborate on this in the next subsection. We believe that these new complexity measures are interesting and important in their own right, and that understanding them better may lead to progress on Conjecture 1.1.

1.2 Random restrictions and graph-theoretic complexity measures

In this subsection we give a high level description of our new complexity measures and perspectives on the sensitivity graph and of how we use them to approach Conjecture 1.3 and prove Theorem 1.2. As both have the same conclusion, namely strong Fourier concentration, we describe both approaches together until they diverge. This leads to analyzing two different graph parameters (as we shall see, the stronger assumption of Theorem 1.2 allows the use of a weaker graph parameter that we can better control).

First we give a precise definition of the *sensitivity graph*: to every Boolean function f we associate a graph G_f whose vertex set is $\{0,1\}^n$ and whose edge set E consists of all edges (x, y) of the hypercube that have $f(x) \neq f(y)$. Each edge is labelled by the coordinate in $[n]$ at which x and y differ. The degree of vertex x is exactly $s(f, x)$, and the maximum degree of G_f is $s(f)$.

The starting point of our approach is to reinterpret the moments of the degree and sensitivity distributions of f in terms of its random restrictions. Let $\mathcal{R}_{k,n}$ denote the distribution over random restrictions that leave exactly k of the n variables unset and set the rest uniformly at random. We first show, in Section 3, that the k^{th} moment of the sensitivity distribution controls the probability that a random restriction $f_{\boldsymbol{\rho}}$ of f, where $\boldsymbol{\rho} \leftarrow \mathcal{R}_{k,n}$, has full sensitivity (Theorem 3.1). Similarly, moments of the Fourier distribution capture the event that $f_{\boldsymbol{\rho}}$ has full degree (Theorem 3.2).[2]

Random restrictions under sensitivity moment bounds

Via Theorems 3.1 and 3.2, Conjecture 1.3 may be rephrased as saying that if a function f has low sensitivity moments, then a random restriction $f_{\boldsymbol{\rho}}$ is unlikely to have full degree. An

[2] We note that Tal has proved a result of a similar flavor; [25, Theorem 3.2] states that strong Fourier concentration of f implies that random restrictions of f are unlikely to have high degree.

intuition supporting this statement is that the sensitivity graphs of functions with full degree should be "complex" (under some suitable complexity measure), whereas the graph of $f_{\boldsymbol{\rho}}$ is unlikely to be "complex" if f has low sensitivity moments. More precisely, the fact that G_f has no (or few) vertices of high degree suggests that structures with many sensitive edges in distinct directions will not survive a random restriction.

Some evidence for this intuition is given by Theorem 3.1, which tells us that if f has low sensitivity moments then $f_{\boldsymbol{\rho}}$ is unlikely to have full sensitivity. If full degree implied full sensitivity then we would be done, but this is false as witnessed e.g. by the three-variable majority function and by composed variants of it. (Conjecture 1.1 asserts that the gap between degree and sensitivity is at most polynomial, but of course we do not want to invoke the conjecture!) This leads us in Section 4 to consider our first relaxation of sensitivity, which we call *tree-sensitivity*. To motivate this notion, note that a vertex with sensitivity k is simply a star with k edges in the sensitivity graph. We relax the star requirement and consider all *sensitive trees*: trees of sensitive edges (i.e. edges in G_f) where every edge belongs to a *distinct* coordinate direction (as is the case, of course, for a star). Analogous to the usual notion of sensitivity, the tree sensitivity of f at x is the size of the largest sensitive tree containing x, and the tree sensitivity of f is the maximum tree sensitivity of f at any vertex.

Theorem 4.11 shows that the sensitivity moments of f control the probability that f_ρ has full tree sensitivity. Its proof crucially uses a result by Sidorenko [24] on counting homomorphisms to trees. Theorem 4.11 would immediately imply Conjecture 1.3 if every function of degree k must have tree sensitivity k. (This is easily verified for $k = 3, 4$, which, as alluded to in the previous subsection, gives Conjecture 1.3 for those values of k.) The best we can prove, though, is a tree sensitivity lower bound of $\Omega(\sqrt{k})$ (Theorem 4.9); the proof uses notions of maximality and "shifting" of sensitive trees that we believe may find further application in the study of tree sensitivity. We conjecture that full degree does imply full tree sensitivity, implying Conjecture 1.3. This is a rare example where having a precise bound between the two complexity measures (rather than a polynomial relationship) seems to be important.

Random restrictions under a max sensitivity bound

Next, we aim to prove *unconditional* moment bounds on the Fourier distribution of functions with low max sensitivity, and thereby obtain Theorem 1.2. Towards this goal, in Section 5 we relax the notion of tree sensitivity and study certain walks in the Boolean hypercube that we call *proper walks*: these are walks such that every time a coordinate direction is explored for the first time, it is along a sensitive edge. We show in Theorem 5.4 that having full decision tree depth implies the existence of a very short (length $O(n)$) proper walk containing sensitive edges along every coordinate. In Lemma 5.6, we analyze random restrictions to show that such a structure is unlikely to survive in the remaining subcube of unrestricted variables. This may be viewed as a "switching lemma for low-sensitivity functions", which again may be independently interesting (note that strictly speaking this result is not about switching from a DNF to a CNF or vice versa, but rather it upper bounds the probability that a restricted function has large decision tree depth, in the spirit of standard "switching lemmas"). It yields Theorem 1.2 via a rather straightforward argument. The analysis requires an upper bound on the maximum sensitivity because we do not know an analogue of Sidorenko's theorem for proper walks.

1.3 Some high-level perspective

An important goal of this work is to motivate a better understanding of the combinatorial structure of the sensitivity graph G_f associated with a Boolean function. In our proofs other notions suggest themselves beyond tree sensitivity and proper walks, most notably the *component dimension* of the graph, which may be viewed as a further relaxation of sensitivity. Better relating these measures to decision tree depths, as well as to each other, remains intriguing, and in our view promising, for making progress on Conjecture 1.1 and Conjecture 1.3 and for better understanding Boolean functions in general. We hope that some of the notions and techniques we introduce in this work will be of use to this goal.

Another high level perspective relates to "switching lemmas". As mentioned above, we prove here a new result of this kind, showing that under random restrictions low sensitivity functions have low decision tree depth with high probability. The classical switching lemma shows the same for small width DNF (or CNF) formulas (and hence for AC^0 circuits as well). Our proof is quite different than the standard proofs, as it is essentially based on the combinatorial parameters of the sensitivity graph. Let us relate the assumptions of both switching lemmas. On the one hand, by the sensitivity Conjecture 1.1 (which we can't use, and want to prove), low sensitivity should imply low degree and hence low decision tree depth and small DNF width. On the other hand, small DNF width (indeed small, shallow circuits) imply (by [18]) low *average* sensitivity, which is roughly the assumption of the robust sensitivity Conjecture 1.3. As it turns out, we can use our combinatorial proof of our switching lemma to derive a somewhat weaker form of the original switching lemma, and also show that the same combinatorial assumption (relating tree sensitivity to decision tree depth) which implies Conjecture 1.3 would yield a nearly tight form of the original switching lemma. This lends further motivation to the study of these graph parameters.

Another conjecture formalizing the maxim that *low sensitivity implies Fourier concentration* is the celebrated Entropy-Influence conjecture of Freidgut and Kalai [10] which posits the existance of a universal constant C such that $H(\mathbf{T}) \leq C\,\mathbb{E}[\boldsymbol{s}]$ where $H(.)$ denotes the entropy function of a random variable.[3] The conjecture states that functions with low sensitivity on average (measured by $\mathbb{E}[\boldsymbol{s}] = \mathbb{E}[\boldsymbol{d}]$) have their Fourier spectrum concentrated on a few coefficients, so that the entropy of the Fourier distribution is low. However, unlike in Conjecture 1.3 the degree of those coefficients does not enter the picture.

Organization

We present some standard preliminaries and notation in Section 2. Section 3 proves Theorems 3.1 and 3.2 which show that degree and sensitivity moments govern the degree and sensitivity respectively of random restrictions. In Section 4 we study tree sensitivity. Section 4.1 relates it to other complexity measures, while Section 4.2 shows how the tree sensitivity of a random restriction is governed by sensitivity moments. We explore some consequences of these results in Section 4.3. Section 5 studies proper walks, and shows how to construct short proper walks. In Section 5.1, we use proper walks to analyze random restrictions of low-sensitivity functions and prove Theorem 1.2. Section 5 uses results from Section 4.1 but is independent of the rest of Section 4.

[3] Recall that the entropy $H(\mathbf{T})$ of the random variable $\mathbf{T}$ is $H(\mathbf{T}) = \sum_{T\subseteq[n]} \Pr[\mathbf{T}=T] \log_2 \frac{1}{\Pr[\mathbf{T}=T]}$.

2 Preliminaries

The Fourier distribution. Let $f : \{0,1\}^n \to \{\pm 1\}$ be a Boolean function. We define the usual inner product on the space of such functions by $\langle f, g\rangle = \mathbb{E}_{\boldsymbol{x}\leftarrow\{0,1\}^n}[f(\boldsymbol{x})g(\boldsymbol{x})]$. For $S \subseteq [n]$ the parity function χ_S is $\chi_S(x) = (-1)^{\sum_{i\in S} x_i}$. The Fourier expansion of f is given by $f(x) = \sum_{S\subset[n]} \hat{f}(S)\chi_S(x)$, where $\hat{f}(S) = \langle f, \chi_S\rangle$. By Parseval's identity we have $\sum_{S\subseteq[n]} \hat{f}(S)^2 = 1$. This allows us to view any Boolean function f as inducing a probability distribution $\mathcal{D}_f$ on subsets $S \subseteq [n]$, given by $\Pr_{\mathbf{R}\leftarrow\mathcal{D}_f}[\mathbf{R} = S] = \hat{f}(S)^2$. We refer to this as the *Fourier distribution*. We define $\mathrm{supp}(f) \subseteq 2^{[n]}$ as $\mathrm{supp}(f) = \{S \subseteq [n] : \hat{f}(S)^2 \neq 0\}$. The Fourier expansion of f can be viewed as expressing S as a multilinear polynomial in $x_1, \ldots, x_n$, so that $\deg(f) = \max_{S\in\mathrm{supp}(f)} |S|$. Viewing $\mathcal{D}_f$ as a probability distribution on $2^{[n]}$, we define the following quantities which we refer to as "influence moments" of f:

$$\mathbb{I}^k[f] = \mathop{\mathbb{E}}_{\mathbf{R}\leftarrow\mathcal{D}_f}\left[|\mathbf{R}|^k\right] = \sum_S \hat{f}(S)^2|S|^k, \tag{1}$$

$$\mathbb{I}^{\underline{k}}[f] = \mathop{\mathbb{E}}_{\mathbf{R}\leftarrow\mathcal{D}_f}\left[\prod_{i=0}^{k-1}(|\mathbf{R}| - i)\right] = \sum_{|S|\geq k} \hat{f}(S)^2 \prod_{i=0}^{k-1}(|S| - i). \tag{2}$$

We write $\deg_\epsilon(f)$ to denote the minimum k such that $\sum_{S\subseteq[n];|S|\geq k} \hat{f}(S)^2 \leq \epsilon$. It is well known that $\deg_\epsilon(f) \leq k$ implies the existence of a degree k polynomial g such that $\mathbb{E}_{\boldsymbol{x}}[(f(\boldsymbol{x}) - g(\boldsymbol{x}))^2] \leq \epsilon$; g is obtained by truncating the Fourier expansion of f to level k.

The sensitivity distribution. We use $d(\cdot,\cdot)$ to denote Hamming distance on $\{0,1\}^n$. The n-dimensional hypercube H_n is the graph with vertex set $V = \{0,1\}^n$ and $\{x,y\} \in E$ if $d(x,y) = 1$. For $x \in \{0,1\}^n$, let $N(x)$ denote its neighborhood in H_n. As described in Section 1, the *sensitivity* of a function f at point x is defined as $s(f,x) = |\{y \in N(x) : f(x) \neq f(y)\}|$, and the (worst-case) sensitivity of f, denoted $s(f)$, is defined as $s(f) = \max_{x\in\{0,1\}^n} s(f,x)$. Analogous to (1) and (2), we define the quantities $s^k(f)$ and $s^{\underline{k}}(f)$ which we refer to as "sensitivity moments" of f:

$$s^k(f) = \mathop{\mathbb{E}}_{\boldsymbol{x}\leftarrow\{0,1\}^n}\left[s(f,\boldsymbol{x})^k\right], \qquad s^{\underline{k}}(f) = \mathop{\mathbb{E}}_{\boldsymbol{x}\leftarrow\{0,1\}^n}\left[\prod_{i=0}^{k-1}(s(f,\boldsymbol{x}) - i)\right]. \tag{3}$$

With this notation, we can restate Conjecture 1.3 (with a small modification) as

▶ **Conjecture** (Conjecture 1.3 restated)**.** *For every k, there exists constants a_k, b_k such that $\mathbb{I}^k(f) \leq a_k s^k(f) + b_k$.*

The reason for the additive constant b_k is that for all non-negative integers x, we have $\prod_{i=0}^{k-1}(x - i) \leq x^k \leq e^k \prod_{i=0}^{k-1}(x - i) + k^k$. Hence allowing the additive factor lets us freely interchange $\mathbb{I}^k$ with $\mathbb{I}^{\underline{k}}$ and s^k with $s^{\underline{k}}$ in the statement of the Conjecture. We note that $\mathbb{I}^1[f] = \mathbb{I}^{\underline{1}}[f] = s^1(f) = s^{\underline{1}}(f)$, and as stated earlier it is not difficult to show that $\mathbb{I}^2[f] = s^2(f)$ (see e.g. Lemma 3.5 of [7]). However, in general $\mathbb{I}^k(f) \neq s^k(f)$ for $k \geq 3$ (as witnessed, for example, by the AND function).

Some other complexity measures. We define $\dim(f)$ to be the number of variables that f depends on and $\mathrm{dt}(f)$ to be the smallest depth of a deterministic decision tree computing f. In particular $f : \{0,1\}^n \to \{\pm 1\}$ has $\dim(f) = n$ iff f is sensitive to every co-ordinate, and has $\mathrm{dt}(f) = n$ iff f is evasive. It is easy to see that $\deg(f) \leq \mathrm{dt}(f) \leq \dim(f)$ and $s(f) \leq \mathrm{dt}(f)$.

3 Random restrictions and moments of degree and sensitivity

We write $\mathcal{R}_{k,n}$ to denote the set of all restrictions that leave exactly k variables live (unset) out of n. A restriction $\rho \in \mathcal{R}_{k,n}$ is viewed as a string in $\{0,1,\star\}^n$ where $\rho_i = \star$ for exactly the k live variables. We denote the set of live variables by $L(\rho)$, and we use $f_\rho : \{0,1\}^{L(\rho)} \to \{\pm 1\}$ to denote the resulting restricted function. We use $C(\rho) \subseteq \{0,1\}^n$ to denote the subcube consisting of all possible assignments to variables in $L(\rho)$. We sometimes refer to "a random restriction $\boldsymbol{\rho} \leftarrow \mathcal{R}_{k,n}$" to indicate that $\boldsymbol{\rho}$ is selected uniformly at random from $\mathcal{R}_{k,n}$.

A random restriction $\boldsymbol{\rho} \leftarrow \mathcal{R}_{k,n}$ can be chosen by first picking a set $\mathbf{K} \subset [n]$ of k co-ordinates to set to $\star$ and then picking $\boldsymbol{\rho}_{\bar{\mathbf{K}}} \in \{0,1\}^{[n]\setminus \mathbf{K}}$ uniformly at random. Often we will pick both $\boldsymbol{x} \in \{0,1\}^n$ and $\mathbf{K} \subset [n]$ of size k independently and uniformly at random. This is equivalent to sampling a random restriction $\boldsymbol{\rho}$ and a random point $\boldsymbol{y}$ within the subcube $C(\rho)$.

The following two theorems show that $\mathbb{I}^{\underline{k}}[f]$ captures the degree of f_ρ, whereas $s^{\underline{k}}(f)$ captures its sensitivity.

▶ **Theorem 3.1.** *Let $f : \{0,1\}^n \to \{\pm 1\}$, $\rho \leftarrow \mathcal{R}_{k,n}$, and $1 \le j \le k$. Then*

$$\frac{s^{\underline{j}}(f)}{n^j} \approx \frac{s^{\underline{j}}(f)}{\prod_{i=0}^{j-1}(n-i)} \le \Pr_{\boldsymbol{\rho} \leftarrow \mathcal{R}_{k,n}}[s(f_{\boldsymbol{\rho}}) \ge j] \le \frac{2^k s^{\underline{j}}(f)\binom{k}{j}}{\prod_{i=0}^{j-1}(n-i)} \approx \frac{2^k s^{\underline{j}}(f)\binom{k}{j}}{n^j}. \tag{4}$$

Proof. Consider the bipartite graph in which the vertices X on the left are all j-edge stars S in G_f, the vertices Y on the right are all restrictions $\rho \in \mathcal{R}_{k,n}$, and an edge connects S and ρ if the star S lies in the subcube $C(\rho)$ specified by the restriction ρ. The desired probability $\Pr_{\rho \in \mathcal{R}_{k,n}}[s(f_\rho) \ge j]$ is the fraction of nodes in Y that are incident to at least one edge.

The number of nodes on the left is equal to

$$|X| = \sum_{x \in \{0,1\}^n} \binom{s(f,x)}{j} = \frac{2^n s^{\underline{j}}(f)}{j!}.$$

The degree of each node S on the left is exactly $\binom{n-j}{k-j}$, since if S is adjacent to ρ then j of the k elements of $L(\rho)$ must correspond to the j edge coordinates of S and the other $k-j$ elements of $L(\rho)$ can be any of the $n-j$ remaining coordinates (note that the non-$\star$ coordinates of ρ are completely determined by S). On the right, a restriction $\rho \in \mathcal{R}_{k,n}$ is specified by a set $L(\rho)$ of k live co-ordinates where $\rho_i = \star$, and a value $\rho_i \in \{0,1\}$ for the other coordinates, so $|Y| = |\mathcal{R}_{k,n}| = \binom{n}{k}2^{n-k}$. We thus have

$$\Pr_{\boldsymbol{\rho} \leftarrow \mathcal{R}_{k,n}}[s(f_{\boldsymbol{\rho}}) \ge j] \le \frac{\text{total \# of edges into } Y}{|Y|} = \frac{\left(\frac{2^n s^{\underline{j}}(f)}{j!}\right) \cdot \binom{n-j}{k-j}}{\binom{n}{k}2^{n-k}} = \frac{2^k s^{\underline{j}}(f)\binom{k}{j}}{\prod_{i=0}^{j-1}(n-i)}.$$

For the lower bound, in order for S to lie in $C(\rho)$ the root of S must belong to $C(\rho)$ (2^k choices) and all edges of S must correspond to elements of $L(\rho)$ ($\binom{k}{j}$ choices), so the maximum degree of any $\rho \in Y$ is $2^k\binom{k}{j}$. Hence we have

$$\Pr_{\boldsymbol{\rho} \leftarrow \mathcal{R}_{k,n}}[s(f_{\boldsymbol{\rho}}) \ge j] \ge \frac{\frac{(\text{total \# of edges into } Y)}{(\text{max degree of any } \rho \in Y)}}{|Y|} = \frac{\left(\frac{2^n s^{(j)}(f)}{j!}\right) \cdot \binom{n-j}{k-j}}{2^k\binom{k}{j} \cdot \binom{n}{k}2^{n-k}} = \frac{s^{\underline{j}}(f)}{\prod_{i=0}^{j-1}(n-i)}.$$

◀

▶ **Theorem 3.2.** [4] *Let* $f : \{0,1\}^n \to \{\pm 1\}$ *and* $\rho \leftarrow \mathcal{R}_{k,n}$. *Then*

$$\frac{\mathbb{I}^{\underline{k}}(f)}{n^k} \approx \frac{\mathbb{I}^{\underline{k}}(f)}{\prod_{i=0}^{k-1}(n-i)} \leq \Pr_{\rho \leftarrow \mathcal{R}_{k,n}}[\deg(f_{\rho}) = k] \leq \frac{2^{2k-2}\,\mathbb{I}^{\underline{k}}(f)}{\prod_{i=0}^{k-1}(n-i)} \approx \frac{2^{2k-2}\,\mathbb{I}^{\underline{k}}(f)}{n^k}. \tag{5}$$

Proof. We first fix $K \subseteq [n]$ and consider the restricted function $f_{\boldsymbol{\rho}}$ that results from a random choice of $\boldsymbol{y} = \boldsymbol{\rho}_{\bar{K}} \in \{0,1\}^{[n]\setminus K}$. The degree k Fourier coefficient of $f_{\boldsymbol{\rho}}$ equals $\hat{f}_{\boldsymbol{\rho}}(K)$ and is given by

$$\hat{f}_{\boldsymbol{\rho}}(K) = \sum_{S \subset [n]\setminus K} \hat{f}(S \cup K)\chi_S(\boldsymbol{y}).$$

Hence we have

$$\mathop{\mathbb{E}}_{\boldsymbol{y}}[\hat{f}_{\boldsymbol{\rho}}(K)^2] = \sum_{S \subset [n]\setminus K} \hat{f}(S \cup K)^2,$$

and hence over a random choice of $\mathbf{K}$, we have

$$\mathop{\mathbb{E}}_{\boldsymbol{\rho}}[\hat{f}_{\boldsymbol{\rho}}(\mathbf{K})^2] = \sum_{S \subset [n]} \mathop{\mathbb{E}}_{\boldsymbol{\rho}}[\mathbf{1}(\mathbf{K} \subseteq S)]\hat{f}(S)^2 = \sum_{S \subset [n]} \frac{\prod_{i=0}^{k-1}(|S|-i)}{\prod_{i=0}^{k-1}(n-i)}\hat{f}(S)^2 = \frac{\mathbb{I}^{\underline{k}}[f]}{\prod_{i=0}^{k-1}(n-i)}. \tag{6}$$

Note that $\deg(f_{\boldsymbol{\rho}}) = k$ iff $\widehat{f_{\boldsymbol{\rho}}}(\mathbf{K})^2 \neq 0$. Further, when it is non-zero $\widehat{f_{\boldsymbol{\rho}}}(\mathbf{K})^2$ lies in the range $[2^{-(2k-2)}, 1]$, since a non-zero Fourier coefficient in a k-variable Boolean function has magnitude at least 2^{-k+1}. Hence we have

$$2^{-2k+2}\Pr_{\boldsymbol{\rho}}[\widehat{f_{\boldsymbol{\rho}}}(\mathbf{K})^2 \neq 0] \leq \mathop{\mathbb{E}}_{\boldsymbol{\rho}}[\widehat{f_{\boldsymbol{\rho}}}(\mathbf{K})^2] \leq \Pr_{\boldsymbol{\rho}}[\widehat{f_{\boldsymbol{\rho}}}(\mathbf{K})^2 \neq 0] \tag{7}$$

which gives the desired bound when plugged into Equation (6). ◀

Conjecture 1.3 revisited: An easy adaptation of the Theorem 3.2 argument gives bounds on $\Pr_{\boldsymbol{\rho} \leftarrow \mathcal{R}_{k,n}}[\deg(f_{\boldsymbol{\rho}}) \geq j]$. Given these bounds, Conjecture 1.3 implies that for any $j \leq k$,

$$\Pr_{\boldsymbol{\rho} \leftarrow \mathcal{R}_{k,n}}[\deg(f_{\boldsymbol{\rho}}) \geq j] \leq a_k \Pr_{\boldsymbol{\rho} \leftarrow \mathcal{R}_{k,n}}[s(f_{\boldsymbol{\rho}}) \geq j] + o_n(1).$$

Indeed, by specifying the $o_n(1)$ term, we can get a reformulation of Conjecture 1.3. This formulation has an intuitive interpretation: *gap examples exhibiting low sensitivity but high degree are not robust to random restrictions.* Currently, we do not know how to upper bound $\deg(f)$ by a polynomial in $s(f)$, indeed we do know of functions f where $\deg(f) \geq s(f)^2$. But the Conjecture implies that if we hit any function f with a random restriction, the probability that the restriction has large degree can be bounded by the probability that it has large sensitivity. Thus the conjecture predicts that these gaps do not survive random restrictions in a rather strong sense.

Implications for AC^0: For functions with small AC^0 circuits, a sequence of celebrated results culminating in the work of Håstad [14] gives upper bounds on $\Pr[\mathrm{dt}(f_{\boldsymbol{\rho}}) \geq j]$. Since $\Pr[\mathrm{dt}(f_{\boldsymbol{\rho}}) \geq j] \geq \Pr[\deg(f_{\boldsymbol{\rho}}) \geq j]$, we can plug these bounds into Theorem 3.2 to get upper

[4] The upper bound in the following theorem is essentially equivalent to Theorem 3.2 of [25], while the lower bound is analogous to [18]. The only difference is in the family of restrictions.

bounds on the Fourier moments, and derive a statement analogous to [18, Lemma 7], [26, Theorem 1.1] on the Fourier concentration of functions in AC^0.

Similarly $\Pr[\mathrm{dt}(f_{\boldsymbol{\rho}}) \geq j] \geq \Pr[s(f_{\boldsymbol{\rho}}) \geq j]$, so via this approach Theorem 3.1 gives upper bounds on the sensitivity moments, and hence sensitivity tail bounds for functions computed by small AC^0 circuits. This can be viewed as an extension of [18, Lemma 12], which bounds the average sensitivity (first moment) of such functions. For depth 2 circuits, such tail bounds are implied by the satisfiability coding lemma [21], but we believe these are the first such bounds for depth 3 and higher. As this is not the focus of our current work, we leave the details to the interested reader.

4 Tree sensitivity

In this section we study the occurrence of trees of various types in the sensitivity graph G_f, by defining a complexity measure called tree sensitivity. We study its relation to other complexity measures like decision tree depth.

▶ **Definition 4.1.** A set $S \subseteq \{0,1\}^n$ *induces a sensitive tree* T in G_f if (i) the points in S induce the (non-trivial) tree T in the Boolean hypercube; (ii) every edge induced by S is a sensitive edge for f, i.e. belongs to $E(G_f)$; and (iii) each induced edge belongs to a distinct co-ordinate direction.

Given a fixed function f, a sensitive tree T is completely specified by the set $V(T)$ of its vertices. We can think of each edge $e \in E(T)$ as being labelled by the coordinate $\ell(e) \in [n]$ along which f is sensitive, so every edge has a distinct label. Let $\ell(T)$ denote the set of all edge labels that occur in T. We refer to $|\ell(T)|$ as the *size* of T, and observe that it lies in $\{1, \ldots, n\}$. We note that $|V(T)| = |\ell(T)| + 1$ by the tree property. Further, any two vertices in $V(T)$ differ on a subset of coordinates in $\ell(T)$. Hence the set $V(T)$ lies in a subcube spanned by coordinates in $\ell(T)$, and all points in $V(T)$ agree on all the coordinates in $\overline{\ell(T)} \stackrel{\text{def}}{=} [n] \setminus \ell(T)$.

▶ **Definition 4.2.** For $x \in \{0,1\}^n$, the *tree-sensitivity of f at x*, denoted $\mathrm{ts}(f, x)$, is the maximum of $|\ell(T)|$ over all sensitive trees T such that $x \in V(T)$. We define the tree-sensitivity of f as $\mathrm{ts}(f) = \max_{x \in \{0,1\}^n} \mathrm{ts}(f, x)$.

Note that a vertex and all its sensitive neighbors induce a sensitive tree (which is a star). Thus one can view tree-sensitivity as a generalization of sensitivity, and hence we have that $\mathrm{ts}(f) \geq \mathrm{s}(f)$. Lemma A.1 will show that $\mathrm{ts}(f)$ can in fact be exponentially larger than both $\mathrm{s}(f)$ and $\mathrm{dt}(f)$ (the decision tree depth of f), and thus it cannot be upper bounded by some polynomial in standard measures like decision tree depth, degree, or block sensitivity. However, Theorem 4.9, which we prove in the next subsection, gives a polynomial lower bound.

4.1 Tree sensitivity and decision tree depth

A sensitive tree T is *maximal* if there does not exist any sensitive tree T' with $V(T) \subsetneq V(T')$. In this subsection we study maximal sensitive trees using a "shifting" technique, introduce the notion of an "orchard" (a highly symmetric configuration of isomorphic sensitive trees that have been shifted in all possible ways along their insensitive coordinates), and use these notions to prove Theorem 4.9, which lower bounds tree sensitivity by square root of decision tree depth.

The *support* of a vector $v \in \{0,1\}^n$, denoted $\mathrm{supp}(v)$, is the set $\{i \in [n] : v_i = 1\}$. For $x, v \in \{0,1\}^n$, $x \oplus v$ denotes the coordinatewise xor. Given a set $S \subseteq \{0,1\}^n$, let $S \oplus v = \{x \oplus v : x \in S\}$.

▶ **Definition 4.3.** Let v be a vector supported on $\overline{\ell(T)}$ where T is a sensitive tree in G_f. We say that T can be *shifted* by v if $f(x) = f(x \oplus v)$ for all $x \in V(T)$.

If T can be shifted by v then $V(T) \oplus v$ also induces a sensitive tree which we denote by $T \oplus v$. Mapping x to $x \oplus v$ gives an isomorphism between T and $T \oplus v$ which preserves both adjacency and edge labels, and in particular we have $\ell(T \oplus v) = \ell(T)$.

We have the following characterization of maximality (both directions follow easily from the definitions of maximality and of shifting by the unit basis vector e_i):

▶ **Lemma 4.4.** *A sensitive tree T is maximal if and only if it can be shifted by e_i for all $i \in \overline{\ell(T)}$ (equivalently, if none of the vertices in $V(T)$ is sensitive to any coordinate in $\overline{\ell(T)}$).*

The notion of maximality allows for a "win-win" analysis of sensitive trees: for each co-ordinate $i \in \overline{\ell(T)}$, we can either increase the size of the tree by adding an edge in direction i, or we can shift by e_i to get an isomorphic copy of the tree. Repeating this naturally leads to the following definition.

▶ **Definition 4.5.** Let T be a sensitive tree that can be shifted by every v supported on $\overline{\ell(T)}$. We refer to the set of all such trees $F = \{T \oplus v\}$ as an *orchard*, and we say that T belongs to the orchard F.

An orchard guarantees the existence of $2^{n-\ell(T)}$ trees that are isomorphic to T in G_f. It is *a priori* unclear that orchards exist in G_f. The following simple but key lemma proves their existence.

▶ **Lemma 4.6.** *Let T be a sensitive tree. Either T belongs to an orchard, or there exists a shift $T \oplus v$ of T which is not maximal.*

Proof. Assume the tree T does not belong to an orchard. Pick the smallest weight vector v' supported on $\overline{\ell(T)}$ such that T cannot be shifted by v' (if there is more than one such vector any one will do). Since T can trivially be shifted by 0^n, we have $\mathrm{wt}(v') \geq 1$. Pick any co-ordinate $i \in \mathrm{supp}(v')$, and let $v = v' \oplus e_i$ so that $\mathrm{wt}(v) = \mathrm{wt}(v') - 1$. By our choice of v', T can be shifted by v, but not by $v' = v \oplus e_i$. This implies that there exists $x \in V(T)$ so that $f(x) = f(x \oplus v) \neq f(x \oplus v')$, hence $T \oplus v$ is not maximal. ◀

This lemma directly implies the existence of orchards for every G_f:

▶ **Corollary 4.7.** *Every sensitive tree T where $|\ell(T)| = \mathrm{ts}(f)$ belongs to an orchard.*

The lemma also gives the following intersection property for orchards. Since any two trees in an orchard F are isomorphic, we can define $\ell(F) = \ell(T)$ to be the set of edge labels for any tree $T \in F$.

▶ **Lemma 4.8.** *Let F_1 and F_2 be orchards. Then $\ell(F_1) \cap \ell(F_2) \neq \emptyset$.*

Proof. Assume for contradiction that $\ell(F_1)$ and $\ell(F_2)$ are disjoint. We choose trees $T_1 \in F_1$ and $T_2 \in F_2$, and $x \in V(T_1), y \in V(T_2)$ such that $f(x) = 1$ and $f(y) = -1$. Now define $z \in \{0,1\}^n$ where z_i equals x_i if $i \in \ell(T_1)$ and z_i equals y_i otherwise. Since z agrees with x on $\ell(T_1) = \ell(F_1)$, it can be obtained by shifting x by $z \oplus x$ which is supported on $\overline{\ell(T_1)}$. Since T_1 belongs to an orchard, we get $f(z) = f(x) = 1$. However, we also have that $z_i = y_i$ for all $i \in \ell(T_2)$. Hence by similar reasoning, $f(z) = f(y) = -1$, which is a contradiction. ◀

We use this intersection property to lower bound tree sensitivity in terms of decision tree depth, via an argument similar to other upper bounds on $\mathrm{dt}(f)$ (such as the well known [5, 27, 13] quadratic upper bound on $\mathrm{dt}(f)$ in terms of certificate complexity).

▶ **Theorem 4.9.** *For any Boolean function* $f : \{0,1\}^n \to \{\pm 1\}$, *we have* $\mathrm{ts}(f) \geq \sqrt{2\,\mathrm{dt}(f)} - 1$.

Proof. We construct a decision tree for f by iterating the following step until we are left with a constant function at each leaf: at the current node in the decision tree, pick the largest sensitive tree T in the (restricted) function and read all the variables in $\ell(T)$.

Let k be the largest number of iterations before we terminate, taken over all paths in the decision tree. Fix a path that achieves k iterations and let f_i be the restriction of f that is obtained, at the end of the i-th iteration (and let $f_0 = f$). We claim that $\mathrm{ts}(f_i) \leq \mathrm{ts}(f) - i$. Note that if f_i is not constant then $\mathrm{ts}(f_i) \geq 1$, hence this claim implies that $k \leq \mathrm{ts}(f)$.

It suffices to prove the case $i = 1$, since we can then apply the same argument repeatedly. Consider all trees in $f_0 = f$ of size $\mathrm{ts}(f)$. Each of them occurs in an orchard by Corollary 4.7 and by Lemma 4.8 any two of them share at least one variable. Hence when we read all the variables in some tree T, we restrict at least one variable in every tree of size $\mathrm{ts}(f)$, reducing the size by at least 1. The size of the other trees cannot increase after restriction, since G_{f_1} is an induced subgraph of G_f. Hence all the sensitive trees in f_1 have size at most $\mathrm{ts}(f) - 1$.

It follows that overall we can bound the depth of the resulting decision tree by

$$\mathrm{dt}(f) \leq \sum_{i=1}^{k} \mathrm{ts}(f_{i-1}) \leq \sum_{i=1}^{k} (\mathrm{ts}(f) - (i-1)) \leq \frac{\mathrm{ts}(f)(\mathrm{ts}(f)+1)}{2}. \qquad ◀$$

It is natural to ask whether $\mathrm{ts}(f)$ is polynomially related to $\mathrm{dt}(f)$ and other standard complexity measures. Lemma A.1 in Appendix A gives an example of a function on n variables where $\mathrm{dt}(f) = \log(n+1)$ whereas $\mathrm{ts}(f) = n$. In the other direction, it is likely that the bound in Theorem 4.9 can be improved further. We conjecture that the following bound should hold:

▶ **Conjecture 4.10.** *For any Boolean function* $f : \{0,1\}^n \to \{\pm 1\}$, *we have* $\mathrm{ts}(f) \geq \mathrm{dt}(f)$.

In addition to being a natural question by itself, we will show in Section 4.3 that Conjecture 4.10 would have interesting consequences via the switching lemma in Section 4.2.

4.2 Tree Sensitivity under Random Restrictions

In this subsection we show that the probability of a random restriction of f having large tree sensitivity is both upper and lower bounded by suitable sensitivity moments of f.

▶ **Theorem 4.11.** *Let* $f : \{0,1\}^n \to \{\pm 1\}$, $\boldsymbol{\rho} \sim \mathcal{R}_{k,n}$ *and* $1 \leq j \leq k$. *Then we have*

$$\frac{s^{\underline{j}}(f)}{n^j} \approx \frac{s^{\underline{j}}(f)}{\prod_{i=0}^{j-1}(n-i)} \leq \Pr_{\rho \in \mathcal{R}_{k,n}}[\mathrm{ts}(f_{\boldsymbol{\rho}}) \geq j] \leq \frac{(2k)^{2k} s^j(f)}{\prod_{i=0}^{j-1}(n-i)} \approx \frac{(2k)^{2k} s^j(f)}{n^j}.$$

The lower bound follows from the fact that $\mathrm{ts}(f) \geq s(f)$ and Theorem 3.1. The key ingedient in the upper bound is Sidorenko's theorem [24], which bounds the number of homomorphisms from a graph G to a tree T with j edges in terms of the j^{th} degree moment of G. For a formal statement of Sidorenko's theorems, we refer the reader to [24, 8]. Below, we state the result we will use in our language. We also present an elegant proof due to Yuval Peres which seems considerably simpler than the known proofs of Sidorenko's theorem (though the lemma follows directly from that theorem).

▶ **Lemma 4.12** ([22]). *Let $\mathcal{S}_j$ denote the set of sensitive trees of size j in G_f. Then we have that*

$$|\mathcal{S}_j| \leq j! \sum_{x\in\{0,1\}^n} s(f,x)^j.$$

Proof. We consider the set $\mathcal{T}$ of all rooted unlabelled trees with j edges. For each tree $t \in \mathcal{T}$ we define a labelling of its vertices as follows: each tree $t \in \mathcal{T}$ has the vertex set $V = \{0,\ldots,j\}$ where 0 is the root. The adjacency structure is specified by a parent function $p_t : \{1,\ldots,j\} \to \{0,\ldots,j-1\}$ where $p_t(i) < i$ is the parent of vertex i. The tree t is completely specified by the function p_t, and hence $|\mathcal{T}| \leq j!$. For a given $t \in \mathcal{T}$, let $\mathcal{S}(t)$ denote the set of sensitive trees $T \in G_f$ whose adjacency structure is given by t.

For conciseness let us write $s_{\text{tot}}(f)$ to denote $\sum_{x\in\{0,1\}^n} s(f,x)$. Let $\mathcal{D}$ denote the distribution on $\{0,1\}^n$ where

$$\Pr_{\mathcal{D}}[x] = \frac{s(f,x)}{s_{\text{tot}}(f)}.$$

Note that $\mathcal{D}$ is supported only on vertices where $s(f,x) \geq 1$. Further $\mathcal{D}$ is a stationary distribution for the simple random walk on G_f: if we sample a vertex from $\mathcal{D}$ and then walk to a random neighbor, it is also distributed according to $\mathcal{D}$.

Fix a tree $t \in \mathcal{T}$ and consider a random walk on G_f which is the following vector $\mathbf{X} = (\mathbf{X}_0,\ldots,\mathbf{X}_j)$ of random variables:

- We sample $\mathbf{X}_0$ from $\{0,1\}^n$ according to $\mathcal{D}$.
- For $i \geq 1$, let $\mathbf{X}_i$ be a a random neighbor of $\mathbf{X}_{i'}$ in G_f where $i' = p_t(i) < i$.

Note that every $\mathbf{X}_i$ is distributed according to $\mathcal{D}$. The vector $\mathbf{X} = (\mathbf{X}_0,\ldots,\mathbf{X}_j)$ is such that $(\mathbf{X}_i, \mathbf{X}_{p_t(i)}) \in E(G_f)$, but it might contain repeated vertices and edge labels (indeed, this proof bounds the number of homomorphisms from G_f to t).

A vector $x = (x_0,\ldots,x_j) \in (\{0,1\}^n)^{j+1}$ will be sampled with probability

$$\begin{aligned}
\Pr[\mathbf{X} = x] &= \Pr[\mathbf{X}_0 = x] \prod_{i=1}^{j} \Pr[\mathbf{X}_i = x_i | \mathbf{X}_0,\ldots,\mathbf{X}_{i-1}] \\
&= \frac{s(f,x_0)}{\sum_{x\in\{0,1\}^n} s(f,x)} \prod_{i=0}^{j-1} \frac{1}{s(f,x_i)} \\
&= \frac{1}{\sum_{x\in\{0,1\}^n} s(f,x)} \prod_{i=1}^{j-1} \frac{1}{s(f,x_i)}.
\end{aligned}$$

Clearly $\mathcal{S}(t)$ lies in the support of $\mathbf{X}$, hence

$$\begin{aligned}
|\mathcal{S}(t)| &\leq \text{supp}(\mathbf{X}) \\
&\leq \mathop{\mathbb{E}}_{\mathbf{X}}\left[\frac{1}{\Pr[\mathbf{X} = x]}\right] \\
&\leq \mathop{\mathbb{E}}_{\mathbf{X}}\left[\sum_{x\in\{0,1\}^n} s(f,x) \prod_{i=1}^{j-1} s(f,\mathbf{X}_i)\right] \\
&= s_{\text{tot}}(f) \mathop{\mathbb{E}}_{\mathbf{X}}\left[\prod_{i=1}^{j-1} s(f,\mathbf{X}_i)\right] \\
&\leq s_{\text{tot}}(f) \mathop{\mathbb{E}}_{\mathbf{Y}\sim\mathcal{D}}\left[s(f,\mathbf{Y})^{j-1}\right] \qquad (8)
\end{aligned}$$

where the last inequality holds since the $\mathbf{X}_i$'s are identically distributed and each $s(f, \mathbf{X}_i)$ is non-negative (or can be derived via the AM-GM inequality). We bound the moment under $\mathcal{D}$ as follows:

$$\begin{aligned}\mathop{\mathbb{E}}_{\mathbf{Y}\sim\mathcal{D}}\left[s(f,\mathbf{Y})^{j-1}\right] &\leq \sum_{y\in\{0,1\}^n} \Pr[\mathbf{Y}=y]s(f,y)^{j-1}\\ &= \sum_{y\in\{0,1\}^n} \frac{s(f,y)}{s_{\text{tot}}(f)} s(f,y)^{j-1}\\ &= \frac{\sum_{y\in\{0,1\}^n} s(f,y)^j}{s_{\text{tot}}(f)}.\end{aligned}$$

Plugging this back into Equation (8) gives

$$|\mathcal{S}(t)| \leq \sum_{y\in\{0,1\}^n} s(f,y)^j$$

Summing over all possibilities for t, we get

$$|\mathcal{S}_j| \leq \sum_{t\in\mathcal{T}} |\mathcal{S}(t)| \leq j! \sum_{y\in\{0,1\}^n} s(f,y)^j.$$

One can save a factor of $(j+1)$, since there are $j+1$ ways to root each tree in S_j. ◀

Theorem 4.11 now follows from an argument similar to Theorem 3.1.

Proof of Theorem 4.11. The lower bound follows from (the lower bound in) Theorem 3.1 and the observation that $\mathrm{ts}(f_\rho) \geq s(f_\rho)$. We now prove the upper bound.

Similar to Theorem 3.1, consider the bipartite graph where the LHS is the set $\mathcal{S}_j$ of all sensitive trees T of size j in G_f, the RHS is the set $\mathcal{R}_{k,n}$ of all restrictions ρ, and (T, ρ) is an edge if the tree T lies in the subcube $C(\rho)$ specified by the restriction ρ. The desired probability $\Pr_{\boldsymbol{\rho}\in\mathcal{R}_{k,n}}[\mathrm{ts}(f_{\boldsymbol{\rho}}) \geq j]$ is the fraction of nodes in $\mathcal{R}_{k,n}$ that are incident to at least one edge.

We first bound the degree of each vertex on the left. To have T lying in $C(\rho)$,

- The edge labels of T must be live variables for ρ.
- The values ρ_i for the fixed coordinates $i \in [n] \setminus L(\rho)$ must be consistent with the values in $V(T)$.

The only choice is of the $(k-j)$ remaining live coordinates. Hence $T \in C(\rho)$ for at most $\binom{n-j}{k-j}$ values of ρ corresponding to choices of the remaining live variables.

The number of vertices in $\mathcal{S}_j$ is bounded using Lemma 4.12 by $|\mathcal{S}_j| \leq j! \sum_{x\in\{0,1\}^n} s(f,x)^j = j!2^n s^j(f)$, so the total number of edges is at most $\binom{n-j}{k-j}2^n j! s^j(f)$. A restriction $\rho \in \mathcal{R}_{k,n}$ is specified by a set $L(\rho)$ of k live co-ordinates where $\rho_i = \star$, and a value $\rho_i \in \{0,1\}$ for the other coordinates, and hence $|\mathcal{R}_{k,n}| = \binom{n}{k}2^{n-k}$. Recall that $\mathrm{ts}(f_\rho) \geq j$ iff $C(\rho)$ contains some tree from $\mathcal{S}_j$. Hence the fraction of restrictions ρ that have an edge incident to them is

$$\Pr_{\rho\in\mathcal{R}_{k,n}}[\mathrm{ts}(f_\rho) \geq j] \leq \frac{\binom{n-j}{k-j}2^n j! s^j(f)}{\binom{n}{k}2^{n-k}} \leq \frac{k^j 2^k s^j(f)}{\binom{n}{j}}.$$ ◀

4.3 Applications

By combining Theorems 4.9 and 4.11, we get upper and lower bounds on the probability that a random restriction of a function has large decision tree depth in terms of its sensitivity moments.

▶ **Corollary 4.13.** *Let $f : \{0,1\}^n \to \{\pm 1\}$, $\boldsymbol{\rho} \sim \mathcal{R}_{k,n}$ and $1 \le j \le k$. Then*

$$\frac{s^{\underline{j}}(f)}{n^j} \approx \frac{s^{\underline{j}}(f)}{\prod_{i=0}^{j-1}(n-i)} \le \Pr_{\boldsymbol{\rho}\in\mathcal{R}_{k,n}}[\mathrm{dt}(f_{\boldsymbol{\rho}}) \ge j] \le \frac{(2k)^{2k} s^{\sqrt{2j}}(f)}{\prod_{i=0}^{\sqrt{2j}-2}(n-i)} \approx \frac{(2k)^{2k} s^{\sqrt{2j}}(f)}{n^{\sqrt{2j}-1}}.$$

Note that the denominator in the lower bound is $n^{\Omega(j)}$ but for the upper bound, it is $n^{\Omega(\sqrt{j})}$. This quadratic gap comes from Theorem 4.9. However, if Conjecture 4.10 stating that $\mathrm{ts}(f) \ge \mathrm{dt}(f)$ were true, it would imply the following sharper upper bound.

▶ **Corollary 4.14.** *Let $f : \{0,1\}^n \to \{\pm 1\}$, $\rho \sim \mathcal{R}_{k,n}$ and $1 \le j \le k$. If Conjecture 4.10 holds, then*

$$\Pr_{\boldsymbol{\rho}\in\mathcal{R}_{k,n}}[\mathrm{dt}(f_{\boldsymbol{\rho}}) \ge j] \le \frac{(2k)^{2k} s^j(f)}{\prod_{i=0}^{j-1}(n-i)} \approx \frac{(2k)^{2k} s^j(f)}{n^j}.$$

The dependence on n here matches that in the lower bound of Corollary 4.13. Conjecture 1.3 follows from this as an easy consequence (indeed showing $\mathrm{ts}(f) \ge \deg(f)$ rather than Conjecture 4.10 suffices):

▶ **Corollary 4.15.** *Conjecture 4.10 implies Conjecture 1.3.*

Proof. We will prove that $\mathbb{I}^{\underline{k}}(f) \le (2k)^{2k} s^k(f)$. Let $\boldsymbol{\rho} \leftarrow \mathcal{R}_{k,n}$ and consider the event that $\deg(f_{\boldsymbol{\rho}}) = k$. By Theorem 3.2, we can lower bound this probability in terms of the Fourier moments of f as

$$\frac{\mathbb{I}^{\underline{k}}(f)}{\prod_{i=0}^{k-1}(n-i)} \le \Pr_{\boldsymbol{\rho}\leftarrow\mathcal{R}_{k,n}}[\deg(f_{\boldsymbol{\rho}}) = k].$$

To upper bound it, by Corollary 4.14, if Conjecture 4.10 holds, then we have

$$\Pr_{\boldsymbol{\rho}\in\mathcal{R}_{k,n}}[\deg(f_{\boldsymbol{\rho}}) \ge k] \le \Pr_{\boldsymbol{\rho}\in\mathcal{R}_{k,n}}[\mathrm{dt}(f_{\boldsymbol{\rho}}) \ge k] \le \frac{(2k)^{2k} s^k(f)}{\prod_{i=0}^{k-1}(n-i)}.$$

The claim follows by comparing the upper and lower bounds. ◀

For $k = 3, 4$, it is an easy exercise to verify that $\mathrm{dt}(f_\rho) = k$ implies $\mathrm{ts}(f_\rho) = k$. This implies that Conjecture 1.3 holds for $k = 3, 4$.

We conclude this section with an application to the class of width-w DNF formulas. In Section 3 we showed how the switching lemma implies sensitivity moment bounds for DNFs (and AC^0). Here we show the converse, how a version of the switching lemma can be derived using sensitivity moment bounds. The Satisfiability Coding Lemma of [21] implies the following moment bounds for DNFs:

▶ **Lemma 4.16.** *[21] There exists a constant c such that if f has a width-w DNF formula, then $s^k(f) \le (ckw)^k$.*

([21] proved tail bounds on the sensitivity of small-width DNFs from which a simple calculation leads to the above moment bound. We refer the reader to [12] for more details, and for an example showing the tightness of this bound.)

If Conjecture 4.10 holds, plugging these bounds into Corollary 4.14 gives that for any width-w DNF f, there exists $c', c'' > 0$ such that

$$\Pr_{\boldsymbol{\rho}\in\mathcal{R}_{k,n}}[\mathrm{dt}(f_{\boldsymbol{\rho}}) \ge k] \le \frac{(c'k)^{3k} w^k}{\prod_{i=0}^{k-1}(n-i)} \approx \left(\frac{c'' k^3 w}{n}\right)^k.$$

This nearly matches the bound one gets from Håstad's switching lemma (with k^3 in place of k). Thus proving Conjecture 4.10 would give a combinatorial proof of the switching lemma for DNFs which seems very different from the known proofs of Håstad [14] and Razborov [23].

5 Proper Walks

Since $s^j(f) \leq (s(f))^j$ for all j, one can trivially bound the sensitivity moments of a function in terms of its max sensitivity. Hence Corollaries 4.14 and 4.15 show that under Conjecture 4.10, low sensitivity functions simplify under random restrictions. In this section we prove this unconditionally. The key ingredient is a relaxation of sensitive trees that we call *proper walks*.

A *walk* W in the n-dimensional Boolean cube is a sequence of vertices $(w_0, w_1, \ldots, w_t)$ such that w_i and w_{i+1} are at Hamming distance precisely 1. We allow walk to backtrack and visit vertices more than once. We say that t is the *length* of such a walk.

Let $\ell(W) \subseteq [n]$ denote the set of coordinates that are flipped by walk W. We define $k = |\ell(W)|$ to be the *dimension* of the walk. We order the coordinates in $\ell(W)$ as $\ell_1, \ldots, \ell_k$ according to the order in which they are first flipped. For each $\ell_i \in \ell(W)$, let x_i denote the first vertex in W at which we flip coordinate i.

▶ **Definition 5.1.** A walk W is a *proper walk* for a Boolean function $f : \{0,1\}^n \to \{\pm 1\}$ if for each $\ell_i \in \ell(W)$, the vertex x_i is sensitive to ℓ_i.

Thus a walk is proper for f if the first edge flipped along a new coordinate direction is always sensitive. This implies that while walking from x_i to x_{i+1}, we are only allowed to flip a subset of the coordinates $\{\ell_1, \ldots, \ell_i\}$, hence $\mathrm{supp}(x_i \oplus x_{i+1}) \subseteq \{\ell_1, \ldots, \ell_i\}$. Hence if there is a proper walk of dimension k then there is one of length at most $k(k+1)/2$, by choosing a shortest path between x_i and x_{i+1} for each i.

In studying proper walks, it is natural to try to maximize the dimension and minimize the length. We first focus on the former. The following lemma states that the obvious necessary condition for the existence of an n-dimensional walk is in fact also sufficient:

▶ **Lemma 5.2.** *Every Boolean function $f : \{0,1\}^n \to \{\pm 1\}$ that depends on all n coordinates has a proper walk of dimension n.*

Proof. Pick $\ell_1 \in [n]$ arbitrarily and let x_1 be any vertex in $\{0,1\}^n$ which is sensitive to coordinate ℓ_1. Let $1 \leq i \leq n$. Inductively we assume we have picked coordinates $L = \{\ell_1, \ldots, \ell_i\}$ and points $X = \{x_1, \ldots, x_i\}$ so that for every $j \leq i$,

1. x_j is sensitive to ℓ_j.
2. For $j \geq 2$, $\mathrm{supp}(x_{j-1} \oplus x_j) \subseteq \{\ell_1, \ldots, \ell_{j-1}\}$.

If we visit $x_1, \ldots, x_i$ in that order and walk from each x_j to x_{j+1} along a shortest path, the resulting walk is a proper walk for f. Let C be the subcube that spans the dimensions in L and contains X.

Case 1: Some vertex in C is sensitive to a coordinate outside of L. Name this vertex x_{i+1} and the sensitive co-ordinate ℓ_{i+1}, and add them to X and L repectively. Note that $x_i \oplus x_{i+1}$ is indeed supported on $\{\ell_1, \ldots, \ell_i\}$, so both conditions (1) and (2) are met.

Case 2: No vertex in C is sensitive to a coordinate outside L. So for any co-ordinate $j \notin L$, we have $f(x) = f(x \oplus e_j)$. But this means that the set of points $X \oplus e_j$ and co-ordinates L also satisfy the inductive hypothesis (specifically conditions (1) and (2) above).

Let d denote the Hamming distance from C to the closest vertex which is sensitive to some coordinate outside L. Let z denote one such closest vertex to C (there could be many) and pick any coordinate j in which z differs from the closest point in C. If we replace X by $X \oplus e_j$, the Hamming distance to z has decreased to $d-1$. We can repeat this till the Hamming distance drops to 0, which puts us in Case (1). ◀

Given this result, it is natural to try to find full dimensional walks of the smallest possible length. The length of the walk constructed above is bounded by $\sum_{i=1}^{n}(i-1) \leq n^2/2$. Lemma A.2 in Appendix A gives an example showing that this is tight up to constants. So while we cannot improve the bound in general, we are interested in the case of functions with large decision tree complexity, where the following observation suggests that better bounds should be possible.

▶ **Lemma 5.3.** *If* $\mathrm{ts}(f) = n$*, then* f *has a proper walk of dimension* n *and length* $2n-1$*.*

The proof is by doing a pre-order traversal of a sensitive tree of dimension n. Thus if Conjecture 4.10 were true, it would imply that functions requiring full decision tree depth have proper walks of length $O(n)$. We now give an unconditional proof of this result (we will use it as an essential ingredient in our "switching lemma" later).

▶ **Theorem 5.4.** *If* $\mathrm{dt}(f) = n$*, then* f *has a proper walk of dimension* n *and length at most* $3n$*.*

Proof. The proof is by induction on n. The base case $n = 2$ is trivial since in this case there exists a proper walk of length 2. Assume the claim holds for all $n' < n$. Let f be a function where $\mathrm{dt}(f) = n$. If $\mathrm{ts}(f) = n$ we are done by Lemma 5.3, so we assume that $\mathrm{ts}(f) = m < n$. By Corollary 4.7, there is an orchard $\{T \oplus v\}$ of sensitive trees where $\dim(T) = m$. Assume by relabeling that $\ell(T) = \{1, \dots, m\}$.

Since $\mathrm{dt}(f) = n$, there exists a setting $t_1, \dots, t_m$ of variables in $[m]$ such that the restriction $f' = f|_{x_1=t_1,\dots,x_m=t_m}$ on $n' = n - m$ variables satisfies $\mathrm{dt}(f') = n - m$. By the inductive hypothesis, there exists a proper walk in f' of dimension $n-m$ and length $3(n-m)$ in the subcube $x_1 = t_1, \dots, x_m = t_m$ which starts at some vertex $s' = (t_1, \dots, t_m, s'_{m+1}, \dots, s'_n)$ and ends at some vertex $t' = (t_1, \dots, t_m, t'_{m+1}, \dots, t'_n)$, which flips all coordinates in $[n] \setminus [m]$.

Consider the tree $T \oplus v$ in the orchard such that the coordinates of $V(T \oplus v)$ in $[n] \setminus [m]$ agree with s'. Our walk can be divided into three phases:

1. By Lemma 5.3, we can visit every vertex in $T \oplus v$ using a proper walk of length at most $2m-1$ that only uses edges in $[m]$. Assume that this walk starts at a and ends at b. By our choice of v we have that $(b_{m+1}, \dots, b_n) = (s'_{m+1}, \dots, s'_n)$.
2. From b, we then walk to the vertex $s = (t_1, \dots, t_m, s'_{m+1}, \dots, s'_n)$. This only requires flipping bits in $[m]$, so it keeps the walk proper and adds only m to its length.
3. The inductive hypothesis applied to f' allows us to construct a proper walk from s to t that only walks along edges in $[n] \setminus [m]$ and has length at most $3(n-m)$.

Thus the total length of the walk is bounded by $2m - 1 + m + 3(n-m) < 3n$. ◀

5.1 Random Restrictions of Low Sensitivity Functions

In this section we prove our "switching lemma for low-sensitivity functions," Lemma 5.6. The high-level idea is to study the existence of (short) proper walks for a random restriction $f_{\boldsymbol{\rho}}$ of f, and use Theorem 5.4 to transfer a bound on the probability that $f_{\boldsymbol{\rho}}$ has such short proper walks, to a bound on the probability that $f_{\boldsymbol{\rho}}$ has full decision tree depth. Similar in spirit to Theorem 4.11, the proof proceeds by grouping walks according to their *topology*, and showing that $f_{\boldsymbol{\rho}}$ is unlikely to contain any of them. We now define the notion of a "walk topology":

▶ **Definition 5.5.** A *walk topology* of dimension k and length ℓ is a sequence $\mathrm{wt} = (\mathrm{wt}_1, \ldots, \mathrm{wt}_\ell)$ of coordinates in $[k]$ (possibly with repetitions), where all elements of $[k]$ appear in wt, and they first appear in the order $1, \ldots, k$.

Given a walk topology wt, a starting point $x_0 \in \{0,1\}^n$, and a sequence $L = (\ell_1, \ldots, \ell_k)$ of distinct coordinates in $[n]$, we get a walk $W = W(x_0, L, \mathrm{wt})$ on the n-dimensional cube by starting at x_0 and associating label i of w with coordinate ℓ_i for $i \in [k]$. The walk W has length ℓ and dimension k. Conversely, every walk W gives a unique triple (x_0, L, wt) corresponding to its starting point, order in which coordinates are first flipped, and its topology.

▶ **Lemma 5.6.** *Let $f : \{0,1\}^n \to \{\pm 1\}$. Then*

$$\Pr_{\boldsymbol{\rho} \in \mathcal{R}_{k,n}}[\mathrm{dt}(f_{\boldsymbol{\rho}}) = k] \leq \frac{(2k^3 s(f))^k}{\prod_{i=0}^{k-1}(n-i)}.$$

Proof. Fix a restriction $\rho \in \mathcal{R}_{k,n}$. Theorem 5.4 implies that if $\mathrm{dt}(f_\rho) = k$, then f_ρ contains a proper walk W of dimension k and length $3k$. Let $\mathcal{TOP}$ denote the set of all walk topologies of dimension k and length $3k$, so that $|\mathcal{TOP}| \leq k^{3k}/k!$. (Observe that of the k^{3k} strings in $[k]^{3k}$, precisely a $1/k!$ fraction of those in which all k elements appear will have them appearing first in the order $1, \ldots, k$.) Let $\mathbb{S}_K$ denote the set of permutations of the live variables K of ρ.

Fix $\mathrm{wt} \in \mathcal{TOP}$. We say that wt is *good for* $\rho \in \mathcal{R}_{k,n}$ if there exists a proper walk for f_ρ with topology wt; in other words there exists $y \in C(\rho)$ and $L \in \mathbb{S}_K$ such that $W = W(y, L, \mathrm{wt})$ is a proper walk for f. To bound the probability of this event, we first show that it suffices to consider the case when y and L are uniformly random: we have that $\Pr_{\boldsymbol{\rho} \in \mathcal{R}_{k,n}}[\mathrm{wt} \text{ is good for } f_{\boldsymbol{\rho}}]$ equals

$$\begin{aligned} &\Pr_{\boldsymbol{\rho} \in \mathcal{R}_{k,n}}[\exists y \in C(\boldsymbol{\rho}), L \in \mathbb{S}_K \text{ s.t. } W = W(y, L, \mathrm{wt}) \text{ is a proper walk for } f_{\boldsymbol{\rho}}] \\ &\leq k! 2^k \Pr_{\boldsymbol{\rho} \leftarrow \mathcal{R}_{k,n}, \boldsymbol{y} \leftarrow C(\boldsymbol{\rho}), \mathbf{L} \leftarrow \mathbb{S}_K}[W(\boldsymbol{y}, \mathbf{L}, \mathrm{wt}) \text{ is a proper walk for } f_{\boldsymbol{\rho}}], \end{aligned} \tag{9}$$

where the inequality holds since for each outcome ρ of $\boldsymbol{\rho}$ there are 2^k points $y \in C(\rho)$ and $k!$ elements $L \in \mathbb{S}_K$.

Sampling the triple $(\boldsymbol{\rho}, \boldsymbol{y}, \mathbf{L})$ is equivalent to independently sampling $\boldsymbol{x} \leftarrow \{0,1\}^n$ and $\mathbf{L} = (\boldsymbol{\ell}_1, \ldots, \boldsymbol{\ell}_k)$ by picking k coordinates uniformly from $[n]$ without replacement. It is easy to see that this determines $(\boldsymbol{\rho}, \boldsymbol{y}, \mathbf{L})$ and hence the walk $\mathbf{W} = W(\boldsymbol{y}, \mathbf{L}, \mathrm{wt})$. We can now define the sequence of points on the walk $\mathbf{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k)$ as described earlier so that $\mathbf{W}$ is proper for $f_{\boldsymbol{\rho}}$ if $\boldsymbol{x}_i$ is sensitive to $\boldsymbol{\ell}_i$. Hence

$$\begin{aligned} \Pr_{\boldsymbol{x}, \mathbf{L}}[\mathbf{W} \text{ is a proper walk for } f_{\boldsymbol{\rho}}] &= \Pr_{\boldsymbol{x}, \mathbf{L}}[\boldsymbol{x}_i \text{ is sensitive to } \boldsymbol{\ell}_i \ \forall i \in [k]] \\ &= \prod_{i \leq k} \Pr_{\boldsymbol{x}, \mathbf{L}}[\boldsymbol{x}_i \text{ is sensitive to } \boldsymbol{\ell}_i \big| \boldsymbol{x}_j \text{ is sensitive to } \boldsymbol{\ell}_j \ \forall j < i]. \end{aligned} \tag{10}$$

Let us first sample $\boldsymbol{x} \leftarrow \{0,1\}^n$, and then sample the elements of $\mathbf{L} = (\boldsymbol{\ell}_1, \ldots, \boldsymbol{\ell}_k)$ one at a time without replacement. Observe that $\boldsymbol{x}_i$ (the first time on the walk $\mathbf{W}$ that we flip

$\boldsymbol{\ell}_i$) is a function of $\boldsymbol{x}, \boldsymbol{\ell}_1, \ldots, \boldsymbol{\ell}_{i-1}$, since $\boldsymbol{x}_1 = \boldsymbol{x}$ and $\mathrm{supp}(\boldsymbol{x}_i \oplus \boldsymbol{x}_1) \subseteq \{\boldsymbol{\ell}_1, \ldots, \boldsymbol{\ell}_{i-1}\}$ (and the exact subset is specified by wt). Hence fixing outcomes $x, \ell_1, \ldots, \ell_{i-1}$ of $\boldsymbol{x}, \boldsymbol{\ell}_1, \ldots, \boldsymbol{\ell}_{i-1}$ fixes the random variables $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_i$ and whether $\boldsymbol{x}_j$ is sensitive to $\boldsymbol{\ell}_j$ for $j < i$ (the events that we condition on in Equation 10). We then sample $\boldsymbol{\ell}_i$ uniformly from the coordinates in $[n] \setminus \{\ell_1, \ldots, \ell_{i-1}\}$; crucially, $\boldsymbol{x}_i$ is sensitive to at most $s(f)$ of these coordinates. Hence

$$\Pr_{\boldsymbol{x},\mathbf{L}}[\boldsymbol{x}_i \text{ is sensitive to } \boldsymbol{\ell}_i | \boldsymbol{x}_j \text{ is sensitive to } \ell_j \ \forall j < i] \leq \frac{s(f)}{n - i + 1}.$$

Plugging this into Equation (10),

$$\Pr_{\boldsymbol{x},\mathbf{L}}[\mathbf{W}\text{is a proper walk for } f_{\boldsymbol{\rho}}] \leq \frac{(s(f))^k}{\prod_{i=0}^{k-1}(n-i)}. \tag{11}$$

Hence by Equation (9),

$$\Pr_{\boldsymbol{\rho} \in \mathcal{R}_{k,n}}[\text{wt is good for } f_{\boldsymbol{\rho}}] \leq \frac{k! 2^k (s(f))^k}{\prod_{i=0}^{k-1}(n-i)}. \tag{12}$$

Taking a union bound over all (at most) $k^{3k}/k!$ possible choices of $\mathrm{wt} \in \mathcal{TOP}$, we get that

$$\Pr_{\rho \in \mathcal{R}_{k,n}}[\mathrm{dt}(f_\rho) = k] \leq \frac{k^{3k} 2^k (s(f))^k}{\prod_{i=0}^{k-1}(n-i)} \leq \frac{(2k^3 s(f))^k}{\prod_{i=0}^{k-1}(n-i)}.$$ ◀

We note that one can prove a similar bound for $\Pr_{\rho \in \mathcal{R}_{k,n}}[\mathrm{dt}(f_\rho) \geq j]$; here we have presented only the case $j = k$ both because it is simpler and because it suffices for the concentration results in Section 5.2.

We would like to replace the $(s(f))^k$ term with $s^k(f)$, the k^{th} sensitivity moment. The above proof does not seem to generalize to that case, because we do not have an analogue of Sidorenko's result on trees for proper walks.

5.2 Fourier tails of low sensitivity functions

We have the necessary pieces in place to give an upper bound on $\mathbb{I}^{\underline{k}}[f]$:

▶ **Lemma 5.7.** *For every $f : \{\pm 1\}^n \to \{\pm 1\}$ and every $k \geq 1$, we have $\mathbb{I}^{\underline{k}}[f] \leq (2k^3 s(f))^k$.*

Proof. By Theorem 3.2 and Lemma 5.6, we have that

$$\frac{\mathbb{I}^{\underline{k}}[f]}{\prod_{i=0}^{k-1}(n-i)} \leq \Pr_{\rho \in \mathcal{R}_{k,n}}[\deg(f_\rho) = k] \leq \Pr_{\rho \in \mathcal{R}_{k,n}}[\mathrm{dt}(f_\rho) = k] \leq \frac{(2k^3 s(f))^k}{\prod_{i=0}^{k-1}(n-i)},$$

which may be rewritten as the claimed bound. ◀

Next we observe that bounding $\mathbb{I}^{\underline{k}}[f]$ yields tail bounds for the Fourier spectrum of f.

▶ **Lemma 5.8.** *For every $f : \{\pm 1\}^n \to \{\pm 1\}$, every $k \geq 1$, and every $\epsilon > 0$, we have*

$$\deg_\epsilon(f) \leq \max\left(k, e\left(\frac{\mathbb{I}^{\underline{k}}[f]}{\epsilon}\right)^{1/k}\right).$$

Proof. We first consider the case when $\mathbb{I}^{\underline{k}}[f]/k! \leq \epsilon$. In this case,

$$\sum_{|S| \geq k} \hat{f}(S)^2 \leq \sum_{|S| \geq k} \hat{f}(S)^2 \binom{|S|}{k} = \frac{\mathbb{I}^{\underline{k}}[f]}{k!} \leq \epsilon$$

so $\deg_\epsilon(f) \leq k$.

So assume that $\mathbb{I}^{\underline{k}}[f]/k! \geq \epsilon$. It suffices to prove that for

$$t_0 = e\left(\frac{\mathbb{I}^{\underline{k}}[f]}{\epsilon}\right)^{1/k}$$

we have

$$\sum_{|S|\geq t_0} \hat{f}(S)^2 = \Pr_{\mathbf{R}\leftarrow\mathcal{D}_f}[|\mathbf{R}| \geq t_0] \leq \epsilon.$$

Since $\binom{t}{k}$ is strictly incresing for $t \geq k$, we have

$$\Pr_{\mathbf{R}\leftarrow\mathcal{D}_f}[|\mathbf{R}| \geq t] = \Pr_{\mathbf{R}\leftarrow\mathcal{D}_f}\left[\binom{|\mathbf{R}|}{k} \geq \binom{t}{k}\right].$$

Now observe that we have

$$\frac{\mathbb{I}^{\underline{k}}[f]}{k!} = \mathop{\mathbb{E}}_{\mathbf{R}\leftarrow\mathcal{D}_f}\left[\binom{|\mathbf{R}|}{k}\right].$$

Hence for any t such that

$$t \geq k\left(\frac{\mathbb{I}^{\underline{k}}[f]}{k!\epsilon}\right)^{1/k} \geq k$$

Markov's inequality gives

$$\Pr_{\mathbf{R}\leftarrow\mathcal{D}_f}[|\mathbf{R}| \geq t] \leq \frac{\mathbb{E}_{\mathbf{R}\leftarrow\mathcal{D}_f}[\binom{|\mathbf{R}|}{k}]}{\binom{t}{k}} \leq \frac{\mathbb{I}^{\underline{k}}[f]}{k!\cdot(t/k)^k} \leq \epsilon,$$

One can check that t_0 satisfies the required bound using Stirling's approximation. ◀

Now we are ready to prove Theorem 1.2:

▶ **Theorem 1.2** (restated). *For any function f and any $\epsilon > 0$, we have $\deg_\epsilon(f) \leq O(s(f)(\log 1/\epsilon)^3)$.*

Proof. Applying Lemma 5.8 and Lemma 5.7, for every $f : \{\pm1\}^n \to \{\pm1\}$, every $k \geq 1$, and every $\epsilon > 0$, we have

$$\deg_\epsilon(f) \leq \max\left\{k, e\frac{2k^3 s(f)}{\epsilon^{1/k}}\right\} \leq 2es(f)\frac{k^3}{\epsilon^{1/k}}.$$

Taking $k = \log(1/\epsilon)$, we get that

$$\deg_\epsilon(f) = O(s(f)\log(1/\epsilon)^3),$$

as claimed. ◀

We note that the relations between influence moments and Fourier concentration that are established in [26, Section 4] can also be used to obtain Theorem 1.2 from Lemma 5.7. [26, Section 4] also shows that bounded k-th influence moments imply bounded Fourier L_1 spectral norm on the k-th level, which in turn implies Fourier concentration on a small number of Fourier coefficients (smaller than the trivial $\binom{n}{k}$ bound on the number of coefficients up to degree k). These results can be used with Lemma 5.7 to establish the corresponding Fourier bounds for functions with bounded max sensitivity.

6 Additional questions and complexity measures

As stated earlier, we hope that this work will stimulate further research on the sensitivity graph G_f and on complexity measures associated with it. Towards this end we conclude with some additional questions and a new complexity measure.

The graph G_f consists of a number of connected components. This component structure naturally suggests another complexity measure:

▶ **Definition 6.1.** For $x \in \{0,1\}^n$, the *component dimension of f at x*, denoted $\mathrm{cdim}(f,x)$, is the dimension of the connected component of G_f that contains x (i.e. the number of coordinates i such that x's component contains at least one edge in the i-th direction). We define $\mathrm{cdim}(f)$ to be $\max_{x\in\{0,1\}^n} \mathrm{cdim}(f,x))$.

It is easy to see that $\mathrm{cdim}(f) \geq \mathrm{ts}(f) \geq s(f)$, and thus a consequence of Conjecture 4.10 is that $\mathrm{cdim}(f) \geq \mathrm{dt}(f)$; however we have not been able to prove a better lower bound for $\mathrm{cdim}(f)$ in terms of $\mathrm{dt}(f)$ than that implied by Theorem 4.9. We note that $\mathrm{cdim}(f)$ and $\mathrm{ts}(f)$ are not polynomially related, since the addressing function shows that the gap between them can be exponential.

Lastly, it is an intriguing open question whether the reverse direction of the robust sensitivity conjecture also holds: for every k, does there exist a'_k, b'_k such that $\mathbb{E}[\boldsymbol{s}^k] \leq a'_k\,\mathbb{E}[\boldsymbol{d}^k] + b'_k$? Can one relate this question to a statement about graph-theoretic (or other) complexity measures?

Acknowledgments. We thank Yuval Peres and Laci Lovasz for pointing us to Sidorenko's theorem and the related literature. We also thank Yuval for useful discussions and for letting us present his proof of Lemma 4.12 here. We thank Gagan Aggarwal for showing us a combinatorial proof of Sidorenko's theorem. We thank Avishay Tal for helpful comments and suggestions.

References

1 Andris Ambainis, Mohammad Bavarian, Yihan Gao, Jieming Mao, Xiaoming Sun, and Song Zuo. Tighter relations between sensitivity and other complexity measures. In *Automata, Languages, and Programming – 41st International Colloquium, ICALP 2014*, pages 101–113, 2014.

2 Andris Ambainis and Krisjanis Prusis. A tight lower bound on certificate complexity in terms of block sensitivity and sensitivity. In *MFCS*, pages 33–44, 2014.

3 Andris Ambainis, Krisjanis Prusis, and Jevgenijs Vihrovs. Sensitivity versus certificate complexity of boolean functions. *CoRR*, abs/1503.07691, 2015.

4 Andris Ambainis and Jevgenijs 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*, pages 122–133, 2015.

5 M. Blum and R. Impagliazzo. Generic oracles and oracle classes. In *Proceedings of the 28th Annual Symposium on Foundations of Computer Science*, pages 118–126, 1987.

6 H. Buhrman and R. de Wolf. Complexity measures and decision tree complexity: a survey. *Theoretical Computer Science*, 288(1):21–43, 2002.

7 S. Chakraborty, R. Kulkarni, S. Lokam, and N. Saurabh. Upper Bounds on Fourier Entropy. ECCC report TR13-052 Revision #1, available at http://eccc.hpi-web.de/report/2013/052/, 2013.

8 P. Csikvári and Z. Lin. Graph homomorphisms between trees. *Electronic Journal of Combinatorics*, 21(4):P4.9, 2014.

9 E. Friedgut. Boolean functions with low average sensitivity depend on few coordinates. *Combinatorica*, 18(1):474–483, 1998.
10 Ehud Friedgut and Gil Kalai. Every monotone graph property has a sharp threshold. *Proceedings of the American mathematical Society*, 124(10):2993–3002, 1996.
11 P. Gopalan, N. Nisan, R. Servedio, K. Talwar, and A. Wigderson. Smooth boolean functions are easy: efficient algorithms for low sensitivity functions. In *ITCS*, pages 59–70, 2016.
12 P. Gopalan, R. Servedio, A. Tal, and A. Wigderson. Degree and Sensitivity: tails of two distributions. to appear, 2016.
13 J. Hartmanis and L.A. Hemachandra. One-way functions, robustness and non-isomorphism of NP-complete classes. *Theor. Comput. Sci*, 81(1):155–163, 1991.
14 J. Håstad. *Computational Limitations for Small Depth Circuits.* MIT Press, Cambridge, MA, 1986.
15 Pooya Hatami, Raghav Kulkarni, and Denis Pankratov. *Variations on the Sensitivity Conjecture.* Number 4 in Graduate Surveys. Theory of Computing Library, 2011. URL: `http://www.theoryofcomputing.org/library.html`, `doi:10.4086/toc.gs.2011.004`.
16 J. Kahn, G. Kalai, and N. Linial. The influence of variables on boolean functions. In *Proc. 29th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 68–80, 1988.
17 Claire Kenyon and Samuel Kutin. Sensitivity, block sensitivity, and l-block sensitivity of Boolean functions. *Information and Computation*, pages 43–53, 2004.
18 N. Linial, Y. Mansour, and N. Nisan. Constant depth circuits, Fourier transform and learnability. *Journal of the ACM*, 40(3):607–620, 1993.
19 N. Nisan and M. Szegedy. On the degree of Boolean functions as real polynomials. *Comput. Complexity*, 4:301–313, 1994.
20 Noam Nisan. CREW PRAMs and decision trees. *SIAM Journal on Computing*, 20(6):999–1007, 1991.
21 R. Paturi, P. Pudlák, and F. Zane. Satisfiability coding lemma. In *38th Annual Symposium on Foundations of Computer Science, FOCS'97*, pages 566–574, 1997.
22 Y. Peres. Personal communication. 2015.
23 Alexander Razborov. Bounded arithmetic and lower bounds in Boolean complexity. In *Feasible Mathematics II*, pages 344–386. Springer, 1995.
24 A. Sidorenko. A partially ordered set of functionals corresponding to graphs. *Discrete Mathematics*, 131(1-3):263–277, 1994.
25 A. Tal. Shrinkage of de Morgan Formulae from Quantum Query Complexity. ECCC report TR14-048, available at `http://eccc.hpi-web.de/report/2014/048/`, 2014.
26 A. Tal. Tight Bounds on The Fourier Spectrum of AC^0. ECCC report TR14-174 Revision #1, available at http://eccc.hpi-web.de/report/2014/174/, 2015.
27 G. Tardos. Query complexity, or why is it difficult to separate $NP^A \cap co-NP^A$ from P^A by a random oracle A? *Combinatorica*, 8(4):385–392, 1989.

A Some Examples

▶ **Lemma A.1.** *Let* $n = 2^k - 1$. *There exists* $f : \{0,1\}^n \to \{\pm 1\}$ *for which* $\mathrm{dt}(f) = \log(n+1)$ *whereas* $ts(f) = n$.

Proof. Take a complete binary tree with n internal nodes and $n+1$ leaves. The leaves are alternately labelled 1 and -1 from left to right, while the internal nodes are labelled with $x_1, \ldots, x_n$ according to an in-order traversal of the tree. The bound on decision tree depth follows from the definition of f. To lower bound $\mathrm{ts}(f)$, we start at the -1^n input and start flipping bits from -1 to 1 in the order $x_1, \ldots, x_n$. It can be verified that every bit flip changes the value of the function. ◀

▶ **Lemma A.2.** *There exists a Boolean function f on n variables such that any proper walk for f has length $\Omega(n^2)$.*

Proof. Assume that n is a power of 2 and fix a Hadamard code of length $n/2$. We define an n-variable function f over variables $x_1, \ldots, x_{n/2}$ and $y_1, \ldots, y_{n/2}$ as follows: if the string $x_1, \ldots, x_{n/2}$ equals the i-th codeword in the Hadamard code of length $n/2$, then the output is y_i, otherwise the output is 0. Note that for any $i \neq j$, if n-bit inputs a, b are sensitive to y_i, y_j respectively then the Hamming distance between a and b must be at least $n/4$. Thus any proper walk must flip at least $n/4$ bits between any two vertices that are sensitive to different y_is, so the minimum length of any proper walk must be at least $n^2/8$. ◀