Pseudorandom Generators for Low-Sensitivity Functions

Pooya Hatami*  
DIMACS  
pooyahat@math.ias.edu

Avishay Tal†  
Institute for Advanced Study  
avishay.tal@gmail.com

Abstract

A Boolean function is said to have maximal sensitivity $s$ if $s$ is the largest number of Hamming neighbors of a point which differ from it in function value. We construct a pseudorandom generator with seed-length $2^{O(\sqrt{s})} \cdot \log(n)$ that fools Boolean functions on $n$ variables with maximal sensitivity at most $s$. Prior to our work, the best pseudorandom generators for this class of functions required seed-length $2^{O(s)} \cdot \log(n)$.

1 Introduction

The sensitivity of a Boolean function $f : \{-1,1\}^n \rightarrow \{-1,1\}$ at a point $x \in \{-1,1\}^n$, denoted $s(f, x)$, is the number of neighbors of $x$ in the Hypercube whose $f$-value is different than $f(x)$. The maximal sensitivity of $f$, denoted $s(f)$, is the maximum over $s(f, x)$ for all $x \in \{-1,1\}^n$. The sensitivity conjecture by Nisan and Szegedy [Nis92, NS94] asserts that low-sensitivity functions (also called “smooth” functions) are “easy”. More precisely, the conjecture states that any Boolean function whose maximal sensitivity is $s$ can be computed by a decision tree of depth $\text{poly}(s)$. The conjecture remains wide open for several decades now, and the state-of-the-art upper bounds on decision tree complexity are merely $\exp(O(s))$.

Assuming the sensitivity conjecture, low-sensitivity functions are not any stronger than low-depth decision trees, which substantially limits their power. Hence, towards settling the conjecture, it is natural to inspect how powerful low-sensitivity functions are. One approach that follows this idea aims to prove limitations of low-sensitivity functions, which follow from the sensitivity conjecture, unconditionally. This line of work was initiated recently by Gopalan et al. [GNS+16], who considered low-sensitivity functions as a complexity class. Denote by $\text{Sens}(s)$ the class of Boolean functions with sensitivity at most $s$. The sensitivity conjecture asserts that $\text{Sens}(s) \subseteq \text{DecTree-depth}(\text{poly}(s))$, which then implies

\[
\text{Sens}(s) \subseteq \text{DecTree-depth}(\text{poly}(s)) \subseteq \text{DNF-size}(2^{\text{poly}(s)}) \subseteq \text{AC}^0\text{-size}(2^{\text{poly}(s)}) \subseteq \text{Formula-depth}(\text{poly}(s)) \subseteq \text{Circuit-size}(2^{\text{poly}(s)}),
\]

*Partially supported by the National Science Foundation under agreement No. CCF-1412958.
†Supported by the Simons Collaboration on Algorithms and Geometry, and by the National Science Foundation grant No. CCF-1412958.
whereas Gopalan et al. [GNS+16] proved that $\text{Sens}(s) \subseteq \text{Formula-depth}(\text{poly}(s))$ unconditionally. It remains open to prove that $\text{Sens}(s)$ is contained in smaller complexity classes such as $\text{AC}^0\text{-size}(2^{\text{poly}(s)})$ or even $\text{TC}^0\text{-size}(2^{\text{poly}(s)})$.

One consequence of the sensitivity conjecture is the existence of pseudorandom generators (PRGs) with short seeds fooling low-sensitivity functions. This follows since $k$-wise independence fools degree $k$ functions and the sensitivity conjecture asserts that $\deg(f) \leq \text{poly}(s(f))$ for any Boolean function $f$. Thus, under the conjecture, the standard construction of $k$-wise distributions gives a PRG with seed length $\deg(f) \cdot \log(n) \leq \text{poly}(s) \cdot \log(n)$ fooling $\text{Sens}(s)$. The goal of our work is to construct PRGs fooling $\text{Sens}(s)$ unconditionally. We fall short of achieving seed length $\text{poly}(s) \cdot \log(n)$ and get the weaker seed length of $2^{O(\sqrt{n})} \cdot \log(n)$. Nonetheless, prior to our work, only seed-length $2^{O(s)} \cdot \log(n)$ was known, which follows from the state of the art upper bounds on degree in terms of sensitivity $\deg(f) \leq 2^{s(1+o(1))}$ [ABG+14].

The paradigm of Hardness vs Randomness, initiated by Nisan and Wigderson [NW94], asserts that PRGs and average-case lower bounds are essentially equivalent, for almost all reasonable complexity classes. For example, the average-case lower bound of Håstad [Hås86] for the parity function by $\text{AC}^0$ circuits implies a pseudorandom generator fooling $\text{AC}^0$ circuits with poly-logarithmic seed-length. This general transformation of hardness to randomness is achieved via the NW-generator, which constructs a PRG based on the hard function. In [GSTW16], it was proved that low-sensitivity functions can be $\varepsilon$-approximated by real polynomials of degree $O(s \cdot \log(1/\varepsilon))$, which implies that the parity function on $n$ variables can only have agreement $1/2 + 2^{-\Omega(n/s)}$ with Boolean functions of sensitivity $s$. In other words, the parity function on $n$ variables is average-case hard for the class $\text{Sens}(s)$. It thus seems very tempting to use the parity function in the NW-generator to construct a PRG fooling $\text{Sens}(s)$, however, the proof does not follow through since the class of low-sensitivity functions is not closed under the transformations made by the analysis of the NW-generator (in particular it is not closed under identifying a set of the input variables with one variable). We do not claim that the NW-generator with the parity function does not fool $\text{Sens}(s)$, but we point out that the argument in the standard proof breaks. (See more details in Appendix A).

### 1.1 Our Results

A function $G : \{-1,1\}^r \rightarrow \{-1,1\}^n$ is said to be a pseudorandom generator with seed-length $r$ that $\varepsilon$-fools a class of Boolean functions $\mathcal{C}$ if for every $f \in \mathcal{C}$:

$$\left| \varepsilon \cdot \mathbb{E}_{z \in \{-1,1\}^r} [f(G(z))] - \mathbb{E}_{x \in \{-1,1\}^n} [f(x)] \right| \leq \varepsilon.$$  

In other words, any $f \in \mathcal{C}$ cannot distinguish (with advantage greater than $\varepsilon$) between an input sampled according to the uniform distribution over $\{-1,1\}^n$ and an input sampled according to the uniform distribution over $\{-1,1\}^r$ and expanded to an $n$-bit string using $G$.

The main contribution of this paper is the first pseudorandom generator for low-sensitivity Boolean functions with subexponential seed length in the sensitivity.

---

$^1$Even under the weaker conjecture $\text{Sens}(s) \subseteq \text{AC}^0\text{-size}(n^{\text{poly}(s)})$, we would get that $\text{poly}(s, \log n)$-wise independence fools $\text{Sens}(s)$ via the result of [Bra10].
Theorem 1.1. There is a distribution $D$ on $\{-1,1\}^n$ with seed-length $2^{O(\sqrt{s+\log(1/\varepsilon)})} \cdot \log(n)$ that $\varepsilon$-fools every $f : \{-1,1\}^n \rightarrow \{-1,1\}$ with $s(f) = s$.

Our construction relies on the following strengthening of Friedgut’s Theorem for low sensitivity functions. (In the following, we denote by $W^k[f] = \sum_{S \subseteq \{1,\ldots,n\} : |S| \geq k} \hat{f}(S)^2$.)

Lemma 1.2. Let $f : \{-1,1\}^n \rightarrow \{-1,1\}$ with $s(f) \leq s$. Let $1 \leq k \leq s/10$. Assume $W^k[f] \leq 2^{-6s}$, and that at most $2^{-6s}$ fraction of the points in $\{-1,1\}^n$ have sensitivity at least $k$. Then, $f$ is a $2^{20k}$-junta.

1.2 Proof Outline

Below we give a sketch of our proof of Theorem 1.1.

Similar to a construction of Trevisan and Xue [TX13], our pseudorandom generator involves repeated applications of “pseudorandom restrictions”. Using Lemma 1.2 and studying the behavior of the Fourier spectrum of low-sensitivity functions under pseudorandom restrictions, we are able to prove the following. Let $f : \{-1,1\}^n \rightarrow \{-1,1\}$ be a Boolean function, let $S \subseteq [n]$ be randomly selected according to a $k$-wise independent distribution such that $|S| \approx pn$, and let $x_S = (x_i)_{i \in S} \in \{-1,1\}^{|S|}$ be selected uniformly at random. Then

$$\Pr_{S,x_S}[f(x_S, \_\_\_) \text{ is not a $2^{20k}$-junta}] \leq O(ps)^k \cdot 2^{6s}. \tag{1}$$

Since every $2^{20k}$-junta is fooled by an almost $2^{20k}$-wise independent distribution, we will fill the $x_S$ coordinates according to efficient constructions of such distributions due to [AGHP92].

The final distribution involves applying the above process repeatedly over the remaining unset variables (i.e. $x_\overline{S}$) until all the coordinates are set, observing that for every $J \subseteq [n]$ and $x_J$, $f(\_\_, x_J)$ has sensitivity at most $s$. The subexponential seed-length is achieved by optimizing the parameters $k$ and $p$ from (1) while making sure that the overall error does not exceed $\varepsilon$.

Discussion

Our overall construction involves a combination of several samples from any $k$-wise independent distribution for an appropriate $k$. It is not clear whether simply one sample from a $k$-wise independent distribution suffices to fool low sensitivity functions (recall that this is a consequence of the sensitivity conjecture with $k = \text{poly}(s)$). If this were true for all $k$-wise independent distributions, then via LP Duality (see the work of Bazzi [Baz09]) we would get that every Boolean function $f$ with sensitivity $s$ has sandwiching real polynomials $f_t$, $f_u$ of degree $k$ such that $\forall x : f_t(x) \leq f(x) \leq f_u(x)$ and $E_x[f_u(x) - f_t(x)] \leq \varepsilon$. We ask if a similar characterization can be obtained for the class of functions fooled by our construction.

2 Preliminaries

We denote by $[n] = \{1,\ldots,n\}$. We denote by $U_n$ the uniform distribution over $\{-1,1\}^n$. We denote by $\log$ and $\ln$ the logarithms in bases 2 and $e$, respectively. For $f : \{-1,1\}^n \rightarrow \mathbb{R}$,
we denote by $\|f\|_p = (E_{x \in \{-1,1\}^n}[|f(x)|^p])^{1/p}$. For $x \in \{-1,1\}^n$, denote by $x \oplus e_i$ the vector obtained from $x$ by changing the sign of $x_i$.

For a Boolean function $f : \{-1,1\}^n \to \{-1,1\}$, denote by $S(f, y)$, the set of sensitive coordinates of $f$ at $y$, i.e.,

$$S(f, y) \triangleq \{i \in [n] : f(y) \neq f(y \oplus e_i)\}.$$  

The sensitivity of $f$, denoted $s(f, x)$, is defined to be the number of sensitive coordinates of $f$, namely $s(f, x) = |S(f, x)|$. For example if $f(x_1, x_2, x_3) = x_1x_2$, then $s(f, 111) = 2$ and $S(f, 111) = \{1, 2\}$. The sensitivity of a Boolean function $f$, denoted $s(f)$ is the maximum $s(f, x)$ over all choices of $x$.

### 2.1 Harper’s Inequality

**Theorem 2.1** (Harper’s Inequality). Let $G = (V, E)$ be the $n$-dimensional hypercube, where $V = \{-1,1\}^n$. Let $A \subseteq V$ be a non-empty set. Then,

$$\frac{|E(A, A^c)|}{|A|} \geq \log_2 \left( \frac{2^n}{|A|} \right).$$

We will use the following simple corollary of Harper’s inequality on multiple occasions:

**Corollary 2.2.** Let $f : \{-1,1\}^n \to \{-1,1\}$ be a non-constant function with $s^1(f) \leq s$. Then, $|f^{-1}(1)| \geq 2^{n-s}$.

**Proof.** Let $A = f^{-1}(1)$. Since $f$ is non-constant, $|A| > 0$. By Harper’s inequality the average sensitivity of $f$ on $A$ is at least $\log(2^n/|A|)$. However the average sensitivity of $f$ on $A$ is at most $s$, hence $\log(2^n/|A|) \leq s$, or equivalently, $|A| \geq 2^{n-s}$.

### 2.2 Restrictions

**Definition 2.3** (Restriction). Let $f : \{-1,1\}^n \to \{-1,1\}$ be a Boolean function. A restriction is a pair $(J, z)$ where $J \subseteq [n]$ and $z \in \{-1,1\}^J$. We denote by $f_{J,z} : \{-1,1\}^n \to \{-1,1\}$ the function $f$ restricted according to $(J, z)$, defined by

$$f_{J,z}(x) = f(y), \quad \text{where} \quad y_i = \begin{cases} x_i, & i \in J \\ z_i, & \text{otherwise} \end{cases}.$$  

**Definition 2.4** (Random Valued Restriction). Let $n \in \mathbb{N}$. A random variable $(J, z)$, distributed over restrictions of $\{-1,1\}^n$ is called random-valued if conditioned on $J$, the variable $z$ is uniformly distributed over $\{-1,1\}^J$.

**Definition 2.5** ($(p,k)$-wise Random Selection). A random variable $J \subseteq [n]$ is said to be a $(p,k)$-wise random selection if the events $\{(1 \in J), (2 \in J), \ldots, (n \in J)\}$ are $k$-wise independent, and each one of them happens with probability $p$.

A $(k,p)$-wise independent restriction is a random-valued restriction in which $J$ is chosen using a $(k,p)$-wise independent selection.
2.3 Fourier Analysis of Boolean Functions

Any function $f : \{-1, 1\}^n \to \mathbb{R}$ has a unique Fourier representation:

$$f(x) = \sum_{S \subseteq [n]} \hat{f}(S) \cdot \prod_{i \in S} x_i,$$

where the coefficients $\hat{f}(S) \in \mathbb{R}$ are given by $\hat{f}(S) = \mathbb{E}_x[f(x) \cdot \prod_{i \in S} x_i]$. Parseval’s identity states that $\sum_{S} \hat{f}(S)^2 = \mathbb{E}[f(x)^2] = \|f\|_2^2$, and in the case that $f$ is Boolean (i.e., $f : \{-1, 1\}^n \to \{-1, 1\}$), all are equal to 1. The Fourier representation is the unique multilinear polynomial, which also equals $\max\{|S| : \hat{f}(S) \neq 0\}$. We denote by

$$W^k[f] \triangleq \sum_{S \subseteq [n], |S| = k} \hat{f}(S)^2$$

the Fourier weight at level $k$ of $f$. Similarly, we denote $W^{\geq k}[f] \triangleq \sum_{S \subseteq [n], |S| \geq k} \hat{f}(S)^2$. For $k \in \mathbb{N}$ we denote the $k$-th Fourier moment of $f$ by

$$\text{Inf}^k[f] \triangleq \sum_{S \subseteq [n]} \hat{f}(S)^2 \cdot \binom{|S|}{k} = \sum_{d=1}^{n} W^d[f] \cdot \binom{d}{k}.$$

We will use the following result of Gopalan et al. [GSTW16].

**Theorem 2.6 ([GSTW16]).** Let $f$ be a Boolean function with sensitivity at most $s$. Then, for all $k$, $\text{Inf}^k[f] \leq (16 \cdot s)^k$.

For more about Fourier moments of Boolean functions see [Tal14, GSTW16]. The following fact relates the Fourier coefficients of $f$ and $f_{J,z}$, where $(J, z)$ is a random valued restriction.

**Fact 2.7** (Proposition 4.17, [O’D14]). Let $f : \{-1, 1\}^n \to \mathbb{R}$, let $S \subseteq [n]$, and let $D$ be a distribution of random valued restrictions. Then,

$$\mathbb{E}_{(J, z) \sim D} \left[ \hat{f}_{J,z}(S) \right] = \hat{f}(S) \cdot \mathbb{Pr}_{(J, z) \sim D}[S \subseteq J]$$

and

$$\mathbb{E}_{(J, z) \sim D} \left[ \hat{f}_{J,z}(S)^2 \right] = \sum_{U \subseteq [n]} \hat{f}(U)^2 \cdot \mathbb{Pr}_{(J, z) \sim D}[J \cap U = S].$$

We include the proof of this fact for completeness.

**Proof.** Let $(J, z) \sim D$. Then, by definition of random valued restriction, given $J$ we have that $z$ is a random string in $\{-1, 1\}^J$.

Fix $J$, and rewrite $f$’s Fourier expansion by splitting the variables to $(J, \bar{J})$.

$$f(x) = \sum_{S \subseteq [n]} \hat{f}(S) \cdot \prod_{i \in S} x_i = \sum_{T \subseteq J} \prod_{i \in T} x_i \cdot \sum_{T' \subseteq \bar{J}} \hat{f}(T \cup T') \cdot \prod_{j \in T'} x_j$$
Hence,
\[ f_{J,z}(x) = \sum_{T \subseteq J} \prod_{i \in T} x_i \cdot \sum_{T' \subseteq J} \hat{f}(T \cup T') \cdot \prod_{j \in T'} z_j \]

So the Fourier coefficient of \( S \) on \( f_{J,z} \) is 0 if \( S \not\subseteq J \) and it is \( \sum_{T' \subseteq J} \hat{f}(S \cup T') \cdot \prod_{j \in T'} z_j \) otherwise. In other words,
\[ \hat{f}_{J,z}(S) = 1_{S \subseteq J} \cdot \sum_{T' \subseteq J} \hat{f}(S \cup T') \cdot \prod_{j \in T'} z_j , \]
and it’s expectation in \( z \) in the case \( S \subseteq J \) is \( \hat{f}(S) \). As for the second moment,
\[
\mathbb{E}[\hat{f}_{J,z}(S)^2] = \mathbb{E}[\mathbb{E}[\hat{f}_{J,z}(S)^2]] = \mathbb{E}[1_{S \subseteq J} \cdot \mathbb{E}[\sum_{T' \subseteq J} \hat{f}(S \cup T') \prod_{j \in T'} z_j]^2]
\]
\[ = \mathbb{E}[1_{S \subseteq J} \cdot \sum_{T' \subseteq J} \hat{f}(T \cup T')^2] = \sum_{U \subseteq [n]} \hat{f}(U)^2 \cdot \Pr[J \cap U = S]. \]

\section{PRGs for Low-Sensitivity Functions}

In this section we prove our main theorem.

\textbf{Theorem 1.1.} There is a distribution \( D \) on \( \{-1,1\}^n \) with seed-length \( 2^{O(\sqrt{s} \log(1/\varepsilon))} \cdot \log(n) \) that \( \varepsilon \)-fools every \( f : \{-1,1\}^n \to \{-1,1\} \) with \( s(f) = s \).

Our main tool will be the following theorem stating that under \( k \)-wise independent random restrictions every low-sensitivity function becomes a junta with high probability. We postpone the proof of Theorem 3.1 to Section 4.

\textbf{Theorem 3.1.} Let \( f : \{-1,1\}^n \to \{-1,1\} \) with \( s(f) = s \). Let \( 1 \leq k \leq s/10 \), and let \( D \) be a distribution of \((k,p)\)-wise independent restrictions. Then,
\[ \Pr_{(J,z) \sim D}[f_{J,z} \text{ is not a } (2^{20k})\text{-junta}] \leq O(ps)^k \cdot 2^{6s} \]

Theorem 3.1 allows us to employ the framework of Trevisan and Xue [TX13] who used a derandomized switching lemma to construct pseudorandom generators for AC0 circuits. In what follows we will make the following choices of parameters

i. \( k := O(\sqrt{s + \log(1/\varepsilon)}) \).

ii. \( p := 2^{-k}/s = 2^{-O(\sqrt{s + \log(1/\varepsilon))})} \)

iii. \( m := O(p^{-1} \cdot \log(s \cdot 4^s/\varepsilon)) = 2^{O(\sqrt{s + \log(1/\varepsilon))})} \)

We select a sequence of disjoint sets \( J_1, \ldots, J_m \) as follows. We pick \( J_i \subseteq [n] \setminus (J_1 \cup \cdots \cup J_{i-1}) \) by letting \( J_i := K_i \setminus (J_1 \cup \cdots \cup J_{i-1}) \) where \( K_i \subseteq [n] \) is drawn from a \((p,k)\)-wise independent selection. For each \( i \), we pick \( x_i \in \{-1,1\}^{|J_i|} \) according to an \( \frac{\varepsilon}{4m} \)-almost \( 2^{20k} \)-wise independent distribution. Finally, we will fix \( x_i := 0 \) for any \( i \in [n] \setminus (J_1 \cup \cdots \cup J_m) \).

To account for the seed-length:
• By a construction of [AB186] each $K_i$ can be selected using $O(k \cdot \log n)$ random bits, and
• By constructions of [AGHP92] each $x_{J_i} \in \{-1, 1\}^{\left|J_i\right|}$ can be selected using $O(2^{20k} + \log \log(n) + \log(1/\varepsilon))$ random bits.

Thus, the total seed-length is

$$O \left( m \cdot \left( 2^{20k} + \log \log(n) + \log(1/\varepsilon) + k \cdot \log(n) \right) \right) \leq 2^{O(\sqrt{s} \log(1/\varepsilon))} \cdot \log(n).$$

To conclude the proof, we show that the above distribution fools sensitivity $s$ Boolean functions. Denote by $\mathcal{D}$ the distribution described above, and suppose $f : \{-1, 1\}^n \rightarrow \{-1, 1\}$ satisfies $s(f) = s$. We first note that by a result of Simon [Sim83], $f$ depends on at most $s \cdot 4^s$ variables, denote this set $S$, so that $|S| \leq s \cdot 4^s$. By our choice of $m$, with probability at least $1 - \frac{\varepsilon}{2}$, $S \subseteq J_1 \cup \cdots \cup J_m$.

We use $x$ to denote a vector drawn from $\mathcal{D}$ and $y$ to denote a vector drawn according to the uniform distribution over $\{-1, 1\}^n$. Moreover, for every $i = 0, 1, \ldots, m$, we let $z_i := (x_{J_1}, \ldots, x_{J_i}, y_{[n] \setminus (J_1 \cup \cdots \cup J_i)})$. Note that $z_0 = y$. We first prove that for every $i = 0, 1, \ldots, m - 1$,

$$\left| \frac{\mathbb{E}}{x \sim \mathcal{D}, y \sim \mathcal{U}} f(z_i) - \frac{\mathbb{E}}{x \sim \mathcal{D}, y \sim \mathcal{U}} f(z_{i+1}) \right| \leq \frac{\varepsilon}{2m}. \quad (2)$$

This holds since by Theorem 3.1, for every fixed choice of $J_1, \ldots, J_i$ and $x_{J_1}, \ldots, x_{J_i}$, we have

$$\Pr_{J_{i+1}, y \sim \mathcal{U}} \left[ f(x_{J_1}, \ldots, x_{J_i}, \cdot, y_{[n] \setminus (J_1 \cup \cdots \cup J_{i+1})}) \right. \left. \text{ is not a } 2^{20k}-\text{junta} \right] \leq O(ps)^k \cdot 2^{5s} \leq \frac{\varepsilon}{4m},$$

and that every $2^{20k}$-junta is $\varepsilon/4m$-fooled by any $\varepsilon/4m$-almost $2^{20k}$-wise independent distribution. By triangle inequality and summing up (2) for all $i$ we get

$$\left| \frac{\mathbb{E}}{y \sim \mathcal{U}} f(y) - \frac{\mathbb{E}}{x \sim \mathcal{D}, y \sim \mathcal{U}} f(z_m) \right| \leq \sum_{i=0}^{m-1} \left| \frac{\mathbb{E}}{x \sim \mathcal{D}, y \sim \mathcal{U}} f(z_i) - \frac{\mathbb{E}}{x \sim \mathcal{U}, y \sim \mathcal{D}} f(z_{i+1}) \right| \leq \frac{\varepsilon}{2}. \quad (3)$$

To finish the proof of Theorem 1.1, note that with probability at least $1 - \varepsilon/2$, $f(x_{J_1}, \ldots, x_{J_m}, \cdot)$ is a constant function (which follows from $S \subseteq J_1 \cup \cdots \cup J_m$), and thus $|\mathbb{E}_{x,y} f(z_m) - \mathbb{E}_x f(x)| \leq \varepsilon/2$. Combining this with Eq. (3) gives $|\mathbb{E}_{y \sim \mathcal{U}} f(y) - \mathbb{E}_{x \sim \mathcal{D}} f(x)| \leq \varepsilon/2 + \varepsilon/2$.

4 Measures of Boolean Functions under $k$-Wise Independent Random Restrictions

Lemma 4.1. Let $t \in \mathbb{R}^+$ and $f : \{-1, 1\}^n \rightarrow \{-1, 1\}$. Let $\mathcal{D}$ be a distribution of $(k, p)$-wise independent restrictions. Then, for any $d \leq k$ we have

$$\mathbb{E}_{(J,z) \sim \mathcal{D}} \left[ W^d[f_{J,z}] \right] \leq p^d \cdot \text{Inf}^d[f]. \quad (4)$$
Theorem 3.1. Let \( S \subseteq \) a given \( y \) where the second to last inequality is due to the following observation. We observe that for \((a distribution of \( k \)-wise independent restrictions. Then, we upper and lower bound the value of \( \text{Pr}_J[S \subseteq J] = p^d \) by the fact that \( J \) is a \( k \)-wise \( p \)-random restriction. We thus have

\[
E_{J,z}[W^{zd}[f|J,z]] \leq \sum_{U \subseteq [n]} \hat{f}(U)^2 \cdot \left( \frac{|U|}{d} \right) \cdot p^d = \text{Inf}^d[f] \cdot p^d.
\]

Lemma 4.2. Let \( f : \{-1, 1\}^n \to \{-1, 1\} \), with \( s(f) = s \). Let \( D \) be a distribution of \((k, p)\)-wise independent restrictions. Then,

\[
E_{(J,z) \sim D} \left[ \text{Pr}_x[s(f|J,z), x] \geq k \right] \leq (ps)^k.
\]

Proof. We expand \( E_{(J,z) \sim D} \left[ \text{Pr}_x[s(f|J,z), x] \geq k \right] \):

\[
E_{J,z} \left[ \text{Pr}_x[s(f|J,z), x] \geq k \right] = E_{J,z} \left[ E_{x \in \{-1, 1\}^n} \left[ \text{Pr}_x[s(f(z), x) \geq k] \right] \right]
\]

\[
= E_{J,z} \left[ E_{x \in \{-1, 1\}^n} \left[ \text{Pr}_y[s(f(y), y) \geq k] \right] \right]
\]

\[
\leq E_{y \in \{-1, 1\}^n} \left[ \text{Pr}_J[J \cap S(f, y) \geq k] \right]
\]

\[
\leq (ps)^k
\]

where the second to last inequality is due to the following observation. We observe that for a given \( y \) and a set \( S = \{i_1, \ldots, i_k\} \) of \( k \) sensitive directions of \( f \) at \( y \), the probability that \( S \subseteq J \) is \( p^k \). We then union-bound over all subsets \( S \) of cardinality \( k \) of \( S(f, y) \).

We are now ready to prove the main theorem of this section (restated next).

Theorem 3.1. Let \( f : \{-1, 1\}^n \to \{-1, 1\} \) with \( s(f) = s \). Let \( 1 \leq k \leq s/10 \), and let \( D \) be a distribution of \((k, p)\)-wise independent restrictions. Then,

\[
\text{Pr}_{(J,z) \sim D}[f|J,z] \text{ is not a } (2^{20k})\text{-junta} \leq O(ps)^k \cdot 2^{6s}
\]

Proof. We upper and lower bound the value of

\[
(*) = E_{(J,z) \sim D} \left[ W^{\geq k}[f|J,z] + \text{Pr}_x[s(f|J,z), x] \geq k] \right].
\]

8
For the upper bound we use Lemma 4.2 to get
\[ \mathbb{E}_{(J,z) \sim \mathcal{D}} \left[ \Pr_x[s(f_{J,z}, x) \geq k] \right] \leq (ps)^k, \]
and Lemma 4.1 and Theorem 2.6 to get
\[ \mathbb{E}_{(J,z) \sim \mathcal{D}} \left[ W^{\geq k}[f_{J,z}] \right] \leq O(ps)^k, \]
which gives \((*) \leq O(ps)^k\).

For the lower bound we use the following lemma, the proof of which we defer to Section 5.

**Lemma 1.2.** Let \( f : \{-1,1\}^n \to \{-1,1\} \) with \( s(f) \leq s \). Let \( 1 \leq k \leq s/10 \). Assume \( W^{\geq k}[f] \leq 2^{-6s} \), and that at most \( 2^{-6s} \) fraction of the points in \( \{-1,1\}^n \) have sensitivity at least \( k \). Then, \( f \) is a \( 2^{20k} \)-junta.

Let \( \mathcal{E} \) be the event that \( f_{J,z} \) is not a \( 2^{20k} \)-junta. Whenever \( \mathcal{E} \) occurs, Lemma 1.2 implies that either \( \Pr_x[s(f_{J,z}, x) \geq k] \leq 2^{-6s} \) or \( W^{\geq k}[f_{J,z}] \geq 2^{-6s} \). In both cases, \( \Pr_x[s(f_{J,z}, x) \geq k] + W^{\geq k}[f_{J,z}] \geq 2^{-6s} \). Thus, we get the lower bound
\[ (*) \geq \Pr[\mathcal{E}] \cdot \mathbb{E}_{(J,z)} \left[ W^{\geq k}[f_{J,z}] + \Pr_x[s(f_{J,z}, x) \geq k] \mid \mathcal{E} \right] \geq \Pr[\mathcal{E}] \cdot 2^{-6s} \]

Comparing the upper and lower bound gives
\[ \Pr_{(J,z) \sim \mathcal{D}}[f_{J,z} \text{ is not a } K\text{-junta}] = \Pr[\mathcal{E}] \leq 2^{6s} \cdot (*) \leq 2^{6s} \cdot O(ps)^k. \]

## 5 A Strengthening of Friedgut’s Theorem for Low Sensitivity Functions

**Theorem 5.1** (Friedgut’s Junta Theorem - [O’D14, Thm 9.28]). Let \( f : \{-1,1\}^n \to \{-1,1\} \). Let \( 0 < \varepsilon \leq 1 \) and \( k \geq 0 \). If \( W^{> k}[f] \leq \varepsilon \), then \( f \) is \( 2\varepsilon \)-close to a \((9^k \cdot \inf |f|^2 / \varepsilon^2)\)-junta.

**Lemma 1.2.** Let \( f : \{-1,1\}^n \to \{-1,1\} \) with \( s(f) \leq s \). Let \( 1 \leq k \leq s/10 \). Assume \( W^{\geq k}[f] \leq 2^{-6s} \), and that at most \( 2^{-6s} \) fraction of the points in \( \{-1,1\}^n \) have sensitivity at least \( k \). Then, \( f \) is a \( 2^{20k} \)-junta.

**Proof.** We first show that \( \inf |f| \leq k \). By Simon’s work [Sim83], \( f \) depends on at most \( 4^s \cdot s \) variables\(^2\). Thus, \( \inf |f| \leq (k - 1) + W^{\geq k}[f] \cdot (4^s \cdot s) \leq (k - 1) + 1 = k \). Apply Friedgut’s theorem with \( \varepsilon = 2^{-6k-1} \geq W^{\geq k}[f] \). We get a \( K \)-junta \( h \), for
\[ K = 9^k \cdot \inf |f|^3 / \varepsilon^2 \leq 9^k \cdot k^3 \cdot 2^{12k+2} < 2^{20k}, \]
that \( 2\varepsilon = 2^{-6k} \) approximates \( f \). Let \( C_1, \ldots, C_N \) be the subcube corresponding to the \( N = 2^k \) different assignments to the junta variables. Without loss of generality, under each \( C_i \), \( h \)

\(^2\)Note that our final goal will be to show that \( f \) actually depends on \( 2^{20k} \) variables, and that \( k \) can be significantly smaller than \( s \).
attains the constant value that is the majority-vote of $f$ on $C_i$. In other words, $f$ and $h$ agree on at least $1/2$ of the points in each subcube $C_i$.

Let $p_i = \frac{|\{x \in C_i : f(x) \neq h(x)\}|}{|C_i|}$, for $i \in [N]$. By the above discussion, $0 \leq p_i \leq 1/2$. In addition, since $f|_{C_i}$ has sensitivity at most $s$, if $p_i > 0$, then $p_i \geq 2^{-s}$ using Corollary 2.2.

Assume towards contradiction that $h \neq f$. We will think of the hamming cube $\{-1, 1\}^n$ as an outer cube of dimension $K$, and an inner cube of dimension $n-K$. Each subcube $C_i$ is an instance of the inner cube $\{-1, 1\}^{n-K}$. The graph of subcubes is an instance of the outer cube $\{-1, 1\}^K$. Call a subcube $C_i$:

- **decisive** if $p_i = 0$,
- **confused** if $2^{-s} \leq p_i < 2^{-k-1}$, or
- **indecisive** if $p_i \geq 2^{-k-1}$.

Denote by $\alpha, \beta, \gamma$ the fraction of decisive, confused and indecisive subcubes correspondingly.

Since we assumed (towards contradiction) that $h \neq f$, at least one subcube is confused or indecisive. Consider the graph $G$ of subcubes, which is isomorphic to $\{-1, 1\}^K$, in which each vertex represents either a decisive, confused or indecisive subcube, and two vertices are adjacent if and only if their corresponding subcubes are adjacent in $\{-1, 1\}^n$. First, we show that at least $2^{-2s}$ fraction of the subcubes are confused or indecisive. Assume otherwise, then by Harper’s inequality (Thm. 2.1) there is a confused or indecisive cube $C_i$ with at least $2s+1$ decisive subcubes as neighbors. As there are points with both $\{-1, 1\}$ values in $C_i$, we may pick a point $x \in C_i$ whose value is the opposite of the majority of the decisive neighbor subcubes of $C_i$, which gives $s(f, x) \geq s+1$, a contradiction. We thus have

$$\beta + \gamma \geq 2^{-2s} \tag{5}$$

Next, we show that $\beta$ is very small and in particular much smaller than $\gamma$. Towards this end, we shall analyze the sensitivity within confused subcubes. If $C_i$ is confused (i.e., $2^{-s} \leq p_i < 2^{-k-1}$), then by Harper’s inequality (inside $C_i$) the average sensitivity on the minority of $f|_{C_i}$ is greater than $k+1$. Since sensitivity ranges between 0 to $s$, at least $1/s$ of the points with minority value in $f|_{C_i}$ have sensitivity at least $k$ (otherwise the average sensitivity among them will be less than $(1/s) \cdot s + k \leq k+1$). As there are at least $2^{-s}$ points with the minority value on the subcube $C_i$, we get that at least $2^{-s}/s \geq 2^{-2s}$ fraction of the points in $C_i$ have sensitivity at least $k$.

If the fraction of confused subcubes is more than $2^{-2s}/(K+1)$, then more than $2^{-4s}/(K+1) \geq 2^{-6s}$ fraction of the points in $\{-1, 1\}^n$ has sensitivity at least $k$, which contradicts one of the assumptions. Thus,

$$\beta \leq 2^{-2s}/(K+1). \tag{6}$$

Furthermore, combining Eq. (5) and (6), we have that the fraction of indecisive subcubes, $\gamma$, is at least

$$\gamma \geq 2^{-2s} \cdot \frac{K}{K+1} \geq K \cdot \beta. \tag{7}$$

Consider again the graph $G$ of subcubes (which is isomorphic to $\{-1, 1\}^K$). Recall that each vertex in the graph $G$ corresponds to a subcube which is either decisive, confused
or indecisive. Call $A$ the set of vertices that correspond to indecisive subcubes. Then, $|A| = \gamma \cdot 2^K$. By the fact that $h$ approximates $f$ with error at most $2^{-6k}$, the size of $A$ is at most $2^{-6k} \cdot 2^{k+1} \cdot 2^K \leq 2^{-4k} \cdot 2^K$, i.e., $\gamma \leq 2^{-4k}$. By Harper's inequality, $|E(A, A)| \geq |A| \cdot (4K - 1)$. There are at most $\beta \cdot 2^K \cdot K \leq \gamma \cdot 2^K = |A|$ edges touching confused nodes, hence there are at least $|A| \cdot (4K - 1)$ edges from $A$ to decisive nodes. As before, the maximal number of edges from a node in $A$ to decisive nodes is at most $2s$, otherwise we get a contradiction to $s(f) \leq s$. This implies that at least $1/2s$ fraction of the nodes in $A$ have at least $(4k - 2)/2 \geq 2k - 1 \geq k$, and the fraction of such points in $C_i$ is at least $2^{-k-1}$. Using Eq. (7) we get that

$$\gamma \cdot \frac{1}{2s} \cdot 2^{-k-1} \geq 2^{-2s} \cdot \frac{K}{K + 1} \cdot \frac{1}{2s} \cdot 2^{-k-1} \geq 2^{-6s}$$

of the points in $\{-1, 1\}^n$ have sensitivity at least $k$, which yields a contradiction.

## Acknowledgements

We would like to thank Li-Yang Tan for bringing the problem to our attention and for stimulating and helpful discussions.

## References


A Does the NW-Generator Fool Low-Sensitivity Functions?

In this section we recall the construction and analysis of the NW-Generator [NW94]. For ease of notation, we treat Boolean functions here as $f : \{0, 1\}^n \to \{0, 1\}$. Suppose we want to construct a pseudorandom generator fooling a class of Boolean functions $\mathcal{C}$. Nisan and Wigderson provide a generic way to construct such PRGs based on the premise that there is some explicit function $f$ which is average-case hard for a class $\mathcal{C}'$ that slightly extends $\mathcal{C}$. Recall that $\textbf{Sens}(s)$ is the class of all Boolean functions with sensitivity at most $s$. In the case $\mathcal{C} = \textbf{Sens}(s)$, the argument may fail, because $\mathcal{C}'$ is not provably similar to $\mathcal{C}$. The difficulty comes from the fact that low-sensitivity functions are not closed under projections as will be explained later.

Let $f : \{0, 1\}^\ell \to \{0, 1\}$ be a function that is average-case hard for class $\mathcal{C}$. Let $S_1, \ldots, S_n \subseteq [r]$ be a design over a universe of size $r$ where $|S_i| = \ell$, and $|S_i \cap S_j| \leq \alpha$ for all $i \neq j \in [n]$ (think of $\alpha$ as much smaller than $\ell$). The NW-generator $G_f : \{0, 1\}^r \to \{0, 1\}^n$ is defined as

$$G_f(x_1, \ldots, x_r) = (f(x_{S_1}), f(x_{S_2}), \ldots, f(x_{S_n}))$$

where $x_{S_i}$ is the restriction of $x$ to the coordinates in $S_i$, for any set $S_i \subseteq [n]$. 


The proof that the NW-generator fools $C$ goes via a contrapositive argument. We assume that there is a distinguisher $c \in C$ such that

$$\left| \mathbb{E}_{z \in \mathbb{R}^n} \left[ c(G_f(z)) \right] - \mathbb{E}_{x \in \mathbb{R}^n} \left[ c(x) \right] \right| \geq \varepsilon,$$  

and prove that $f$ can be computed on more than $1/2 + \Omega(\varepsilon)/n$ fraction of the inputs by some function $c''$ which is not much more complicated than $c$. First, by Yao’s next-bit predictor lemma, there exists an $i \in [n]$ and constants $a_i, \ldots, a_n, b \in \{0, 1\}$ such that

$$\Pr_{x \in \{0, 1\}^n} \left[ c(f(x_{i}), f(x_{i+1}), \ldots, f(x_{i+k})), a_i, \ldots, a_n) \oplus b = f(x_{i}) \right] \geq \frac{1}{2} + \frac{\Omega(\varepsilon)}{n}.$$  

Since the class of function with sensitivity $s$ is closed under restrictions (i.e., fixing the input variables to constant values) and negations we have that $c'(z_1, \ldots, z_{i-1}) := c(z_1, \ldots, z_{i-1}, a_i, \ldots, a_n) \oplus b$ is of sensitivity at most $s$. We get

$$\Pr_{x \in \{0, 1\}^n} \left[ c'(f(x_{i}), f(x_{i+1}), \ldots, f(x_{i+k})) = f(x_{i}) \right] \geq \frac{1}{2} + \frac{\Omega(\varepsilon)}{n}.$$  

Next, we wish to fix all values in $[r] \setminus S_i$. By averaging there exists an assignment $y$ to the variables in $[r] \setminus S_i$ such that

$$\Pr_{x \in \{0, 1\}^S_i} \left[ c'(f((x \circ y)_{S_1}), f((x \circ y)_{S_2}), \ldots, f((x \circ y)_{S_{i-1}})) = f(x_{i}) \right] \geq \frac{1}{2} + \frac{\Omega(\varepsilon)}{n}.$$  

Note that for $j = 1, \ldots, i-1$, the value of $f((x \circ y)_{S_j})$ depends only on the variables in $S_j \cap S_i$ and there aren’t too many such variables (at most $\alpha$). The next step is to consider $c'' : \{0, 1\}^S_i \rightarrow \{0, 1\}$, defined by $c''(x) = c'(f((x \circ y)_{S_1}), f((x \circ y)_{S_2}), \ldots, f((x \circ y)_{S_{i-1}}))$, that have agreement at least $1/2 + \Omega(\varepsilon)/n$ with $f(x_{S_i})$. If $c''$ is a “simple” function then we get a contradiction as $f$ is average-case hard.

It seems that $c''$ is simple, since it is the composition of $c'$ with $\alpha$-juntas. However, the point that we want to make is that even if $c'$ is low-sensitivity and even if $\alpha = 1$, we are not guaranteed that $c''$ is of low-sensitivity.

To see this, suppose that $\alpha = 1$, i.e., all $|S_j \cap S_i| \leq 1$ for $j < i$. This means that as a function of $x$, each $f((x \circ y)_{S_j})$ depends on at most one variable, i.e., $f((x \circ y)_{S_j}) = a_j \cdot x_{k_j} \oplus b_j$ for some index $k_j \in S_i$ and some constants $a_j, b_j \in \{0, 1\}$. We get that

$$c''(x) = c'(a_1 \cdot x_{k_1} \oplus b_1, a_2 \cdot x_{k_2} \oplus b_2, \ldots, a_{i-1} \cdot x_{k_{i-1}} \oplus b_{i-1}).$$  

Next, we argue that $c''$ could potentially have very high sensitivity. To see that, observe that flipping one bit $x_i$ in the input to $c''$ results in changing a block of variables in the input to $c'$, as there may be several $j$ for which $k_j = i$. In the worst-case scenario, the sensitivity of $c''$ could be as big as the block sensitivity of $c'$. However, the best known bound is only $bs(f) \leq 2^{s(f) \cdot (1+o(1))}$ for any Boolean function $f$ [ABG+14]. This means that we can only guarantee that $s(c'') \leq bs(c') \leq 2^{s(1+o(1))}$, and we do not have average-case hardness for such high-sensitivity functions.
**Remark:** The above argument shows that the standard analysis of the Nisan-Wigderson generator applied to low-sensitivity Boolean functions breaks, but it does not mean that the generator does not ultimately fool $\text{Sens}(s)$. Indeed, assuming the sensitivity conjecture, the argument will follow through.