Low degree functions which are almost Boolean

Yuval Filmus

28 July 2023

1 Introduction

If $f\colon \{0,1\}^n \to \{0,1\}$ has degree 1 then f is a dictator. Friedgut, Kalai and Naor [FKN02] showed that a similar result holds when f is only assumed to be *close* to Boolean:

Theorem 1 (FKN). *If a degree 1 function $f\colon \{0,1\}^n \to \mathbb{R}$ is ϵ-close to Boolean then f is $O(\epsilon)$-close to a dictator.*

Here the assumption is that $\mathbb{E}[\operatorname{dist}(f,\{0,1\})^2] \le \epsilon$, and the conclusion is that $\mathbb{E}[(f-g)^2] = O(\epsilon)$ for some dictator g.

Kindler and Safra [Kin03] extended this to degree d functions:

Theorem 2 (Kindler–Safra). *If a degree d function $f\colon \{0,1\}^n \to \mathbb{R}$ is ϵ-close to Boolean then f is $O(\epsilon)$-close to a Boolean degree d function.*

The original statement of Kindler and Safra approximates f by a Boolean junta g which doesn't necessarily have degree d. We deduce this version by observing that if g doesn't have degree d then it must be somewhat far from f since the latter does have degree d.

In contrast to the case $d = 1$, in general we do not have a list of all Boolean degree d function (though such a list does exist for $d = 2$). Nevertheless, we can say that all such functions are juntas, and moreover can be represented as decision trees of depth $poly(d)$.

Up to now, we have implicitly been considering the uniform measure on the hypercube $\{0,1\}^n$. Next we explore what happens on the biased hypercube and on the symmetric group.

2 FKN theorems

2.1 Biased hypercube

The biased hypercube is $(\{0,1\}^n, \mu_p)$, where $\mu_p(x) = p^{|x|}(1-p)^{n-|x|}$. We assume throughout that $p \le 1/2$.

One might expect that the FKN theorem extends literally to the p-biased setting. However, it turns out that there do exist degree 1 functions which are close to Boolean but not to any dictator. Consider the following function:

$$f = x_1 + \cdots + x_{\delta/p}.$$

This is a degree 1 function whose distribution is roughly Poisson with expectation δ. In particular, $\Pr[f \notin \{0,1\}] = \Theta(\delta^2)$, and it turns out that f is also $\Theta(\delta^2)$-close to Boolean.

The FKN theorem for the biased hypercube [Fil16] states that (up to negation), these are the only degree 1 functions which are close to Boolean:

Theorem 3 (Biased FKN). *If a degree 1 function $f\colon (\{0,1\}^n, \mu_p) \to \mathbb{R}$ (where $p \le 1/2$) is ϵ-close to Boolean then either f or $1-f$ is $O(\epsilon)$-close to*

$$x_{i_1} + \cdots + x_{i_m},$$

where $m \le \max\bigl(1, O(\sqrt{\epsilon}/p)\bigr)$.

Furthermore, all such functions are $O(\epsilon)$*-close to Boolean.*

The "furthermore" part shows that the structure promised by the theorem is tight.

A direct corollary of the biased FKN theorem shows that a degree 1 function which is close to Boolean is essentially constant:

Corollary 4. *If a degree* 1 *function* $f\colon (\{0,1\}^n, \mu_p) \to \mathbb{R}$ *(where* $p \le 1/2$*) is* ϵ*-close to Boolean then* f *is* $O(\sqrt{\epsilon} + p)$*-close to constant (i.e., to the constant function* b *for some* $b \in \{0,1\}$*).*

Furthermore, f *is* $O(\sqrt{\epsilon})$*-close to a dictator.*

The exponent of ϵ is tight since when $p \ll \sqrt{\epsilon}$, the function $x_1 + \cdots + x_{\sqrt{\epsilon}/p}$ is $\Theta(\epsilon)$-close to Boolean and $\Theta(\sqrt{\epsilon})$-close to constant.

2.2 Symmetric group

A degree 1 function on the symmetric group S_n is a function $f\colon S_n \to \mathbb{R}$ of the form

$$f(x) = \sum_{i,j=1}^{n} c_{ij} x_{ij},$$

where x_{ij} indicates that i gets mapped to j. More generally, a degree d on the symmetric group is one that can be expressed as a degree d polynomial in the x_{ij}'s.

Ellis, Friedgut and Pilpel [EFP11] determined all Boolean degree 1 functions on the symmetric group:

- For some $i \in [n]$ and $J \subseteq [n]$, the function
$$\sum_{j \in J} x_{ij}.$$
- For some $j \in [n]$ and $I \subseteq [n]$, the function
$$\sum_{i \in I} x_{ij}.$$

In the first case, $x(i)$ dictates the value of the function (thinking of x as a permutation), and in the second case, it is $x^{-1}(j)$ which dictates. We can think of these functions as the analogs of dictators in the symmetric group.

As in the case of the biased cube, there exist degree 1 functions other than dictators which are close to Boolean. Let $K \subseteq [n] \times [n]$, and consider the function

$$f = \sum_{(i,j) \in K} x_{ij}.$$

We say that two pairs $(i_1, j_1), (i_2, j_2) \in K$ *conflict* if $i_1 \ne i_2$ and $j_1 \ne j_2$. If no two pairs in K conflict, then f is a dictatorship. If $(i_1, j_1), (i_2, j_2) \in K$ conflict then f is not Boolean, since with probability $\frac{1}{n(n-1)}$, a random permutation satisfies $x_{i_1 j_1} = x_{i_2 j_2} = 1$. If there are m conflicting pairs then the probability that f is not Boolean is bounded by (roughly) m/n^2, and consequently, if $m = O(\epsilon n^2)$ then f is Boolean with probability $1 - O(\epsilon)$.

The FKN theorem for the symmetric group [Fil21] states that (up to negation), these are the only degree 1 functions which are close to Boolean:

Theorem 5 (FKN on S_n)**.** *If a degree* 1 *function* $f\colon S_n \to \mathbb{R}$ *is* ϵ*-close to Boolean then either* f *or* $1 - f$ *is* $O(\epsilon)$*-close to a function of the form*

$$\sum_{(i,j) \in K} x_{ij},$$

where $|K| = O(n)$ *and* K *contains* $O(\epsilon n^2)$ *conflicting pairs.*

Furthermore, all such functions are $O(\epsilon)$*-close to Boolean.*

A simple corollary of this FKN theorem shows that a degree 1 function which is close to Boolean is close to some dictator, once we allow a larger error parameter:

Corollary 6. *If a degree* 1 *function* $f\colon S_n \to \mathbb{R}$ *is* ϵ*-close to Boolean then* f *is* $O(\sqrt{\epsilon})$*-close to a dictator. If furthermore* $\delta \leq \mathbb{E}[f] \leq 1-\delta$*, where* $\delta \geq \sqrt{\epsilon}$*, then* f *is* $O(\epsilon/\delta)$*-close to a dictator.*

The exponent of ϵ in the first statement is tight for the function $f = x_{11} + \cdots + x_{\sqrt{\epsilon n}\sqrt{\epsilon n}}$, where $1/n \ll \sqrt{\epsilon}$.

3 Biased Kindler–Safra theorem

The Kindler–Safra theorem also extends to the p-biased setting [DFH23]. In contrast to the biased FKN theorem, we no longer have an explicit expression such as $x_i + \cdots + x_{\sqrt{\epsilon/p}}$. Instead, our description is more similar to the FKN theorem for the symmetric group:

Theorem 7 (Biased Kindler–Safra)**.** *If a degree* d *function* $f\colon (\{0,1\}^n, \mu_p)$ *(where* $p \leq 1/2$*) is* ϵ*-close to Boolean then* f *is* $O(\epsilon)$*-close to a degree* d *polynomial* g *satisfying the following properties:*

- *If* $|x| \leq d$ *then* $g(x) \in \{0,1\}$*.*
- *For every* T*, the "link of* T*" contains* $O(1/p^e)$ *monomials for every* $e \leq d - |T|$*.*
- *The number of "minimally non-Boolean" inputs of weight* e *is* $O(\epsilon/p^e)$*.*

Conversely, every such function is $O(\epsilon)$*-close to Boolean.*

We need to explain what is the "link of T" and what are "minimally non-Boolean" inputs:

- We identify g with its support hypergraph, which contains a hyperedge S whenever the coefficient of $\prod_{i\in S} x_i$ in g is non-zero. The *link of* T consists of all hyperedges containing T, with the set T removed.
- An input x is *minimally non-Boolean* if $g(x) \notin \{0,1\}$ but $g(y) \in \{0,1\}$ for all inputs $y \neq x$ obtained from x by changing 1s to 0s.

Let us see what these properties mean in the case $d = 1$, both for the p-biased cube and for the symmetric group (for which the theorem is not yet known). In the case of the symmetric group, we take $p = 1/n$, which is the bias of each variable x_{ij}.

- Either $g = \sum_{i\in K} x_i$ or $g = 1 - \sum_{i\in K} x_i$, where $K \subseteq [n]$.
  In the case of the symmetric group, we similarly get that $g = \sum_{(i,j)\in K} x_{ij}$ or $g = 1 - \sum_{(i,j)\in K} x_{ij}$.
- $|K| = O(1/p)$.
  In the case of the symmetric group, this property implies that $|K| = O(n)$.
- A minimal non-Boolean input is $e_i + e_j$ for some $i, j \in K$. The number of minimal non-Boolean inputs of weight 2 is $\binom{|K|}{2}$, and so either $|K| \leq 1$ or $|K|^2 = O(\epsilon/p^2)$, implying that $|K| = O(\sqrt{\epsilon}/p)$.
  In the case of the symmetric group, a minimal non-Boolean input is the same as a conflicting pair, and so the number of conflicting pairs is at most $O(\epsilon n^2)$.

When $d = 1$, there is a single type of non-Boolean input. In general, there are many, but all of them have weight $O(2^d)$, and in particular, there are finitely many types.

Perhaps the most mysterious property here is the second. For the empty link $T = \emptyset$, it states that the expected number of monomials in the support of g evaluating to 1 is bounded, hence the term *sparse junta*. For general T, it implies that the same holds even assuming that $x_{i_1} = \cdots = x_{i_m} = 1$ for some constant m.

3.1 Monotone version

When the function f is monotone, we can choose g to be monotone as well:

Theorem 8 (Monotone biased Kindler–Safra). *If $f\colon (\{0,1\}^n, \mu_p) \to \{0,1\}$ (where $p \leq 1/2$) is ϵ-close to degree d then f is $O(\epsilon)$-close to a width w monotone DNF satisfying the following properties:*

- *For every T, the link of T contains $O(1/p^e)$ terms for every $e \leq d - |T|$.*
- *The number of minimally non-degree-d inputs of weight e is $O(\epsilon/p^e)$.*

Here the assumption is that f is ϵ-close to some degree d function, and the conclusion is that f is $O(\epsilon)$-close to a monotone DNF in which all terms have size at most w and satisfies the two stated properties. An input $x = 1_S$ is *minimally non-degree-d* if the restriction of g to S (obtained by zeroing out all other variables) doesn't have degree d, but the restriction of g to any proper subset of S does have degree d. Such inputs have bounded weight.

When $d = 1$, we obtain that f is close to a DNF $x_{i_1} \vee \cdots \vee x_{i_m}$, where $m \leq \max\bigl(1, O(\sqrt{\epsilon}/p)\bigr)$.

3.2 Constant and junta approximations

As in the preceding settings, the biased Kindler–Safra theorem implies that the function in question is close to a constant and to a junta.

Corollary 9. *If a degree d function $f\colon (\{0,1\}^n, \mu_p) \to \mathbb{R}$ (where $p \leq 1/2$) is ϵ-close to Boolean then f is $O(\epsilon^{c_d} + p)$-close to a constant function.*

Furthermore, f is $O(\epsilon^{C_d})$-close to a junta.

We can determine the value of the constant c_d, which is obtained by the following construction. Let P be a degree d polynomial such that $P(0) \neq P(1)$ and $P(0), \ldots, P(t-1) \in \{0,1\}$, and consider the function

$$f = P(x_1 + \cdots + x_{\delta/p}).$$

The output of f has the same distribution as $P(\text{Poisson}(\delta))$. In particular, f is not Boolean with probability $O(\delta^t)$. Since $P(0) \neq P(1)$, the function f is $\Omega(\delta)$-far from constant. Consequently, $c_d \leq 1/t$.

Lemma 10. *The optimal value of the constant c_d of Corollary 9 is $1/t$ for the maximal t such that there exists a degree d polynomial P satisfying $P(0) \neq P(1)$ and $P(0), \ldots, P(t-1) \in \{0,1\}$.*

A variant of the same parameter was studied by von zur Gathen and Roche [vzGR97]. In their version, the condition $P(0) \neq P(1)$ is replaced by the condition that P be non-constant.

When $d = 1$, the optimal polynomial is $P(x) = x$, and so $c_1 = 1/2$. When $d = 2$, it is $P(x) = x - \binom{x}{2}$, and so $c_2 = 1/4$. In general, it is always the case that $1/c_d \leq 2d$.

Unfortunately we don't have a similar formula for the other constant C_d.

4 Proof at a glance

The main idea behind the proofs of Theorems 3, 5, 7 and 8 is a reduction to the unbiased Boolean hypercube.

Let us start with the case of the p-biased hypercube, where $p \leq 1/2$. We can sample a point according to $\mu_p(\{0,1\}^n)$ using the following two-step process:

1. Sample a set $S \sim \mu_{2p}(\{0,1\}^n)$, and zero out all coordinates outside of S.
2. Sample the coordinates in S according to the uniform distribution over $\{0,1\}^S$.

Suppose that f is a degree d function which is ϵ-close to Boolean. On average, the restriction of f to $\{0,1\}^S$ is ϵ-close to Boolean, and so close to a dictator (when $d = 1$) or to a junta (when $d > 1$) by the unbiased FKN and Kindler–Safra theorems. By "pasting together" these dictators or juntas (in the latter case, using an agreement theorem), we obtain the approximating function g.

By construction, g has degree d. It is $O(\epsilon)$-close to f (at least in probability) due to the FKN or Kindler–Safra theorem. Regarding the other properties:

- If $|x| \leq d$ then $g(x) \in \{0,1\}$. This property is inherited from the FKN and Kindler–Safra theorems. The function directly obtained by pasting together the local approximating juntas can have a few bad inputs, but these can be corrected to obtain this property.
- For every T, the link of T contains $O(1/p^e)$ monomials. This follows from the local approximating functions being juntas, due to the fact that the coefficients of g are obtained from the local approximations by "majority decoding".
- The number of minimally non-Boolean inputs of weight e is $O(\epsilon/p^e)$. This follows from g being $O(\epsilon)$-close to f and so $O(\epsilon)$-close to Boolean.
  We obtained the bound $O(\epsilon/p^e)$ using a simple union bound. The preceding property (sparsity of all links) implies that we can "reverse" the union bound in this case, that is, that the union bound is tight up to a constant factor.

The FKN theorem for the symmetric group is obtained using a similar approach, via the following two-step process for sampling a random permutation in S_n:

1. Choose two random permutations $a_1, \ldots, a_n$ and $b_1, \ldots, b_n$.
2. Sample $\lfloor n/2 \rfloor$ random bits $c_1, \ldots, c_{\lfloor n/2 \rfloor}$.
   For every $i \in [\lfloor n/2 \rfloor]$, if $c_i = 0$ then map a_{2i-1}, a_{2i} to b_{2i-1}, b_{2i}, and if $c_i = 1$ then map a_{2i-1}, a_{2i} to b_{2i}, b_{2i-1}. If n is odd then map a_n to b_n.

5 Future directions

The most obvious next step is proving a Kindler–Safra theorem for the symmetric group. Such a theorem would have two parts:

- A result analogous to Theorem 7, giving an approximating function up to error $O(\epsilon)$.
- A result analogous to Corollary 9, stating that the function is $O(\epsilon^{k_d})$-close to a Boolean degree d function.

In contrast to the setting of the hypercube, a Boolean degree d function on the symmetric group need not be a junta. For example, the function

$$x_{12}x_{21} + \cdots + x_{1n}x_{n1}$$

has degree 2 but depends on all "coordinates". Nevertheless, it can still be represented as a decision tree of depth $poly(d)$ whose queries are of the form $x(i) = ?$ and $x^{-1}(j) = ?$ [DFL+21].

Another interesting extension is to other domains, such as the Grassmann scheme and the bilinear scheme. Even the case $d = 1$ is not known. In fact, a complete characterization of all Boolean degree 1 functions on these domains is also missing, though some special cases are known [FI19].

References

[DFH23] Irit Dinur, Yuval Filmus, and Prahladh Harsha. Sparse juntas on the biased hypercube, 2023.

[DFL+21] Neta Dafni, Yuval Filmus, Noam Lifshitz, Nathan Lindzey, and Marc Vinyals. Complexity measures on the symmetric group and beyond. In *12th Innovations in Theoretical Computer Science*, volume 185 of *LIPIcs*, pages Art. No. 87, 5. Schloss Dagstuhl, 2021.

[EFP11] David Ellis, Ehud Friedgut, and Haran Pilpel. Intersecting families of permutations. *J. Amer. Math. Soc.*, 24(3):649–682, 2011.

[FI19] Yuval Filmus and Ferdinand Ihringer. Boolean degree 1 functions on some classical association schemes. *J. Combin. Theory Ser. A*, 162:241–270, 2019.

[Fil16] Yuval Filmus. Friedgut-Kalai-Naor theorem for slices of the Boolean cube. *Chic. J. Theoret. Comput. Sci.*, pages Art. 14, 17, 2016.

[Fil21] Yuval Filmus. Boolean functions on S_n which are nearly linear. *Discrete Anal.*, pages 25:1–25:27, 2021.

[FKN02] Ehud Friedgut, Gil Kalai, and Assaf Naor. Boolean functions whose Fourier transform is concentrated on the first two levels. *Adv. in Appl. Math.*, 29(3):427–437, 2002.

[Kin03] Guy Kindler. *Property Testing, PCP and Juntas.* PhD thesis, Tel Aviv University, 2003.

[vzGR97] Joachim von zur Gathen and James R. Roche. Polynomials with two values. *Combinatorica*, 17(3):345–362, 1997.