A simple proof of the Kindler–Safra theorem

Yuval Filmus

March 27, 2022

1 Introduction

A Boolean degree 1 function (on a Boolean cube $\{0,1\}^n$) is a dictator. Friedgut, Kalai and Naor [FKN02] showed that if a Boolean function is *close* to degree 1, then it is *close* to a Boolean dictator.

Nisan and Szegedy [NS94] showed that a Boolean degree d function is an $O(d2^d)$-junta. The size of the junta was later improved to $O(2^d)$, which is optimal up to the hidden constant [CHS20, Wel20]. Kindler and Safra [KS02, Kin03] shows that if a Boolean degree d function is *close* to degree d, then it is *close* to a Boolean degree d function.

In this note, we give a simple proof of the Kindler–Safra theorem using the invariance principle. Our proof works in the more general setting of A*-valued functions*, which are functions whose output lies in some finite set A; this generalizes the Boolean setting, which corresponds to $A = \{0,1\}$.

2 Nisan–Szegedy

We start by showing that A-valued degree d functions are juntas. While this can be proved by reduction to the Boolean case,[1] we give a direct proof which relies only on hypercontractivity. The main idea is the following dichotomy:

Lemma 2.1. *Let B be a finite set and let $d \geq 1$. If f is a B-valued degree d function on $\{0,1\}^n$, then either $f = 0$ or $\|f\|^2 = \Omega(1)$.*

Here $\|f\|^2 = \mathbb{E}[f^2]$, where the underlying distribution is the uniform distribution, and the hidden constant in $\Omega(1)$ depends on B, d.

Proof. First observe that if $y \in B$ then

$$y^2 \leq My^4, \text{ where } M = \max_{0 \neq b \in B} \frac{1}{b^2}.$$

[1] Given an A-valued function f, write $f = \sum_{a \in A} a f_a$, where $f_a = \prod_{b \neq a} \frac{f-b}{a-b}$ is Boolean.

Since $\deg f \leq d$, hypercontractivity shows that $\mathbb{E}[f^4] \leq 9^d \mathbb{E}[f^2]^2$. On the other hand, the observation above shows that $\mathbb{E}[f^2] \leq M \mathbb{E}[f^4]$. Therefore $\mathbb{E}[f^2] \leq 9^d M \mathbb{E}[f^2]^2$. Consequently, either $\mathbb{E}[f^2] = 0$ or $\mathbb{E}[f^2] \geq 1/(9^d M)$. □

In order to deduce that an A-valued degree d function f must be a junta, we apply Lemma 2.1 not to f itself but to its *Laplacians* $L_i f(x) = \frac{f(x) - f(x^{\oplus i})}{2}$. Recall that $L_i f = \sum_{i \in S} \hat{f}(S) \chi_S$ and $\|f_i\|^2 = \mathrm{Inf}_i[f]$. The Laplacians are B-valued for

$$B = \frac{A - A}{2} := \left\{ \frac{a_1 - a_2}{2} : a_1, a_2 \in A \right\}.$$

Theorem 2.2. *Let A be a finite set and let $d \geq 1$. If f is an A-valued degree d function on $\{0,1\}^n$, then f depends on $O(1)$ coordinates.*

Proof. Let $B = \frac{A-A}{2}$. For every $i \in [n]$, the function $L_i f$ is a B-valued degree d function satisfying $\|L_i f\|^2 = \mathrm{Inf}_i[f]$, and so Lemma 2.1 shows that either $\mathrm{Inf}_i[f] = 0$ or $\mathrm{Inf}_i[f] = \Omega(1)$. Since $\sum_i \mathrm{Inf}_i[f] \leq d\|f\|^2 \leq d \max_{a \in A} a^2$, it follows that at most $O(1)$ many coordinates can have non-zero influence. The function f only depends on those coordinates. □

3 Kindler–Safra

The starting point of our proof of the Kindler–Safra theorem is a version of Lemma 2.1 for functions which are close to degree d.

Lemma 3.1. *Let B be a finite set and let $d \geq 1$. If f is a B-valued function on $\{0,1\}^n$ satisfying $\|f^{>d}\|^2 = \epsilon$ then either $\|f\|^2 = O(\epsilon)$ or $\|f^{\leq d}\|^2 = \Omega(1)$.*

Proof. As in the proof of Lemma 2.1, we can find M such that $b^2 \leq M b^4$ for all $b \in B$. Given β, we want to bound $(b-\beta)^2$ in terms of b^4 and β^2. If $|\beta| \leq |b|/2$ then $\rho = (b-\beta)/b$ satisfies $|\rho| \geq 1/2$, and so

$$(b-\beta)^2 = \rho^2 b^2 \leq M \rho^2 b^4 = M \rho^{-2} (b-\beta)^4 \leq 4M(b-\beta)^4.$$

In contrast, if $|\beta| \geq |b|/2$ then

$$(b-\beta)^2 \leq 9\beta^2.$$

Therefore for all β we have

$$(b-\beta)^2 \leq 4M(b-\beta)^4 + 9\beta^2.$$

Hypercontractivity shows that $\mathbb{E}[(f^{\leq d})^4] \leq 9^d \mathbb{E}[(f^{\leq d})^2]^2$. On the other hand, applying the observation above to $b = f$ and $\beta = f^{>d}$ shows that $\mathbb{E}[(f^{\leq d})^2] \leq \mathbb{E}[f^2] \leq 4M \mathbb{E}[(f^{\leq d})^4] + 9\epsilon$. Altogether, this gives

$$\mathbb{E}[(f^{\leq d})^2] \leq 4M 9^d \mathbb{E}[(f^{\leq d})^2]^2 + 9\epsilon.$$

If $\mathbb{E}[(f^{\leq d})^2] \leq 18\epsilon$ then we are done, since

$$\|f\|^2 = \|f^{\leq d}\|^2 + \|f^{>d}\|^2 \leq 19\epsilon.$$

In contrast, if $\mathbb{E}[(f^{\leq d})^2] \geq 18\epsilon$ then

$$\frac{1}{2}\mathbb{E}[(f^{\leq d})^2] \leq \mathbb{E}[(f^{\leq d})^2] - 9\epsilon \leq 4M9^d\,\mathbb{E}[(f^{\leq d})^2]^2,$$

and so $\|f^{\leq d}\|^2 \geq 1/(8M9^d)$. □

Now suppose that f is an A-valued function on $\{0,1\}^n$ satisfying $\|f^{>d}\|^2 = \epsilon$. As in the proof of Theorem 2.2, we apply Lemma 3.1 to the Laplacians $L_i f$. However, the conclusion is slightly different.

Lemma 3.2. *Let A be a finite set and let $d \geq 1$. If f is an A-valued function on $\{0,1\}^n$ such that $\|f^{>d}\|^2 = \epsilon$ then we can find a set J of $O(1)$ coordinates such that for each $z \in \{0,1\}^J$, the function f_z on $\{0,1\}^{\overline{J}}$ obtained by substituting $x|_J = z$ satisfies*

$$\mathrm{Inf}_i[f_z] = O(\epsilon) \text{ for every } i \in \overline{J}.$$

Proof. Let $B = \frac{A-A}{2}$. For every $i \in \overline{J}$, the function $L_i f$ is a B-valued function satisfying

$$\|(L_i f)^{>d}\|^2 = \sum_{\substack{i\in S\\|S|>d}} \hat{f}(S)^2 \leq \sum_{|S|>d} \hat{f}(S)^2 = \|f^{>d}\|^2 = \epsilon.$$

Therefore Lemma 3.1 shows that either $\mathrm{Inf}_i[f] = O(\epsilon)$ or $\mathrm{Inf}_i[f^{\leq d}] = \Omega(1)$. Since

$$\sum_{i=1}^n \mathrm{Inf}_i[f^{\leq d}] \leq d\|f^{\leq d}\|^2 \leq d\|f\|^2 \leq d\max_{a\in A} a^2,$$

at most $O(1)$ many variables can satisfy $\mathrm{Inf}_i[f^{\leq d}] = \Omega(1)$. Put all these variables in a set J. If $i \notin J$ then $\mathrm{Inf}_i[f] = O(\epsilon)$, and so for each assignment $z \in \{0,1\}^J$,

$$\mathrm{Inf}_i[f_z] = \mathbb{E}[(L_i f_z)^2] = \mathbb{E}[(L_i f)^2 \mid x_J = z] \leq 2^{|J|}\,\mathbb{E}[(L_i f)^2] = 2^{|J|}\,\mathrm{Inf}_i[f] = O(\epsilon). \qquad \square$$

In order to complete the proof, we would like to show that each f_z is nearly constant, in the sense that it has low variance. The first step is to apply Lemma 3.1.

Lemma 3.3. *Assume the setting of Lemma 3.2.*

For every $z \in \{0,1\}^J$, either $\mathrm{Var}[f_z] = O(\epsilon)$ or $\mathrm{Var}[f_z] = \Omega(1)$.

Proof. We define a function g_z on $\{0,1\}^{\overline{J}} \times \{0,1\}^{\overline{J}}$ as follows:

$$g_z(x,y) = f_z(x) - f_z(y).$$

Since f_z is A-valued, g_z is $(A-A)$-valued. Also,

$$\|g_z^{>d}\|^2 = 2\|f_z^{>d}\|^2 \leq 2^{|J|+1}\|f^{>d}\|^2 = O(\epsilon).$$

Applying Lemma 3.1, either $\|g_z\|^2 = O(\epsilon)$ or $\|g_z\|^2 \geq \|g_z^{\leq d}\|^2 = \Omega(1)$. The lemma now follows from $\|g_z\|^2 = 2\,\mathrm{Var}[f_z]$. □

We rule out the case $\mathrm{Var}[f_z] = \Omega(1)$ using the invariance principle.

Lemma 3.4. *Assume the setting of Lemma 3.2.*

For every $z \in \{0,1\}^J$ we have $\mathrm{Var}[f_z] = O(\epsilon)$.

Proof. According to Lemma 3.3, either $\mathrm{Var}[f_z] = O(\epsilon)$ or $\mathrm{Var}[f_z] = \Omega(1)$. If $\mathrm{Var}[f_z] = O(\epsilon)$ then we are done, so suppose that $\mathrm{Var}[f_z] = \Omega(1)$.

The invariance principle [MOO10] (see also [O'D14, Theorem 11.71]) implies that if g is a degree d function on $\{0,1\}^n$ with variance 1 and all influences at most δ then for every u,

$$\left| \Pr_{x \sim \{0,1\}^n}[g(x) \le u] - \Pr_{w \sim N(0,1)}[g(w) \le u] \right| = O(\delta^{1/(4d+1)}).$$

For every $\gamma > 0$, this implies that

$$\left| \Pr_{x \sim \{0,1\}^n}[g(x) \in (u-\gamma, u]] - \Pr_{w \sim N(0,1)}[g(w) \in (u-\gamma, u]] \right| = O(\delta^{1/(4d+1)}).$$

Since $g(w)$ is a continuous random variable, if we take the limit $\gamma \to 0$ then we obtain

$$\Pr[g = u] = O(\delta^{1/(4d+1)}).$$

Applying this to $g = f_z/\sqrt{\mathrm{Var}[f_z]}$, in which all influences are at most $O(\epsilon)/\Omega(1) = O(\epsilon)$, we deduce that for all t,

$$\Pr[f_z = t] = O(\epsilon^{1/(4d+1)}).$$

On the other hand, since f_z is A-valued, we can find $a \in A$ such that $\Pr[f_z = a] \ge 1/|A|$. Thus $1/|A| = O(\epsilon^{1/(4d+1)})$, and so $\epsilon = \Omega(1/|A|^{4d+1}) = \Omega(1)$.

Finally, $\mathrm{Var}[f_z] \le \|f_z\|^2 \le \max_{a \in A} a^2 = O(1)$. Since $\epsilon = \Omega(1)$, it follows that $\mathrm{Var}[f_z] = O(\epsilon)$. □

At this point we can show that f is close to a J-junta.

Lemma 3.5. *Assume the setting of Lemma 3.2.*

There is a function g on $\{0,1\}^n$, depending only on the coordinates in J, such that $\|f - g\|^2 = O(\epsilon)$.

Proof. Let G be the function on $\{0,1\}^J$ given by $G(z) = \mathbb{E}[f_z]$. We define $g(x) = G(x|_J)$. Then

$$\|f - g\|^2 = \mathop{\mathbb{E}}_{z \in \{0,1\}^J} \|f_z - \mathbb{E}[f_z]\|^2 = \mathop{\mathbb{E}}_{z \in \{0,1\}^J} [\mathrm{Var}[f_z]] = O(\epsilon),$$

using Lemma 3.4. □

We can now state and prove the A-valued Kindler–Safra theorem.

Theorem 3.6. *Let A be a finite set and let $d \ge 1$. If f is an A-valued function on $\{0,1\}^n$ such that $\|f^{>d}\|^2 = \epsilon$ then there exists an A-valued degree d function h in $\{0,1\}^n$ such that* $\Pr[f \ne h] = O(\epsilon)$.

According to Theorem 2.2, the function h depends on $O(1)$ coordinates.

Proof. Lemma 3.5 gives a function g, depending on $O(1)$ coordinates, such that $\|f-g\|^2 = O(\epsilon)$. Let $h(x)$ be obtained by rounding $g(x)$ to the closest element of A. For every x we have $|h(x)-g(x)| \leq |f(x)-g(x)|$ and so $|h(x)-f(x)| \leq |h(x)-g(x)|+|g(x)-f(x)| \leq 2|f(x)-g(x)|$. Consequently, $\|h-f\|^2 \leq 4\|g-f\|^2 = O(\epsilon)$.

Since f and h are both A-valued, for each x either $f(x) = h(x)$ or $|h(x) - f(x)| = \Omega(1)$. Consequently, $\mathbb{E}[(h-f)^2] = \Omega(\Pr[h \neq f])$, and so $\Pr[h \neq f] = O(\|h-f\|^2) = O(\epsilon)$.

Finally, suppose that h does not have degree d. Then $\hat{h}(S)^2 \neq 0$ for some $|S| > d$. Since h depends on $M = O(1)$ coordinates, $\hat{h}(S) = \mathbb{E}[h\chi_S]$ is a non-zero value which is the average of 2^M elements from $A \cup -A$, and consequently $\hat{h}(S)^2 = \Omega(1)$, implying that $\|h^{>d}\|^2 = \Omega(1)$. On the other hand,

$$\|h^{>d}\|^2 \leq 2\|f^{>d}\|^2 + 2\|h^{>d} - f^{>d}\|^2 = 2\epsilon + 2\|(h-f)^{>d}\|^2 \leq 2\epsilon + \|h-f\|^2 = O(\epsilon).$$

This shows that $\epsilon = \Omega(1)$. Therefore $\|f-h\|^2 \leq \max_{a_1,a_2 \in A}(a_1-a_2)^2 = O(\epsilon)$. □

References

[CHS20] John Chiarelli, Pooya Hatami, and Michael Saks. An asymptotically tight bound on the number of relevant variables in a bounded degree Boolean function. *Combinatorica*, 40(2):237–244, 2020.

[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.

[KS02] Guy Kindler and Shmuel Safra. Noise-resistant Boolean functions are juntas. (manuscript), 2002.

[MOO10] Elchanan Mossel, Ryan O'Donnell, and Krzysztof Oleszkiewicz. Noise stability of functions with low influences: invariance and optimality. *Ann. of Math. (2)*, 171(1):295–341, 2010.

[NS94] Noam Nisan and Márió Szegedy. On the degree of Boolean functions as real polynomials. volume 4, pages 301–313. 1994. Special issue on circuit complexity (Barbados, 1992).

[O'D14] Ryan O'Donnell. *Analysis of Boolean functions.* Cambridge University Press, New York, 2014.

[Wel20] Jake Wellens. Relationships between the number of inputs and other complexity measures of Boolean functions. *arXiv*, abs/2005.00566, 2020.