The Keller–Klein Proof of the Kindler–Safra Theorem

Yuval Filmus

July 28, 2021

The Kindler–Safra theorem states, in one formulation, that if a Boolean function on the (unbiased) Boolean cube is close to degree k, then it is close to a Boolean degree k function; the latter are known to be juntas depending on $O(2^k)$ variables, or expressible as decision trees of depth $poly(k)$. In this note we give a proof of this theorem due to Nathan Keller and Ohad Klein [KK20].

We will deduce the Kindler–Safra theorem from a more general version, for integer-valued functions. In this result, $\epsilon_k > 0$ is a constant depending on k, which arises from the proof; in fact, $\epsilon_k = e^{-O(k)}$.

Theorem 1. *Let $f\colon \{\pm 1\}^n \to \mathbb{Z}$ satisfy $\|f^{>k}\|^2 \le \epsilon$, where $\epsilon \le \epsilon_k$. There exists a degree k function $g\colon \{\pm 1\}^n \to \mathbb{Z}$ such that $\|f - g\|^2 \le 2\epsilon$.*

We can easily deduce the Boolean version.

Corollary 2. *Let $f\colon \{\pm 1\}^n \to \{0,1\}$ satisfy $\|f^{>k}\|^2 \le \epsilon$, where $\epsilon \le \epsilon_k$. There exists a degree k function $g\colon \{\pm 1\}^n \to \{0,1\}$ such that $\|f - g\|^2 \le 2\epsilon$.*

Proof. Applying Theorem 1 gives a function g satisfying $\|f-g\|^2 \le 2\epsilon$. This implies that $\Pr[f \neq g] \le 2\epsilon$ and so $\Pr[g \notin \{0,1\}] \le 2\epsilon$. Therefore $h = g^2 - g$ is a degree $2k$ integer-valued function satisfying $\Pr[h \neq 0] \le 2\epsilon$. On the other hand, a simple induction (given below) shows that either $h = 0$ or $\Pr[h \neq 0] \ge 2^{-2k}$. Choosing ϵ_k accordingly completes the proof.

We now prove by induction that if $\deg h \le d$ then either $h = 0$ or $\Pr[h \neq 0] \ge 2^{-d}$. Each step of the induction either decreases d or decreases n. When $d = 0$, h is constant, and so either $h = 0$ or $\Pr[h \neq 0] = 1$. For $d > 0$, let $h = x_n h_1 + h_2$, so that $\deg h_1 \le d - 1$. If $h_1 = 0$ then $h = h_2$ depends on only $n - 1$ variables, and so the claim follows by induction. Otherwise, $\Pr[h_1 \neq 0] \ge 2^{-(d-1)}$, and when this happens, $h \neq 0$ with probability at least $1/2$, since either $h_1 + h_2 \neq 0$ or $h_2 \neq 0$. Therefore $\Pr[h_1 \neq 0] \ge 2^{-(d-1)} \cdot 1/2 = 2^{-d}$. □

We use the following notation: for a set $I \subseteq [n]$,

$$L_I f = \sum_{S \supseteq I} \hat{f}(S) x_S,$$

where $x_S = \prod_{i \in S} x_i$.

Below, we also use the following observation: if $f \in \mathbb{Z}$ then $L_I f \in 2^{-|I|}\mathbb{Z}$. This is because

$$L_I f(x) = 2^{-|I|} \sum_{J \subseteq I} (-1)^{|J|} f(x \oplus J),$$

where $x \oplus J$ is obtained from x by flipping the coordinates in J.

We prove Theorem 1 via the following main lemma.

Lemma 3. *Let $f\colon \{\pm 1\}^n \to \mathbb{Z}$ satisfy $\|f^{>k}\|^2 \le \epsilon$, where $\epsilon \le \epsilon_k$, and let $\ell \le k$. Suppose that for every $I \subseteq [n]$ of size $\ell + 1$ we have $\|L_I f\|^2 \le 2\epsilon$. Then there exists a function $g\colon \{\pm 1\}^n \to \mathbb{Z}$ of degree at most ℓ such that for every $J \subseteq [n]$ of size ℓ we have $\|L_J(f - g)\|^2 \le 2\epsilon$.*

Given this lemma, here is how to prove the theorem.

Proof of Theorem 1. If $|I| = k+1$ then $\|L_I f\|^2 \le \|f^{>k}\|^2 \le \epsilon$. This shows that f itself satisfies the requirements of the lemma when $\ell = k$.

Let now $f_{k+1} = f$, and for $\ell = k, \ldots, 0$, let g_ℓ be the function constructed by Lemma 3 for $f_{\ell+1}$ and the given value of ℓ, and define $f_\ell = f_{\ell+1} - g_\ell$; the lemma guarantees that f_ℓ satisfies the requirements of the lemma for $\ell - 1$.

Let $g = g_k + \cdots + g_0$. Then g is an integer-valued function of degree at most k satisfying $\|f - g\|^2 = \|f_0\|^2 \le 2\epsilon$. □

The main step in the proof of Lemma 3 is the following, in whose proof we determine ϵ_k.

Lemma 4. *Under the conditions of Lemma 3, for every $J \subseteq [n]$ of size ℓ we have* $\mathrm{Var}(L_J f) = O(\epsilon)$.

Given this, we prove Lemma 3 by "coefficient matching". But first, we need a simple consequence of hypercontractivity.

Lemma 5. *Suppose that $f\colon \{\pm 1\}^n \to 2^{-k}\mathbb{Z}$ satisfies $\|f^{>k}\|^2 \le \epsilon$. Either $\|f\|^2 \le 2\epsilon$ or $\|f\|^2 \ge 36^{-k}/4$.*

Proof. We have
$$\|f\|^2 \le \epsilon + \|f^{\le k}\|^2 \le \epsilon + 3^k \|T_{1/\sqrt{3}} f\|^2 \le \epsilon + 3^k \|f\|_{4/3}^2.$$
Now $\|f\|_{4/3}^{4/3} = \mathbb{E}[|f|^{4/3}] \le (2^k)^{2-4/3}\,\mathbb{E}[|f|^2] = (2^k)^{2/3}\|f\|^2$, and so
$$\|f\|^2 \le \epsilon + 6^k \|f\|^3,$$
implying that
$$(1 - 6^k\|f\|)\|f\|^2 \le \epsilon.$$
If $\|f\| \le 6^{-k}/2$ then it follows that $\|f\|^2 \le 2\epsilon$. □

We will also need the following folklore fact.

Lemma 6. *Let $M > 0$. If $a \in \mathbb{Z}/M$ and $b \in \mathbb{R}$ then*
$$|a - \mathrm{round}(b, \mathbb{Z}/M)| \le 2|a - b|.$$

Proof. Let $b = B + \delta$, where $B \in \mathbb{Z}/M$ and $|\delta| \le 1/(2M)$. The left-hand side is at most $|a - b| + |\delta|$. Now $|\delta|$ is the minimum distance between b and $\mathbb{Z}/M$. Since $a \in \mathbb{Z}/M$, this implies that $|\delta| \le |a - b|$, and so the left-hand side is at most $2|a - b|$. □

We can now prove Lemma 3 given Lemma 4.

Proof of Lemma 3. For $|J| = \ell$, let $c_J = \mathrm{round}(\hat{f}(J), 2^{-\ell}\mathbb{Z})$. We claim that c_J is close to $\hat{f}(J)$. Indeed, since $L_J f(x) \in 2^{-\ell}\mathbb{Z}$, we have $|L_J f(x) - c_J x_J| \le 2|L_J f(x) - \hat{f}(J) x_J|$ by Lemma 6, which implies that
$$\|L_J f - c_J x_J\|^2 \le 4\|L_J f - \hat{f}(J) x_J\|^2 = 4\,\mathrm{Var}(L_J f) = O(\epsilon),$$
and so
$$(c_J - \hat{f}(J))^2 = \|c_J x_J - \hat{f}(J) x_J\|^2 \le 2\|L_J f - c_J x_J\|^2 + 2\|L_J f - \hat{f}(J) x_J\|^2 = O(\epsilon).$$

We can now define the function g:
$$g = \sum_{|J|=\ell} 2^\ell c_J \prod_{j \in J} \frac{1 + x_j}{2}.$$
By construction, g is an integer-valued function of degree ℓ satisfying $\hat{g}(J) = c_J$. Therefore
$$\|L_J(f - g)\|^2 \le \mathrm{Var}(L_J f) + (\hat{f}(J) - c_J)^2 = O(\epsilon).$$
Applying Lemma 5, we obtain that in fact $\|L_J(f - g)\|^2 \le 2\epsilon$, for small enough ϵ_k. □

It remains to prove Lemma 4.

Proof of Lemma 4. Without loss of generality, suppose that $J = \{n-\ell+1, \ldots, n\}$. Let $F = L_J f \in 2^{-k}\mathbb{Z}$, which satisfies $\|F^{>k}\|^2 \le \|f^{>k}\|^2 \le \epsilon$. Our goal is to show that $\mathrm{Var}(F) = O(\epsilon)$ under the assumption that $\|L_j F\|^2 \le 2\epsilon$ for all $j \notin J$.

We will use an inductive argument. Let

$$\mathrm{Var}_r(F) = \sum_{S \cap [r] \neq \emptyset} \hat{F}(S)^2,$$

so that $\mathrm{Var}_0(F) = 0$ and $\mathrm{Var}_{[n-\ell]}(F) = \mathrm{Var}(F)$. We will show that $\mathrm{Var}_r(F) \le K\epsilon$ by induction on r, for an appropriate constant $K > 0$.

For $y \in \{\pm 1\}^{n-r}$, let $F_y(x) = F(x,y)$ be obtained from F by fixing the final $n-r$ coordinates. Then

$$\mathop{\mathbb{E}}_y[\mathrm{Var}(F_y)] = \mathop{\mathbb{E}}_{x,y}[(F(x,y) - \mathop{\mathbb{E}}_x[F(x,y)])^2] = \mathrm{Var}_r(F),$$

since the Fourier expansion of $\mathbb{E}_x[F(x,y)]$ is obtained from that of F by removing all Fourier coefficients intersecting $[r]$. Accordingly, from now on we concentrate on bounding $\mathrm{Var}(F_y)$.

The basic idea is to use Lemma 5. To this end, we round $\mathbb{E}_x[F(x,y)]$ to $2^{-k}\mathbb{Z}$, and note that

$$\mathrm{Var}(F_y) \le \mathbb{E}[(F_y - \mathrm{round}(\mathbb{E}[F_y], 2^{-k}\mathbb{Z}))^2] \le 4\,\mathrm{Var}(F_y).$$

The first inequality holds since $\mathbb{E}[(F_y - a)^2]$ is minimized at $a = \mathbb{E}[F_y]$, and the second inequality holds by Lemma 6, since $F_y \in 2^{-k}\mathbb{Z}$.

The function $F_y - \mathrm{round}(\mathbb{E}[F_y], 2^{-k}\mathbb{Z})$ has degree k and takes values in $2^{-k}\mathbb{Z}$, and so according to Lemma 5,

$$\mathbb{E}[(F_y - \mathrm{round}(\mathbb{E}[F_y], 2^{-k}\mathbb{Z}))^2] \le 2\|F_y^{>k}\|^2 \text{ or } \mathbb{E}[(F_y - \mathrm{round}(\mathbb{E}[F_y], 2^{-k}\mathbb{Z}))^2] \ge 36^{-k}/4,$$

In terms of $\mathrm{Var}(F_y)$,

$$\mathrm{Var}(F_y) \le 2\|F_y^{>k}\|^2 \text{ or } \mathrm{Var}(F_y) \ge 36^{-k}/16.$$

Accordingly, letting $\delta_k = 36^{-k}/16$,

$$\mathrm{Var}_r(F) \le 2 \mathop{\mathbb{E}}_y[\|F_y^{>k}\|^2] + \mathop{\mathbb{E}}_y[\mathrm{Var}(F_y) \cdot [\![\mathrm{Var}(F_y) \ge \delta_k]\!]], \tag{1}$$

where $[\![E]\!]$ is the indicator variable corresponding to the event E.

The first term in this inequality is easy to bound. Recall that $F_y^{>k}$ is obtained from F by fixing the last $n-r$ coordinates, and then retaining only monomials of degree k of larger. Since fixing coordinates cannot increase the degree of a monomial, we can obtain $F_y^{>k}$ by fixing the last $n-r$ coordinates of $F^{>k}$ and then removing monomials of degree smaller than k. Undoing the very last bit can only increase the norm, and so

$$\mathop{\mathbb{E}}_y[\|F_y^{>k}\|^2] \le \|F^{>k}\|^2 \le \epsilon. \tag{2}$$

It remains to bound the second term in (1). For starters, let $G(x,y) = F_y(x) - \mathbb{E}_x[F_y]$, so that $\mathrm{Var}_r(F) = \|G\|^2$. Let $G_y = F_y - \mathbb{E}[F_y]$ be its restriction to a specific setting of y. We decompose $G_y = G^{\le k}|_y + G^{>k}|_y$, where $G^{\le k}|_y(x) = G^{\le k}(x,y)$ and similarly $G^{>k}|_y(x) = G^{>k}(x,y)$. This decomposition is not the same as the decomposition $G_y = G_y^{\le k} + G_y^{>k}$, which is obtained by first substituting y and then separating according to levels; instead, in this part of the proof we first separate according to levels and then substitute y.

Since $\mathrm{Var}(F_y) = \|G_y\|^2$ and $G_y = G^{\le k}|_y + G^{>k}|_y$, we can bound

$$\mathrm{Var}(F_y) \le 2\|G^{\le k}|_y\|^2 + 2\|G^{>k}|_y\|^2.$$

If $\|G^{\leq k}|_y\|^2 \geq \|G^{>k}|_y\|^2$ then $\|G^{\leq k}|_y\|^2 \geq \operatorname{Var}(F_y)/4$, and so assuming that $\operatorname{Var}(F_y) \neq 0$, we can bound $\|G^{\leq k}|_y\|^2 \leq \frac{4}{\operatorname{Var}(F_y)}\|G^{\leq k}|_y\|^4$. Considering also the case $\|G^{\leq k}|_y\|^2 \leq \|G^{>k}|_y\|^2$, this shows that

$$\operatorname{Var}(F_y) \leq \frac{8}{\operatorname{Var}(F_y)}\|G^{\leq k}|_y\|^4 + 4\|G^{>k}|_y\|^2.$$

This shows that

$$\mathop{\mathbb{E}}_y[\operatorname{Var}(F_y) \cdot [\![\operatorname{Var}(F_y) \geq \delta_k]\!]] \leq \frac{8}{\delta_k} \mathop{\mathbb{E}}_y[\|G^{\leq k}|_y\|^4] + 4 \mathop{\mathbb{E}}_y[\|G^{>k}|_y\|^2]. \tag{3}$$

The second term is at most $4\|G^{>k}\|^2 \leq 4\|F^{>k}\|^2 \leq 4\epsilon$, since G is obtained from F by deleting some of the Fourier coefficients. As for the first term, convexity of t^2 shows that

$$\mathop{\mathbb{E}}_y[\|G^{\leq k}|_y\|^4] = \mathop{\mathbb{E}}_y\big[\mathop{\mathbb{E}}_x[(G^{\leq k})^2]^2\big] \leq \mathop{\mathbb{E}}_{x,y}[(G^{\leq k})^4].$$

Rewriting the right-hand side as an L_4 norm and applying hypercontractivity, we bound

$$\mathop{\mathbb{E}}_y[\|G^{\leq k}|_y\|^4] \leq \|G^{\leq k}\|_4^4 \leq 9^k\|G\|^4 = 9^k \operatorname{Var}_r(F)^2.$$

(At this point it is crucial that we decomposed $G_y = G^{\leq k}|_y + G^{>k}|_y$ rather than $G_y = G_y^{\leq k} + G_y^{>k}$. The latter decomposition would result in a bound of $\mathbb{E}_y[\|G_y\|^2]$, which is potentially much larger.) Substituting this in (3) and combining with (1) and (2), we deduce

$$\operatorname{Var}_r(F) \leq e^{O(k)} \operatorname{Var}_r(F)^2 + O(\epsilon). \tag{4}$$

In order to complete the proof, it remains to show that $\operatorname{Var}_r(F)$ is somewhat small:

$$\operatorname{Var}_r(F) = \sum_{S\cap[r]\neq\emptyset} \hat{F}(S)^2 \leq \sum_{S\cap[r-1]\neq\emptyset} \hat{F}(S)^2 + \sum_{r\in S}\hat{F}(S)^2 = \operatorname{Var}_{r-1}(F) + \|L_rF\|^2 \leq (K+2)\epsilon,$$

using the induction hypothesis and the assumption of the lemma. Substituting this in (4), for $K \geq 1$ we get

$$\operatorname{Var}_r(F) \leq e^{Ck}K^2\epsilon^2 + D\epsilon,$$

for some constants $C, D > 0$. Choosing $K = D + 1$ and $\epsilon_k = e^{-Ck}/K^2$ completes the proof. □

References

[KK20] Nathan Keller and Ohad Klein. A structure theorem for almost low-degree functions on the slice. *Isr. J. Math.*, 2020.