Friedgut–Kalai–Naor theorem for slices of the Boolean cube

Yuval Filmus*

Institute for Advanced Study, Princeton, NJ

October 15, 2014

Abstract

The Friedgut–Kalai–Naor theorem states that if a Boolean function $f\colon \{0,1\}^n \to \{0,1\}$ is close (in L^2-distance) to an affine function $\ell(x_1,\ldots,x_n) = c_0 + \sum_i c_i x_i$, then f is close to a *Boolean* affine function (which necessarily depends on at most one coordinate). We prove a similar theorem for functions defined over $\binom{[n]}{k} = \{(x_1,\ldots,x_n) \in \{0,1\}^n : \sum_i x_i = k\}$.

1 Introduction

Suppose $f\colon \{0,1\}^n \to \{0,1\}$ is a Boolean function which is an affine combination of its inputs:

$$f(x_1,\ldots,x_n) = c_0 + \sum_{i=1}^n c_i x_i.$$

An easy calculation shows that f must be one of the following functions: one of $0,1$ or one of $x_i, 1-x_i$ for some $1 \leqslant i \leqslant n$, and in particular f is either constant or depends on exactly one coordinate (input variable). The seminal Friedgut–Kalai–Naor (FKN) theorem [3] is a stability version of this result: if $f\colon \{0,1\}^n \to \{0,1\}$ is ϵ-close (in L^2-distance) to an affine function, then f is $O(\epsilon)$-close to one of the functions $0,1,x_i,1-x_i$.

The FKN theorem has been extended in many directions. Kindler and Safra [5] proved a similar theorem for almost low-degree functions, the original theorem corresponding to the almost linear case. Jendrej, Oleszkiewicz and Wojtaszczyk [4] and Rubinstein and Safra [6] have extended the FKN theorem to Boolean functions which are close to a sum of general independent random variables. Ellis, Filmus and Friedgut [1, 2] extended the FKN theorem to functions on the symmetric group.

In this article we prove an FKN theorem for functions defined on a slice of the Boolean cube,

$$\binom{[n]}{k} = \{(x_1,\ldots,x_n) \in \{0,1\}^n : \sum_{i=1}^n x_i = k\}.$$

*Research conducted at the Simons Institute for the Theory of Computing during the 2013 fall semester on Real Analysis in Computer Science. This material is based upon work supported by the National Science Foundation under agreement No. DMS-1128155. Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors, and do not necessarily reflect the views of the National Science Foundation.

Our main theorem states that for $2 \leqslant k \leqslant n-2$, if a function $f\colon \binom{[n]}{k} \to \{0,1\}$ is ϵ-close to an affine function for $\epsilon = O(p^2)$, then f is $O(\epsilon)$-close to one of the functions $0, 1, x_i, 1-x_i$. For larger ϵ, we prove that either f or $1-f$ is $O(\epsilon)$-close to $\max_{i \in S} x_i$ for some $|S| = O(\sqrt{\epsilon}/p)$, assuming $2n^{3/4} \leqslant k \leqslant n - 2n^{3/4}$.

Acknowledgements I thank Guy Kindler, Elchanan Mossel and Karl Wimmer for helpful discussions, and Manh Tuan Tran for pointing out a mistake in an earlier version.

2 Preliminaries

The following simple fact will be useful.

Fact 1. *If $x \in \{0,1\}$, y is arbitrary, and Y results from rounding y to $\{0,1\}$, then $|x-Y| \leqslant 2|x-y|$.*

Throughout this article, a *Boolean function* is a function whose range is $\{0,1\}$. An *affine function* is a function of the form

$$\ell(x_1, \ldots, x_n) = c_0 + \sum_{i=1}^{n} c_i x_i.$$

Every function on the Boolean cube $\{0,1\}^n$ has a (unique) Fourier expansion

$$f(x_1, \ldots, x_n) = \sum_{S \subseteq [n]} \hat{f}(S) \chi_S,$$

where $[n] = \{1, \ldots, n\}$ and $\chi_S = \prod_{i \in S} (-1)^{x_i}$. The L^2-distance between two functions f, g on the Boolean cube is defined with respect to the uniform measure:

$$\|f - g\|^2 = \mathop{\mathbb{E}}_{x \in \{0,1\}^n} (f(x) - g(x))^2.$$

If $\|f-g\|^2 \leqslant \epsilon$ then we say that f, g are *ϵ-close*.

We will use the FKN theorem in the following form.

Theorem 2.1 (Friedgut–Kalai–Naor [3])**.** *Suppose $f\colon \{0,1\}^n \to \{0,1\}$ is ϵ-close to an affine function. Then for some function $g \in \{0, 1, x_1, 1-x_i, \ldots, x_n, 1-x_n\}$, we have $\|f-g\|^2 = O(\epsilon)$.*

See for example Wojtaszczyk [7, Theorem 3.1]. Using the notation $\operatorname{dist}(x, S) = \min_{y \in S} |x-y|$ and the fact that the affine function closest to f is $\hat{f}(\varnothing) + \sum_{i=1}^n \hat{f}(\{i\}) x_i$, an easy corollary is:

Corollary 2.2. *Suppose $f\colon \{0,1\}^n \to \{0,1\}$ is ϵ-close to an affine function $\ell\colon \{0,1\}^n \to \mathbb{R}$. Then*

$$\sum_{i=1}^{n} \operatorname{dist}(2\hat{\ell}(\{i\}), \{0, \pm 1\})^2 = O(\epsilon).$$

Proof. Theorem 2.1 shows that f is $O(\epsilon)$-close to some function $g \in \{0, 1, x_i, 1-x_i\}$. The Fourier expansions of these functions are, respectively, $0, 1, (1-x_i)/2, (1+x_i)/2$, and we conclude that $\hat{g}(\{i\}) \in \{0, \pm 1/2\}$ for all $i \in [n]$. On the one hand, $\|g-\ell\|^2 \leqslant 2\|g-f\|^2 + 2\|f-\ell\|^2 = O(\epsilon)$. On the other hand,

$$\|2(g-\ell)\|^2 \geqslant \sum_{i=1}^{n} (2\hat{g}(\{i\}) - 2\hat{\ell}(\{i\}))^2 \geqslant \sum_{i=1}^{n} \operatorname{dist}(2\hat{\ell}(\{i\}), \{0, \pm 1\})^2.$$

The corollary follows by combining both bounds. □

Slices of the Boolean cube For integers $n \geqslant 2$ and $1 \leqslant k \leqslant n-1$, define the *slice* $\binom{[n]}{k}$ by

$$\binom{[n]}{k} = \left\{(x_1, \dots, x_n) \in \{0,1\}^n : \sum_{i=1}^n x_i = k.\right\}$$

Every affine function on the slice has a unique representation of the form

$$\ell(x_1, \dots, x_n) = \sum_{i=1}^n c_i x_i.$$

The L^2-distance between two functions f, g on the slice $\binom{[n]}{k}$ is defined with respect to the uniform measure on the slice.

We associate with each vector $(x_1, \dots, x_n) \in \binom{[n]}{k}$ a subset $S \subseteq [n]$ of size k as follows: $S = \{i \in [n] : x_i = 1\}$. This forms a bijection between $\binom{[n]}{k}$ and the set of subsets of $[n]$ of size k. We use this interpretation of $\binom{[n]}{k}$ freely in the sequel.

3 Main theorem

In this section we state and prove our main theorem, a version of Theorem 2.1 for functions on a slice.

Theorem 3.1. *Suppose $f\colon \binom{[n]}{k} \to \{0,1\}$ is ϵ-close to an affine function, where $\min(k, n-k) \geqslant 2n^{3/4}$. Define $p \triangleq \min(k, n-k)/n$. Then either f or $1-f$ is $O(\epsilon)$-close to $\max_{i \in S} x_i$ for some set S of size at most $\max(1, O(\sqrt{\epsilon}/p))$.*

Furthermore, when $\epsilon < p/128$, it is enough to assume that $\min(k, n-k) \geqslant 2$.

Note that for some constant C, if $\epsilon < Cp^2$ then we are guaranteed that $|S| \leqslant 1$, and so f can be approximated by a function of one of the forms $0, 1, x_i, 1 - x_i$.

The theorem is tight in two senses. First, the condition $\min(k, n-k) \geqslant 2$ is necessary since when $k = 1$, every Boolean function is affine (and can be trivially written in the form $\max_{i \in S} x_i$ for some set S). The bound on the size of S is also optimal: for $ps^2 = o(1)$, the function $\max(x_1, \dots, x_s)$ is $O((ps)^2)$-close to the linear function $x_1 + \cdots + x_s$, but cannot be approximated using a smaller maximum without incuring an error of $\Omega(p) = \omega((ps)^2)$. On the other hand, our arguments show that the condition $\min(k, n-k) \geqslant 2n^{3/4}$ can be relaxed up to $n^{2/3+\delta}$, and the theorem might remain true even beyond that.

Amusingly, we can derive Theorem 2.1 from Theorem 3.1. Suppose we are given a Boolean function $f\colon \{0,1\}^n \to \{0,1\}$ which is ϵ-close to an affine function ℓ, for small enough ϵ. Extend the function f to a function f_N on $2N \geqslant n$ coordinates, and consider its restriction g_N to $\binom{[2N]}{N}$. Both functions f_N, g_N depend only on the first n coordinates. For a random set $S \in \binom{[2N]}{N}$ and large N, the distribution of $S \cap [n]$ is very close to uniform, and so for large enough N, g_N is 2ϵ-close to ℓ. Theorem 3.1 implies that g_N is $O(\epsilon)$-close to one of the functions $0, 1, x_i, 1 - x_i$, and we conclude that f_N and f are also close to this function.

As a warm-up, we prove that the only Boolean affine functions on the slice are $0, 1, x_i, 1 - x_i$.

Lemma 3.2. *Suppose $f\colon \binom{[n]}{k} \to \{0,1\}$ is affine, where $2 \leqslant k \leqslant n-2$. Then $f \in \{0,1\}$ or $f \in \{x_i, 1 - x_i\}$ for some i.*

Proof. Let $f = \sum_{i=1}^n c_i x_i$. Without loss of generality, suppose that $c_1 = \min(c_1, \ldots, c_n)$. For any $i \neq 1$, let $S \in \binom{[n]}{k}$ be some set containing 1 but not i. Since $f(S \triangle \{1, i\}) - f(S) = c_i - c_1$ and f is Boolean, we conclude that $c_i \in \{c_1, c_1 + 1\}$. If for all $i \neq 1$ we have $c_i = c_1$, then $f \in \{0, 1\}$. So we can assume that $I_0 = \{i \in [n] : c_i = c_1\}$ and $I_1 = \{i \in [n] : c_i = c_1 + 1\}$ are both non-empty.

We claim that either $|I_0| = 1$ or $|I_1| = 1$. Otherwise, suppose without loss of generality that $1, 2 \in I_0$ and $3, 4 \in I_1$. Let $S \in \binom{[n]}{k}$ be some set containing both $1, 2$ but neither of $3, 4$. Then $f(S \triangle \{1, 2, 3, 4\}) - f(S) = c_3 + c_4 - c_1 - c_2 = 2$, contradicting the fact that f is Boolean. This shows that either $|I_0| = 1$ or $|I_1| = 1$. If $I_0 = \{1\}$ then

$$f = c_1 x_1 + \sum_{i=2}^n (c_1 + 1) x_i = (c_1 + 1) \sum_{i=1}^n x_i - x_1 = (c_1 + 1)k - x_1,$$

and since f is Boolean, necessarily $f = 1 - x_1$. Similarly, if $I_1 = \{i\}$ then we get $f = x_i$. □

We proceed with the proof of Theorem 3.1. Since every function on $\binom{[n]}{k}$ is equivalent to a function on $\binom{[n]}{n-k}$, and the equivalence preserves affine functions, it suffices to consider the case $k \leqslant n/2$.

For the rest of this section, we make the assumption that $2 \leqslant k \leqslant n/2$ and fix the following notation:

- $p = k/n \leqslant 1/2$.
- $f \colon \binom{[n]}{k} \to \{0, 1\}$ is a Boolean function.
- $\ell = \sum_{i=1}^n c_i x_i$ is an affine function satisfying $\|f - \ell\|^2 \leqslant \epsilon$.

We first explain the proof of Theorem 3.1 in the slightly easier case $\epsilon < p/128$. Extending Lemma 3.2, we show that the coefficients $c_1, \ldots, c_n$ are close to two values $x, x+1$, say most of them close to x. We define $d_i = c_i$ or $d_i = c_i - 1$ in such a way that $d_1, \ldots, d_n$ are all close to x, and let $r = \sum_i d_i x_i$. Note that $h = \ell - r$ is of the form $\sum_{i \in S} x_i$. Applying the classical Friedgut–Kalai–Naor theorem to a random sub-cube of $\binom{[n]}{k}$ (the second equality below), we deduce that

$$k \mathop{\mathbb{E}}_{i \neq j} (d_i - d_j)^2 = k \mathop{\mathbb{E}}_{i \neq j} \operatorname{dist}(c_i - c_j, \{0, \pm 1\})^2 = O(\epsilon).$$

A simple calculation shows that $\mathbb{V} r \leqslant k \mathbb{E}_{i \neq j} (d_i - d_j)^2 = O(\epsilon)$, and so, putting $m = \mathbb{E}\, r$, we get that

$$\|f - (h + m)\|^2 \leqslant 2\|f - \ell\|^2 + 2\|r - m\|^2 = O(\epsilon).$$

This means that f is close to the function H obtained from rounding $h + m$ to $\{0, 1\}$. The proof is complete by showing that the only way a function of the form $h + m$ is close to a Boolean function is when $H = \max_{i \in S} x_i$ or $H = 1$.

When $\epsilon \geqslant p/128$ we cannot deduce that $d_i - d_j = \operatorname{dist}(c_i - c_j, \{0, \pm 1\})^2$. Instead, we start by showing that all but an $O(\epsilon)$ fraction of the coefficients $c_1, \ldots, c_n$ are concentrated on *three* values $x - 1, x, x + 1$. This allows us to approximate f by a function of the form $\sum_{i \in S_+} x_i - \sum_{i \in S_-} x_i + m$. Further arguments show that throwing one of the first two summands results in a loss of only $O(\epsilon)$, and the proof is completed as before.

We start by showing that for small ϵ/p, the coefficients $c_1, \ldots, c_n$ are all concentrated around two values $x, x + 1$.

Lemma 3.3. *There exist c, d satisfying $|c-d| = 1$ and a subset $S \subseteq [n]$ of size $|S| \leqslant n/2$ such that for all $j \in S$, $|c_j - c|^2 \leqslant 8\epsilon/p$, and for all $j \notin S$, $|c_j - d|^2 \leqslant 8\epsilon/p$.*

Proof. Suppose, without loss of generality, that $c_1 = \min(c_1, \ldots, c_n)$, and fix $i \neq 1$. A random $T \in \binom{[n]}{k}$ contains 1 but not i with probability $\frac{k(n-k)}{n(n-1)} \geqslant p(1-p) \geqslant p/2$. For any such set T, define

$$\begin{aligned}\Delta &= \operatorname{dist}(\ell(T), \{0,1\})^2 + \operatorname{dist}(\ell(T \triangle \{1,i\}), \{0,1\})^2 \\ &= \operatorname{dist}(\ell(T), \{0,1\})^2 + \operatorname{dist}(\ell(T) + c_i - c_1, \{0,1\})^2.\end{aligned}$$

Let $r = \ell(T)$ and $\delta = c_i - c_1 \geqslant 0$. Suppose that $\ell(T)$ is closer to $a \in \{0,1\}$ and $\ell(T \triangle \{1,i\})$ is closer to $b \in \{0,1\}$, where $a \leqslant b$. Then

$$\Delta = (r-a)^2 + (r+\delta-b)^2 = 2r^2 + 2(\delta - b - a)r + a^2 + (\delta - b)^2.$$

Minimizing the quadratic (the minimum of $\alpha r^2 + \beta r + \gamma$ for $\alpha > 0$ is $\frac{\alpha\gamma - (\beta/2)^2}{\alpha}$), we deduce that

$$\Delta \geqslant \frac{2a^2 + 2(\delta-b)^2 - (\delta-b-a)^2}{2} = \frac{(\delta - b + a)^2}{2}.$$

Since $b - a \in \{0,1\}$, we deduce that

$$\operatorname{dist}(\ell(T), \{0,1\})^2 + \operatorname{dist}(\ell(T \triangle \{1,i\}), \{0,1\})^2 \geqslant \frac{1}{2} \operatorname{dist}(c_i - c_1, \{0,1\})^2.$$

Taking expectation over random T, we conclude that

$$\frac{1}{2} \operatorname{dist}(c_i - c_1, \{0,1\})^2 \Pr[1 \in T, i \notin T] \leqslant 2\epsilon,$$

and so $\operatorname{dist}(c_i - c_1, \{0,1\})^2 \leqslant 8\epsilon/p$.

For $x \in \{0,1\}$, let I_x be the set of indices i such that $c_i - c_1$ is closer to x. The lemma now follows by either taking $S = I_0$, $c = c_1$ and $d = c_1 + 1$ or $S = I_1$, $c = c_1 + 1$ and $d = c_1$. ☐

Unconditionally, the values c_i are on average either close to one another or at distance roughly 1. We show this by taking a random sub-cube of dimension k, and applying the classical Friedgut–Kalai–Naor theorem to the restrictions of f and ℓ to the sub-cube.

Lemma 3.4. *We have*

$$k \mathop{\mathbb{E}}_{i \neq j} \operatorname{dist}(c_i - c_j, \{0, \pm 1\})^2 = O(\epsilon).$$

Proof. Let $a_1, b_1, \ldots, a_k, b_k$ be $2k$ distinct random indices taken from $[n]$, and define

$$D = \{a_1, b_1\} \times \cdots \times \{a_k, b_k\} \subseteq \binom{[n]}{k}.$$

Clearly

$$\mathop{\mathbb{E}}_{D} \|f|_D - \ell|_D\|^2 = \|f - \ell\|^2 \leqslant \epsilon.$$

Using the mapping $\{a_1, b_1\} \times \cdots \times \{a_k, b_k\} \approx \{0,1\} \times \cdots \times \{0,1\} = \{0,1\}^k$, we can think of D as a k-dimensional Boolean cube. Under this encoding, $f|_D$ is a Boolean function $\{0,1\}^k \to \{0,1\}$, and

$$\ell|_D(y_1, \ldots, y_k) = \sum_{i=1}^{n} c_{a_i} + \sum_{i=1}^{n} (c_{b_i} - c_{a_i}) y_i.$$

Since $y_i = (1-(-1)^{y_i})/2$, we see that $2\widehat{\ell|_D}(\{i\}) = c_{a_i} - c_{b_i}$. Corollary 2.2 therefore shows that

$$\sum_{i=1}^{k} \operatorname{dist}(c_{b_i} - c_{a_i}, \{0, \pm 1\})^2 = O(\|f|_D - \ell|_D\|^2).$$

The lemma now follows by taking the expectation over the choice of D. □

An estimate of the type given by Lemma 3.4 is useful since it can potentially bound the variance of ℓ, as the following lemma shows.

Lemma 3.5. *For $r = \sum_i d_i x_i$ we have*

$$\mathbb{V} r \leqslant k \mathop{\mathbb{E}}_{i \neq j} (d_i - d_j)^2.$$

Proof. By shifting all coefficients d_i, we can assume without loss of generality that $\mathbb{E} r = 0$ and so $\sum_{i=1}^n d_i = 0$ (this does not affect the quantities $d_i - d_j$). For every $i \neq j$ we have $\mathbb{E} x_i = \mathbb{E} x_i^2 = p$ and $\mathbb{E} x_i x_j = \frac{k-1}{n-1} p$. Therefore

$$\begin{aligned}\mathbb{E} r^2 &= p \sum_i d_i^2 + \frac{k-1}{n-1} p \sum_i d_i \sum_{j \neq i} d_j \\ &= p \left(1 - \frac{k-1}{n-1}\right) \sum_i d_i^2 \leqslant p \sum_i d_i^2.\end{aligned}$$

On the other hand,

$$\begin{aligned}\mathop{\mathbb{E}}_{i \neq j} (d_i - d_j)^2 &= \frac{2}{n} \sum_i d_i^2 - \frac{2}{n(n-1)} \sum_i d_i \sum_{j \neq i} d_j \\ &= \left(1 - \frac{1}{n-1}\right) \frac{2}{n} \sum_i d_i^2 \geqslant \frac{1}{n} \sum_i d_i^2.\end{aligned}$$

This completes the proof. □

As a corollary, we obtain the following criterion for approximating f by an affine function.

Corollary 3.6. *Suppose $d_1, \ldots, d_n$ are coefficients satisfying $k \mathbb{E}_{i \neq j} (d_i - d_j)^2 = O(\epsilon)$, and define $g = \sum_i (c_i - d_i) x_i$. Then for some m, $\|f - (g + m)\|^2 = O(\epsilon)$.*

Proof. Define $r = \ell - g = \sum_i d_i x_i$. Lemma 3.5 shows that $\mathbb{V} r = O(\epsilon)$, and so

$$\|f - (g + \mathbb{E} r)\|^2 \leqslant 2\|f - \ell\|^2 + 2\|r - \mathbb{E} r\|^2 = O(\epsilon).$$ □

As an application of the corollary, we show that when $\epsilon < p/128$, the function f can be approximated by a function of the form $\pm \sum_{i \in S} x_i + m$. The condition $\epsilon < p/128$ ensures that the estimate of Lemma 3.3 is strong enough to deduce $|d_i - d_j| = \operatorname{dist}(c_i - c_j, \{0, \pm 1\})$ for appropriate d_i chosen according to the lemma.

Lemma 3.7. *If $\epsilon < p/128$ then there exist $\delta \in \{\pm 1\}$, real m, and a subset $S \subseteq [n]$ of size at most $n/2$, such that $\|f - (\delta \sum_{i \in S} x_i + m)\|^2 = O(\epsilon)$.*

Proof. Lemma 3.3 shows that for some c, d satisfying $|c-d| = 1$ there exists a subset $S \subseteq [n]$ of size at most $n/2$ such that for $i \in S$, $|c_i - c|^2 \leqslant 8\epsilon/p$, and for $i \notin S$, $|c_i - d|^2 < 8\epsilon/p$. Let $\delta = d - c \in \{\pm 1\}$, and define $r = \ell + \delta \sum_{i \in S} x_i$. Note that $r = \sum_i d_i x_i$, where $d_i = c_i + \delta$ for $i \in S$ and $d_i = c_i$ for $i \notin S$. In both cases, $|d_i - d| \leqslant \sqrt{8\epsilon/p} < 1/4$, and so $|d_i - d_j| < 1/2$ for all i, j. Since $d_i - d_j = c_i - c_j + \kappa$ for some $\kappa \in \{0, \pm 1\}$, this shows that $\operatorname{dist}(c_i - c_j, \{0, \pm 1\}) = |d_i - d_j|$, and so Lemma 3.4 implies that

$$k \mathop{\mathbb{E}}_{i \neq j} (d_i - d_j)^2 = O(\epsilon),$$

The lemma now follows from Corollary 3.6. □

The next step is to determine when a function of the form $\pm \sum_{i \in S} x_i + m$ can be close to Boolean. The idea is to analyze the hypergeometric random variable $\sum_{i \in S} x_i$.

Lemma 3.8. *There exists a constant $\gamma_0 > 0$ such that the following holds. Consider the random variable $X = \sum_{i \in [t]} x_i$. If $t \leqslant n/2$ and $\Pr[X \in \{m, m+1\}] \geqslant 1 - \gamma_0$ for some m then $\Pr[X = 0] = \Omega(1)$ and $t \leqslant (3/2)p^{-1}$.*

Proof. The distribution of X is given by

$$\Pr[X = s] = \frac{\binom{t}{s}\binom{n-t}{k-s}}{\binom{n}{k}}.$$

If $t \leqslant 3$ then $\Pr[X = 0] = \Omega_t((1-p)^t) = \Omega_t(1)$ and $t \leqslant (3/2)2 \leqslant (3/2)p^{-1}$. Similarly, if $k \leqslant 3$ then $\Pr[X = 0] = \Omega_k((1 - t/n)^k) = \Omega_k(1)$ and $t \leqslant n/2 \leqslant (3/2)(n/3) \leqslant (3/2)p^{-1}$. We can therefore assume that $t, k \geqslant 4$.

In view of showing that the mode of X, given by the classical formula $s_0 = \lfloor \frac{(k+1)(t+1)}{n+2} \rfloor$, is attained at zero, assume that $s_0 \geqslant 1$. Note that $s_0 + 2 \leqslant k$ since otherwise $s_0 \geqslant k - 1$ and so $(k+1)(t+1) \geqslant (k-1)(n+2)$, implying $t + 1 \geqslant (3/5)(n+2) \geqslant (6/5)(t+1)$, which is impossible. Similarly, $s_0 + 2 \leqslant t$.

A simple calculation shows that $\rho_s \triangleq \frac{\Pr[X=s+1]}{\Pr[X=s]} = \frac{(t-s)(k-s)}{(s+1)(n-t-k+s+1)}$. Therefore

$$\begin{aligned} \frac{\rho_{s+1}}{\rho_s} &= \frac{t-s-1}{t-s} \frac{k-s-1}{k-s} \frac{s+1}{s+2} \frac{n-t-k+s+1}{n-t-k+s+2} \\ &= \left(1 - \frac{1}{t-s}\right)\left(1 - \frac{1}{k-s}\right)\left(1 - \frac{1}{s+2}\right)\left(1 - \frac{1}{n-t-k+s+2}\right). \end{aligned}$$

Since $t - s_0 \geqslant 2$, $k - s_0 \geqslant 2$, $s_0 + 2 \geqslant 2$ and $n - t - k + s_0 + 2 \geqslant 2$ (using $t, k \leqslant n/2$), we conclude that $\rho_{s_0+1}/\rho_{s_0} \geqslant 1/16$. This shows that $\Pr[X = s_0 + 2] = \rho_{s_0+1}\rho_{s_0} \Pr[X = s_0] \geqslant \frac{\rho_{s_0}^2}{16} \Pr[X = s_0]$.

If $\Pr[X = s_0] \leqslant 1/3$ then $\Pr[X \in \{m, m+1\}] \leqslant 2/3$ and so $\gamma \geqslant 1/3$. We can therefore assume that $\Pr[X = s_0] \geqslant 1/3$, and so $\Pr[X = s_0 + 2] \geqslant (\rho_{s_0}^2/16)(1/3) = \Omega(\rho_{s_0}^2)$. Since $\{m, m+1\}$ cannot contain both s_0 and $s_0 + 2$, this shows that $\gamma = \Omega(\rho_{s_0}^2)$, implying that we can assume that $\rho_{s_0} < \tau_0$ for some small τ_0.

Suppose now that $s_0 \geqslant 1$. Then $s_0 > \frac{(k+1)(t+1)}{n+2} - 1$, and so

$$\begin{aligned} \rho_{s_0} &\geqslant \frac{(t - \frac{(k+1)(t+1)}{n+2})(k - \frac{(k+1)(t+1)}{n+2})}{\frac{(k+1)(t+1)}{n+2}(n - t - k + \frac{(k+1)(t+1)}{n+2})} \\ &\geqslant \Omega(1) \frac{(t+1)(1 - \frac{k+1}{n+2})(k+1)(1 - \frac{t+1}{n+2})}{\frac{(k+1)(t+1)}{n+2} \cdot O(n)} = \Omega(1). \end{aligned}$$

By choosing τ_0 (and so γ_0) appropriately, we can conclude that $s_0 = 0$, which shows that $1 > \frac{(k+1)(t+1)}{n+2} > pt$, and so $t < p^{-1}$. Moreover, $\Pr[X = 0] = \Pr[X = s_0] \geqslant 1/3$. □

We can now determine which functions of the form $\pm \sum_{i\in S} x_i + m$ are close to Boolean. It is enough to consider the case $\sum_{i\in S} x_i + m$, the other case following by considering the function $1 - f$ instead.

Lemma 3.9. *There exists a constant $\gamma_1 > 0$ such that the following holds. If $\gamma \triangleq \|f - (\sum_{i\in S} x_i + m)\|^2 \leqslant \gamma_1$, where $|S| \leqslant n/2$, then for some $\mu \in \{0, 1\}$, $\|f - (\sum_{i\in S} x_i + \mu)\|^2 = O(\gamma)$. Furthermore, $|S| \leqslant (3/2)p^{-1}$ and $|m - \mu| = O(\sqrt{\gamma})$.*

Proof. Let $X = \sum_{i\in S} x_i$, and define μ to be the integer closest to m. Since $\Pr[X \notin \{-\mu, 1 - \mu\}] \leqslant \|f - (\sum_{i\in S} x_i + m)\|^2 \leqslant \gamma_0$, Lemma 3.8 shows that $\Pr[X = 0] = \Omega(1)$ and $|S| \leqslant (3/2)p^{-1}$. This implies that $\mu \in \{0, 1\}$, since otherwise $\gamma = \Omega(\Pr[X = 0]) = \Omega(1)$. Furthermore, $\gamma = \Omega(|m - \mu|^2)$ and so $|m - \mu| = O(\sqrt{\gamma})$. For any non-zero integer z, we have $|z - \mu| \leqslant |z - m| + |m - \mu| \leqslant 2|z - m|$, since μ is the integer closest to m. This shows that $\|f - (\sum_{i\in S} x_i + \mu)\|^2 \leqslant 4\gamma$. □

Putting Lemma 3.7 and Lemma 3.9 together, we get that f or $1 - f$ is $O(\epsilon)$-approximated by a function of the form $\max_{i\in S} x_i$, where $|S| = O(p^{-1})$. (When $\mu = 1$, f is close to a constant.) In order to improve the bound on $|S|$, we estimate the probability that $\sum_{i\in S} x_i \geqslant 2$.

Lemma 3.10. *Let S be a subset of $[n]$ of size $t \triangleq |S| \leqslant (3/2)p^{-1}$. If $t \geqslant 2$ then*

$$\Pr\left[\sum_{i\in S} x_i \geqslant 2\right] = \Omega((pt)^2).$$

Proof. Let $p' = \frac{k-1}{n-1} \geqslant p/2$ (using $k \geqslant 2$) and $p'' = \frac{k-2}{n-2} \leqslant p$. The inclusion-exclusion principle shows that

$$\begin{aligned}\Pr\left[\sum_{i\in S} x_i \geqslant 2\right] &\geqslant \binom{t}{2}\Pr[x_1 = x_2 = 1] - \binom{t}{3}\Pr[x_1 = x_2 = x_3 = 1] \\ &= \binom{t}{2} pp' \left(1 - \frac{tp''}{3}\right) \\ &\geqslant \frac{t^2}{2}\frac{p^2}{2}\frac{1}{2} = \frac{(tp)^2}{8}.\end{aligned}$$

□

We can now prove Theorem 3.1 when $\epsilon < p/128$.

Lemma 3.11. *Suppose $f\colon \binom{[n]}{k} \to \{0, 1\}$ is ϵ-close to an affine function, and let $p = \min(k/n, 1 - k/n)$. If $2 \leqslant k \leqslant n - 2$ and $\epsilon < p/128$ then either f or $1 - f$ is $O(\epsilon)$-close to $\max_{i\in S} x_i$ for some set S of size at most $\max(1, \sqrt{\epsilon}/p)$.*

Proof. Lemma 3.7 shows that for some real m and set S of size at most $n/2$ we have $\|f - (\delta \sum_{i\in S} x_i + m)\|^2 = O(\epsilon)$. For simplicity, assume that $\delta = 1$ (when $\delta = -1$, consider $1 - f$ instead of f). Lemma 3.9 then implies that $\|f - (\sum_{i\in S} x_i + \mu)\|^2 = O(\epsilon)$ for some $\mu \in \{0, 1\}$, assuming ϵ is small enough (otherwise the lemma is trivially true), and moreover $|S| \leqslant (3/2)p^{-1}$.

Suppose first that $\mu = 0$. In this case, if $|S| \geqslant 2$ then Lemma 3.10 implies that $(p|S|)^2 = O(\epsilon)$ and so $|S| = O(\sqrt{\epsilon}/p)$. The function $g = \max_{i\in S} x_i$ results from rounding $h \triangleq \sum_{i\in S} x_i$ to Boolean, and so Fact 1 implies that $\|f - g\|^2 = O(\|f - h\|^2) = O(\epsilon)$. When $\mu = 1$, we similarly get $\|f - 1\|^2 = O(\epsilon)$. □

We move on to the case $\epsilon = \Omega(1/p)$. In this case the analog of Lemma 3.7 states that f can be approximated by a function of the form $\sum_{i\in S_+} x_i - \sum_{i\in S_-} x_i + m$, where at least one of S_+, S_- is small.

Lemma 3.12. *There exist real m and two subsets $S_+, S_- \subseteq [n]$ satisfying $\frac{|S_+|}{n}\frac{|S_-|}{n} = O(\epsilon/k)$ such that $\|f - (\sum_{i\in S_+} x_i - \sum_{i\in S_-} x_i + m)\|^2 = O(\epsilon)$.*

Proof. Lemma 3.4 shows that $k\,\mathbb{E}_{j\neq i} \operatorname{dist}(c_i - c_j, \{0, \pm 1\})^2 = O(\epsilon)$. This implies that for some $i_0 \in [n]$,

$$k \mathop{\mathbb{E}}_{j\neq i_0} \operatorname{dist}(c_{i_0} - c_j, \{0, \pm 1\})^2 = O(\epsilon).$$

We partition the coordinates in $[n]$ into four sets. For $\delta \in \{0, \pm 1\}$, we let $S_\delta = \{j \in [n] : |c_j - c_{i_0} - \delta| < 1/4\}$, and we put the rest of the coordinates in a set R. Since $\mathbb{E}_{j\neq i_0} \operatorname{dist}(c_{i_0} - c_j, \{0, \pm 1\})^2 = \Omega(\frac{|R|}{n})$, we conclude that $\frac{|R|}{n} = O(\epsilon/k)$. Since $\mathbb{E}_{i\neq j}(c_i - c_j)^2 = \Omega(\frac{|S_{-1}|}{n}\frac{|S_{+1}|}{n})$, we conclude that $\frac{|S_{-1}|}{n}\frac{|S_{+1}|}{n} = O(\epsilon/k)$.

Define now $d_i = c_i - 1$ for $i \in S_{+1}$, $d_i = c_i + 1$ for $i \in S_{-1}$, and $d_i = c_i$ otherwise. When $i, j \in S_0 \cup S_{+1}$ or $i, j \in S_0 \cup S_{-1}$, we get $|d_i - d_j| < 1/2$, and since $d_i - d_j = c_i - c_j + \kappa$ for some $\kappa \in \{0, \pm 1\}$, we conclude that $\operatorname{dist}(c_i - c_j, \{0, \pm 1\}) = |d_i - d_j|$. For all i, j we claim that $(d_i - d_j)^2 \leqslant 7 \operatorname{dist}(c_i - c_j, \{0, \pm 1\})^2 + 16$. Indeed, for some $\kappa \in \{0, \pm 1, \pm 2\}$ we have $d_i - d_j = c_i - c_j + \kappa$. If $|c_i - c_j| \leqslant 2$ then $(d_i - d_j)^2 \leqslant 16$, whereas if $|c_i - c_j| \geqslant 2$, say $c_i - c_j \geqslant 2$, then $c_i - c_j - 1 \leqslant (c_i - c_j - 1)^2$ and so

$$\begin{aligned}(d_i - d_j)^2 &= ((c_i - c_j - 1) + (\kappa + 1))^2 \\ &= \operatorname{dist}(c_i - c_j, \{0, \pm 1\})^2 + 2(c_i - c_j - 1)(\kappa + 1) + (\kappa + 1)^2 \\ &\leqslant 7 \operatorname{dist}(c_i - c_j, \{0, \pm 1\})^2 + 9.\end{aligned}$$

Call a pair of indices i, j *good* if $i, j \in S_0 \cup S_{+1}$ or $i, j \in S_0 \cup S_{-1}$, and note that the probability that i, j is bad (not good) is at most $2\frac{|R|}{n} + 2\frac{|S_{+1}|}{n}\frac{|S_{-1}|}{n} = O(\epsilon/k)$. This shows that

$$k \mathop{\mathbb{E}}_{i\neq j} (d_i - d_j)^2 \leqslant 7k \mathop{\mathbb{E}}_{i\neq j} \operatorname{dist}(c_i - c_j, \{0, \pm 1\})^2 + 16k \Pr[i, j \text{ bad}] = O(\epsilon).$$

Corollary 3.6 now completes the proof. □

An argument similar to the one in Lemma 3.11 completes the proof of the theorem.

Proof of Theorem 3.1. If $\epsilon \leqslant p/128$ then the result follows from Lemma 3.11, so we can assume that $\epsilon > p/128$, and in particular we can assume that $p \leqslant 1/3$, since otherwise the result is trivial. We can also assume that n is large enough. Indeed, for any fixed n, since there are only finitely many functions on at most n coordinates, considering the best approximation ℓ, either $\epsilon = 0$, in which case the result follows from Lemma 3.2, or $\epsilon = \Omega(1)$, in which case the result is trivial. In several other places in the proof we also assume that ϵ is small enough.

Lemma 3.12 shows that for some real m and sets S_+, S_- we have $\|f - (\sum_{i\in S_+} x_i - \sum_{i\in S_-} x_i + m)\|^2 = O(\epsilon)$, where the sets S_+, S_- satisfy $\frac{|S_+|}{n}\frac{|S_-|}{n} = O(\epsilon/k)$. In particular, one of them, without loss of generality S_-, has size $O(\sqrt{\epsilon/k}n) = O(n^{5/8})$. Note that S_- could be empty.

Consider some setting of the variables in S_- which sets w of them to 1. This setting reduces the original slice to a slice $\binom{[n']}{k'}$, where $n' = n - |S_-|$ and $k' = k - w$. The corresponding $p' = \frac{k'}{n'}$

satisfies $\frac{k-|S_-|}{n-|S_-|} \leqslant p' \leqslant \frac{k}{n-|S_-|}$. Note that $|S_-| = O(\sqrt{1/k}k^{4/3}) = O(k^{5/6})$, and so for large enough n we get $p/2 \leqslant p' \leqslant 1/2$ and $2|S_-| \leqslant n'$. Lemma 3.9 then shows that for each such setting, either $\|f - (\sum_{i \in S_+} x_i - \sum_{i \in S_-} x_i + m)\|^2 = \Omega(1)$ or $\operatorname{dist}(m - \sum_{i \in S_-} x_i, \{0,1\}) \leqslant 1/4$. In the latter case, we say that w is good.

If no w is good then $\epsilon = \Omega(1)$, so we can assume that some w is good. The condition on m shows that at most two values $w_0, w_0 + 1$ can be good, and so $\Pr[\sum_{i \in S_-} x_i \in \{w_0, w_0 + 1\}] = 1 - O(\epsilon)$. Lemma 3.8 shows that $\Pr[\sum_{i \in S_-} x_i = 0] = \Omega(1)$ and so we can assume that $w_0 = 0$. Furthermore, if 1 is good then $\operatorname{dist}(m, \{0,1\}), \operatorname{dist}(m-1, \{0,1\}) \leqslant 1/4$, showing that $|m-1| \leqslant 1/4$.

Assume first that $|S_+| \leqslant n'/2$. Let $\||\cdot\||$ denote the norm restricted to inputs in which $\sum_{i \in S_-} x_i = 0$. Since $\Pr[\sum_{i \in S_-} x_i = 0] = \Omega(1)$, we deduce that $\||f - (\sum_{i \in S_+} x_i + m)\||^2 = O(\epsilon)$. Lemma 3.9 shows that $\||f - (\sum_{i \in S_+} x_i + \mu)\||^2 = O(\epsilon)$ for some $\mu \in \{0,1\}$ satisfying $|m - \mu| \leqslant 1/4$ and $|S_+| \leqslant (3/2)p'^{-1}$, where $p' = \frac{k}{n'} \geqslant p$.

Suppose first that $\mu = 0$. Since $|m| \leqslant 1/4$, we deduce that 1 is not good and conclude that $\Pr[\sum_{i \in S_-} x_i = 0] = 1 - O(\epsilon)$. Since $\||f - \sum_{i \in S_+} x_i\||^2 = O(\epsilon)$, Lemma 3.10 then implies that $|S_+| \leqslant \max(1, O(\sqrt{\epsilon/p'})) = \max(1, O(\sqrt{\epsilon}/p))$. Take now the function $g = \max_{i \in S_+} x_i$, which results from rounding $\sum_{i \in S_+} x_i$ to Boolean. Since $\||f - g\||^2 = O(\||f - \sum_{i \in S_+} x_i\||^2) = O(\epsilon)$ and $\Pr[\sum_{i \in S_-} x_i = 0] = 1 - O(\epsilon)$, we can conclude that $\|f - g\|^2 = O(\epsilon)$, completing the proof in this case.

Suppose next that $\mu = 1$. This implies that $\Pr[\sum_{i \in S_+} x_i = 0 | \sum_{i \in S_-} x_i = 0] \geqslant 1 - O(\epsilon)$, and so also $\Pr[\sum_{i \in S_+} x_i = 0] \geqslant 1 - O(\epsilon)$. Lemma 3.10 shows that $|S_+| \leqslant \max(1, O(\sqrt{\epsilon}/p)) = O(n^{1/4})$, and so $\frac{k}{n-|S_+|} \leqslant 1/2$ for large enough n. Defining $\||\cdot\||'$ to be the norm restricted to inputs in which $\sum_{i \in S_+} x_i = 0$, we conclude that for $m' = 1 - m$, $\||1 - f - (\sum_{i \in S_-} x_i + m')\||'^2 = O(\epsilon)$. If $|S_-| \leqslant (n - |S_+|)/2$ then Lemma 3.9 shows that $\||1 - f - (\sum_{i \in S_-} x_i + \mu')\||'^2 = O(\epsilon)$ for some $\mu' \in \{0,1\}$ and that $|\mu'| \leqslant (3/2)\frac{n-|S_+|}{k}$. When $\mu' = 0$, Lemma 3.10 shows that $|S_-| \leqslant \max(1, O(\sqrt{\epsilon}/p))$ and as before $\|1 - f - \max_{i \in S_-} x_i\|^2 = O(\epsilon)$. When $\mu' = 1$, we similarly get $\|1 - f - 1\|^2 = O(\epsilon)$. If $|S_-| > (n - |S_+|)/2$ then $\||f - (\sum_{i \in [n] \setminus (S_- \cup S_+)} x_i + m'')\||'^2 = O(\epsilon)$ for $m'' = 1 - m' - k$, and we can reason as before.

Finally, suppose that $|S_+| > n'/2$. Then $\||1 - f - (\sum_{i \in [n] \setminus (S_- \cup S_+)} x_i + m')\||^2 = O(\epsilon)$ for $m' = 1 - m - k$, and so Lemma 3.9 shows that $\||1 - f - (\sum_{i \in [n] \setminus (S_- \cup S_+)} x_i + \mu)\||^2 = O(\epsilon)$ for some $\mu \in \{0,1\}$, and again we can reason as before. □

References

[1] David Ellis, Yuval Filmus, and Ehud Friedgut. A quasi-stability result for dictatorships in S_n. *Combinatorica.*

[2] David Ellis, Yuval Filmus, and Ehud Friedgut. A stability result for balanced dictatorships in S_n. *Random Structures and Algorithms.*

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

[4] Jacek Jendrej, Krzysztof Oleszkiewicz, and Jakub Onufry Wojtaszczyk. On some extensions of the FKN theorem. Manuscript, 2013.

[5] Guy Kindler and Shmuel Safra. Noise-resistant Boolean functions are juntas. *Manuscript.*

[6] Aviad Rubinstein and Muli Safra. Boolean functions whose Fourier transform is concentrated on pairwise disjoint subsets of the inputs. Manuscript, 2013.

[7] Jakub Onufry Wojtaszczyk. Sums of independent variables approximating a boolean function. Manuscript, 2010.