FKN theorems for the biased cube and the slice

Yuval Filmus

October 23, 2021

Abstract

The classical Friedgut–Kalai–Naor theorem describes the structure of linear functions on the Boolean cube which are almost Boolean. We describe extensions of this theorem to functions on the biased Boolean cube and on the slice, simplifying earlier work by the author.

1 Introduction

A function $f\colon \{0,1\}^n \to \mathbb{R}$ on the Boolean cube is *linear* if it can be written in the form

$$f(x) = c + \sum_{i=1}^n c_i x_i.$$

The function f is *Boolean* if it is $\{0,1\}$-valued. It is easy to check that the Boolean linear functions are

$$0, 1, x_1, 1-x_1, \ldots, x_n, 1-x_n.$$

We call such functions *dictators*. Our dictators are always Boolean, and we consider constant functions to be dictators.

Can f be drastically different if it is only *approximately* Boolean? The answer depends on the notion of approximation being considered. Motivated by Boolean function analysis [O'D14], we consider L_2-approximation, saying that f is *ϵ-close to Boolean* if it satisfies

$$\mathop{\mathbb{E}}_{x\sim\mu}[\mathrm{dist}(f(x),\{0,1\})^2] \le \epsilon,$$

where μ is some probability distribution on the Boolean cube. (In the sequel, we will write the expectation more succinctly as $\mathbb{E}[\mathrm{dist}(f,\{0,1\})^2]$, suppressing the argument x.)

The most natural probability distribution μ to consider is the uniform distribution on the cube. In this case, the Friedgut–Kalai–Naor (FKN) theorem [FKN02] states that a linear function which is ϵ-close to Boolean must be $O(\epsilon)$-close to a dictator; here two functions $f,g\colon \{0,1\}^n \to \mathbb{R}$ are *δ-close* if $\mathbb{E}[(f-g)^2] \le \delta$.

There are two other natural distributions which show up in applications:

1. The product distribution μ_p, in which each coordinate is independently distributed $\mathrm{Ber}(p)$. In other words, $\mu_p(x) = p^{\sum_i x_i}(1-p)^{\sum_i (1-x_i)}$.
2. The slice distribution ν_k, which is the uniform distribution over all vectors in $\{0,1\}^n$ whose Hamming weight is exactly k.

Roughly speaking, the distributions ν_k and $\mu_{k/n}$ have similar behavior.

When p or k/n is balanced (bounded away from 0 and 1), the FKN theorem still holds. More interesting behavior arises when p or k/n is small. The extreme example is $k = 1$, in which case *all* functions have degree 1, and nothing at all can be said (however, when $2 \leq k \leq n-2$, a simple argument shows that the only Boolean linear functions are the dictators). A more generic example is

$$x_1 + \cdots + x_m,$$

for small enough m.

In previous work [Fil16], we extended the FKN theorem to slice distributions ν_k, and deduced an FKN theorem for product distributions μ_p; subsequently we extended the FKN theorem to the multislice, an analog of the slice for $[m]^n$ [Fil20]. In this paper we present a streamlined version of the proof in [Fil16], also slightly generalizing it to arbitrary product distributions, allowing different biases for different coordinates.

Our work will repeatedly use the "L_2^2 triangle inequality",

$$(a+b)^2 \leq 2a^2 + 2b^2.$$

2 FKN theorem for the unbiased cube

The FKN theorem has several equivalent formulations. Here is the original one:

Theorem 2.1 ([FKN02])**.** *If $f\colon \{0,1\}^n \to \{0,1\}$ satisfies $\|f^{>1}\|^2 \leq \epsilon$ then f is $O(\epsilon)$-close to a dictator.*

(Recall that $\|g\|^2 = \mathbb{E}[g^2]$, and $f^{>1} = f - f^{\leq 1}$, where $f^{\leq 1}$ is the orthogonal projection of f onto the space of linear functions.)

We prefer the following simple corollary:

Corollary 2.2. *If a linear function $f\colon \{0,1\}^n \to \mathbb{R}$ satisfies $\mathbb{E}[\operatorname{dist}(f,\{0,1\})^2] \leq \epsilon$ then f is $O(\epsilon)$-close to a dictator.*

Proof. Let $F = \operatorname{round}(f,\{0,1\})$, so that

$$\|F^{>1}\|^2 = \mathbb{E}[(F - F^{\leq 1})^2] \leq \mathbb{E}[(F-f)^2] = \mathbb{E}[\operatorname{dist}(f,\{0,1\})^2] \leq \epsilon.$$

Applying the theorem, we find that F is $O(\epsilon)$-close to some dictator g. It follows that f is also $O(\epsilon)$-close to g, since

$$\mathbb{E}[(f-g)^2] \leq 2\,\mathbb{E}[(f-F)^2] + 2\,\mathbb{E}[(F-g)^2] = 2\,\mathbb{E}[\operatorname{dist}(f,\{0,1\})^2] + O(\epsilon) = O(\epsilon). \qquad \square$$

We find it useful to think of the FKN theorem in the following way: if a linear function is close to Boolean, then the coefficients of its Fourier expansion (given explicitly below) are close to $\{0,\pm 1\}$, and moreover, most of them are close to 0.

Lemma 2.3. *Let $f\colon \{0,1\}^n \to \mathbb{R}$ be a linear function, given by the formula*

$$f(x) = c + \sum_{i=1}^n c_i x_i,$$

and let $d_i = \operatorname{round}(c_i,\{0,\pm 1\})$.

If f is ϵ-close to Boolean then

$$\sum_{i=1}^n (c_i - d_i)^2 = O(\epsilon),$$

and furthermore,

$$\#\{i \in [n] : d_i \neq 0\} \leq 1 + O(\epsilon).$$

Proof. We have $x_i = (1-(-1)^{x_i})/2$, and conversely, $(-1)^{x_i} = 1 - 2x_i$. Therefore the Fourier expansion of f is

$$f = c' + \sum_{i=1}^{n} \frac{-c_i}{2} (-1)^{x_i},$$

for some constant c'.

According to Corollary 2.2, there is a dictator g such that $\mathbb{E}[(f-g)^2] = O(\epsilon)$. Since the functions $1, (-1)^{x_1}, \ldots, (-1)^{x_n}$ are orthonormal, Parseval's identity implies that

$$\sum_{i=1}^{n} \bigl(-c_i/2 - \hat{g}(\{i\})\bigr)^2 = O(\epsilon),$$

where $\hat{g}(\{i\})$ is the coefficient of $(-1)^{x_i}$ in the Fourier expansion of g. This, in turn, implies that

$$\sum_{i=1}^{n} \bigl(c_i + 2\hat{g}(\{i\})\bigr)^2 = O(\epsilon).$$

The Fourier expansion of g is given by an expression of one of the following types:

$$0, 1, \frac{1-(-1)^{x_I}}{2}, \frac{1+(-1)^{x_I}}{2}.$$

In all cases, $\hat{g}(\{i\}) \in \{0, \pm 1/2\}$ for all i and so $-2\hat{g}(\{i\}) \in \{0, \pm 1\}$ for all i. Thus $(c_i - d_i)^2 \le (c_i + 2\hat{g}(\{i\}))^2$, and the first statement of the lemma follows.

To see the second statement, notice that

$$\sum_{i \neq I} c_i^2 = O(\epsilon),$$

where we choose I arbitrarily when $g \in \{0,1\}$. If $d_i \neq 0$ then $c_i^2 \ge 1/4$, implying the second statement. □

3 FKN theorem for the biased cube

For a vector $p \in [0,1]^n$, let μ_p be the product distribution on $\{0,1\}^n$ given by[1]

$$\mu_p(x) = \prod_{i \colon x_i = 1} p_i \prod_{i \colon x_i = 0} (1-p_i).$$

We can assume, without loss of generality, that $p_i \le 1/2$ for all i (since we can flip coordinates with $p_i > 1/2$).

Here is the FKN theorem for μ_p:

Theorem 3.1. *Let $p \in (0,1/2]^n$. If $f \colon \{0,1\}^n \to \mathbb{R}$ is a linear function which is ϵ-close to Boolean with respect to μ_p, then f is $O(\epsilon)$-close to a function of the form*

$$x_{i_1} + \cdots + x_{i_m} \text{ or } 1 - x_{i_1} - \cdots - x_{i_m},$$

where all i_t's are distinct, and

$$\sum_{1 \le s < t \le m} p_{i_s} p_{i_t} = O(\epsilon). \tag{$*$}$$

(If $p_i = q$ for all $i \in [n]$, then this condition simplifies to $m \le 1$ or $m = O(\sqrt{\epsilon}/q)$.)

Conversely, any function of this form is $O(\epsilon)$-close to Boolean.

[1]This notation extends the notation μ_p appearing in the introduction by identifying p with the constant vector $(p, \ldots, p)$.

To see where the condition on the p_{i_t} comes from, notice that

$$\sum_{1 \le s < t \le m} p_{i_s} p_{i_t}$$

is an upper bound on $\Pr[h \ge 2]$, where $h = x_{i_1} + \cdots + x_{i_m}$. We can actually say more: $\mathbb{E}[\mathrm{dist}(h, \{0,1\})^2] = O(\epsilon)$, which implies the converse direction of Theorem 3.1. The trick is the easy observation that if $z \in \mathbb{N}$ then $\mathrm{dist}(z, \{0,1\})^2 \le z(z-1)$. Indeed, if $z \in \{0,1\}$ then both sides vanish, and otherwise

$$\mathrm{dist}(z, \{0,1\})^2 = (z-1)^2 \le z(z-1).$$

This observation implies that

$$\mathbb{E}[\mathrm{dist}(h, \{0,1\})^2] \le \mathbb{E}[h^2 - h] = \mathbb{E}\left[\sum_{1 \le s < t \le m} x_{i_s} x_{i_t}\right] = \sum_{1 \le s < t \le m} p_{i_s} p_{i_t} = O(\epsilon).$$

We pause to mention that if $p_1 = \cdots = p_n = q$ then $(*)$ becomes

$$\binom{m}{2} q^2 \le C\epsilon,$$

for some $C > 0$, and so either $m \le 1$ or $C\epsilon \ge m(m-1)q^2/2 \ge (mq/2)^2$, and so $m \le 2\sqrt{C\epsilon}/q = O(\sqrt{\epsilon}/q)$. This explains the parenthetical remark in the statement of the theorem.

Before proceeding with the proof of the theorem, let us note the following curious corollary:

Corollary 3.2. *Let $p \in (0,1/2]^n$, and let $q = \max(p_1, \ldots, p_n)$. If $f \colon \{0,1\}^n \to \mathbb{R}$ is a linear function which is ϵ-close to Boolean with respect to μ_p, then f is $O(q + \sqrt{\epsilon} + \epsilon)$-close to a constant function (either 0 or 1).*

Proof. Suppose without loss of generality that f is $O(\epsilon)$-close to $h = x_{i_1} + \cdots + x_{i_m}$, and $(*)$ holds. The distance from h to the constant zero function is

$$\mathbb{E}[h^2] = \sum_{t=1}^m \mathbb{E}[x_{i_t}^2] + 2 \sum_{1 \le s < t \le m} \mathbb{E}[x_{i_s} x_{i_t}] = \sum_{t=1}^m p_{i_t} + \sum_{1 \le s < t \le m} p_{i_s} p_{i_t}.$$

The second sum is $O(\epsilon)$ by assumption. In order to bound the first sum, which we denote by P, we consider its square:

$$P^2 = \sum_{t=1}^m p_{i_t}^2 + 2 \sum_{1 \le s < t \le m} p_{i_s} p_{i_t} \le qP + O(\epsilon).$$

Writing the left-hand side as $P^2/2 + P^2/2$, we see that either $P^2 \le 2qP$ or $P^2 = O(\epsilon)$. In the first case, $P \le 2q$, and in the second case, $P = O(\sqrt{\epsilon})$. In both cases, we conclude that $\mathbb{E}[h^2] = O(q + \sqrt{\epsilon} + \epsilon)$. $\square$

Finally, let us note that the theorem is trivial if ϵ is large, say at least some constant $\epsilon_0 > 0$. Indeed, in this case the distance from f to the constant zero function is

$$\mathbb{E}[f^2] \le 2\,\mathbb{E}[(f - \mathrm{round}(f, \{0,1\}))^2] + 2\,\mathbb{E}[\mathrm{round}(f, \{0,1\})^2] \le 2\,\mathbb{E}[\mathrm{dist}(f, \{0,1\})^2] + 2 \le 2(\epsilon + 1) \le 2(1 + 1/\epsilon_0)\epsilon = O(\epsilon).$$

Therefore, in the proof we can assume that ϵ is "small enough", that is, smaller than some unspecified positive constant ϵ_0.

3.1 The proof

The main idea of the proof is a two-step process for sampling a point according to μ_p. First, we sample a subset $S \subseteq [n]$ by putting i into S with probability $2p_i$ independently (this is why we need $p_i \leq 1/2$), a distribution we denote by $\mu_{2p}([n])$ (identifying Boolean vectors and subsets). We fix coordinates outside of S to zero, leaving us with a copy of $\{0,1\}^S$, within which we sample a uniformly random point. This results in a μ_p-distributed point, since the probability that $x_i = 1$ is precisely $2p_i \cdot \frac{1}{2} = p_i$.

The starting point is the given linear function

$$f = c + \sum_{i=1}^{n} c_i x_i,$$

which is ϵ-close to Boolean.

Given $S \subseteq [n]$, let $f|_S$ be the restriction of f to $\{0,1\}^S$ obtained by zeroing out coordinates outside S. Thus

$$f|_S = c + \sum_{i \in S} c_i x_i.$$

If we denote $\epsilon_S = \mathbb{E}_{\mu_{1/2}}[\mathrm{dist}(f|_S, \{0,1\})^2]$, then

$$\mathop{\mathbb{E}}_{S \sim \mu_{2p}([n])}[\epsilon_S] = \mathop{\mathbb{E}}_{\substack{S \sim \mu_{2p}([n]) \\ x \sim \mu_{1/2}(\{0,1\}^S)}}[\mathrm{dist}(f(x), \{0,1\})^2] = \mathop{\mathbb{E}}_{x \sim \mu_p(\{0,1\}^n)}[\mathrm{dist}(f(x), \{0,1\})^2] = \epsilon.$$

In other words, on average $f|_S$ is close to Boolean.

Defining $d_i = \mathrm{round}(c_i, \{0, \pm 1\})$ and applying Lemma 2.3, we obtain

$$\mathop{\mathbb{E}}_{S \sim \mu_{2p}([n])}\left[\sum_{i \in S} (c_i - d_i)^2\right] = O(\epsilon), \tag{3.1}$$

as well as

$$\mathop{\mathbb{E}}_{S \sim \mu_{2p}([n]}[\#\{i \in S : d_i \neq 0\}] \leq 1 + O(\epsilon). \tag{3.2}$$

If $S \sim \mu_{2p}([n])$ then every i belongs to S with probability $2p_i$. Consequently, (3.1) implies that

$$\sum_{i=1}^{n} p_i (c_i - d_i)^2 = O(\epsilon). \tag{3.3}$$

Similarly, since

$$\mathop{\mathbb{E}}_{S \sim \mu_{2p}([n]}[\#\{i \in S : d_i \neq 0\}] = \mathop{\mathbb{E}}_{S \sim \mu_{2p}([n]}\left[\sum_{i=1}^{n} 1_{i \in S \text{ and } d_i \neq 0}\right] = \sum_{i\colon d_i = 0} 2p_i,$$

Equation (3.2) implies that

$$\sum_{i\colon d_i \neq 0} p_i \leq \frac{1}{2} + O(\epsilon) \leq \frac{3}{4}, \tag{3.4}$$

assuming that ϵ is small enough.

The idea now is to replace each c_i with the corresponding d_i, capitalizing on (3.3). If we want the resulting function to be close to f, we have to do it carefully, rewriting f as

$$f = c' + \sum_{i=1}^{n} \sigma_i c_i \frac{x_i - p_i}{\sigma_i},$$

where $\sigma_i = \sqrt{p_i(1-p_i)}$ and c' is some constant. This expresses f as a linear combination of *orthonormal* functions. Consequently, if we define

$$g' = c' + \sum_{i=1}^n \sigma_i d_i \frac{x_i - p_i}{\sigma_i} = d' + \sum_{i=1}^n d_i x_i,$$

where d' is some constant, then

$$\mathbb{E}[(f-g')^2] = \sum_{i=1}^n \sigma_i^2 (c_i - d_i)^2 \leq \sum_{i=1}^n p_i (c_i - d_i)^2 = O(\epsilon),$$

and so $\mathbb{E}[\mathrm{dist}(g', \{0,1\})^2] = O(\epsilon)$ by the L_2^2 triangle inequality.

Equation (3.4) implies that $\Pr[g' = d'] \geq 1/4$, and so $\mathrm{dist}(d', \{0,1\})^2 = O(\epsilon)$. Consequently, if we define $d = \mathrm{round}(d, \{0,1\})$ and

$$g = d + \sum_{i=1}^n d_i x_i,$$

then $\mathbb{E}[(g'-g)^2] = O(\epsilon)$, which implies that $\mathbb{E}[(f-g)^2] = O(\epsilon)$ and so $\mathbb{E}[\mathrm{dist}(g, \{0,1\})^2] = O(\epsilon)$.

We are almost there: the major remaining step is showing that if $d = 0$ then most non-zero d_i's are equal to 1, and if $d = 1$ then most of them are equal to -1.

Assume for concreteness that $d = 0$, let $A = \{i \in [n] : d_i \neq 0\}$ be the indices of non-zero d_i, and let $B = \{i \in [n] : d_i = -1\}$ be the indices of "bad" d_i. We would like to show that g is $O(\epsilon)$-close to the function obtained by removing the indices in B, namely,

$$h = \sum_{i \notin B} d_i x_i = \sum_{i \in A \setminus B} x_i.$$

Applying (3.4), for each $i \in B$ we have

$$\Pr[x_i = 1 \text{ and } x_j = 0 \text{ for all } j \in A \setminus \{i\}] \geq \frac{p_i}{4}.$$

If this event happens then $g(x) = -1$, and so $\mathrm{dist}(g(x), \{0,1\})^2 = 1$. Since these events are disjoint for different $i \in B$, it immediately follows that

$$\sum_{i \in B} p_i = O(\epsilon).$$

Therefore

$$\mathbb{E}[(g-h)^2] = \mathbb{E}\left[\left(\sum_{i \in B} x_i\right)^2\right] = \sum_{i \in B} p_i + \sum_{\substack{i,j \in B \\ i \neq j}} p_i p_j \leq$$

$$\sum_{i \in B} p_i + \sum_{i,j \in B} p_i p_j = \sum_{i \in B} p_i + \left(\sum_{i \in B} p_i\right)^2 = O(\epsilon + \epsilon^2) = O(\epsilon).$$

It follows that $\mathbb{E}[(f-h)^2] = O(\epsilon)$ and so $\mathbb{E}[\mathrm{dist}(h, \{0,1\})^2] = O(\epsilon)$.

It remains to show that $(*)$ holds. Let $A \setminus B = \{i_1, \ldots, i_{m'}\}$. Applying (3.4), if $1 \leq s < t \leq m'$ then

$$\Pr[x_{i_s} = x_{i_t} = 1 \text{ and } x_{i_r} = 0 \text{ for all } r \neq s, t] \geq \frac{p_{i_s} p_{i_t}}{4}.$$

If this event happens then $h(x) = 2$ and so $\mathrm{dist}(h(x), \{0,1\})^2 = 1$. These events are disjoint, and so $\mathbb{E}[\mathrm{dist}(h, \{0,1\})^2] = O(\epsilon)$ directly implies $(*)$.

4 FKN theorem for the slice

The slice $\binom{[n]}{k}$ is the subset of $\{0,1\}^n$ consisting of all vectors containing exactly k many 1s. We denote the uniform distribution over the slice variously by $\nu_{n,k}, \nu_k, \nu$, depending on the context.

In many respects, the distribution $\nu_{n,k}$ is similar to the distribution μ_p, for $p = k/n$. For example, the two distributions have the same marginals:

$$\Pr_{x\sim\nu_{n,k}}[x_i = 1] = \Pr_{x\sim\mu_p}[x_i = 1] = p.$$

When looking at several coordinates at once, differences arise, though they are slight unless k is very extreme (either very close to 0 or very close to n). For example, whereas the probability that $x_i = x_j = 1$ under μ_p is p^2, under $\nu_{n,k}$ this probability is

$$\frac{k(k-1)}{n(n-1)} = \left(1 - \frac{n-k}{k(n-1)}\right) p^2, \tag{4.1}$$

which is $\Theta(p^2)$ unless $k \leq 1$.

Accordingly, the FKN theorem for $\nu_{n,k}$ is very similar to its counterpart for μ_p, where once again we assume, without loss of generality, that $k/n \leq 1/2$:

Theorem 4.1. *Let $2 \leq k \leq n/2$, and define $p = k/n$. If $f\colon \binom{[n]}{k} \to \mathbb{R}$ is a linear function which is ϵ-close to Boolean with respect to the uniform distribution ν, then f is $O(\epsilon)$-close to a function of the form*

$$x_{i_1} + \cdots + x_{i_m} \text{ or } 1 - x_{i_1} - \cdots - x_{i_m},$$

where all i_t's are distinct, and either $m \leq 1$ or $m = O(\sqrt{\epsilon}/p)$.

Conversely, any function of this form is $O(\epsilon)$-close to Boolean.

We should mention at this point that while it is possible to define linear functions on the slice exactly the same as on the Boolean cube, it is sometimes more convenient to define them as functions of the form

$$\sum_{i=1}^{n} c_i x_i.$$

There is no need for the constant coefficient since it can be expressed via the identity $\sum_{i=1}^{n} x_i = k$, which holds for all points on the slice. As a bonus, this representation is unique.

Why do we ask that $k \geq 2$? We want to rule out $k = 0$ since then $p = 0$. When $k = 1$, *every* function is linear, and so the theorem isn't true: for example, the function $x_1 + \cdots + x_{\lfloor n/2 \rfloor}$ is Boolean but is not a dictator.

Before proceeding with the proof, let us explain in brief how to prove the converse part of the theorem, as well as the following counterpart of Corollary 3.2:

Corollary 4.2. *Let $2 \leq k \leq n/2$, and define $p = k/n$. If $f\colon \binom{[n]}{k} \to \mathbb{R}$ is a linear function which is ϵ-close to Boolean with respect to the uniform distribution ν, then f is $O(p + \sqrt{\epsilon} + \epsilon)$-close to a constant function (either 0 or 1).*

The proofs of both the converse part of the theorem and of the corollary closely follow their counterparts in Section 3, using the estimate $\mathbb{E}[x_i x_j] \leq p^2$ following from (4.1). Also, just as in Section 3, for the proof of the main part of Theorem 4.1, we can assume that ϵ is small enough.

4.1 The proof

As in the proof of Theorem 4.1, the idea is to use a two-step process for sampling a point according to $\nu_{n,k}$. This time the reduction is a bit different. First, we choose a random permutation $a = a_1, \ldots, a_n$ of $1, \ldots, n$. Consider the subset $\mathcal{D}_a$ of the slice which consists of all vectors satisfying

$$x_{a_1} + x_{a_2} = x_{a_3} + x_{a_4} = \cdots = x_{a_{2k-1}} + x_{a_{2k}} = 1,$$

the rest of the coordinates necessarily vanishing. We identify this subset with a copy of $\{0,1\}^k$ as follows: we map $y \in \{0,1\}^k$ to the point $x \in \mathcal{D}_a$ given by

$$\begin{aligned} &x_{a_1} = y_1, && x_{a_3} = y_2, && \ldots, && x_{a_{2k-1}} = y_k, \\ &x_{a_2} = 1 - y_1, && x_{a_4} = 1 - y_2, && \ldots, && x_{a_{2k}} = 1 - y_k, \end{aligned}$$

the remaining coordinates being set to zero.

The starting point is the given linear function

$$f = \sum_{i=1}^{n} c_i x_i,$$

which is ϵ-close to Boolean.

The restriction of f to $\mathcal{D}_a$ takes the form

$$\begin{aligned} f|_{\mathcal{D}_a} &= c_{a_1} y_1 + c_{a_2}(1 - y_1) + \cdots + c_{a_{2k-1}} y_k + c_{a_{2k}}(1 - y_k) \\ &= c + (c_{a_1} - c_{a_2}) y_1 + \cdots + (c_{a_{2k-1}} - c_{a_{2k}}) y_k, \end{aligned}$$

for some constant c. Choosing a at random and applying Lemma 2.3, we obtain

$$\mathop{\mathbb{E}}_{a}\left[\sum_{\ell=1}^{k} (c_{a_{2\ell-1}} - c_{a_{2\ell}} - d_{a_{2\ell-1},a_{2\ell}})^2\right] = O(\epsilon), \tag{4.2}$$

as well as

$$\mathop{\mathbb{E}}_{a}\left[\#\{\ell \in [k] : d_{a_{2\ell-1},a_{2\ell}} \neq 0\}\right] \leq 1 + O(\epsilon), \tag{4.3}$$

where $d_{i,j} = \mathrm{round}(c_i - c_j, \{0, \pm 1\})$.

If $i \neq j$ are both in $[n]$, then the probability that $(a_{2\ell-1}, a_{2\ell}) = (i, j)$ for some $\ell \in [k]$ is $\frac{k}{n(n-1)}$. Therefore

$$k \mathop{\mathbb{E}}_{i \neq j} (c_i - c_j - d_{i,j})^2 = O(\epsilon) \tag{4.4}$$

and

$$k \Pr_{i \neq j}[d_{i,j} \neq 0] \leq 1 + O(\epsilon). \tag{4.5}$$

In both cases, the expectation and probability go over all $n(n-1)$ unordered pairs $i, j \in [n]$ such that $i \neq j$.

At this point of the proof, we need a new idea, which will allow us to convert the information we get on pairs of coefficients to information on single coefficients. The trick is very simple. If we choose j at random, then (3.3) still holds, where now only i varies. In particular, with probability at least $2/3$ over the choice of j, (3.3) holds with a hidden constant three times as big. Similarly, with probability at least $2/3$ over the choice of j, (3.4) holds with the right-hand side replaced by $3 + O(\epsilon)$. Both of these bounds hold with probability at least $1/3$, and in particular, for some choice $j = J$ we have

$$k \mathop{\mathbb{E}}_{i \neq J} (c_i - c_J - d_{i,J})^2 = O(\epsilon) \tag{4.6}$$

and (for small enough ϵ)

$$k \Pr_{i \neq J}[d_{i,J} \neq 0] \leq 3 + O(\epsilon) \leq 4, \tag{4.7}$$

For brevity, let us write $d_i = d_{i,J}$, and define $d_J = 0$.

This suggests replacing c_i with $c_J + d_i$. As in Section 3.1, we have to do it carefully: first we write

$$f = c + \sum_{i=1}^{n} c_i (x_i - p)$$

for some constant c, recalling that $p = k/n$, and then we take

$$g' = c + \sum_{i=1}^n (c_J + d_i)(x_i - p) = d' + \sum_{i=1}^n d_i x_i;$$

we swallowed $\sum_{i=1}^n c_J x_i = c_J k$ inside the constant d'.

While the functions $x_i - p$ are not orthogonal, they are almost orthogonal: if $i \neq j$ then

$$\mathbb{E}[(x_i - p)(x_j - p)] = \mathbb{E}[x_i x_j] - p^2 = -\frac{k(n-k)}{n^2(n-1)},$$

using (4.1). Since $\mathbb{E}[(x_i - p)^2] = p(1-p)$ (just as for μ_p), we have

$$\mathbb{E}[(f - g')^2] = p(1-p) \sum_{i=1}^n (c_i - c_J - d_i)^2 - \frac{k(n-k)}{n^2(n-1)} \sum_{i \neq j} (c_i - c_J - d_i)(c_j - c_J - d_j). \tag{4.8}$$

The first term in (4.8) is at most

$$\frac{k}{n} \sum_{i=1}^n (c_i - c_J - d_i)^2 \leq k \mathop{\mathbb{E}}_{i \neq J} (c_i - c_J - d_i)^2 = O(\epsilon),$$

since $c_i = c_J + d_i$ when $i = J$.

In order to bound the second term in (4.8), we use the inequality $|zw| \leq \frac{z^2 + w^2}{2}$:

$$- \frac{k(n-k)}{n^2(n-1)} \sum_{i \neq j} (c_i - c_J - d_i)(c_j - c_J - d_j) \leq \frac{k}{n(n-1)} \sum_{i \neq j} \frac{(c_i - c_J - d_i)^2 + (c_j - c_J - d_j)^2}{2} =$$

$$\frac{k}{n(n-1)} \sum_i \sum_{j \neq i} \frac{(c_i - c_J - d_i)^2}{2} + \frac{k}{n(n-1)} \sum_j \sum_{i \neq j} \frac{(c_j - c_J - d_j)^2}{2} = \frac{k}{n} \sum_{i=1}^n (c_i - c_J - d_i)^2 = O(\epsilon),$$

using the bound on the first term. Over all, this shows that $\mathbb{E}[(f - g')^2] = O(\epsilon)$ and so $\mathbb{E}[\text{dist}(g', \{0,1\})^2] = O(\epsilon)$.

At this point in the proof of Theorem 3.1, we appealed to the bound $\sum_{i \colon d_i \neq 0} p_i \leq 3/4$, which implies that $\Pr[g' = d'] \geq 1/4$. Here a similar bound holds, unless k is very small; we handle that case separately later. Let us start by observing that (4.7) implies that

$$m := \#\{i : d_i \neq 0\} \leq 4\frac{n-1}{k} \leq \frac{4}{p}.$$

Now suppose that $k \geq 9$. Then $m \leq \frac{4}{9}n$, and so

$$\Pr[x_i = 0 \text{ whenever } d_i \neq 0] = \left(1 - \frac{k}{n}\right) \cdots \left(1 - \frac{k}{n-m+1}\right) \geq \left(1 - \frac{k}{n-m+1}\right)^m.$$

Now, $n - m + 1 \geq \frac{5}{9}n$, and so $\frac{k}{n-m+1} \leq \frac{k}{n} \cdot \frac{9}{5} \leq \frac{9}{10}$, hence $1 - \frac{k}{n-m+1} = \exp -\Theta(\frac{k}{n-m+1})$, and so

$$\Pr[x_i = 0 \text{ whenever } d_i \neq 0] \geq \exp -O\left(\frac{km}{n-m}\right) = \exp -O(pm) = \Omega(1).$$

This implies that $\Pr[g' = d'] = \Omega(1)$, and so $\text{dist}(d', \{0,1\})^2 = O(\epsilon)$. Defining $d = \text{round}(d', \{0,1\})$ and

$$g = d + \sum_{i=1}^n d_i x_i,$$

we have $\mathbb{E}[(g'-g)^2] = O(\epsilon)$, implying that $\mathbb{E}[(f-g)^2] = O(\epsilon)$ and so $\mathbb{E}[\text{dist}(g,\{0,1\})^2 = O(\epsilon)$.

The rest of the proof (assuming $k \geq 9$) is very similar to the proof in Section 3.1. Assume, without loss of generality, that $d = 0$. Let A be the indices of non-zero d_i, let B the indices where $d_i = -1$, and define

$$h = \sum_{i \notin B} d_i x_i = \sum_{i \in A \setminus B} x_i.$$

For each $i \in B$, we have

$$\Pr[x_i = 1 \text{ and } x_j = 0 \text{ for all } j \in A \setminus \{i\}] = p\left(1 - \frac{k-1}{n-1}\right) \cdots \left(1 - \frac{k-1}{n-m+1}\right) = \Omega(p).$$

In each such event, $g(x) = -1$ and so $\text{dist}(g(x),\{0,1\})^2 = 1$. Since the events are disjoint, we immediately get $|B| = O(\epsilon/p)$. Therefore as in Section 3.1,

$$\mathbb{E}[(g-h)^2] = \mathbb{E}\left[\left(\sum_{i \in B} x_i\right)^2\right] \leq p|B| + p^2|B|^2 = O(\epsilon),$$

which implies that $\mathbb{E}[(f-h)^2] = O(\epsilon)$ and $\mathbb{E}[\text{dist}(h,\{0,1\})^2] = O(\epsilon)$.

It remains to bound the size of $A \setminus B$. If $i,j \in A \setminus B$ are different then

$$\Pr[x_i = x_j = 1 \text{ and } x_\ell = 0 \text{ for all } \ell \in A \setminus (B \cup \{i,j\})] = \Omega(p^2)\left(1 - \frac{k-2}{n-2}\right) \cdots \left(1 - \frac{k-2}{n-m+1}\right) = \Omega(p^2),$$

using (4.1). In each such event, $h(x) = 2$ and so $\text{dist}(h(x),\{0,1\})^2 = 1$. These events are disjoint, and so $\mathbb{E}[\text{dist}(h,\{0,1\})^2] = O(\epsilon)$ implies that the number of pairs of elements in $A \setminus B$ is $O(\epsilon/p^2)$. Hence either $|A \setminus B| \leq 1$ or $|A \setminus B| = O(\sqrt{\epsilon}/p)$.

Now let us take care of the case $2 \leq k \leq 8$. Let $\Delta \in \{0,\pm 1\}$ be the most common value of $d_1,\ldots,d_n$, so that at least $n/3$ many of the d_i are equal to Δ. Then

$$g' = d'' + \sum_{i=1}^n (d_i - \Delta) x_i$$

for some constant d'', using $\sum_{i=1}^n x_i = k$. Denote the number of indices such that $d_i \neq \Delta$ by $M \leq 2n/3$. If $n/3 \geq 2k$ then

$$\Pr[x_i = 0 \text{ whenever } d_i \neq \Delta] = \left(1 - \frac{k}{n}\right) \cdots \left(1 - \frac{k}{n-M+1}\right) \geq \exp -O\left(\frac{Mk}{n-M}\right) = \Omega(1),$$

and so we can repeat the previous proof, with one difference: the coefficients $d_i - \Delta$ are no longer guaranteed to belong to $\{0,\pm 1\}$, but they do belong to $\{0,\pm 1,\pm 2\}$, and this suffices for the proof.

If $n/3 \leq 2k$ then $n \leq 6k < 50$. In this case, if g' is not a dictator that it is not Boolean and so $\mathbb{E}[\text{dist}(g',\{0,1\})^2] = \Omega(1)$, since n is bounded. Assuming that ϵ is small enough, we can rule out this case, and so g' is a dictator, completing the proof.

References

[Fil16] Yuval Filmus. Friedgut–Kalai–Naor theorem for slices of the Boolean cube. *Chicago Journal of Theoretical Computer Science*, pages 14:1–14:17, 2016.

[Fil20] Yuval Filmus. FKN theorem for the multislice, with applications. *Combin. Probab. Comput.*, 29(2):200–212, 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.

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