Boolean Functions whose Fourier Transform is Concentrated on the First Two Levels

Ehud Friedgut, Gil Kalai and Assaf Naor
Institute of Mathematics,
Hebrew University, Jerusalem, Israel

Abstract

In this note we describe Boolean functions $f(x_1, x_2, \ldots, x_n)$ whose Fourier coefficients are concentrated on the lowest two levels. We show that such a function is close to a constant function or to a function of the form $f = x_k$ or $f = 1 - x_k$. This result implies a "stability" version of a classical discrete isoperimetric result and has an application in the study of neutral social choice functions. The proofs touch on interesting harmonic analysis issues.

1 Introduction

A fundamental result in extremal combinatorics asserts that for every subset A of the discrete cube $\Omega_n = \{0,1\}^n$ the number of edges from A to its compliment it at least $\min(|A|, 2^n - |A|)$. The only examples of equality are $A = \emptyset$, $A_j = \{(x_1, x_2, \ldots, x_n) \in \Omega_n : x_j = 0$, for some $j, 1 \leq j \leq n$ and their complements. The precise minimum of the number of edges from A to its complement as a function of $|A|$ is also known, see Hart (1976). The main result of this paper is that if

the number of edges between A and its complement is "close" to $|A|$ then the set A is "close" to the examples described above. Equivalently, we describe Boolean functions whose Fourier transform is concentrated on the lowest two levels. This result touches on interesting issues in harmonic analysis and is used in Kalai (2001), to derive a "stability" version of Arrow's theorem for neutral social choice functions. We will use the terminology of Kalai (2001). Our main result is:

Theorem 1.1. *If f is a Boolean function, $\|f\|_2^2 = p$ and if $\sum_{|S|>1} \widehat{f}^2(S) \leq \delta$ then either $p < K'\delta$ or $p > 1 - K'\delta$ or $\|f(x_1, x_2, \ldots, x_n) - x_i\|_2^2 \leq K\delta$ for* *some i or $\|f(x_1, x_2, \ldots, x_n) - (1 - x_i)\|_2^2 \leq K\delta$ for some i. Here, K' and K are absolute constants.*

For $A \subset \Omega_n$ let $E(A)$ denotes the number of edges between A and its complement in Ω_n.

Corollary 1.2. *There exists an absolute constant K with the following property. Let $A \subset \Omega_n$ and suppose that $|A| \leq 2^{n-1}$. If $E(A) \leq (1+\epsilon) \cdot |A|$ then for some i, $1 \leq i \leq n$ $\mathbf{P}(A \triangle A_j) \leq K \cdot \epsilon$ or $\mathbf{P}(A \triangle \bar{A}_j) \leq K \cdot \epsilon$.*

Suppose that $|A| \leq 2^{n-1}$ and let $p = \|\chi_A\|_2 = |A|/2^{n-1}$. Let $I(A) = E(A)/2^{n-1}$. $I(A)$ is refered to as the *total influence* of A, as the *average sensitivity* of A and as the *edge boundary* or *edge-expansion constant* of A. It is well known that if $f = \chi_A$,

$$I(A) = 4 \sum_{S \subset [n]} \widehat{f}^2(S)|S|.$$

Since $\sum_{S \subset [n], S \neq \emptyset} \widehat{f}^2(S) = 2p$, we obtain that $E(A) \leq |A|$ with equality only of $p = 1/2$ and $\widehat{f}(S) = 0$ if $|S| > 1$. By the same argument we see how Theorem 1.1 implies Corollary 1.2. Corollary 1.2 complements a result by Friedgut (1998). Friedgut showed that if $I(S)$ is bounded from above by a constant K then $f = \chi_A$ is close to a "Junta" namely to a function which depends on a bounded number of variables. We show here that if $I(A) \leq 1+\epsilon$ then χ_A is close to "dictatorship".

We describe now briefly how Theorem 1.1 is applied in [6]. We consider a society consisting of n individuals, which is faced with a finite set of alternatives. A *profile* is a list of n linear orders, which we think of as the list of preferences of each individual. A social choice function is a function which given such profile, yields an asymmetric relation on the alternatives. We think of a social choice function as a method with which the society makes a decision between each two alternatives, based on the preferences of each individual. The social choice (i.e. the value of the social choice function) is called rational if it is an order relation on the alternatives. Social choice functions are assumed to be independent of irrelevant alternatives, i.e. for every two alternatives a and b the society's choice between a and b depends only on the individual preferences between these two alternatives. A social choice function is called a dictatorship if it depends only on the preferences of one particular individual (such an individual is called a dictator. Note that a dictator might be ill fortuned in the sense that the society always chooses the exact opposite of his preference). Finally, a social choice function is called *neutral* if it is invariant under permutations of the alternatives.

In [6] Theorem 1.1 is applied to show that if the outcome of a neutral social choice function for random profiles is almost surely rational then the social choice is approximately a dictatorship:

Theorem 1.3. *There exists an absolute constant K such that the following assertion holds. For every $\epsilon > 0$ and for every neutral social choice function, if the probability that the social choice is irrational is less than ϵ then there is an individual such that for every pair of alternatives the probability that the social choice is different from his choice is less that $K \cdot \epsilon$.*

We refer to [6] for a proof of Theorem 1.3. We also refer to [6] and the references therein for a detailed description of related concepts and results concerning the social choice problem.

We will present two proofs for Theorem 1.1:

2 Proof I

2.1 Reduction to weighted majority

The proof is based on a reduction to the case of weighted majority functions followed by a further reduction to an auxiliary result Proposition 2.2. A simple proof of Proposition 2.2 based on a recent theorem by König, Schütt and Tomczak-Jaegermann (2000) gives best estimates for large values of δ. We present a completely elementary proof which gives best results for small values of δ.

First, we state the following simple proposition:

Proposition 2.1. *Let $t_1, t_2, \dots, t_n$ be real numbers such that $\sum t_i^2 = p > 0$. Put $t = (t_1, t_2, \dots, t_n)$ and $T = \sum t_i$. For a Boolean function f let $U_t(f) = \sum \widehat{f}(\{i\}) t_i$. Then $U_t(f)$ attains its maximum value when f is the following weighted majority function: $f(S) = 1$ iff $\sum_{i \in S} t_i < T/2$.*

Proof: The proof follows at once from the following simple relation:

$$U_t(f) = \sum_{S \subset [n]} 2^{-n} f(S) \left(\sum_{i \notin S} t_i - \sum_{i \in S} t_i \right) = \sum_{S \subset [n]} 2^{-n} \left(T - 2 \sum_{i \in S} t_i \right) f(S).$$

In order to maximize $U_t(f)$ we need to set $f(S) = 1$ when $T - 2\sum_{i \in S} t_i$ is positive and $f(S) = 0$ when $T - 2\sum_{i \in S} t_i$ is negative. □

It follows from Proposition 2.1 that if g is an arbitrary Boolean function then $\sum_{i=1}^{n} \widehat{f}(\{i\})\widehat{g}(\{i\})$ attained is maximum when f is a weighted majority function.

Remark: Proposition 2.1 can be extended to identify the Boolean functions with maximum inner product with $\sum \alpha_S u_S$. This may have some further applications.

We now return to the proof of Theorem 1.1. Let $t_i = \widehat{f}(\{i\})$, $i = 1, 2, \ldots, n$ and $t = (t_1, t_2, \ldots, t_n)$. Consider now the vector $\mathcal{T} = (\mathcal{T}_S : S \subset [n])$ of length 2^n whose entries are given by

$$\mathcal{T}_S = \left(\sum_{i \notin S} t_i - \sum_{i \in S} t_i \right).$$

Note that

$$\|\mathcal{T}\|_2^2 = 2^{-n} \sum_{S \subset [n]} \left(\sum_{i \in S} t_i - \sum_{i \notin S} t_i \right)^2 = \sum t_i^2 \leq p - p^2.$$

Note also that

$$U_t(f) \leq \frac{1}{2} \sum_{S \subset [n]} 2^{-n} \left| \sum_{i \notin S} t_i - \sum_{i \in S} t_i \right| = \|\mathcal{T}\|_1 \leq \|\mathcal{T}\|_2 =$$

$$= \frac{1}{2} \left(\sum_{S \subset [n]} 2^{-n} \left(\sum_{i \notin S} t_i - \sum_{i \in S} t_i \right)^2 \right)^{1/2} \leq \frac{\sqrt{p - p^2}}{2}.$$

Here is a slight correction to the journal version: Using Guy Kindler's argument which is presented on page 8 we may assume that $p = 1/2$, and hence... if

$$U_t(f) \geq (1 - \delta)(p - p^2)$$

then the 2-norm of the vector $\mathcal{T}$ is close to its 1-norm.

Recall the following elementary fact: If $v = (a_1, a_2, \ldots, a_n)$ is a vector (say positive) such that $\|v\|_2 = 1$ and $\|v\|_1 > 1 - \delta$, then the variance of $a_1, \ldots, a_n$ is at most 2δ.

It follows that the vector $\mathcal{T}$ is at a distance of at most 2δ from a vector $\mathcal{S}$ such that the absolute value of its coordinates are the same.

At this point it remains to show that this can only occur if for the original vector t all coordinates but one are close to zero.

Proposition 2.2. *Let $v = (a_1, a_2, ..., a_n)$ be a unit vector of length n and suppose that $a_1 \geq a_2 \geq a_3 \ldots \geq 0$. Consider all the 2^n sums $S(\varepsilon) = \varepsilon_1 a_2 + \varepsilon_2 a_2 + ... + \varepsilon_n a_n$ where ε_i is +1 or -1. If the variance of $|S(\varepsilon)|$ is smaller than δ, then $a_1 > 1 - K'\delta$.*

2.2 Proposition 2.2: first proof

We begin by giving a short proof of Proposition 2.2 which is based on a recent result of König, Schütt and Tomczak-Jaegermann (2000). Later we present a self-contained proof of the proposition which gives a better estimate for K' for small δ. König, Schütt and Tomczak-Jaegermann's theorem can be stated as follows:

Theorem 2.3. *Let $a_1, a_2, \ldots, a_n$ be real numbers and let E be the expected value of the 2^n expressions $|\varepsilon_1 a_1 + \varepsilon_2 a_2 + \cdots + \varepsilon_n a_n|$ where each ε_i is +1 or -1. Then*

$$\left| E - \sqrt{\frac{2}{\pi}} \sqrt{\sum_{i=1}^{n} a_i^2} \right| \leq \left(1 - \sqrt{\frac{2}{\pi}} \right) \max_{1 \leq i \leq n} |a_i|. \tag{2.1}$$

Recall that the expectation of $S(\varepsilon)^2$ is 1 and therefore the variance V of $|S(\varepsilon)|$ is $(1 - E^2) = (1 - E)(1 + E)$. Using Theorem 2.3 we obtain

$$1 - E = 1 - \sqrt{\frac{2}{\pi}} + \sqrt{\frac{2}{\pi}} - E \geq \left(1 - \sqrt{\frac{2}{\pi}} \right) (1 - a_1).$$

This completes the proof of Proposition 2.2.

This result can be further improved as follows: By Szarek's inequality (Khinchine's inequality with best constant. See Szarek (1976) or Latała and Oleszkiewicz (1994) for a short proof), $E \geq 1/\sqrt{2}$. On the other hand, if we define $X = a_2 \varepsilon_2 + \ldots + a_n \varepsilon_n$, then since X is symmetric,

$$E = \mathbb{E}|a_1 \varepsilon_1 + X| = \mathbb{E} \left(\frac{|a_1 \varepsilon_1 + X| + |a_1 \varepsilon_1 - X|}{2} \right) \geq a_1.$$

Hence,

$$V \geq \left(1 - \sqrt{\frac{2}{\pi}} \right) (1 - a_1) \left(1 + \max \left\{ \frac{1}{\sqrt{2}}, a_1 \right\} \right) \geq$$

$$\geq \left(1 - \sqrt{\frac{2}{\pi}} \right) \max \left\{ (1 - a_1^2), \left(1 + \frac{1}{\sqrt{2}} \right) (1 - a_1) \right\},$$

and this yields our theorem with the best constants that we managed to obtain for large δ.

Remark: It is worth noting that Theorem 2.3 implies that $Var(S^2) \sim (\mathbb{E}S^2)Var(|S|)$ where the constants of equivalence are independent of n and $a_1, \ldots, a_n$ (this can be viewed as a second-order Khinchine inequality and it would be interesting to derive similar statements for other powers of S). In order to prove this, note that the inequality $Var(X^2) \geq (\mathbb{E}X^2)Var(X)$ holds for any non-negative random variable X with finite fourth moment. Indeed, we can assume that $\mathbb{E}X^2 = 1$ and by Hölders inequality:

$$1 = \mathbb{E}X^2 = \mathbb{E}X^{\frac{2}{3}+\frac{1}{3}\cdot 4} \leq (\mathbb{E}X)^{2/3}(\mathbb{E}X^4)^{1/3}.$$

Hence, $\mathbb{E}X^4 \geq 1/(\mathbb{E}X)^2$ so that

$$(\mathbb{E}X)^2 + \mathbb{E}X^4 \geq (\mathbb{E}X)^2 + \frac{1}{(\mathbb{E}X)^2} =$$

$$= 2 + \left(\mathbb{E}X - \frac{1}{\mathbb{E}X}\right)^2 \geq 2,$$

which gives that $\mathbb{E}X^4 - 1 \geq 1 - (\mathbb{E}X)^2$ as required. To prove the reverse inequality assume that $\mathbb{E}S^2 = 1$, in which case:

$$Var(S^2) = \mathbb{E}\left(\sum_{i=1}^{n} a_i\varepsilon_i\right)^4 - 1 = \sum_{i=1}^{n} a_i^4 + 6\sum_{i\neq j} a_i^2 a_j^2 - 1 =$$

$$= 5\left(1 - \sum_{i=1}^{n} a_i^4\right) \leq 5(1 - a_1^4) \leq 10(1 - a_1^2).$$

On the other hand, we have seen that $Var(|S|) \geq (1 - \sqrt{2/\pi})(1 - a_1^2)$ so that:

$$Var(S^2) \leq \frac{10}{1 - \sqrt{2/\pi}} \cdot (\mathbb{E}S^2)Var(|S|).$$

We do not know what the best constants are in the above inequalities.

2.3 Proposition 2.2: An elementary proof

We now give an elementary proof with better estimates for K' as long as δ is sufficiently small. We continue to use the notation $X = a_2\varepsilon_2 + \ldots + a_n\varepsilon_n$. We begin by noting that the following identity holds:

$$E = a_1 + \mathbb{E}(|X| - a_1)\mathbf{1}_{\{|X|\geq a_1\}}.$$

Indeed, using the fact that X and ε_1 are independent and the fact that X is symmetric, we obtain:

$$\begin{aligned}
E = \mathbb{E}|a_1\varepsilon_1 + X| = \frac{1}{2}\mathbb{E}|X + a_1| + \frac{1}{2}\mathbb{E}|X - a_1| = \mathbb{E}|X - a_1| = \\
= \mathbb{E}(X - a_1)\mathbf{1}_{\{X \geq a_1\}} + \mathbb{E}(a_1 - X)\mathbf{1}_{\{X < a_1\}} = \\
= \mathbb{E}|X|\mathbf{1}_{\{|X| \geq a_1\}} + a_1 P(|X| < a_1) = \\
= a_1 + \mathbb{E}(|X| - a_1)\mathbf{1}_{\{|X| \geq a_1\}}.
\end{aligned}$$

By the Chernoff bound, for every $u > 0$:

$$\begin{aligned}
P(|X| \geq u) = P\left(\frac{|X|}{\sqrt{1 - a_1^2}} \geq \frac{u}{\sqrt{1 - a_1^2}}\right) \leq \\
\leq 2\exp\left(-\frac{1}{4} \cdot \frac{u^2}{1 - a_1^2}\right).
\end{aligned}$$

Now,

$$\begin{aligned}
E \leq a_1 + \mathbb{E}|X|\mathbf{1}_{\{|X| \geq a_1\}} = \\
= a_1 + \int_{a_1}^{\infty} P(|X| \geq u)du \leq \\
\leq a_1 + 2\int_{a_1}^{\infty} \exp\left(-\frac{1}{4} \cdot \frac{u^2}{1 - a_1^2}\right) du = \\
= a_1 + 4\sqrt{1 - a_1^2}\int_{a_1/(2\sqrt{1 - a_1^2})}^{\infty} e^{-v^2} dv \leq \\
\leq a_1 + 4\sqrt{1 - a_1^2}\int_{a_1/(2\sqrt{1 - a_1^2})}^{\infty} \frac{v}{a_1/(2\sqrt{1 - a_1^2})} e^{-v^2} dv = \\
= a_1 + 4 \cdot \frac{1 - a_1^2}{a_1} \exp\left(-\frac{1}{4} \cdot \frac{a_1^2}{1 - a_1^2}\right).
\end{aligned}$$

Finally:

$$\delta \geq V = 1 - E^2 \geq 1 - a_1^2\left[1 + \frac{4(1 - a_1^2)}{a_1^2} \exp\left(-\frac{1}{4} \cdot \frac{a_1^2}{1 - a_1^2}\right)\right]^2.$$

This directly implies that:

$$a_1 \geq 1 - \frac{\delta}{2} - O(\delta^2).$$

□.

3 Proof II

This proof relies on the Bonamie-Beckner inequality.

Lemma 3.1. *Let* $f : \{0,1\}^n \to \{0,1\}$ *. Assume that*

$$\sum_{|T|>1} \widehat{f}^2(T) = \epsilon.$$

Then there exists i *and* D_i*, which is a linear function of* x_i*, such that*

$$\|f - D_i\|_2^2 < \left(1 + \frac{17496}{1/4 - \epsilon}\right)\epsilon.$$

Proof: First, as pointed out to us by Guy Kindler, we may assume without loss of generality that f is balanced, i.e. that $Pr(f = 1) = 1/2$. To see this let $f(x) = f(x_1, \ldots, x_n)$ and define a Boolean function $g(x_1, \ldots, x_n, y)$ as follows:

$$g(x, 0) = f(x), g(x, 1) = 1 - f(1 - x)$$

where $1-x$ is the vector $(1-x_1, \ldots 1-x_n)$. Clearly g is balanced, $\sum_{|S|>1} \widehat{g}^2(S) = \sum_{|S|>1|} hatf^2(S)$ and if $g \approx D_i$ then $f \approx D_i$. Let

$$S = \sum_{|T|\leq 1} \widehat{f}(T) u_T$$

and

$$L = \sum_{|T|>1} \widehat{f}(T) u_T.$$

(S and L stand for small and large) and let

$$R = S^2 - S.$$

We will show that since S is close to being Boolean, R is typically close to 0, and we will deduce some information on the Fourier coefficients of f. It is straightforward to compute the Fourier coefficients of R: First note that from the orthogonality of L and S we have

$$E(S^2) + E(L^2) = E(f^2) = 1/2.$$

Hence

$$E(S^2) = 1/2 - \epsilon$$

and

$$E(S) = 1/2.$$

Therefore,

$$\widehat{R}(\emptyset) = E(R) = -\epsilon.$$

Next, for any i $\widehat{R}(i) = 0$. Finally, for $i \neq j$

$$\widehat{R}(ij) = 2\widehat{f}(i)\widehat{f}(j).$$

This yields

$$R = -\epsilon + 2\sum \widehat{f}(i)\widehat{f}(j)u_{ij}.$$

Lemma 3.2.

$$E(R^2) \leq 8748\epsilon.$$

Corollary 3.3. *There exists i such that* $\widehat{f}^2(i) \geq 1/4 - \left(1 + \frac{17496}{(1/4-\epsilon)}\right)$.

Clearly this corollary implies Lemma 3.1: We know that $\widehat{f}(\emptyset) = 1/2$ and that $\sum \widehat{f}(T)^2 = 1/2$, hence

$$\|f - 1/2 + \widehat{f}(i)u_i\|_2^2 = \sum_{T \neq i, \emptyset} \widehat{f}^2(T) \leq \left(1 + \frac{17496}{1/4 - \epsilon}\right).$$

Proof of corollary: Since $E(R^2) = \sum \widehat{R}(T)^2$ we have by summing only over the sets $\{ij\}$

$$\sum \widehat{f}^2(i)\widehat{f}^2(j) \leq 8748\epsilon.$$

On the other hand, since

$$E(S^2) = \widehat{f}(\emptyset)^2 + \sum \widehat{f}(i)^2 = 1/2 - \epsilon$$

we have

$$\sum \widehat{f}(i)^2 = 1/4 - \epsilon.$$

Hence, if $\max \widehat{f}^2(i) = \delta$

$$(1/4 - \epsilon)^2 \leq 2 \cdot 8748\epsilon + (1/4 - \epsilon)\delta$$

which gives the desired bound on δ.

□

Proof of Lemma 3.2 : The proof consists of two parts: first we will show that typically R obtains values close to 0. Then we will use a hypercontractive estimate due to Beckner and Bonami to bound higher moments of R in terms of its second moment showing that its tail decays fast enough.

Lemma 3.4. *Let $0 < \alpha < 1/4$ be a constant to be chosen later. Let*

$$p = Prob(|R| > \alpha).$$

Then

$$p \leq \frac{16\epsilon}{\alpha^2}.$$

We defer the proof of this lemma for the moment.

Lemma 3.5.

$$E(R^2) \leq \frac{\alpha^2}{1 - 36\sqrt{\epsilon}/\alpha}.$$

Choosing the optimal value of α (which is $54\sqrt{\epsilon}$) immediately proves Lemma 3.2. So, to finish the proof we now present the proofs of Lemmas 3.4 and 3.5.

Proof of Lemma 3.4: Recall that $S = f - L$ and that $E(L^2) = \epsilon$. Using the fact that $f^2 = f$ yields

$$R = (f - L)^2 - (f - L) = L^2 + L(1 - 2f).$$

A simple analysis of the different possible cases shows that if $|L| < \alpha/4$, then $|R| < \alpha$. Hence by Markov's inequality

$$Pr[|R| > \alpha] \leq Pr[L^2 > \alpha^2/16] \leq \frac{16\epsilon}{\alpha^2}.$$

□

Proof of Lemma 3.5: For convenience of notation let $X = E(R^2)$ and $Y = E(R^4)$. We now need some information on the relative values of X and Y. To this end we use a powerful hypercontractive estimate proven independently by Beckner and Bonami.

Lemma 3.6 (Beckner (1975), Bonami (1970)). *Let $f : \{0,1\}^n \to \mathbb{R}$ be a function that is a linear combination of $\{u_T : |T| \leq k\}$. Let $p > 2$. Then*

$$\|f\|_p \leq (\sqrt{p-1})^k \|f\|_2.$$

In our particular case, taking $p = 4$ and $k = 2$ gives

$$Y \le 81X^2.$$

Using this we obtain

$$\begin{aligned} X = E(R^2) = (1-p)E(R^2|R^2 \le \alpha^2) + pE(R^2|R^2 > \alpha^2) \le \\ (1-p)\alpha^2 + p\sqrt{E(R^4|R^2 > \alpha^2)} \le \\ (1-p)\alpha^2 + p\sqrt{\frac{Y}{p}} \le \\ (1-p)\alpha^2 + \sqrt{p}9X \le \\ \alpha^2 + \frac{4\sqrt{\epsilon}}{\alpha}9X. \end{aligned}$$

This yields

$$X \le \frac{\alpha^2}{1 - 36\sqrt{\epsilon}/\alpha}.$$

□

4 Acknowledgments

We wish to thank Guy Kindler for various remarks.

References

[1] Beckner, W. (1975) Inequalities in Fourier analysis, *Annals of Math.* **102**, 159–182.

[2] Bonami, A. (1970), Etude des coefficients Fourier des fonctiones de $L^p(G)$, *Ann. Inst. Fourier* **20**, 335–402.

[3] Friedgut, E. (1998), Boolean functions with low average sensitivity, *Combinatorica* 18, 27–35.

[4] Hart, S., A note on the edges of the n-cube. Discrete Math. 14 (1976), no. , 157–163.

[5] Kahn, J., G. Kalai and N. Linial (1988), The influence of variables on boolean functions, *Proc. 29-th Ann. Symp. on Foundations of Comp. Sci.*, (1988), 68–80.

[6] Kalai, G.(2001), A Fourier-Theoretic Perspective for the Condorcet Paradox and Arrow's theorem, Adv. Appl. Math., this issue.

[7] König, H., C. Schütt, N. Tomczak-Jaegermann (1999), Projection constants of symmetric spaces and variants of Khintchine's inequality. *J. Reine Angew. Math.* **511**, 1–42.

[8] Latała, R. and K. Oleszkiewicz (1994), On the best constant in the Khinchin-Kahane inequality. *Studia Math.* **109** (1994), no. 1, 101-104.

[9] Szarek, S. J. (1976), On the best constant in the Khinchin inequality. *Studia Math.* **58**, no. 2, 197-208.