THEORY OF COMPUTING, Volume 11 (18), 2015, pp. 445–469
www.theoryofcomputing.org

SPECIAL ISSUE: ANALYSIS OF BOOLEAN FUNCTIONS

On Some Extensions of the FKN Theorem

Jacek Jendrej   Krzysztof Oleszkiewicz*   Jakub O. Wojtaszczyk

Received January 19, 2013; Revised September 19, 2015; Published December 29, 2015

Abstract: Let $S = a_1 r_1 + a_2 r_2 + \cdots + a_n r_n$ be a weighted Rademacher sum. Friedgut, Kalai, and Naor have shown that if $\mathrm{Var}(|S|)$ is much smaller than $\mathrm{Var}(S)$, then the sum is largely determined by one of the summands. We provide a simple and elementary proof of this result, strengthen it, and extend it in various ways to a more general setting.

ACM Classification: G.3

AMS Classification: 60E15, 42C10

Key words and phrases: independent random variables, Rademacher variables, absolute value variation, Fourier expansion, Irit Dinur PCP proof

1 Introduction

Consider a family of independent random variables $(X_i)_{i=1}^n$. It is easy to prove that if the distribution of their sum is supported on a set of cardinality 2, then all the X_i's but one are constant almost surely. In our paper we investigate stability of this phenomenon. Namely, we prove that if the distribution of the sum is concentrated around a two-point set, then there exists $k \in \{1, 2, \ldots, n\}$ such that $\sum_{i:i \neq k} X_i$ is concentrated around some point. We provide various strict quantitative variants of this heuristic statement. One of them is the following theorem:

Theorem 1.1. *Let $\tau \geq 1$. Let $(X_i)_{i=1}^n$ be a sequence of independent square-integrable random variables. Assume that $\mathrm{Var}(X_i) \leq \tau \cdot (\mathbb{E}|X_i - \mathbb{E}X_i|)^2$ for $1 \leq i \leq n$. Then for some $k \in \{1, 2, \ldots, n\}$ we have*

$$\mathrm{Var}\Big(\sum_{i \leq n: i \neq k} X_i\Big) \leq K(\tau) \cdot \inf_{x \in \mathbb{R}} \mathrm{Var}\Big(\Big|x + \sum_{i \leq n} X_i\Big|\Big),$$

where $K(\tau)$ is a constant which depends only on τ.

*Partially supported by Polish MNiSW grant N N201 397437.

 DOI: 10.4086/toc.2015.v011a018

Note that always $\mathrm{Var}(X_i) - (\mathbb{E}|X_i - \mathbb{E}X_i|)^2 = \mathrm{Var}(|X_i - \mathbb{E}X_i|) \geq 0$, with equality only when X_i is either constant almost surely, or uniformly distributed on a two-point set. For such X_i's, the comparison of moments assumption is thus satisfied with $\tau = 1$. In a sense, the closer τ is to one, the more $\sum_i X_i$ must resemble a (shifted) weighted Rademacher sum.

This result for weighted Rademacher sums (i. e., in the case when for each i the random variable X_i is symmetric and takes values $\pm a_i$, where $a_1, a_2, \ldots, a_n$ are some real numbers) was proved in [4] by E. Friedgut, G. Kalai, and A. Naor, and was a part of the proof of their theorem on Boolean functions on the discrete cube with Fourier coefficients concentrated at the first two levels.

FKN Theorem ([4], Theorem 1.1)**.** There exists an absolute constant K such that for any

$$f : \{-1,1\}^n \to \{-1,1\} \qquad \text{and} \qquad \rho = \sqrt{\sum_{|A|\geq 2} |\hat{f}(A)|^2}$$

one of the following holds:

$$\begin{aligned} &\|f(x_1,x_2,\ldots,x_n) - 1\|_2 \leq K\rho, \qquad \text{or} \\ &\|f(x_1,x_2,\ldots,x_n) + 1\|_2 \leq K\rho, \qquad \text{or} \\ &\|f(x_1,x_2,\ldots,x_n) - x_i\|_2 \leq K\rho \text{ for some } i \in \{1,2,\ldots,n\}, \qquad \text{or} \\ &\|f(x_1,x_2,\ldots,x_n) + x_i\|_2 \leq K\rho \text{ for some } i \in \{1,2,\ldots,n\}. \end{aligned} \tag{1.1}$$

Remark 1.2. Note that (1.1) can be equivalently formulated as follows. $\mathbb{P}(f \neq 1) \leq K^2\rho^2/4$ or $\mathbb{P}(f \neq -1) \leq K^2\rho^2/4$, or $\mathbb{P}(f(x) \neq x_i) \leq K^2\rho^2/4$ for some i, or $\mathbb{P}(f(x) \neq -x_i) \leq K^2\rho^2/4$ for some i.

They gave two proofs of the result concerning Rademacher sums. One was a direct application of a theorem of König et al. [7]. The other used a more elementary approach (Chernoff's inequality), but contained an omission—it worked only under the additional assumption that we already know that $\mathrm{Var}(X_k) \geq C\,\mathrm{Var}(\sum_i X_i)$ for some C close to 1. This minor gap is well known by now as well as some ways to fix it, for example by the use of the Berry–Esseen theorem—in fact, it has been fixed already by Kindler and Safra [6], whose proof also yielded better asymptotic estimates than [4]. (Although [6] was not formally published, as far as we know, it was widely circulated; the proof appeared also in Kindler's Ph. D. thesis [5].) The FKN theorem is a direct application of the above variance bound for Rademacher sums. It was originally devised for applications in discrete combinatorics and social science, but turned out to be useful also in theoretical computer science. In particular, the theorem is used in analyzing the Long Code Test in the celebrated expander proof of the PCP theorem by Irit Dinur ([3]). With that in mind, we hope that our easy, self-contained proof will simplify understanding of the PCP theorem's background. Hence we set out to give an elementary proof of Theorem 1.1 for weighted Rademacher sums which does not refer to intricate results such as the Berry–Esseen inequality, [7], [2] or [1] (a proof based on the Bonami–Beckner hypercontractive bounds was also known).

We think it is also interesting (although not very surprising) that the inequality still holds if we replace the Rademacher variables by variables satisfying a moment comparison condition. Note that, in contrast to the weighted Rademacher setting described above, in our results the sums do not need to be linear combinations of an i. i. d. sequence (however, in the discrete cube setting we actually prove a stronger, and essentially optimal, FKN type estimate in Theorem 5.3 by the use of yet another method; Ryan O'Donnell obtained the same bound independently by a slightly different approach—see [11, Theorem 5.33]).

We also provide the following analogous result for symmetric random variables with no additional assumption about moment comparison.

Theorem 1.3. *Let $(X_i)_{i=1}^n$ be a sequence of independent symmetric square-integrable random variables. Then for some $k \in \{1, 2, \ldots, n\}$ we have*

$$\operatorname{Var}\Big(\sum_{i \le n: i \ne k} X_i\Big) \le C \cdot \inf_{x \in \mathbb{R}} \operatorname{Var}\Big(\Big|x + \sum_{i \le n} X_i\Big|\Big),$$

where C is a universal constant. The result holds true with $C = (7+\sqrt{17})/2$.

For the sake of clarity, we start by showing in Section 2 that if a sum of independent random vectors is concentrated around a two-point set, then by removing just one term we may make the sum of remaining vectors concentrate around a single point. Then, in Section 3, we demonstrate how to use this observation in the real-valued case. However, our results and methods can be quite easily adapted to a Banach space setting, with concentration around a finite set of points, which leads to some nice geometric considerations. We present them in Section 6. Only very basic knowledge of Banach space theory is needed, which can be found, e. g., in [15].

Since we tried to make our proofs as transparent and "low-tech" as possible, in many cases our estimates can be easily improved upon some natural optimization.

Readers interested only in the Rademacher case may find it useful to restrict their attention to Section 4, in which Theorem 1.3 is proved, and first two subsections of Section 5, in which the strengthening of the FKN Theorem is described.

2 Splitting of the sum

We begin by analyzing the concentration in terms of probability rather than variance. In what follows, we denote by μ_Z the distribution of a random variable Z. Readers not comfortable with the Banach space formulation may simply replace V by $\mathbb{R}$, and $\|\cdot\|$ by $|\cdot|$.

Lemma 2.1. *Let X, Y be independent random variables with values in a real separable Banach space V. Assume that for $\delta \ge 0$ and $a, b \in V$ we have*

$$\mathbb{P}\left(\|X+Y-a\| \le \varepsilon \text{ or } \|X+Y-b\| \le \varepsilon\right) \ge 1-\delta, \tag{2.1}$$

where $0 \le \varepsilon < \|b-a\|/6$. Then there exists some vector $c \in V$ such that

$$\mathbb{P}\left(\|X-c\| > \varepsilon\right) \le \sqrt{\delta}+\delta \text{ or } \mathbb{P}\left(\|Y-c\| > \varepsilon\right) \le \sqrt{\delta}+\delta.$$

Proof. Let $v = b-a$ and

$$A_y = \{x \in V : \|x+y-a\| \le \varepsilon\} \cup \{x \in V : \|x+y-b\| \le \varepsilon\}$$

for $y \in V$. Then from (2.1) and independence of variables by the Fubini theorem we get

$$1-\delta \leq \int_{V\times V} \mathbf{1}_{A_y}(x)\mathrm{d}\mu_{(X,Y)}(x,y) = \int_V \mu_X(A_y)\,\mathrm{d}\mu_Y(y)\,,$$

so in particular $\mu_X(A_y) \geq 1-\delta$ for some $y \in V$, which means

$$\exists c_1, c_2 \in V, c_2 - c_1 = v : \mathbb{P}(\|X-c_1\| \leq \varepsilon \text{ or } \|X-c_2\| \leq \varepsilon) \geq 1-\delta\,.$$

Similarly we prove that

$$\exists d_1, d_2 \in V, d_2 - d_1 = v : \mathbb{P}(\|Y-d_1\| \leq \varepsilon \text{ or } \|Y-d_2\| \leq \varepsilon) \geq 1-\delta\,.$$

Let $\alpha = \mathbb{P}(\|X-c_1\| \leq \varepsilon)$, $\beta = \mathbb{P}(\|X-c_2\| \leq \varepsilon)$, $\gamma = \mathbb{P}(\|Y-d_1\| \leq \varepsilon)$, $\eta = \mathbb{P}(\|Y-d_2\| \leq \varepsilon)$. Thus

$$\mathbb{P}(\|(X+Y)-(c_1+d_1)\| \leq 2\varepsilon) \geq \alpha\gamma,$$
$$\mathbb{P}(\|(X+Y)-(c_2+d_2)\| \leq 2\varepsilon) \geq \beta\eta\,.$$

Let $\bar{B}(x,r)$ denote the closed ball with center x and radius r. If $\alpha,\gamma,\beta,\eta > \sqrt{\delta}$, then $\alpha\gamma, \beta\eta > \delta$, and (2.1) would imply that

$$\bar{B}(c_1+d_1, 2\varepsilon) \cap (\bar{B}(a,\varepsilon) \cup \bar{B}(b,\varepsilon)) \neq \emptyset \quad \text{and} \quad \bar{B}(c_2+d_2, 2\varepsilon) \cap (\bar{B}(a,\varepsilon) \cup \bar{B}(b,\varepsilon)) \neq \emptyset\,.$$

Since $(c_2+d_2)-(c_1+d_1) = (c_2-c_1)+(d_2-d_1) = 2v$, the diameter of $\bar{B}(a,\varepsilon) \cup \bar{B}(b,\varepsilon)$ would satisfy the inequalities

$$2\|v\| - 4\varepsilon \leq \operatorname{diam}(\bar{B}(a,\varepsilon) \cup \bar{B}(b,\varepsilon)) \leq \|v\| + 2\varepsilon\,,$$

contradicting the assumption $\varepsilon < \|v\|/6$.

Without loss of generality we may therefore assume that $\alpha \leq \sqrt{\delta}$. Since $\alpha+\beta \geq 1-\delta$, this implies $\beta \geq 1-\sqrt{\delta}-\delta$, i. e., $\mathbb{P}(\|X-c_2\| > \varepsilon) \leq \sqrt{\delta}+\delta$. □

Lemma 2.2. *Let $X_1, X_2, \ldots, X_n$ be independent random variables with values in a separable Banach space V. Let $S = X_1 + \cdots + X_n$, $S_i = S - X_i$ for $1 \leq i \leq n$. Assume that for $\delta \geq 0$ and some vectors $a, b \in V$ we have*

$$\mathbb{P}(\|S-a\| \leq \varepsilon \text{ or } \|S-b\| \leq \varepsilon) \geq 1-\delta\,, \tag{2.2}$$

where $\varepsilon < \|b-a\|/6$. Then there exist $k \in \{1,\ldots,n\}$ and $c \in V$ such that

$$\mathbb{P}(\|S_k - c\| > 2\varepsilon) \leq 2(\sqrt{\delta}+\delta) = O(\sqrt{\delta}) \text{ as } \delta \to 0^+\,.$$

Additionally, if $\delta < 1/9$, then

$$\mathbb{P}(\|X_k-(a-c)\| > 3\varepsilon \text{ and } \|X_k-(b-c)\| > 3\varepsilon) \leq \delta/(1-2\sqrt{\delta}-2\delta) = O(\delta) \text{ as } \delta \to 0^+.$$

Proof. Let I be a minimal (in the sense of inclusion) subset of $\{1,\dots,n\}$ such that

$$\forall c \in V : \mathbb{P}\left(\left\|\sum_{i\in I} X_i - c\right\| > \varepsilon\right) > \sqrt{\delta}+\delta$$

(if there is no such I, there is nothing to prove). Of course $I \neq \emptyset$. Let $k \in I$. We have $S = \sum_{i\in I} X_i + \sum_{i\notin I} X_i$, and the two sums are obviously independent, so by our assumption about I, (2.2) and Lemma 2.1, for some $c_1 \in V$ we get

$$\mathbb{P}\left(\left\|\sum_{i\notin I} X_i - c_1\right\| > \varepsilon\right) \leq \sqrt{\delta}+\delta\,. \tag{2.3}$$

But I was minimal, so for some $c_2 \in V$

$$\mathbb{P}\left(\left\|\sum_{i\in I\setminus\{k\}} X_i - c_2\right\| > \varepsilon\right) \leq \sqrt{\delta}+\delta\,. \tag{2.4}$$

$S_k = \sum_{i\notin I} X_i + \sum_{i\in I\setminus\{k\}} X_i$, so that (2.3), (2.4) and the triangle inequality yield

$$\mathbb{P}\left(\|S_k - (c_1+c_2)\| > 2\varepsilon\right) \leq 2(\sqrt{\delta}+\delta)\,.$$

The second assertion of the lemma follows easily upon recalling that S_k and X_k are independent and $S = S_k + X_k$:

$$\begin{aligned}(1-2\sqrt{\delta}-2\delta)\cdot \mathbb{P}\big(&\|X_k-(a-c)\| > 3\varepsilon \text{ and } \|X_k-(b-c)\| > 3\varepsilon\big)\\ &\leq \mathbb{P}\big(\|S_k - c\| \leq 2\varepsilon \text{ and } \|X_k-(a-c)\| > 3\varepsilon \text{ and } \|X_k-(b-c)\| > 3\varepsilon\big)\\ &\leq \mathbb{P}\big(\|S-a\| > \varepsilon \text{ and } \|S-b\| > \varepsilon\big) \leq \delta\,.\end{aligned}$$

For $\delta \geq 1/9$ we have $1-2\sqrt{\delta}-2\delta \leq \delta$, which makes the arising probability bound trivial. □

Remark 2.3. Both bounds are of optimal order for $\delta \to 0^+$ even for $V=\mathbb{R}$, as indicated by the following example. Fix $a=0$, $b=1$, $\varepsilon = 1/7$ and some $n \geq 2$. Let $X_i \sim \mathrm{Pois}(\sqrt{\delta}/n)$, so that $S \sim \mathrm{Pois}(\sqrt{\delta})$ and

$$\mathbb{P}(S\in\{0,1\}) = e^{-\sqrt{\delta}}(1+\sqrt{\delta}) \geq 1-\delta\,.$$

Hence the assumptions of Lemma 2.2 are satisfied. Since for every $k \leq n$ there is

$$S_k \sim \mathrm{Pois}\left(\frac{n-1}{n}\sqrt{\delta}\right),$$

we have $\mathbb{P}(S_k=0)/\sqrt{\delta} \to \infty$ and $\mathbb{P}(S_k=1)/\sqrt{\delta} \to (n-1)/n$ as $\delta \to 0^+$ which shows that the $O(\sqrt{\delta})$ bound cannot be improved. Also, for every $k \leq n$ we have

$$\mathbb{P}(X_k=0)/\delta \to \infty\,, \quad \mathbb{P}(X_k=1)/\delta \to \infty\,, \quad \text{and} \quad \mathbb{P}(X_k=2)/\delta \to \frac{1}{2n^2}$$

as $\delta \to 0^+$. Hence, for δ small enough, for every set A which is a union of two intervals of length $6/7$ each, there is

$$\mathbb{P}(X_k \notin A) \geq \frac{\delta}{3n^2}$$

which proves the optimality of the $O(\delta)$ bound.

3 Proof of Theorem 1.1 (comparison of moments assumption)

We will now show how to use the facts from the previous paragraph to give a proof of Theorem 1.1. Concentration bounds in terms of probability will be translated into statements about variances by the use of a Paley–Zygmund type inequality. We need a few simple and standard lemmas.

Lemma 3.1 (Khinchine inequality)**.** *Let $r_1, r_2, \ldots$ be independent symmetric ± 1 random variables. There exists a universal constant κ such that for every $a_1, a_2, \ldots, a_m \in \mathbb{R}$ there is*

$$\kappa \cdot \mathbb{E}\left|\sum_{i=1}^{m} a_i r_i\right| \geq \left(\sum_{i=1}^{m} a_i^2\right)^{1/2}.$$

Proof. The estimate with the optimal constant $\kappa = \sqrt{2}$ was proved by Szarek, [16] (see [8] for a simpler proof). For the reader's convenience we provide a well-known simple argument which yields $\kappa = \sqrt{3}$. Let $S = \sum_{i=1}^{m} a_i r_i$, so that $\mathbb{E}S^2 = \sum_{i=1}^{m} a_i^2$. Since

$$\mathbb{E}S^4 = \sum_{i=1}^{m} a_i^4 + 6 \sum_{i,j:1\leq i<j\leq m} a_i^2 a_j^2 \leq 3\Big(\sum_{i=1}^{m} a_i^2\Big)\Big(\sum_{j=1}^{m} a_j^2\Big) = 3\left(\mathbb{E}S^2\right)^2,$$

by Hölder's inequality we get

$$\mathbb{E}S^2 = \mathbb{E}\left(|S|^{4/3} \cdot |S|^{2/3}\right) \leq \left(\mathbb{E}S^4\right)^{1/3} (\mathbb{E}|S|)^{2/3} \leq 3^{1/3} \left(\mathbb{E}S^2\right)^{2/3} (\mathbb{E}|S|)^{2/3}.$$

Hence $\sqrt{3}\,\mathbb{E}|S| \geq \left(\mathbb{E}S^2\right)^{1/2} = \left(\sum_{i=1}^{m} a_i^2\right)^{1/2}$. □

Lemma 3.2. *Let $Y_1, Y_2, \ldots, Y_m$ be independent symmetric integrable random variables. Then*

$$\kappa \cdot \mathbb{E}\left|\sum_{i=1}^{m} Y_i\right| \geq \mathbb{E}\left(\sum_{i=1}^{m} Y_i^2\right)^{1/2},$$

where κ is the universal constant from Lemma 3.1.

Proof. Let $(r_i)_{i=1}^m$ be a sequence of independent symmetric ± 1 random variables, independent of $(Y_i)_{i=1}^m$, so that $(Y_i \cdot r_i)_{i=1}^m$ has the same distribution as $(Y_i)_{i=1}^m$. Hence

$$\mathbb{E}\Big|\sum_{i=1}^{m} Y_i\Big| = \mathbb{E}\Big|\sum_{i=1}^{m} Y_i r_i\Big| = \mathbb{E}\mathbb{E}\left(\Big|\sum_{i=1}^{m} Y_i r_i\Big| \,\middle|\, Y_1, Y_2, \ldots, Y_m\right) \geq \kappa^{-1} \cdot \mathbb{E}\Big(\sum_{i=1}^{m} Y_i^2\Big)^{1/2},$$

where we have used Lemma 3.1. □

The following result may be traced back to the work of Marcinkiewicz and Zygmund [9] (see Théorème 2 therein).

Corollary 3.3. *Let $\rho \geq 1$. Let $Y_1, Y_2, \ldots, Y_m$ be independent symmetric square-integrable random variables such that $\mathbb{E}Y_i^2 \leq \rho \cdot (\mathbb{E}|Y_i|)^2$ for $i \leq m$. Set $Z = \sum_{i=1}^{m} Y_i$. Then $\mathbb{E}Z^2 \leq \kappa^2 \rho \cdot (\mathbb{E}|Z|)^2$, where κ is the universal constant from Lemma 3.1.*

Proof. Indeed, by Lemma 3.2 we have

$$\begin{aligned}\kappa\cdot\mathbb{E}|Z| = \kappa\cdot\mathbb{E}\Big|\sum_{i=1}^m Y_i\Big| &\ge \mathbb{E}\Big(\sum_{i=1}^m Y_i^2\Big)^{1/2} = \mathbb{E}\Big\|\big(|Y_1|,|Y_2|,\ldots,|Y_m|\big)\Big\|_{l_2^m} \\ &\ge \Big\|\big(\mathbb{E}|Y_1|,\mathbb{E}|Y_2|,\ldots,\mathbb{E}|Y_m|\big)\Big\|_{l_2^m} = \Big(\sum_{i=1}^m(\mathbb{E}|Y_i|)^2\Big)^{1/2} \\ &\ge \rho^{-1/2}\cdot\Big(\sum_{i=1}^m \mathbb{E}Y_i^2\Big)^{1/2} = \rho^{-1/2}\cdot(\mathbb{E}Z^2)^{1/2},\end{aligned}$$

where we have used Jensen's inequality for the convex function $y\mapsto\|y\|_{l_2^m}$. □

Proof of Theorem 1.1. Obviously, by considering $x+X_1$ instead of X_1 we may reduce our task to proving that

$$\min_{k\le n}\mathrm{Var}\Big(\sum_{i\le n: i\ne k} X_i\Big) \le K(\tau)\cdot\mathrm{Var}\Big(\Big|\sum_{i\le n} X_i\Big|\Big).$$

Let $(X_i')_{i=1}^n$ be an independent copy of $(X_i)_{i=1}^n$. Let $S=\sum_{i=1}^n X_i$ and $S'=\sum_{i=1}^n X_i'$, so that S' is an independent copy of S. Note that random variables $Y_i = X_i - X_i'$ ($i\le n$) are independent and symmetric. By Jensen's inequality,

$$\mathbb{E}|Y_i| = \mathbb{E}\mathbb{E}(|X_i-X_i'|\,|\,X_i) \ge \mathbb{E}|\mathbb{E}(X_i-X_i'|X_i)| = \mathbb{E}|X_i-\mathbb{E}X_i'| = \mathbb{E}|X_i-\mathbb{E}X_i|\,.$$

Since $\mathbb{E}Y_i^2 = 2\mathrm{Var}(X_i)$, we have $\mathbb{E}Y_i^2\le 2\tau\cdot(\mathbb{E}|Y_i|)^2$ for $i=1,2,\ldots,n$. Let $\varepsilon = 18\kappa^2\tau(\mathrm{Var}(|S|))^{1/2}$, where κ is the universal constant from Lemma 3.1. Let $\delta=\kappa^{-4}\tau^{-2}/324$, $a=\mathbb{E}|S|$, and $b=-a=-\mathbb{E}|S|$. We consider two cases:

Case 1. Assume $\varepsilon<|a-b|/6$. By Chebyshev's inequality,

$$\mathbb{P}\big(|S-a|\le\varepsilon \text{ or } |S-b|\le\varepsilon\big)\ge 1-\mathrm{Var}(|S|)\varepsilon^{-2} = 1-\delta\,,$$

so by Lemma 2.2 there exist $c\in\mathbb{R}$ and $k\le n$ such that

$$\mathbb{P}(|S_k-c|\le 2\varepsilon)\ge 1-2\sqrt{\delta}-2\delta\,,$$

where $S_k=\sum_{i\le n:i\ne k}X_i$. Let $Z=\sum_{i\le n:i\ne k}Y_i$. Let $S_k'=\sum_{i\le n:i\ne k}X_i'$, so that S_k' is an independent copy of S_k. Obviously,

$$\mathbb{E}Z^2 = \mathbb{E}(S_k-S_k')^2 = 2\mathrm{Var}(S_k)\,.$$

Note that

$$\begin{aligned}\mathbb{P}(|Z|\le 4\varepsilon) &= \mathbb{P}(|S_k-S_k'|\le 4\varepsilon)\ge\mathbb{P}(|S_k-c|\le 2\varepsilon \text{ and } |S_k'-c|\le 2\varepsilon) \\ &= \mathbb{P}(|S_k-c|\le 2\varepsilon)\cdot\mathbb{P}(|S_k'-c|\le 2\varepsilon)\ge(1-2\sqrt{\delta}-2\delta)^2\ge 1-4\sqrt{\delta}\,,\end{aligned}$$

so that $\mathbb{P}(|Z| > 4\varepsilon) \le 4\sqrt{\delta}$. Now we may apply a Paley–Zygmund type estimate:

$$\begin{aligned}\kappa^{-1}\tau^{-1/2}\cdot(\mathrm{Var}(S_k))^{1/2} &= \kappa^{-1}(2\tau)^{-1/2}(\mathbb{E}Z^2)^{1/2} \le \mathbb{E}|Z| = \mathbb{E}|Z|\mathbf{1}_{|Z|\le 4\varepsilon} + \mathbb{E}|Z|\mathbf{1}_{|Z|>4\varepsilon} \\ &\le 4\varepsilon + (\mathbb{E}Z^2)^{1/2}\cdot(\mathbb{P}(|Z|>4\varepsilon))^{1/2} \le 4\varepsilon + \Big(2\mathrm{Var}(S_k)\cdot 4\sqrt{\delta}\Big)^{1/2} \\ &= 72\kappa^2\tau\cdot(\mathrm{Var}(|S|))^{1/2} + \frac{2}{3}\kappa^{-1}\tau^{-1/2}\cdot(\mathrm{Var}(S_k))^{1/2}\end{aligned}$$

—we have used Corollary 3.3 (with $m = n-1$ and $\rho = 2\tau$) in the first inequality, and the second one follows from the Cauchy–Schwarz inequality. Obvious cancellations yield $\mathrm{Var}(S_k) \le (6\kappa)^6\tau^3\cdot\mathrm{Var}(|S|)$.

Case 2. If $\varepsilon \ge |a-b|/6$, then $\mathbb{E}|S| \le 3\varepsilon$ and

$$\begin{aligned}\mathrm{Var}(S_1) &\le \mathrm{Var}(S) \le \mathbb{E}S^2 = \mathrm{Var}(|S|) + (\mathbb{E}|S|)^2 \\ &\le \mathrm{Var}(|S|) + 9\varepsilon^2 = (1+2976\kappa^4\tau^2)\cdot\mathrm{Var}(|S|) \le (6\kappa)^6\tau^3\cdot\mathrm{Var}(|S|)\end{aligned}$$

because $\kappa,\tau \ge 1$. We have proved the theorem with $K(\tau) = (6\kappa)^6\tau^3$. □

Corollary 3.4. *Let ξ be a square-integrable random variable which is not constant a. s., and let ξ_1, $\xi_2, \ldots$ be its i. i. d. copies. Then there exists a constant K_ξ, depending only on distribution of ξ, such that for any real numbers $a_1, a_2, \ldots, a_n$ there is some $k \le n$ for which*

$$\sum_{i\le n: i\ne k} a_i^2 \le K_\xi \cdot \inf_{x\in\mathbb{R}} \mathrm{Var}\Big(\Big|x + \sum_{i\le n} a_i\xi_i\Big|\Big).$$

Proof. It suffices to apply Theorem 1.1 with $\tau = \mathrm{Var}(\xi)/(\mathbb{E}|\xi - \mathbb{E}\xi|)^2$. Setting $K_\xi = K(\tau)/\mathrm{Var}(\xi)$ ends the proof. □

4 Proof of Theorem 1.3 (symmetric variant)

Now we will prove an analogue of Theorem 1.1 for real symmetric random variables (but with no constrains on moments). It is possible to do it in a similar way as in the proof of Theorem 1.1, i. e., by using Lemma 2.2, but to get better estimates we will adopt another, more direct approach. We will need a lemma.

Lemma 4.1. *Let X and Y be independent square-integrable random variables, at least one of them symmetric. Then*

$$\min\Big(\mathrm{Var}(X), \mathrm{Var}(Y)\Big) \le \frac{7+\sqrt{17}}{4}\cdot\mathrm{Var}(|X+Y|).$$

Proof. Obviously, $\mathbb{E}|X+Y|^2 = \mathbb{E}X^2 + \mathbb{E}Y^2$ because $\mathbb{E}XY = \mathbb{E}X\cdot\mathbb{E}Y = 0$. Since $|X+Y|$ and $|X-Y|$ have the same distribution, there is

$$\mathbb{E}|X+Y| = (\mathbb{E}|X+Y| + \mathbb{E}|X-Y|)/2 = \mathbb{E}(|X+Y|+|X-Y|)/2 = \mathbb{E}\max(|X|,|Y|).$$

Hence

$$\begin{aligned}
\mathrm{Var}(|X+Y|) &= \mathbb{E}(X^2+Y^2) - (\mathbb{E}|X+Y|)^2 \\
&= \mathbb{E}\Big((\max(|X|,|Y|))^2 + (\min(|X|,|Y|))^2 \Big) - (\mathbb{E}\max(|X|,|Y|))^2 \\
&= \mathrm{Var}(\max(|X|,|Y|)) + \mathrm{Var}(\min(|X|,|Y|)) + (\mathbb{E}\min(|X|,|Y|))^2 \\
&\geq \frac{1}{2}\mathrm{Var}\Big(\max(|X|,|Y|) + \min(|X|,|Y|) \Big) + (\mathbb{E}\min(|X|,|Y|))^2 \\
&= \frac{1}{2}\mathrm{Var}(|X|+|Y|) + (\mathbb{E}\min(|X|,|Y|))^2 \\
&= \frac{1}{2}\Big(\mathrm{Var}(|X|) + \mathrm{Var}(|Y|) \Big) + (\mathbb{E}\min(|X|,|Y|))^2 \,.
\end{aligned}$$

We have used the fact that for any square-integrable random variables V and W there is

$$2\mathrm{Var}(V) + 2\mathrm{Var}(W) = \mathrm{Var}(V+W) + \mathrm{Var}(V-W) \geq \mathrm{Var}(V+W)\,.$$

Thus, for $\sigma = (\mathrm{Var}(|X+Y|))^{1/2}$ and $s = \mathrm{Var}(|X|) + \mathrm{Var}(|Y|)$, we have $s \leq 2\sigma^2$ and

$$\mathbb{E}\min(|X|,|Y|) \leq \left(\sigma^2 - \frac{1}{2}s \right)^{1/2}.$$

The identity $a+b-2\min(a,b) = |a-b|$ yields a pointwise bound:

$$\begin{aligned}
|X|+|Y|-2\min(|X|,|Y|) = \Big| |X|-|Y| \Big| &\leq \Big| \mathbb{E}|X| - \mathbb{E}|Y| \Big| + \Big| (|X|-\mathbb{E}|X|) - (|Y|-\mathbb{E}|Y|) \Big| \\
&= \mathbb{E}|X| + \mathbb{E}|Y| - 2\min(\mathbb{E}|X|,\mathbb{E}|Y|) + \Big| (|X|-|Y|) - \mathbb{E}(|X|-|Y|) \Big| \,.
\end{aligned}$$

By taking expectations of both sides we arrive at

$$\begin{aligned}
\min(\mathbb{E}|X|,\mathbb{E}|Y|) &\leq \mathbb{E}\min(|X|,|Y|) + \frac{1}{2}\mathbb{E}\Big| (|X|-|Y|) - \mathbb{E}(|X|-|Y|) \Big| \\
&\leq \mathbb{E}\min(|X|,|Y|) + \frac{1}{2}\Big(\mathrm{Var}(|X|-|Y|) \Big)^{1/2} \leq \left(\sigma^2 - \frac{1}{2}s \right)^{1/2} + \frac{1}{2}\sqrt{s}\,,
\end{aligned}$$

where we have bounded the L^1 norm by the L^2 norm and used the independence of $|X|$ and $|Y|$. Therefore

$$\begin{aligned}
\min(\mathrm{Var}(X),\mathrm{Var}(Y)) &\leq \min(\mathbb{E}X^2,\mathbb{E}Y^2) = \min\Big(\mathrm{Var}(|X|) + (\mathbb{E}|X|)^2, \mathrm{Var}(|Y|) + (\mathbb{E}|Y|)^2 \Big) \\
&\leq s + (\min(\mathbb{E}|X|,\mathbb{E}|Y|))^2 \leq \sigma^2 + \frac{3}{4}s + \left(\sigma^2 s - \frac{1}{2}s^2 \right)^{1/2} \\
&\leq \sigma^2 + \sup_{t\in[0,2\sigma^2]} \left(\frac{3}{4}t + (\sigma^2 t - t^2/2)^{1/2} \right) = \frac{7+\sqrt{17}}{4} \cdot \sigma^2 \,. \qquad \square
\end{aligned}$$

Remark 4.2. An example of X with distribution

$$\frac{1}{8}\delta_{-2} + \frac{3}{4}\delta_0 + \frac{1}{8}\delta_2$$

and ± 1 symmetric Y indicates that the constant $(7+\sqrt{17})/4 \approx 2.78$ in Lemma 4.1 cannot be replaced by any number less than $16/7 \approx 2.29$.

Proof. Proof of Theorem 1.3 We prove the theorem with $C = (7+\sqrt{17})/2 \approx 5.56$. For $x \in \mathbb{R}$ let $\xi_1 = x + X_1$, and $\xi_i = X_i$ for $i \geq 2$. Set $S = \sum_{i=1}^{n} \xi_i$. For $J \subseteq \{1,2,\ldots,n\}$ let $S_J = \sum_{i\in J} \xi_i$. Let I be a minimal (in the sense of inclusion) subset of $[n] = \{1,2,\ldots,n\}$ such that

$$\mathrm{Var}(S_I) > \frac{7+\sqrt{17}}{4} \cdot \mathrm{Var}(|S|)$$

(if no subset satisfies this condition, then the assertion follows trivially). Obviously, $I \neq \emptyset$. Choose any $k \in I$. Certainly,

$$\mathrm{Var}(S_{I\setminus\{k\}}) \leq \frac{7+\sqrt{17}}{4} \cdot \mathrm{Var}(|S|)\,.$$

Since S_I and $S_{[n]\setminus I}$ are independent and at least one of them is symmetric, Lemma 4.1 yields that

$$\min\Big(\mathrm{Var}(S_I), \mathrm{Var}(S_{[n]\setminus I})\Big) \leq \frac{7+\sqrt{17}}{4} \cdot \mathrm{Var}(|S|)\,,$$

so that

$$\mathrm{Var}(S_{[n]\setminus I}) \leq \frac{7+\sqrt{17}}{4} \cdot \mathrm{Var}(|S|)\,.$$

Therefore

$$\mathrm{Var}\Big(\sum_{i\leq n: i\neq k} X_i\Big) = \mathrm{Var}(S_{I\setminus\{k\}}) + \mathrm{Var}(S_{[n]\setminus I}) \leq \frac{7+\sqrt{17}}{2} \cdot \mathrm{Var}(|S|)\,.$$

Thus we have proved that

$$\min_{k\leq n} \mathrm{Var}\Big(\sum_{i\leq n: i\neq k} X_i\Big) \leq \frac{7+\sqrt{17}}{2} \cdot \inf_{x\in\mathbb{R}} \mathrm{Var}\Big(\Big|x + \sum_{i\leq n} X_i\Big|\Big)\,. \qquad \square$$

Remark 4.3. A simple example of $n = 3$ and X_1, X_2, X_3 i. i. d. symmetric ± 1 random variables indicates that the constant C in Theorem 1.3 cannot be less than $8/3 \approx 2.67$ (it suffices to check it for $x = 0$).

5 Harmonic analysis on product spaces

Below, we introduce assumptions and notation which will be used throughout Section 5. This is a natural and convenient setting for harmonic analysis on product spaces, e. g., on the discrete cube with the uniform measure, in which case $(\pi_i)_{i=1}^n$ is the standard Rademacher system. This language will allow us to state FKN type results on product spaces other than the discrete cube.

5.1 Assumptions and notation (A & N)

Let $\xi_1, \xi_2, \ldots, \xi_n$ be independent random variables satisfying $\mathbb{E}\xi_i = 0$ and $\mathbb{E}\xi_i^2 = 1$ for all $i \leq n$. We consider a Hilbert space $L^2 = L^2(\mathbb{R}^n, \mu)$, where $\mu = \mu_{\xi_1} \otimes \mu_{\xi_2} \otimes \cdots \otimes \mu_{\xi_n}$ is the joint distribution of the random vector $\xi = (\xi_1, \xi_2, \ldots, \xi_n)$. It will be convenient to set $\xi_0 \equiv 1$, so that $(\xi_i)_{i=0}^n$ is an orthonormal system in L^2. Let $\mathcal{A}$ be the linear (finite-dimensional and thus closed) subspace of L^2 consisting of all affine real-valued functions on $\mathbb{R}^n$. We define coordinate projection functions $\pi_i : \mathbb{R}^n \longrightarrow \mathbb{R}$ by $\pi_i(x) = x_i$ for $1 \le i \le n$, and $\pi_0 \equiv 1$. Let $\mathcal{A}_\pi = \{\pi_0, -\pi_0, \pi_1, -\pi_1, \dots, \pi_n, -\pi_n\}$. For a Boolean Borel (i. e., $\{-1,1\}$-valued and such that the preimage of $\{1\}$ is a Borel set) function f on $\mathbb{R}^n$, by $f_{\mathcal{A}} : \mathbb{R}^n \longrightarrow \mathbb{R}$ we will denote its orthogonal projection in L^2 onto the subspace $\mathcal{A}$:

$$f_{\mathcal{A}}(x) = a_0 + a_1x_1 + \dots + a_nx_n, \quad \text{i. e.,} \quad f_{\mathcal{A}} = \sum_{i=0}^{n} a_i\pi_i\,,$$

where $a_i = \langle f, \pi_i \rangle_{L^2}$. We may and will use the same notation for a Borel Boolean function f defined only on the support of μ, since obviously it may be extended to a Borel Boolean function F on the whole $\mathbb{R}^n$, and $F_{\mathcal{A}}$ does not depend on the choice of the extension. Let us define the sign function in a slightly non-standard way as $\mathbf{1}_{[0,\infty)} - \mathbf{1}_{(-\infty,0)}$, to make the function Boolean (setting $\operatorname{sign}(0) = -1$ would work as well). Let $\rho = \operatorname{dist}_{L^2}(f, \mathcal{A}) = \inf_{g\in\mathcal{A}} \|f-g\|_{L^2}$ and $d = \operatorname{dist}_{L^2}(f, \mathcal{A}_\pi) = \inf_{g\in\mathcal{A}_\pi} \|f-g\|_{L^2}$. Note that $d \le \sqrt{2}$ because

$$\|f-\pi_1\|_{L^2}^2 + \|f+\pi_1\|_{L^2}^2 = 2\|f\|_{L^2}^2 + 2\|\pi_1\|_{L^2}^2 = 2+2 = 4\,,$$

so that

$$\min\Big(\|f-\pi_1\|_{L^2}, \|f+\pi_1\|_{L^2}\Big) \le \sqrt{2}\,.$$

Obviously, $\rho \le \|f\|_{L^2} = 1$. Finally, let us define random variables $S = f_{\mathcal{A}}(\xi)$ and $R = (f - f_{\mathcal{A}})(\xi)$.

5.2 Symmetric case

We will start with a theorem which recovers and extends the main result of [4] with a quite good, explicit constant.

Theorem 5.1. *Under* **A & N**, *if $\xi_1, \xi_2, \dots, \xi_n$ are additionally symmetric, then there exists some $k \in \{0,1,\dots,n\}$ such that*

$$a_k^2 \ge 1 - \frac{9+\sqrt{17}}{2} \cdot \rho^2.$$

Also,

$$\rho \le d \le (9+\sqrt{17})^{1/2} \cdot \rho\,, \quad \textit{and} \quad d \le \Big(\frac{9+\sqrt{17}}{2}\Big)^{1/2} \cdot \rho + o(\rho)$$

as $\rho \to 0^+$ (uniformly over Boolean functions).

Proof. Since $|f| \equiv 1$, the triangle inequality yields a pointwise bound

$$1 - |f - f_{\mathcal{A}}| \le |f_{\mathcal{A}}| \le 1 + |f - f_{\mathcal{A}}|, \quad \text{i. e.,} \quad \Big||S|-1\Big| \le |R|\,,$$

so $\operatorname{Var}(|S|) = \mathbb{E}(|S|-1)^2 - (\mathbb{E}|S|-1)^2 \le \mathbb{E}R^2 = \|f - f_{\mathcal{A}}\|_{L^2}^2 = \rho^2$. Let us consider independent random variables $(X_i)_{i=0}^n$ given by $X_i = a_i\xi_i$ for $1 \le i \le n$, and with X_0 being symmetric $\pm a_0$ random variable. The sum $|\sum_{i=0}^n X_i|$ has the same distribution (and thus the same variance) as $|S|$, so by using Theorem 1.3

(strictly speaking, its reformulation for $n+1$ instead of n summands) with $x=0$ we infer that for some $k \in \{0,1,\dots,n\}$ there is

$$\sum_{i\in\{0,1,\dots,n\}\setminus\{k\}} a_i^2 = \mathrm{Var}\Big(\sum_{i\in\{0,1,\dots,n\}\setminus\{k\}} X_i\Big) \le \frac{7+\sqrt{17}}{2}\cdot\rho^2.$$

Since by orthogonality we have

$$\sum_{i=0}^n a_i^2 = \|f_{\mathcal{A}}\|_{L^2}^2 = \|f\|_{L^2}^2 - \|f-f_{\mathcal{A}}\|_{L^2}^2 = 1-\rho^2,$$

this ends the proof of the first assertion.

The inequality $\rho \le d$ follows immediately from $\mathcal{A}_\pi \subseteq \mathcal{A}$. Observe that

$$\begin{aligned} d^2 &\le \|f-\mathrm{sign}(a_k)\pi_k\|_{L^2}^2 = \|f\|_{L^2}^2 + \|\pi_k\|_{L^2}^2 - 2\mathrm{sign}(a_k)\langle f,\pi_k\rangle_{L^2} \\ &= 1+1-2\mathrm{sign}(a_k)a_k = 2(1-|a_k|) = 2(1-a_k^2)/(1+|a_k|) \le (9+\sqrt{17})\rho^2/(1+|a_k|)\,. \end{aligned}$$

Thus

$$d \le \Big(9+\sqrt{17}\Big)^{1/2}\cdot\rho\,.$$

The remaining assertion also follows easily because the first assertion implies $|a_k| \ge 1-O(\rho^2)$, so that

$$d^2 \le \frac{9+\sqrt{17}}{2}\cdot\rho^2 + o(\rho^2)\,. \qquad \square$$

Corollary 5.2. *Under assumptions of Theorem 5.1 (**A & N**, and $\xi_1, \xi_2, \dots, \xi_n$ are symmetric) there is some $k \in \{0,1,\dots,n\}$ such that*

$$\min\Big(\|f-(\mathrm{sign}\circ\pi_k)\|_{L^2}, \|f+(\mathrm{sign}\circ\pi_k)\|_{L^2}\Big) \le 2d \le 2\Big(9+\sqrt{17}\Big)^{1/2}\cdot\rho.$$

Proof. Let k be such that $\|f-\pi_k\|_{L^2} = d$ (or, alternatively, such that $\|f+\pi_k\|_{L^2} = d$). Note that for any $s \in \{-1,1\}$ and $u \in \mathbb{R}$ there is $|s-u| \ge |\mathrm{sign}(u)-u|$ (and $|s+u| \ge |\mathrm{sign}(u)-u|$). Hence $|f-\pi_k| \ge |(\mathrm{sign}\circ\pi_k)-\pi_k|$ (and $|f+\pi_k| \ge |(\mathrm{sign}\circ\pi_k)-\pi_k|$) pointwise, so that $\|(\mathrm{sign}\circ\pi_k)-\pi_k\|_{L^2} \le d$. The assertion follows by the triangle inequality. $\square$

Now let us see how to strengthen the result of Friedgut, Kalai and Naor. For a function f defined on the discrete cube $\{-1,1\}^n$ we consider its standard Walsh-Fourier expansion $\sum_A \hat{f}(A)w_A$, where $w_A(x) = \prod_{i\in A} x_i$.

Theorem 5.3. *There exists a universal constant $L>0$ with the following property. For $f: \{-1,1\}^n \to \{-1,1\}$, let*

$$\rho = \Big(\sum_{A\subseteq[n]:|A|\ge 2} |\hat{f}(A)|^2\Big)^{1/2}.$$

Then there exists some $B \subseteq [n]$ with $|B| \le 1$ such that

$$\sum_{A\subseteq[n]:|A|\le 1, A\ne B} |\hat{f}(A)|^2 \le L\cdot\rho^4\ln(2/\rho)$$

and $|\hat{f}(B)|^2 \ge 1-\rho^2-L\cdot\rho^4\ln(2/\rho)$.

Proof. Let $\xi_1, \xi_2, \ldots, \xi_n$ be independent symmetric ± 1 random variables, so that the definition of ρ is consistent with the one from **A & N**. We also have $a_i = \langle f, \pi_i \rangle_{L^2} = \hat{f}(\{i\})$ for $i \in [n]$, and $a_0 = \hat{f}(\emptyset)$. Let us put

$$\theta = \Big(4 \log_2(2/d) - 1\Big)^{-1}.$$

Because $d \leq \sqrt{2}$, always $\theta \in (0,1]$.

Let $k \in \{0,1,\ldots,n\}$ be such that $d = \|f - \pi_k\|_{L^2}$ (if the point of $\mathcal{A}_\pi$ closest to f is of the form $-\pi_k$, then a similar reasoning works). Hence $d^2 = \|f\|_{L^2}^2 + \|\pi_k\|_{L^2}^2 - 2\langle f, \pi_k \rangle_{L^2} = 2(1 - a_k)$. Since $h = f - \pi_k$ is a $\{-2,0,2\}$-valued function, we get

$$\mathbb{P}_\mu(h \neq 0) = \mu(\{x \in \{-1,1\}^n : h(x) \neq 0\}) = \frac{1}{4}\|h\|_{L^2}^2 = (d/2)^2.$$

Therefore

$$\begin{aligned} d^4/2 = 4(d/2)^{\frac{4}{1+\theta}} = 4\Big(\mathbb{P}_\mu(h \neq 0)\Big)^{\frac{2}{1+\theta}} = \|h\|_{L^{1+\theta}}^2 &\overset{B-B}{\geq} \sum_{A \subseteq [n]} \theta^{|A|} \cdot |\hat{h}(A)|^2 \\ \geq \theta \cdot \sum_{A \subseteq [n]: |A| \leq 1} |\hat{h}(A)|^2 = \theta \cdot \Big((1-a_k)^2 + \sum_{i \in \{0,1,\ldots,n\} \setminus \{k\}} a_i^2\Big) &= \theta \cdot \Big(\frac{d^4}{4} + \sum_{i \in \{0,1,\ldots,n\} \setminus \{k\}} a_i^2\Big), \end{aligned}$$

so that

$$\sum_{i \in \{0,1,\ldots,n\} \setminus \{k\}} a_i^2 \leq (2\theta^{-1} - 1)d^4/4 \leq 2d^4 \log_2(2/d). \tag{5.1}$$

The inequality $\overset{B-B}{\geq}$ is the classical $L^2 - L^{1+\theta}$ hypercontractive Bonami–Beckner estimate ([2], [1]). Now we have

$$\begin{aligned} \sum_{i=0}^n a_i^2 = \Big(1 - \frac{d^2}{2}\Big)^2 + \sum_{i \in \{0,1,\ldots,n\} \setminus \{k\}} a_i^2 &\leq \Big(1 - \frac{d^2}{2}\Big)^2 + \frac{1}{4}(2\theta^{-1} - 1)d^4 \\ &= 1 - d^2 + \frac{1}{2}\theta^{-1} d^4 \leq 1 - d^2 + 2d^4 \log_2(2/d) \end{aligned}$$

and thus (recall that $|f| \equiv 1$, so $\sum_{A \subseteq [n]} |\hat{f}(A)|^2 = \|f\|_{L^2}^2 = 1$) we get

$$\rho^2 = \sum_{A \subseteq [n]: |A| \geq 2} |\hat{f}(A)|^2 = 1 - \sum_{i=0}^n a_i^2 \geq d^2 - 2d^4 \log_2(2/d). \tag{5.2}$$

We finish the proof by observing that (5.1) and (5.2) yield

$$\begin{aligned} \sum_{i \in \{0,1,\ldots,n\} \setminus \{k\}} a_i^2 &\leq 2d^4 \log_2(2/d) \leq 2\Big(\rho^2 + 2d^4 \log_2(2/d)\Big)^2 \log_2(2/d) \\ &\leq 2\Big(\rho^2 + 2d^4 \log_2(2/\rho)\Big)^2 \log_2(2/\rho) = 2\rho^4 \log_2(2/\rho) + o(\rho^5), \end{aligned}$$

uniformly, as $\rho \to 0^+$. We have used the $d = O(\rho)$ bound of Theorem 5.1.

For $\rho \le 1/3$, by Theorem 5.1 we have

$$d \le \frac{1}{3}\Big(9+\sqrt{17}\Big)^{1/2} \le 2\mathrm{e}^{-1/2}$$

and we may choose

$$\theta = \Big(4\ln(2/d)-1\Big)^{-1} \quad \text{instead of} \quad \Big(4\log_2(2/d)-1\Big)^{-1}$$

which yields, essentially with the same proof, a slightly better (asymptotically) bound of $\mathrm{e}\rho^4\ln(2/\rho)+o(\rho^5)$ as $\rho \to 0^+$. □

Remark 5.4. The bound $O\big(\rho^4\ln(2/\rho)\big)$ in Theorem 5.3 is better than $O(\rho^2)$ obtained in [4] and $\rho^2+o(\rho^2)$ of [6] (see also Corollary 15.2 of [5]). Also, it is of the optimal order. Indeed, for $2 \le m \le n$ consider the negated OR function

$$f(x) = 1-\frac{1}{2^{m-1}}\prod_{i=1}^{m}(1+x_i)\,.$$

Then $\rho \le 2\cdot 2^{-m/2}$, so that $\rho^4\log_2(2/\rho) \le 8m\cdot 4^{-m}$, whereas $|\hat{f}(A)|^2 \ge 4\cdot 4^{-m}$ for every one-element set A contained in $\{1,2,\dots,m\}$.

5.3 Absolute first moment assumption

Without the symmetry assumption we cannot hope to get the same assertion as in Theorem 5.1—if $n=1$, $\mathbb{P}(\xi_1=-2)=1/5$, $\mathbb{P}(\xi_1=1/2)=4/5$ and $f(x)=(3+4x)/5$, then $f \in \mathcal{A}\setminus\mathcal{A}_\pi$ even though f is Boolean. However, under an additional moment assumption, we still may prove that any Boolean function which is close in L^2 to $\mathcal{A}$ must be also at a comparable L^2-distance from an affine function of a single coordinate.

Theorem 5.5. *There exists a function $C:(0,\infty)\longrightarrow(0,\infty)$ with the following property. Under* **A & N**, *let $\eta=\min_{i\in[n]}\mathbb{E}|\xi_i|$. Then there is some $k\in[n]$ such that $\|f-(a_0+a_k\pi_k)\|_{L^2}\le C(\eta)\rho$ and $\|f-\operatorname{sign}(a_0+a_k\pi_k)\|_{L^2}\le 2C(\eta)\rho$.*

Proof. As in the proof of Theorem 5.1 we observe first that $\operatorname{Var}(|S|)\le\rho^2$, where $S=a_0+\sum_{i=1}^n a_i\xi_i$. By Theorem 1.1 there is some $k\in[n]$ such that $\sum_{i\in[n]\setminus\{k\}}a_i^2\le K(\eta^{-2})\rho^2$. Thus

$$\|f-(a_0+a_k\pi_k)\|_{L^2}^2 = \|f-f_{\mathcal{A}}\|_{L^2}^2+\Bigg\|\sum_{i\in[n]\setminus\{k\}}a_i\pi_i\Bigg\|_{L^2}^2 \le \big(1+K(\eta^{-2})\big)\rho^2$$

which proves the first assertion with $C(\eta)=\sqrt{1+K(\eta^{-2})}$. The second assertion follows from the first one in the same way in which Corollary 5.2 follows from Theorem 5.1. □

5.4 General case

We will need two auxiliary lemmas.

Lemma 5.6. *Let X and Y be independent square-integrable random variables. Assume $\mathbb{E}(|X+Y|-1)^2 \leq \rho^2$ for some $\rho \in [0,1]$. Then* $\mathrm{Var}(X) \leq 25\rho$ *or* $\mathrm{Var}(Y) \leq 25\rho$.

Proof. Let (X',Y') be an independent copy of the pair (X,Y). For a square-integrable random variable Z let $\|Z\|_2 = \left(\mathbb{E}Z^2\right)^{1/2}$. Also, let us define $\phi : \mathbb{R} \longrightarrow \{-2,0,2\}$ by $\phi(u) = 2$ for $u > 1$, $\phi(u) = -2$ for $u < -1$, and $\phi(u) = 0$ otherwise. Note that $|u - \phi(u)| = \mathrm{dist}(u, \{-2,0,2\})$ for any $u \in \mathbb{R}$. Finally, let

$$\alpha = \mathbb{P}(X - X' > 1) = \mathbb{P}(X - X' < -1) \quad \text{and} \quad \beta = \mathbb{P}(Y - Y' > 1) = \mathbb{P}(Y - Y' < -1)\,.$$

By the assumptions of the lemma we have $\|\mathrm{dist}(X+Y, \{-1,1\})\|_2 \leq \rho$. Since $X+Y$ and $X'+Y$ have the same distribution, we obtain $\|\mathrm{dist}(X'+Y, \{-1,1\})\|_2 \leq \rho$. Thus the pointwise bound

$$\mathrm{dist}(X - X', \{-2,0,2\}) \leq \mathrm{dist}(X+Y, \{-1,1\}) + \mathrm{dist}(X'+Y, \{-1,1\})$$

implies

$$\|(X-X') - \phi(X-X')\|_2 = \|\mathrm{dist}(X-X', \{-2,0,2\})\|_2 \leq 2\rho\,. \tag{5.3}$$

In a similar way we prove that

$$\|(Y-Y') - \phi(Y-Y')\|_2 = \|\mathrm{dist}(Y-Y', \{-2,0,2\})\|_2 \leq 2\rho\,. \tag{5.4}$$

The pointwise bound

$$\mathrm{dist}(X+Y-X'-Y', \{-2,0,2\}) \leq \mathrm{dist}(X+Y, \{-1,1\}) + \mathrm{dist}(X'+Y', \{-1,1\})$$

implies $\|\mathrm{dist}\left((X-X')+(Y-Y'), \{-2,0,2\}\right)\|_2 \leq 2\rho$. Hence by (5.3), (5.4), and the triangle inequality we obtain

$$\left\|\mathrm{dist}\left(\phi(X-X') + \phi(Y-Y'), \{-2,0,2\}\right)\right\|_2 \leq 2\rho + 2\rho + 2\rho = 6\rho\,,$$

so that

$$\alpha\beta \leq \frac{1}{4}\mathbb{E}\mathrm{dist}^2(\phi(X-X') + \phi(Y-Y'), \{-2,0,2\})\mathbf{1}_{X-X',Y-Y'>1} \leq 9\rho^2\,. \tag{5.5}$$

Assume $\mathrm{Var}(X), \mathrm{Var}(Y) > 25\rho$, so that $\|X-X'\|_2, \|Y-Y'\|_2 > 7\sqrt{\rho}$. Then by (5.3) we have

$$2\sqrt{2\alpha} = \|\phi(X-X')\|_2 > 7\sqrt{\rho} - 2\rho > 5\sqrt{\rho},$$

so that $\alpha > 3\rho$. In a similar way from (5.4) we get $\beta > 3\rho$, and therefore $\alpha\beta > 9\rho^2$, which contradicts (5.5). □

Lemma 5.7. *Let $X_1, X_2, \ldots, X_n$ be independent square-integrable random variables and let $S = \sum_{i=1}^n X_i$. Assume $\mathbb{E}(|S|-1)^2 \leq \rho^2$ for some $\rho \in [0,1]$. Then there exists some $k \in [n]$ such that* $\mathrm{Var}(S - X_k) \leq 50\rho$.

Proof. Lemma 5.7 follows from Lemma 5.6 in a way similar to that in which we have deduced Theorem 1.3 from Lemma 4.1: we look for a minimal $I \subseteq [n] = \{1,2,\ldots,n\}$ such that $\mathrm{Var}\left(\sum_{i\in I} X_i\right) > 25\rho$, then we choose any $k \in I$ and from Lemma 5.6 we infer that $\mathrm{Var}\left(\sum_{i\in[n]\setminus I} X_i\right) \le 25\rho$. By the minimality of I there is also $\mathrm{Var}\left(\sum_{i\in I\setminus\{k\}} X_i\right) \le 25\rho$, which ends the proof. □

Now we may finally state a result which does not use any additional properties of the marginal distributions.

Theorem 5.8. *Under* **A & N**, *there exists some* $k \in [n]$ *such that*

$$\|f-(a_0+a_k\pi_k)\|_{L^2} \le 8\sqrt{\rho} \quad \textit{and} \quad \|f-\mathrm{sign}(a_0+a_k\pi_k)\|_{L^2} \le 16\sqrt{\rho}\,.$$

Proof. Let $X_i = a_i\xi_i$ for $i \in [n]$, so that $S = a_0 + \sum_{i=1}^n X_i$. Then

$$\mathbb{E}(|S|-1)^2 = \||f_{\mathcal{A}}|-1\|_{L^2}^2 \le \|f-f_{\mathcal{A}}\|_{L^2}^2 = \rho^2$$

because of the pointwise inequality $||f_{\mathcal{A}}|-1| \le |f-f_{\mathcal{A}}|$. Lemma 5.7 easily implies that there is some $k \in [n]$ such that $\sum_{i\in[n]\setminus\{k\}} a_i^2 \le 50\rho$. Thus

$$\|f-(a_0+a_k\pi_k)\|_{L^2}^2 = \|f-f_{\mathcal{A}}\|_{L^2}^2 + \Big\|\sum_{i\in[n]\setminus\{k\}} a_i\pi_i\Big\|_{L^2}^2 \le \rho^2+50\rho \le 64\rho$$

(recall that ρ is always in $[0,1]$ under **A & N**), which proves the first assertion. The second assertion of the theorem follows from the first one in the same way in which Corollary 5.2 follows from Theorem 5.1. □

Under **A & N**, let $n=2$ and $\mathbb{P}(\xi_i = \sqrt{\beta/\alpha}) = \alpha$, $\mathbb{P}(\xi_i = -\sqrt{\alpha/\beta}) = \beta$ for $i \in \{1,2\}$, where $\alpha \in (0,1)$ and $\beta = 1-\alpha$. Then $\beta - \sqrt{\alpha\beta}\xi_i$'s are $\{0,1\}$-valued random variables, so that $f(x)$ given by

$$2(\beta-\sqrt{\alpha\beta}x_1)(\beta-\sqrt{\alpha\beta}x_2)-1 = (2\beta^2-1)-2\beta^{3/2}\alpha^{1/2}(x_1+x_2)+2\alpha\beta x_1x_2$$

is a Boolean function on $\{-\sqrt{\alpha/\beta},\sqrt{\beta/\alpha}\}^2$ equipped with the measure $\mu_{\xi_1}\otimes\mu_{\xi_2}$. A simple analysis shows that $\rho = \mathrm{dist}_{L^2}(f,\mathcal{A}) = 2\alpha\beta$ while the L^2-distance from f to any function of a single coordinate is not less that $2\beta^{3/2}\alpha^{1/2} = \Theta(\rho^{1/2})$ as $\alpha \to 0^+$ (and, consequently, $\beta \to 1^-$ and $\rho \to 0^+$). Thus the $O(\sqrt{\rho})$ general bound of Theorem 5.8 cannot be improved even on two-dimensional biased discrete cube. On the other hand, some FKN-type bounds were obtained in [6] and [5] for the biased discrete cube of arbitrary dimension, in terms of the bias parameter. The approach used in the proof of Theorem 5.3 can be effectively adapted to the case of the biased discrete cube if the Bonami–Beckner estimate is replaced by the hypercontractive bounds of [12], see [10].

5.5 Boolean versus bounded

It is natural to look for an analogue of the FKN theorem for $[-1,1]$-valued functions; however, there is no hope for estimates as good as in the Boolean case. Recall that the FKN theorem states that a Boolean function on the discrete cube (equipped with the uniform measure) which is close in L^2 to $\mathcal{A}$ must be be also at a comparable L^2-distance from a function which is both Boolean and affine (i. e., some function from $\mathcal{A}_\pi$).

Let $\mathcal{A}_{[-1,1]}$ denote the set of $[-1,1]$-valued affine functions on the cube:

$$\mathcal{A}_{[-1,1]} = \left\{ \sum_{i=0}^{n} b_i \pi_i;\ \sum_{i=0}^{n} |b_i| \le 1 \right\}.$$

Let $\psi(t) = 1$ for $t > 1$, $\psi(t) = -1$ for $t < -1$, and $\psi(t) = t$ for $t \in [-1,1]$, and let us define functions $f, g : \{-1,1\}^n \longrightarrow \mathbb{R}$ by $g(x) = s^{-1} n^{-1/2} \sum_{i=1}^{n} x_i$ and $f(x) = \psi(g(x))$ for some positive parameter s. Note that

$$\begin{aligned}\operatorname{dist}_{L^2}\left(g, \mathcal{A}_{[-1,1]}\right)^2 &\ge \inf_{\sum |b_i| \le 1} \sum_{i=1}^{n} (s^{-1} n^{-1/2} - b_i)^2 \\ &\ge \inf_{\sum |b_i| \le 1} \sum_{i=1}^{n} (s^{-2} n^{-1} - 2 s^{-1} n^{-1/2} b_i) \ge s^{-2} - 2 s^{-1} n^{-1/2}\end{aligned}$$

and $\operatorname{dist}_{L^2}(g, \mathcal{A}_{[-1,1]}) \le \|g - 0\|_{L^2} = s^{-1}$, so that $\lim_{n\to\infty} \operatorname{dist}_{L^2}(g, \mathcal{A}_{[-1,1]}) = s^{-1}$. Thus for the $[-1,1]$-valued function f there is

$$\lim_{n\to\infty} \operatorname{dist}_{L^2}(f, \mathcal{A}) \le \lim_{n\to\infty} \|f - g\|_{L^2} = \sqrt{\mathbb{E}(|s^{-1}G| - 1)^2 \mathbf{1}_{|G| \ge s}} = O(\mathrm{e}^{-s^2/4})$$

as $s \to \infty$, where G denotes the standard $\mathcal{N}(0,1)$ Gaussian variable, while we have only

$$\operatorname{dist}_{L^2}(f, \mathcal{A}_{[-1,1]}) \ge \operatorname{dist}_{L^2}(g, \mathcal{A}_{[-1,1]}) - \|f - g\|_{L^2},$$

so that $\lim_{n\to\infty} \operatorname{dist}_{L^2}(f, \mathcal{A}_{[-1,1]}) = \Theta(s^{-1})$ as $s \to \infty$.

The above example, demonstrating the gap between $O(\mathrm{e}^{-s^2/4})$ and $\Theta(s^{-1})$, is as bad as it gets—in [10] Nayar proved that for $s > 0$ and **every** function $f : \{-1,1\}^n \to [-1,1]$ with $\operatorname{dist}_{L^2}(f, \mathcal{A}) = \mathrm{e}^{-s^2/4}$ we have $\operatorname{dist}_{L^2}(f, \mathcal{A}_{[-1,1]}) \le 36 s^{-1}$.

6 Banach space setting

The main result of this section, Theorem 6.10, is an advanced extension of Lemma 2.2. To prove it, we will need to develop some new tools.

In what follows, $(V, \|\cdot\|)$ denotes a separable real Banach space, its continuous dual space (of bounded linear functionals on V) is denoted by V', and $+$ stands for Minkowski addition. By dist we will mean the distance in the norm $\|\cdot\|$. Readers unfamiliar with Banach spaces may find it convenient to assume additionally that V is a finite-dimensional normed vector space and think about Euclidean spaces instead of Hilbert spaces.

For a finite $A \subset V$, positive $\Delta, \varepsilon, \delta$ and a V-valued random vector X we will say that:

- A is Δ-*separated* if either $|A| \ge 2$ and for any two distinct $x, y \in A$ there is $\|x - y\| \ge \Delta$, or $|A| = 1$; we define the *separation constant* of A by $\Delta(A) = \min\{\|x - y\|; x, y \in A \text{ and } x \ne y\}$. Clearly, $\Delta(A) = \infty$ if $|A| = 1$;

- X is (ε,δ)-*close* to A if $\mathbb{P}(\operatorname{dist}(X,A)>\varepsilon)\le\delta$;
- X is (ε,δ)-*present* around A if there is $\mathbb{P}(\|X-a\|\le\varepsilon)>\delta$ for every $a\in A$.

Note that if X is (ε,δ)-close to A, then it is (ε,δ)-close to any finite set B containing A. Similarly, if X is (ε,δ)-present around A, then it is (ε,δ)-present around any finite set B contained in A.

We need some simple lemmas. The first of them is obvious.

Lemma 6.1. *Let X be a V-valued random vector which is (ε,δ)-close to some nonempty finite $A\subset V$ with some $\varepsilon>0$ and $\delta\in[0,1]$. Then there is some $a\in A$ such that $\mathbb{P}(\|X-a\|\le\varepsilon)\ge(1-\delta)/|A|$.*

Lemma 6.2. *Let $A,B\subset V$ be finite and Δ-separated for some $\Delta>0$ and assume that X is a V-valued random vector which is both (ε_1,δ)-close to A and (ε_2,δ)-present around B for some $\delta>0$ and $\varepsilon_1,\varepsilon_2>0$ such that $\varepsilon_1+\varepsilon_2<\Delta/2$. Then $|B|\le|A|$.*

Proof. Let $b\in B$. Since $\mathbb{P}(\|X-b\|\le\varepsilon_2)>\delta$ and $\mathbb{P}(\operatorname{dist}(X,A)>\varepsilon_1)\le\delta$ there is $\mathbb{P}(\|X-b\|\le\varepsilon_2$ and $\operatorname{dist}(X,A)\le\varepsilon_1)>0$, so that $\operatorname{dist}(b,A)\le\varepsilon_1+\varepsilon_2$. Note that there is only one $a_b\in A$ such that $\|b-a_b\|\le\varepsilon_1+\varepsilon_2$ because we assume that A is Δ-separated and $\varepsilon_1+\varepsilon_2<\Delta/2$. For a similar reason ($B$ is also Δ-separated), the mapping $b\mapsto a_b$ is injective. This ends the proof. □

Lemma 6.3. *Let $X_1,X_2,\dots,X_m$ be independent V-valued random vectors and let $S=\sum_{i=1}^m X_i$. Assume that S is (ε,δ)-close to a nonempty finite $A\subset V$ for some $\varepsilon,\delta>0$. Then there exist vectors $v_1,v_2,\dots,v_m\in V$ such that X_i is (ε,δ)-close to v_i+A for every $i\in[m]$.*

Proof. This follows from the Fubini theorem (see the beginning of the proof of Lemma 2.1). □

Lemma 6.4. *Let $\Delta>0$ and let A and B be finite Δ-separated subsets of V with $|A|,|B|\ge2$. Then there exists some $C\subseteq A+B=\{a+b;a\in A,b\in B\}$ with $|C|>\max(|A|,|B|)$ which is also Δ-separated.*

Proof. Without loss of generality we may assume that $|A|\ge|B|$. Let b and b' be arbitrary distinct elements of B. Let $\varphi\in V'$ be such that $\|\varphi\|_{V'}=1$ and $\varphi(b'-b)=\|b'-b\|$ (the existence of such a functional is guaranteed by the Hahn-Banach theorem). Finally, let $\hat a=\arg\max_{a\in A}\varphi(a)$ (any maximizer will do if there is more than one). Then $C=(A+b)\cup\{\hat a+b'\}$ has more elements than A and is Δ-separated. These two facts follow since, for $a\in A$,

$$\|(\hat a+b')-(a+b)\|\ge\varphi(\hat a+b'-a-b)=\varphi(\hat a)-\varphi(a)+\varphi(b'-b)\ge\|b'-b\|\ge\Delta. \qquad \square$$

By an obvious induction we obtain the following corollary.

Corollary 6.5. *Let $\Delta>0$ and let $A_1,\dots,A_m$ be Δ-separated finite subsets of V with $|A_i|\ge2$ for $i\in[m]$. Then there exists some $C\subseteq A_1+A_2+\dots+A_m$ with $|C|>m$ which is also Δ-separated.*

The next corollary easily follows (we leave it as an exercise).

Corollary 6.6. *Let $\Delta,\varepsilon_1,\dots,\varepsilon_m,\delta_1,\dots,\delta_m>0$ and let $A_1,\dots,A_m$ be Δ-separated finite subsets of V with $|A_i|\ge2$ for $i\in[m]$. Assume that $X_1,\dots,X_k$ are independent V-valued random vectors and X_i is (ε_i,δ_i)-present around A_i for all $i\in[m]$. Then there exists some Δ-separated set $C\subseteq A_1+\dots+A_m$ with $|C|>m$ such that $\sum_{i=1}^m X_i$ is $(\sum_{i=1}^m\varepsilon_i,\prod_{i=1}^m\delta_i)$-present around C.*

Lemma 6.7. *Let $\Delta > 0$ and let A and B be finite Δ-separated subsets of a Hilbert space $\mathcal{H}$. Then there exists some $C \subseteq A+B$ with $|C| \geq |A|+|B|-1$ which is also Δ-separated.*

Proof. Without loss of generality we may assume that $0 \in A$. Let $A = \{0, u_1, \ldots, u_m\}$ and $B = \{v_1, \ldots, v_n\}$. For $u \in \mathcal{H}$ let $b(u) = \arg\max_{v \in B} \langle u, v\rangle$ (any choice will do if there is more than one maximizer). Then we can just take $C = \{v_1, \ldots, v_n, u_1 + b(u_1), \ldots, u_m + b(u_m)\}$. Obviously, $C \subseteq A+B$. By definition of $b(u_j)$ for any $i \in [n]$ and $j \in [m]$ we have $\langle u_j, b(u_j)\rangle \geq \langle u_j, v_i\rangle$, so that

$$\|u_j\|\|u_j + b(u_j) - v_i\| \geq \langle u_j, u_j + b(u_j) - v_i\rangle \geq \langle u_j, u_j\rangle = \|u_j\|^2$$

and therefore $\|(u_j + b(u_j)) - v_i\| \geq \|u_j\| \geq \Delta$.

Similarly, for any distinct $j, j' \in [m]$ we have $\langle u_j, b(u_j)\rangle \geq \langle u_j, b(u_{j'})\rangle$ and $\langle u_{j'}, b(u_{j'})\rangle \geq \langle u_{j'}, b(u_j)\rangle$, so that $\langle u_j - u_{j'}, b(u_j) - b(u_{j'})\rangle \geq 0$. Hence

$$\begin{aligned}\|u_j - u_{j'}\|\|(u_j + b(u_j)) - (u_{j'} + b(u_{j'}))\| &\geq \langle u_j - u_{j'}, u_j - u_{j'} + b(u_j) - b(u_{j'})\rangle \\ &\geq \langle u_j - u_{j'}, u_j - u_{j'}\rangle = \|u_j - u_{j'}\|^2\end{aligned}$$

and thus $\|(u_j + b(u_j)) - (u_{j'} + b(u_{j'}))\| \geq \|u_j - u_{j'}\| \geq \Delta$. □

Again, an obvious induction yields the following corollary.

Corollary 6.8. *Let $\Delta > 0$ and $A_1, \ldots, A_m$ be nonempty, Δ-separated and finite subsets of a Hilbert space. Then there exists some Δ-separated $C \subseteq A_1 + \cdots + A_m$ with $|C| > \sum_{i=1}^m (|A_i| - 1)$.*

The next corollary easily follows (we also leave it as an exercise).

Corollary 6.9. *Let $\Delta, \varepsilon_1, \ldots, \varepsilon_m, \delta_1, \ldots, \delta_m > 0$ and let $A_1, \ldots, A_m$ be finite Δ-separated subsets of a Hilbert space $\mathcal{H}$. Assume that $X_1, \ldots, X_m$ are independent $\mathcal{H}$-valued random vectors and X_i is $(\varepsilon_i, \delta_i)$-present around A_i for all $i \in [m]$. Then there exists some Δ-separated set $C \subseteq A_1 + \cdots + A_m$ with $|C| > \sum_{i=1}^m (|A_i| - 1)$ such that $\sum_{i=1}^m X_i$ is $(\sum_{i=1}^m \varepsilon_i, \prod_{i=1}^m \delta_i)$-present around C.*

Now we are in position to prove a structural theorem.

Theorem 6.10. *Let A be a finite subset of V with $|A| \geq 2$. Let ε and δ be positive numbers satisfying*

$$\varepsilon < \frac{\Delta(A)}{2(|A|+1)} \quad \textit{and} \quad \delta \leq |A|^{-|A|}.$$

Assume that $\xi_1, \ldots, \xi_n$ are independent V-valued random vectors such that $S = \sum_{i=1}^n \xi_i$ is (ε, δ)-close to A. For $I \subseteq [n]$ let $S_I = \sum_{i \in I} \xi_i$. Then there exists a nonnegative integer $k < |A|$ and $\{i_1, \ldots, i_k\} \subseteq [n]$ such that

$$\mathbb{P}\left(\left\|S_{[n]\setminus\{i_1,\ldots,i_k\}} - v\right\| \leq |A|\varepsilon\right) \geq \left(1 - \delta - (|A|-1)\delta^{1/|A|}\right)^{|A|}$$

for some $v \in V$. Consequently,

$$\mathbb{P}\left(\left\|S_{[n]\setminus\{i_1,\ldots,i_k\}} - v\right\| > |A|\varepsilon\right) \leq |A|(|A|-1)\delta^{1/|A|} + |A|\delta \leq |A|^2\delta^{1/|A|}.$$

Moreover, if V is a Hilbert space and $k > 0$, then there are vectors $v_1, \ldots, v_k \in V$ and nonempty sets $B_1, \ldots, B_k \subseteq A$ with $\sum_{l=1}^k (|B_l| - 1) < |A|$ such that ξ_{i_l} is $(\varepsilon, |A|\delta^{1/k})$-close, and thus also $(\varepsilon, |A|\delta^{1/(|A|-1)})$-close, to $v_l + B_l$ for every $l \in [k]$.

Proof. We will call $I \subseteq [n]$ *relevant* if there exist some two points $x, y \in V$ such that $\|x-y\| \geq \Delta(A)$, $\mathbb{P}(\|S_I - x\| \leq \varepsilon) > \delta^{1/|A|}$ and $\mathbb{P}(\|S_I - y\| \leq \varepsilon) > \delta^{1/|A|}$. Since $\varepsilon < \Delta(A)/2$, all relevant sets are nonempty. We will inductively construct a sequence of relevant sets $I_1, I_2, \ldots$. Let I_1 be a minimal (in the sense of inclusion) relevant subset of $[n]$—any minimizer will do if there is more than one. Then, having defined $I_1, \ldots, I_s$, we choose for I_{s+1} a minimal relevant subset of $[n] \setminus \bigcup_{l=1}^{s} I_l$. We end up with a collection of pairwise disjoint relevant sets $I_1, \ldots, I_k$, and we know that $J = [n] \setminus \bigcup_{l=1}^{k} I_l$ does not contain any relevant set and thus is not relevant.

Let us assume for a while that $k \geq |A|$. By expressing S as

$$\left(\sum_{l=1}^{|A|} S_{I_l}\right) + S_{[n]\setminus \bigcup_{l=1}^{|A|} I_l}$$

and using Lemma 6.3 (with $m = 2$) we deduce that $\sum_{l=1}^{|A|} S_{I_l}$ is (ε, δ)-close to some shift of A, i. e., to a $\Delta(A)$-separated set of cardinality $|A|$. On the other hand, I_l's are relevant so that S_{I_l}'s are $(\varepsilon, \delta^{1/|A|})$-present around some Δ-separated sets of cardinality greater than 1. Thus, by Corollary 6.6 the sum $\sum_{l=1}^{|A|} S_{I_l}$ is $(|A|\varepsilon, \delta)$-present around some $\Delta(A)$-separated set of cardinality greater than $|A|$ and hence Lemma 6.2 (used for $\varepsilon_2 = |A|\varepsilon$) implies that $|A| < |A|$. This contradiction proves that $k < |A|$.

Select arbitrary $i_1 \in I_1, \ldots, i_k \in I_k$. The way in which we chose I_l's guarantees that sets $I_1 \setminus \{i_1\}, \ldots, I_k \setminus \{i_k\}$ are not relevant. By expressing S as $S_J + S_{I_1 \setminus \{i_1\}} + \cdots + S_{I_k \setminus \{i_k\}} + \xi_{i_1} + \cdots + \xi_{i_k}$ and using again Lemma 6.3 (now with $m = 2k+1$) we prove that there exist vectors $w, w_1, \ldots, w_k, v_1, \ldots, v_k$ such that S_J is (ε, δ)-close to $w + A$ and $S_{I_l \setminus \{i_l\}}$ is (ε, δ)-close to $w_l + A$, and ξ_{i_l} is (ε, δ)-close to $v_l + A$ for every $l \in [k]$. On the other hand, $I_l \setminus \{i_l\}$'s are not relevant and the shifts of A are $\Delta(A)$-separated, so that for each $l \in [k]$ there exists at most one $a_l \in A$ such that $\mathbb{P}\left(\left\|S_{I_l \setminus \{i_l\}} - w_l - a_l\right\| \leq \varepsilon\right) > \delta^{1/|A|}$. Thus

$$\mathbb{P}\left(\left\|S_{I_l \setminus \{i_l\}} - (w_l + a_l)\right\| > \varepsilon\right) \leq (|A|-1)\delta^{1/|A|} + \delta$$

for all $l \in [k]$ and similarly we prove that there exists some $a \in A$ such that

$$\mathbb{P}\left(\|S_J - (w+a)\| > \varepsilon\right) \leq (|A|-1)\delta^{1/|A|} + \delta\,.$$

Hence for $S_{[n]\setminus\{i_1,\ldots,i_k\}} = S_J + \sum_{l=1}^{k} S_{I_l \setminus \{i_l\}}$ and $v = (w+a) + \sum_{l=1}^{k}(w_l + a_l)$ we have

$$\mathbb{P}\left(\left\|S_{[n]\setminus\{i_1,\ldots,i_k\}} - v\right\| \leq (k+1)\varepsilon\right) \geq \left(1 - (|A|-1)\delta^{1/|A|} - \delta\right)^{k+1}$$

and the first assertion of the theorem easily follows from $k < |A|$. Note that $\delta \leq |A|^{-|A|}$ implies $(|A| - 1)\delta^{1/|A|} + \delta < 1$.

Now let us additionally assume that $k > 0$ and V is a Hilbert space. For $l \in [k]$ let

$$B_l = \left\{a \in A : \mathbb{P}\left(\|\xi_{i_l} - (a + v_l)\| \leq \varepsilon\right) > \delta^{1/k}\right\}.$$

Note that $\delta \leq |A|^{-|A|}$ implies $(1-\delta)/|A| \geq \delta^{1/(|A|-1)} \geq \delta^{1/k}$, so that by Lemma 6.1 we have $B_l \neq \emptyset$. We also easily see that ξ_{i_l} is $(\varepsilon, \delta + (|A|-1)\delta^{1/k})$-close and thus also $(\varepsilon, |A|\delta^{1/k})$-close to $v_l + B_l$. Since ξ_{i_l} is

$(\varepsilon, \delta^{1/(|A|-1)})$-present around $v_l + B_l$ for every $l \in [k]$, by Corollary 6.9 the sum $\sum_{l=1}^{k} \xi_{i_l}$ is $(k\varepsilon, \delta^{k/(|A|-1)})$-present, and thus also $((|A|-1)\varepsilon, \delta)$-present around some $\Delta(A)$-separate $C \subseteq B_1 + \cdots + B_k + \sum_{l=1}^{k} v_l$ with $|C| > \sum_{l=1}^{k}(|B_l|-1)$. From Lemma 6.2 used for $\varepsilon_2 = (|A|-1)\varepsilon$ we get $|A| \geq |C|$ which ends the proof. $\square$

Remark 6.11. Interestingly, only a slightly worse quantitative description of the structure of random variables ξ_{i_l}'s is possible also without assuming that V is a Hilbert space—we will briefly sketch the argument. First replace ξ_{i_l} by

$$Y_l = \sum_{a \in A} \mathbf{1}_{\|\xi_{i_l} - (v_l + a)\| \leq \varepsilon} \cdot (v_l + a)\,.$$

Note that Y_l's are independent again and $\mathbb{P}(\|\xi_{i_l} - Y_l\| > \varepsilon) \leq \delta$. Now observe that all Y_l's take values in a finite-dimensional space W which is a linear span of the set A and vectors v_l's. Thus $\dim(W) \leq |A| + k < 2|A|$. Now it suffices to recall the classical theorem of F. John which states that the Banach-Mazur distance to l_2^N of any N-dimensional Banach space is not greater than $\sqrt{N}$, deal with Y_l's in an appropriate Hilbert space as before, and then transfer the obtained bounds back to the Banach space setting.

We finish by posing a problem which we were not able to solve.

Question 6.12. Is Lemma 6.7 valid in an arbitrary Banach (not only Hilbert) space?

7 Addendum

Most of the results and proofs contained in the present paper were presented at several conferences and seminars, including the second-named author's lecture at the *Analysis of Boolean Functions: New Directions and Applications* Simons Symposium, February 5–11, 2012. During the second-named author's Fall 2013 visit to UC Berkeley's Simons Institute for the Theory of Computing, he learned of new results obtained independently by Aviad Rubinstein and Shmuel (Muli) Safra: Rubinstein's MSc thesis [13], written under the supervision of Safra, and their joint research paper [14]. Let us briefly explain how their results relate to ours. We will need the following estimates.

Lemma 7.1. *Let X and Y be independent square-integrable real random variables such that* $\mathrm{Var}(X+Y) > 0$. *Assume that, for some* $\rho \in [0,1]$, $\mathbb{E}(|X+Y|-1)^2 \leq \rho^2$. *Then* $\mathrm{Var}(X) \leq 800\rho^2/\mathrm{Var}(X+Y)$ *or* $\mathrm{Var}(Y) \leq 800\rho^2/\mathrm{Var}(X+Y)$.

Proof. Let us adapt the proof of Lemma 5.6. Using the same notation as there, let us also set $\sigma = \sqrt{\mathrm{Var}(X+Y)}$. Since $\mathrm{Var}(X) + \mathrm{Var}(Y) = \sigma^2$, without loss of generality we may and will assume that $\mathrm{Var}(X) \geq \sigma^2/2$. Now it suffices to show that $\mathrm{Var}(Y) \leq 800\rho^2\sigma^{-2}$.

Indeed, if $\sigma < 6\sqrt{\rho}$, then

$$\mathrm{Var}(Y) = \sigma^2 - \mathrm{Var}(X) \leq \sigma^2/2 \leq 6^4\rho^2\sigma^{-2}/2 \leq 800\rho^2\sigma^{-2}\,.$$

If, on the other hand, $\sigma \geq 6\sqrt{\rho} \geq 6\rho$, then by (5.3) we have

$$2\sqrt{2\alpha} = \|\phi(X-X')\|_2 \geq \|X-X'\|_2 - 2\rho = \sqrt{2\mathrm{Var}(X)} - 2\rho \geq \sigma - 2\rho \geq \frac{2}{3}\sigma\,,$$

so that $\alpha \ge \sigma^2/18$. Thus, by (5.5), $\beta \le 9\rho^2/\alpha \le 162\rho^2\sigma^{-2}$. Since, by (5.4),

$$\sqrt{2\mathrm{Var}(Y)} = \|Y-Y'\|_2 \le 2\rho + \|\phi(Y-Y')\|_2 = 2\rho + 2\sqrt{2\beta}\,,$$

we arrive at

$$\sqrt{2\mathrm{Var}(Y)} \le 2\rho + 36\rho\sigma^{-1} \le 40\rho\sigma^{-1}\,,$$

the last inequality following from the fact that

$$\sigma \le \|X+Y\|_2 \le 1 + \Big\| |X+Y| - 1 \Big\|_2 \le 1+\rho \le 2\,.$$

Therefore, $\mathrm{Var}(Y) \le 800\rho^2\sigma^{-2}$, and the proof is finished. □

Lemma 7.2. *Let $X_1, X_2, \ldots, X_n$ be independent square-integrable random variables and let $S = \sum_{i=1}^n X_i$. Assume that $\mathbb{E}(|S|-1)^2 \le \rho^2$ for some $\rho \in [0,1]$, and that $\mathrm{Var}(S) > 0$. Then there exists some $k \in [n]$ such that $\mathrm{Var}(S - X_k) \le 1600\rho^2/\mathrm{Var}(S)$.*

Proof. Lemma 7.2 can be deduced from Lemma 7.1 in the same way in which Lemma 5.7 follows from Lemma 5.6. □

Now we are in position to state the following refinement of Theorem 5.8—note that in the case $\mathrm{Var}(f) \le \rho$ the bound of Theorem 5.8 is trivial since always

$$\big\|f - a_0 - a_k\pi_k\big\|_{L^2} \le \big\|f - a_0\big\|_{L^2} = \sqrt{\mathrm{Var}(f)}\,,$$

while for $\mathrm{Var}(f) \ge \rho$ we have $\rho/\sqrt{\mathrm{Var}(f)} \le \sqrt{\rho}$.

Corollary 7.3. *Under* **A & N** *(see Subsection 5.1), there is some $k \in [n]$ such that*

$$\|f - (a_0 + a_k\pi_k)\|_{L^2} \le 41\rho/\sqrt{\mathrm{Var}(f)}, \quad \text{and} \quad \|f - \mathrm{sign}(a_0 + a_k\pi_k)\|_{L^2} \le 82\rho/\sqrt{\mathrm{Var}(f)}\,.$$

Proof. Just as in the proof of Theorem 5.8, we arrive at

$$\begin{aligned}\|f - (a_0 + a_k\pi_k)\|_{L^2}^2 &= \|f - f_{\mathcal{A}}\|_{L^2}^2 + \Big\|\sum_{i\in[n]\setminus\{k\}} a_i\pi_i\Big\|_{L^2}^2 \\ &\le \rho^2 + 1600\rho^2/\mathrm{Var}(S) \le 1601\rho^2/\mathrm{Var}(S)\,,\end{aligned}$$

where we have used Lemma 7.2 and the fact that $\mathrm{Var}(S) = \sum_{i=1}^n a_i^2 \le 1$. The second assertion of the corollary follows from the first one in the same way in which Corollary 5.2 follows from Theorem 5.1. □

The above corollary should be compared to the main result of Rubinstein's MSc thesis, [13, Corollary 10]. Note that the setting introduced in Subsection 5.1 is slightly incompatible with Rubinstein's "pair-wise disjoint subsets of the inputs"—one would need to consider the product measure μ on a more abstract probability space than just $\mathbb{R}^n$ to recover [13, Corollary 10] as a special case of Corollary 7.3. Still, [13, Corollary 10] may be easily deduced from Lemma 7.2 by an obvious modification of the proof of Corollary 7.3, in which $a_i\pi_i$'s should be replaced by Rubinstein's restrictions f_i's and a_0 turned into $\hat{f}(\emptyset)$.

Acknowledgements

Research of the second-named author was partially carried out during his stay at the Fields Institute in Toronto. He gratefully acknowledges the hospitality of the institute. The authors thank Prof. Stanisław Kwapień for providing the reference to the work of Marcinkiewicz and Zygmund.

References

[1] WILLIAM BECKNER: Inequalities in Fourier analysis. *Ann. of Math.*, 102(1):159–182, 1975. [doi:10.2307/1970980] 446, 457

[2] ALINE BONAMI: Étude des coefficients Fourier des fonctions de $L^p(G)$. *Ann. Inst. Fourier*, 20(2):335–402, 1970. EuDML. 446, 457

[3] IRIT DINUR: The PCP theorem by gap amplification. *J. ACM*, 54(3:12), 2007. Preliminary versions in STOC'06 and ECCC. [doi:10.1145/1236457.1236459] 446

[4] 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. [doi:10.1016/S0196-8858(02)00024-6] 446, 455, 458

[5] GUY KINDLER: *Property Testing, PCP and Juntas*. Ph. D. thesis, Tel Aviv University, 2002. Available at author's website. 446, 458, 460

[6] GUY KINDLER AND SHMUEL SAFRA: Noise-resistant Boolean functions are juntas. *Preprint*, 2002. Available at author's website. 446, 458, 460

[7] HERMANN KÖNIG, CARSTEN SCHÜTT, AND NICOLE TOMCZAK-JAEGERMANN: Projection constants of symmetric spaces and variants of Khintchine's inequality. *J. Reine Angew. Math.*, 1999(511):1–42, 1999. [doi:10.1515/crll.1999.511.1] 446

[8] RAFAŁ LATAŁA AND KRZYSZTOF OLESZKIEWICZ: On the best constant in the Khinchin-Kahane inequality. *Studia Math.*, 109(1):101–104, 1994. EuDML. 450

[9] JÓZEF MARCINKIEWICZ AND ANTONI ZYGMUND: Sur les fonctions indépendantes. *Fund. Math.*, 29(1):60–90, 1937. EuDML. 450

[10] PIOTR NAYAR: FKN theorem on the biased cube. *Colloq. Math.*, 137(2):253–261, 2014. [doi:10.4064/cm137-2-9, arXiv:1311.3179] 460, 461

[11] RYAN O'DONNELL: *Analysis of Boolean Functions*. Cambridge University Press, 2014. 447

[12] KRZYSZTOF OLESZKIEWICZ: On a nonsymmetric version of the Khinchine-Kahane inequality. In *Stochastic Inequalities and Applications*, volume 56 of *Progress in Probability*, pp. 157–168. Springer/Birkhäuser, 2003. [doi:10.1007/978-3-0348-8069-5_11] 460

[13] AVIAD RUBINSTEIN: Boolean functions whose Fourier transform is concentrated on pair-wise disjoint subsets of the inputs. Master's thesis. School of Computer Science, Tel-Aviv University, October 2012. 465, 466

[14] AVIAD RUBINSTEIN AND MULI SAFRA: Boolean functions whose Fourier transform is concentrated on pairwise disjoint subsets of the input, 2015. [arXiv:1512.09045] 465

[15] WALTER RUDIN: *Functional Analysis*. McGraw-Hill, Inc., New York, 2nd edition, 1991. 447

[16] STANISŁAW J. SZAREK: On the best constants in the Khinchin inequality. *Studia Math.*, 58(2):197–208, 1976. EuDML. 450

AUTHORS

Jacek Jendrej
Ph. D. student
Centre de Mathématiques Laurent Schwartz
Ecole Polytechnique
jacek.jendrej@polytechnique.edu
http://jacek.jendrej.perso.math.cnrs.fr/

Krzysztof Oleszkiewicz
Professor
Institute of Mathematics
University of Warsaw
koles@mimuw.edu.pl
Web page at U. Warsaw

Jakub O. Wojtaszczyk
Software engineer
Google
onufry@google.com
http://research.google.com/pubs/author58121.html

ABOUT THE AUTHORS

JACEK JENDREJ is a Ph. D. student at Ecole Polytechnique advised by Yvan Martel and Frank Merle. He is interested in nonlinear evolution partial differential equations, especially in near-soliton dynamics for critical nonlinearities, energy estimates and compactness methods.

KRZYSZTOF OLESZKIEWICZ is a professor at the University of Warsaw's Institute of Mathematics, where he received his Ph. D. in 1997 under the supervision of Stanisław Kwapień. His research interests lie in probability theory and analysis, especially in various inequalities, including tail and moment estimates for sums of independent random variables, concentration bounds and hypercontractivity. He has no idea about Computer Science.

JAKUB ONUFRY WOJTASZCZYK is a software engineer at Google. Besides developing cluster-level operating systems, he has some knowledge of algorithms and of probability theory.