

# A structure theorem for Boolean functions with small total influences

By HAMED HATAMI

*Dedicated to the memory of Avner Magen*

## Abstract

We show that on every product probability space, Boolean functions with small total influences are essentially the ones that are almost measurable with respect to certain natural sub-sigma algebras. This theorem in particular describes the structure of monotone set properties that do not exhibit sharp thresholds.

Our result generalizes the core of Friedgut’s seminal work on properties of random graphs to the setting of arbitrary Boolean functions on general product probability spaces and improves the result of Bourgain in his appendix to Friedgut’s paper.

## 1. Introduction

We call a function *Boolean* if its range is  $\{0, 1\}$ . The influence of a variable on a Boolean function measures the sensitivity of the function with respect to the changes in that variable. This basic notion arises naturally in many different areas such as statistical physics and probability theory (e.g., phase transition, percolation), computer science (e.g., complexity theory, hardness of approximation, machine learning), combinatorics (e.g., set systems, products of graphs), economics (e.g., social choice). In many instances, when a Boolean function satisfies some nice properties or fits a description, it is possible to bound the influences of its variables. Therefore, Boolean functions with small total influences arise frequently in various contexts, and they are studied for different purposes.

One of the major motivations for studying Boolean functions with small total influences is a profound connection to the *threshold phenomenon*, discovered by Margulis [Mar74] and Russo [Rus81]. The threshold behavior is the quick transition of a property from being very unlikely to hold to being very likely to hold as a certain parameter  $p$  increases. This behavior occurs in various settings, and it is an instance of the phenomenon of phase transition in statistical physics which explains the rapid change of behavior in many physical processes. One of the main questions that arises in studying phase transitions

is: “How sharp is the threshold?” That is, how short is the interval in which the transition occurs? Margulis [Mar74] and Russo [Rus81] observed that the sharpness of the threshold is controlled by the total influence of the indicator function of the property. Hence in order to characterize the properties that do not exhibit sharp thresholds, one needs to understand the structure of Boolean functions that have small total influences.

Due to these motivations, Boolean functions with small total influences are studied extensively, and some remarkable results about their structure are discovered, e.g., by Kalai, Kahn and Linial [KKL88], Talagrand [Tal94], Bourgain and Kalai [BK97], Friedgut [Fri98], [Fri99], and Bourgain [Bou99]. The KKL inequality [KKL88] and Friedgut’s threshold theorem [Fri99] are both mentioned in [Bou00] as notable consequences of the interaction between harmonic analysis and combinatorics. Let us also mention the more recent work of Mossel, O’Donnell, and Oleszkiewicz [MOO10] that studies Boolean functions where all variables have small influences and now is one of the main tools in the study of the hardness of approximation in theoretical computer science.

The purpose of this article is to essentially characterize Boolean functions with small total influences by showing that every such function is almost measurable with respect to a certain sub-sigma algebra.

*1.1. Notations and definitions.* For every statement  $P$ , we define  $1_{[P]} := 1$  if  $P$  is true, and  $1_{[P]} := 0$  otherwise. For every natural number  $n$ , denote  $[n] := \{1, \dots, n\}$ , and for every set  $S$ , let  $\mathcal{P}(S)$  denote the set of all subsets of  $S$ . Consider a probability space  $X = (\Omega, \mathcal{F}, \mu)$ , and let  $X^n$  denote the corresponding product space endowed with the product probability measure  $\mu^n$ . Throughout this article  $X$  is always a probability space and  $n$  is a positive integer, and all asymptotics are meant for  $n \rightarrow \infty$ .

As usual, we use  $O(X)$  to denote a quantity bounded in magnitude by  $CX$  for some absolute constant  $C$ ; if we need  $C$  to depend on a parameter, we will indicate this by subscripts. Thus, for instance,  $O_I(1)$  is a quantity bounded in magnitude by some expression  $C_I$  depending on  $I$ . We use  $o(X)$  to denote a quantity  $Y$  with  $\lim_{n \rightarrow \infty} Y/X = 0$ .

For every  $x = (x_1, \dots, x_n) \in X^n$ , let  $x_S := (x_i : i \in S) \in X^S$  denote the restriction of  $x$  to the coordinates in  $S$ . For two disjoint sets  $S, T \subseteq [n]$ , and elements  $x \in X^S$  and  $y \in X^T$ , let  $(x, y)$  denote the unique element  $z \in X^{S \cup T}$  with  $z_S = x$  and  $z_T = y$ .

Consider a subset  $S \subseteq [n]$ . In the sequel, by a slight abuse of notation, we will view functions on  $X^S$  as also being functions on  $X^n$ . More precisely, we identify every function  $g : X^S \rightarrow \mathbb{R}$  with the function on  $X^n$  that maps every  $x \in X^n$  to  $g(x_S)$ .

Let  $f : X^n \rightarrow \{0, 1\}$  be a measurable function. For  $1 \leq j \leq n$ , the *influence* of the  $j$ -th variable on  $f$  is defined as

$$(1) \quad I_f(j) = \Pr[f(x_1, \dots, x_{j-1}, x_j, x_{j+1}, \dots, x_n) \neq f(x_1, \dots, x_{j-1}, y_j, x_{j+1}, \dots, x_n)],$$

where  $x_1, \dots, x_n, y_1, \dots, y_n$  are independent and identically distributed random variables taking values in  $X$  according to its probability measure. The *total influence* of  $f$ , denoted by  $I_f$ , is defined as

$$(2) \quad I_f := \sum_{j=1}^n I_f(j).$$

*Remark 1.1.* Let  $X = (\Omega, \mathcal{F}, \mu)$ , and let  $f : X^n \rightarrow \{0, 1\}$  be measurable. It follows from the measurability of  $f$  that the set

$$\{(x_1, \dots, x_n, y_j) : f(x_1, \dots, x_{j-1}, x_j, x_{j+1}, \dots, x_n) \neq f(x_1, \dots, x_{j-1}, y_j, x_{j+1}, \dots, x_n)\} \subseteq X^{n+1}$$

is measurable, and thus  $I_f(j)$  is well defined. Moreover in many situations, to prove an statement about influences, one can assume that  $\Omega$  is a finite set. Indeed since  $f$  is measurable, it is possible to find a *finite* sub- $\sigma$ -algebra  $\mathcal{G}$  of  $\mathcal{F}$  and a function  $g : \Omega^n \rightarrow \{0, 1\}$ , measurable with respect to the product  $\sigma$ -algebra generated by  $\mathcal{G}$ , such that  $\Pr[f(x) \neq g(x)]$  is arbitrarily small. Since  $|I_f(j) - I_g(j)| \leq 2\Pr[f(x) \neq g(x)]$ , the differences between the influences are also small.

A particular case of interest is when  $X$  is defined by the Bernoulli distribution  $\mu_p$  on  $\{0, 1\}$  with parameter  $0 < p < 1$ , i.e.,  $\mu_p(\{1\}) = p$  and  $\mu_p(\{0\}) = 1 - p$ . We refer to the corresponding product probability distribution  $\mu_p^n$  on  $\{0, 1\}^n$  as the  *$p$ -biased distribution*.

A function  $f : \{0, 1\}^n \rightarrow \{0, 1\}$  is called *increasing* if  $f(x) \leq f(y)$ , for every  $x, y \in \{0, 1\}^n$  satisfying  $x_i \leq y_i$  for every  $i \in [n]$ . For a set  $A \subseteq [n]$ , we say that a function  $f : X^n \rightarrow \mathbb{R}$  depends only on the coordinates in  $A$  if  $f(x) = f(y)$  for every  $x, y \in X^n$  with  $x_A = y_A$ .

## 2. Main results

Margulis [Mar74] and Russo [Rus81] observed that for every increasing function  $f : \{0, 1\}^n \rightarrow \{0, 1\}$ , we have

$$(3) \quad 2p(1 - p) \frac{d\mu_p(f)}{dp} = I_f,$$

where  $\mu_p(f) := \int f(x) d\mu_p^n(x)$  and  $I_f$  is defined according to the  $p$ -biased distribution. This shows that the rate at which  $\mu_p(f)$  increases with respect to the increase in  $p$  is controlled by  $I_f$ . Consequently, in order to characterize the

properties that do not exhibit sharp thresholds, one needs to understand the structure of Boolean functions that have small total influences. We refer the reader to [Fri99] for more details.

The problem of finding general conditions under which a sharp threshold does not occur is first investigated by Russo [Rus81], [Rus82]. Later, Talagrand [Tal94], extending the work of Russo, showed that for  $p$  that is not too small, if the total influence of  $f$  is small, then there are variables with large influences.

In the setting of the  $p$ -biased distribution, for sufficiently large  $p$ , the works of Talagrand [Tal94], Friedgut and Kalai [FK96], Bourgain and Kalai [BK97], and Friedgut [Fri98] provide a good understanding of the situation. Intuitively these results say that the total influence of  $f$  is large, unless the value of  $f(x)$  is determined only by “local information” about  $x$ , e.g., by a few number of coordinates. As the following simple example shows, these results turn out however not to be useful when  $p$  is small, in particular when  $\log \frac{1}{p} \sim \log n$ , often the case in applications.

*Example 2.1.* Set  $p = n^{-1}$ , and let  $f : \{0, 1\}^n \rightarrow \{0, 1\}$  be defined as  $f(x) = 1$  if and only if  $x \neq (0, \dots, 0)$ . Then  $I_f(1) = \dots = I_f(n) \leq 2p$ , and so  $I_f \leq 2$ . However note that there is no variable with large influence, and also  $f$  does not depend only on a small set of its coordinates. Indeed for every constant size set  $A \subseteq [n]$ , we have

$$\mathbb{E}[f(x)|x_A = (0, \dots, 0)] = 1 - (1 - p)^{n-|A|} = 1 - \frac{1}{e} \pm o(1),$$

where  $x$  is a random variable taking values in  $\{0, 1\}^n$  according to the  $p$ -biased distribution. Since  $\Pr[x_A = (0, \dots, 0)] \geq 1 - |A|p = 1 - o(1)$ , we conclude that for every function  $g$  that depends only on the coordinates in  $A$ , we have  $\|f - g\|_1 \geq \frac{1}{e} - o(1)$ .

No essential progress on the case of small  $p$  was made until the breakthrough work of Friedgut [Fri99]. He gave a satisfactory description for functions with small total influences when they correspond to graph or hypergraph properties. Friedgut’s theorem is now an important tool in the study of the threshold behavior of graph properties (see [Fri05]). A graph property on  $n$ -vertex graphs is modeled by a function  $f : \{0, 1\}^{\binom{n}{2}} \rightarrow \{0, 1\}$ , where each of the  $\binom{n}{2}$  coordinates correspond to the presence of one of the  $\binom{n}{2}$  possible edges. Since a graph property is invariant under graph isomorphisms, every such function  $f$  is invariant under several permutations of its coordinates. Various steps of Friedgut’s proof rely heavily on these symmetry assumptions and do not extend to the more general settings. However he conjectured [Fri99] that his theorem holds without requiring any symmetry assumptions.

In Theorem 2.6 we generalize that the core of Friedgut’s work [Fri99] form graph properties on their corresponding  $p$ -biased probability space to general properties (with no symmetry assumptions) on general product probability spaces. Previously, in this general setting, the situation was only partially understood by a result of Bourgain.

**THEOREM 2.2** ([Bou99, Prop. 1]). *Let  $f : (\{0, 1\}^n, \mu_p^n) \rightarrow \{0, 1\}$  be an increasing function with  $\int f = \alpha > 0$ , and suppose  $p = o_{I_f}(1)$ . There exist  $\delta(\alpha), C(\alpha) > 0$  and a subset  $S \subseteq [n]$  with  $|S| \leq C(\alpha)I_f$  such that*

$$\mathbb{E}[f(x)|x_S = (1, \dots, 1)] \geq \alpha + \delta(\alpha),$$

where  $x$  is a random variable taking values in  $\{0, 1\}^n$  according to the  $p$ -biased distribution.

Let  $\mathcal{J} = \{J_S\}_{S \subseteq [n]}$  be a collection of measurable functions  $J_S : X^S \rightarrow \{0, 1\}$ . Define the map  $J_{\mathcal{J}} : X^n \rightarrow \mathcal{P}([n])$  as  $J_{\mathcal{J}} : x \mapsto \bigcup_{S \subseteq [n], J_S(x)=1} S$ , and note that the map  $x \mapsto (J_{\mathcal{J}}(x), x_{J_{\mathcal{J}}(x)})$  is measurable. Let  $\mathcal{F}_{\mathcal{J}}$  be the sub- $\sigma$ -algebra on  $X^n$  induced by the map  $x \mapsto (J_{\mathcal{J}}(x), x_{J_{\mathcal{J}}(x)})$ ; i.e., it is the coarsest  $\sigma$ -algebra on  $X^n$  that makes this map measurable.

*Definition 2.3.* For a constant  $k > 0$ , a  $k$ -pseudo-junta is a function  $f : X^n \rightarrow \{0, 1\}$  that is measurable with respect to  $\mathcal{F}_{\mathcal{J}}$  for some  $\mathcal{J}$  satisfying

$$\int |J_{\mathcal{J}}(x)|dx \leq k.$$

The functions that depend on a small number of coordinates are usually referred to as *juntas*. The following example explains the choice of the name “pseudo-junta” in Definition 2.3.

*Example 2.4.* Consider a set  $A \subseteq [n]$  with  $|A| \leq k$ . Every measurable function  $f : X^n \rightarrow \{0, 1\}$  that depends only on coordinates in  $A$  is a  $k$ -pseudo-junta. Indeed let  $\mathcal{J}$  be defined by setting  $J_S \equiv 1$  if  $S = A$ , and  $J_S \equiv 0$  otherwise. Then  $J_{\mathcal{J}} \equiv A$ , and hence  $f$  is measurable with respect to  $\mathcal{F}_{\mathcal{J}}$ . Furthermore  $\int |J_{\mathcal{J}}(x)|dx = k$ .

The following simple proposition shows that  $k$ -pseudo-juntas have small total influences.

**PROPOSITION 2.5.** *Let  $f : X^n \rightarrow \{0, 1\}$  be a  $k$ -pseudo-junta. Then  $I_f \leq 2k$ .*

*Proof.* Since  $f$  is a  $k$ -pseudo-junta, it is measurable with respect to  $\mathcal{F}_{\mathcal{J}}$  for some  $\mathcal{J}$  with  $\int |J_{\mathcal{J}}(x)| \leq k$ . Let  $x_1, \dots, x_n, y_1, \dots, y_n$  be independent and identically distributed random variables taking values in  $X$  according to its probability measure, and define the random variables  $x = (x_1, \dots, x_n)$  and  $x^{(j)} = (x_1, \dots, x_{j-1}, y_j, x_{j+1}, \dots, x_n)$  for every  $j \in [n]$ . From the definition of

$J_{\mathcal{J}}$  and the assumption that  $f$  is measurable with respect to  $\mathcal{F}_{\mathcal{J}}$ , it follows that  $j \in J_{\mathcal{J}}(x) \cup J_{\mathcal{J}}(x^{(j)})$  if  $f(x) \neq f(x^{(j)})$ . Hence

$$\begin{aligned} I_f &= \sum_{j \in [n]} \Pr[f(x) \neq f(x^{(j)})] \leq \sum_{j \in [n]} \Pr[j \in J_{\mathcal{J}}(x) \cup J_{\mathcal{J}}(x^{(j)})] \\ &\leq 2 \sum_{j \in [n]} \Pr[j \in J_{\mathcal{J}}(x)] \leq 2 \int |J_{\mathcal{J}}(x)| dx \leq 2k. \quad \square \end{aligned}$$

Our main result is an inverse theorem that essentially says that the inverse of Proposition 2.5 is also true.

**THEOREM 2.6 (Main Theorem).** *Consider a measurable function  $f : X^n \rightarrow \{0, 1\}$ . For every  $\epsilon > 0$ , there exists a  $e^{10^{15}\epsilon^{-3}\lceil I_f \rceil^3}$ -pseudo-junta  $h : X^n \rightarrow \{0, 1\}$  such that  $\|f - h\|_1 \leq \epsilon$ .*

The key point in Theorem 2.6 is that the bound  $e^{10^{15}\epsilon^{-3}\lceil I_f \rceil^3}$  has no dependence on  $n$ .

*Example 2.7.* Consider the function  $f$  in Example 2.1. Define  $\mathcal{J} = \{J_S\}_{S \subseteq [n]}$  as follows. For every  $S \subseteq [n]$ , let  $J_S : \{0, 1\}^S \rightarrow \{0, 1\}$  be defined as

$$J_S : x \mapsto \begin{cases} 1 & x_S = (1, \dots, 1), \\ 0 & \text{otherwise.} \end{cases}$$

Note that  $J_{\mathcal{J}} : x \mapsto \{i : x_i = 1\}$ , and so  $\int |J_{\mathcal{J}}(x)| dx = pn = 1$ . Furthermore  $\mathcal{F}_{\mathcal{J}}$  is the original discrete  $\sigma$ -algebra on  $\{0, 1\}^n$ , and hence  $f$  is measurable with respect to it. So for this example, in Theorem 2.6 one can take  $h = f$ .

Due to the connection to the threshold phenomenon, the case of the  $p$ -biased distribution in Theorem 2.6 is of particular interest. For this probability space, the proof of Theorem 2.6 can be simplified significantly since many steps in the proof are to address the difficulties that arise in dealing with general probability spaces. Furthermore this simpler proof provides some improvement over the bound  $e^{10^{15}\epsilon^{-3}\lceil I_f \rceil^3}$  of Theorem 2.6. Therefore we first state and prove the main theorem in the case of the  $p$ -biased measure, and then in Section 5 we present the more complicated proof of the general case.

**THEOREM 2.8 (Main Theorem, the  $p$ -biased case).** *Consider a function  $f : (\{0, 1\}^n, \mu_p^n) \rightarrow \{0, 1\}$ . For every  $\epsilon > 0$ , there exists a  $e^{10^{10}\epsilon^{-2}\lceil I_f \rceil^2}$ -pseudo-junta  $h : (\{0, 1\}^n, \mu_p^n) \rightarrow \{0, 1\}$  such that  $\|f - h\|_1 \leq \epsilon$ .*

*Remark 2.9.* In Theorem 2.8 the function  $h$  is a pseudo-junta, and so it is measurable with respect to a  $\sigma$ -algebra  $\mathcal{F}_{\mathcal{J}}$  for a certain collection  $\mathcal{J}$  of functions  $J_S : \{0, 1\}^S \rightarrow \{0, 1\}$ . It follows from the proof of Theorem 2.8 that for  $p \leq 1/2$ , it is possible to assume that  $J_S$  are increasing. Similarly for  $p \geq 1/2$ , it is possible to assume that  $J_S$  are decreasing.

2.1. *Increasing set properties.* Let  $p = o(1)$  and  $f : (\{0, 1\}^n, \mu_p^n) \rightarrow \{0, 1\}$  be an *increasing* function with  $\int f = \alpha > 0$ . Bourgain’s result (Theorem 2.2) roughly speaking says that when  $I_f$  is small, it is possible to assign the value 1 to a small set of coordinates such that restricting to this assignment increases the expected value of  $f$  in a nonnegligible amount. The next corollary to Theorem 2.8 improves this result, as it shows that it is possible to increase the expected value to arbitrarily close to 1.

COROLLARY 2.10. *Let  $f : (\{0, 1\}^n, \mu_p^n) \rightarrow \{0, 1\}$  be an increasing function with  $\alpha = \int f$  and  $\epsilon > 0$  be a constant. If  $p \leq \frac{1}{2}$ , then there exists a subset  $S \subseteq [n]$  with  $|S| \leq e^{10^{12} \lceil I_f \rceil^2 \alpha^{-2} \epsilon^{-2}}$  such that*

$$\mathbb{E}[f(x)|x_S = (1, \dots, 1)] \geq 1 - \epsilon,$$

where  $x$  is a random variable taking values in  $\{0, 1\}^n$  according to the  $p$ -biased distribution.

*Proof.* Theorem 2.8 shows that there exists a collection  $\mathcal{J} = \{J_S\}_{S \subseteq [n]}$  with

$$\int |J_{\mathcal{J}}(x)| \leq e^{10^{11} \lceil I_f \rceil^2 \alpha^{-2} \epsilon^{-2}},$$

and a function  $h : \{0, 1\}^n \rightarrow \{0, 1\}$ , measurable with respect to  $\mathcal{F}_{\mathcal{J}}$ , such that  $\|f - h\|_1 \leq \frac{\alpha \epsilon}{2}$ . Set  $k := e^{10^{12} \lceil I_f \rceil^2 \alpha^{-2} \epsilon^{-2}}$ , and set

$$\beta := \max_{\substack{S \subseteq [n]: |S| \leq k \\ y \in \{0, 1\}^S}} \Pr[f(x) = 1 | x_S = y \text{ and } J_{\mathcal{J}}(x) = S],$$

where  $x$  is a random variable taking values in  $\{0, 1\}^n$  according to the  $p$ -biased distribution. Since  $f$  and  $h$  are Boolean and  $h$  is measurable with respect to  $\mathcal{F}_{\mathcal{J}}$ , we have

$$\begin{aligned} \|f - h\|_1 &= \int 1_{[f \neq h]} \geq \int 1_{[|J_{\mathcal{J}}| \leq k]} 1_{[f=0]} 1_{[h=1]} + \int 1_{[|J_{\mathcal{J}}| \leq k]} 1_{[f=1]} 1_{[h=0]} \\ &\geq (1 - \beta) \int 1_{[|J_{\mathcal{J}}| \leq k]} 1_{[h=1]} + \int 1_{[|J_{\mathcal{J}}| \leq k]} 1_{[f=1]} 1_{[h=0]} \\ &\geq (1 - \beta) \int 1_{[|J_{\mathcal{J}}| \leq k]} 1_{[f=1]} \geq (1 - \beta) \left( \int f - \int 1_{[|J_{\mathcal{J}}| > k]} \right) \\ &\geq (1 - \beta) \left( \alpha - k^{-1} \int |J_{\mathcal{J}}| \right) \geq \frac{3}{4} (1 - \beta) \alpha, \end{aligned}$$

which together with  $\|f - h\|_1 \leq \frac{\alpha \epsilon}{2}$  implies that  $\beta \geq 1 - \epsilon$ . Hence there exists a subset  $S \subseteq [n]$  with  $|S| \leq k$  and an element  $y_0 \in \{0, 1\}^S$  such that

$$(4) \quad \mathbb{E}[f(x)|x_S = y_0 \text{ and } J_{\mathcal{J}}(x) = S] \geq 1 - \epsilon,$$

where  $x$  is a random variable taking values in  $\{0, 1\}^n$  according to the  $p$ -biased distribution. Define  $h_1, h_2 : \{0, 1\}^{[n] \setminus S} \rightarrow \{0, 1\}$  as  $h_1 : x \mapsto f(y_0, x)$  and

$h_2 : x \mapsto 1_{[J_{\mathcal{J}}(y_0, x)=S]}$ . We can rewrite (4) as

$$(5) \quad \mathbb{E}[h_1(z)|h_2(z) = 1] \geq 1 - \epsilon,$$

where  $z$  is a random variable taking values in  $\{0, 1\}^{[n] \setminus S}$  according to the  $p$ -biased distribution. Since  $f$  is increasing,  $h_1$  is increasing. Furthermore by Remark 2.9, we can assume that the functions  $J_S$  are increasing. Thus  $h_2$  is decreasing, and then it follows from the classical FKG inequality (see [AS00]) that

$$\begin{aligned} \mathbb{E}[f(x)|x_S = (1, \dots, 1)] &\geq \mathbb{E}[f(x)|x_S = y_0] \\ &= \mathbb{E}[h_1(z)] \geq \mathbb{E}[h_1(z)|h_2(z) = 1] \geq 1 - \epsilon. \end{aligned} \quad \square$$

The following example shows that Corollary 2.10 fails to be valid when the function  $f$  is not necessarily increasing.

*Example 2.11.* Set  $p = n^{-1}$ , and let  $f : \{0, 1\}^n \rightarrow \{0, 1\}$  be defined as  $f(x) = 1$  if and only if  $\sum_{i=1}^n x_i \equiv 0 \pmod{2}$ . Similar to Example 2.1,  $I_f(1) = \dots = I_f(n) \leq 2p$ , and so  $I_f \leq 2$ . Now consider a constant size  $A \subseteq [n]$  and any  $y \in \{0, 1\}^A$ . Let  $a \in \{0, 1\}$  be such that  $a \equiv \sum_{i \in A} y_i \pmod{2}$ . Then we have

$$\begin{aligned} \frac{1}{e} - o(1) &\leq (1-p)^{n-|A|} \leq \mathbb{E}[f(x) = a|x_A = y] \\ &\leq 1 - (n-|A|)p(1-p)^{n-|A|-1} \leq 1 - \frac{1}{e} + o(1), \end{aligned}$$

where  $x$  is a random variable taking values in  $\{0, 1\}^n$  according to the  $p$ -biased distribution. So for every sufficiently large  $n$ , not only is  $f$  far from being determined by the coordinates in  $A$ , but also for every  $y \in \{0, 1\}^A$ , both  $\mathbb{E}[f(x) = 0|x_A = y]$  and  $\mathbb{E}[f(x) = 1|x_A = y]$  are well separated from 0 and 1 as they both belong to the interval  $[\frac{1}{2e}, 1 - \frac{1}{2e}]$ .

### 3. Generalized Walsh expansion

In this short section we review some basic facts about the generalized Walsh expansions that are first defined by Hoeffding in [Hoe48] (see also [ES81]). Let  $L_2(X^n)$  denote the set of functions  $f : X^n \rightarrow \mathbb{C}$  that satisfy  $\int |f(x)|^2 dx < \infty$ . Consider a subset  $S \subseteq [n]$  and a function  $f \in L_2(X^n)$ . Then  $\int f(x) dx_S := \int f(x) \prod_{i \in S} dx_i$  denotes the integral with respect to the coordinates in  $S$ .

*Definition 3.1.* The *generalized Walsh expansion* of a function  $f \in L_2(X^n)$  is the unique expansion  $f = \sum_{S \subseteq [n]} F_S$  that satisfies the following two properties:

- (i) For every  $S \subseteq [n]$ , the function  $F_S$  depends only on the coordinates in  $S$ , i.e.,  $F_S(x) = F_S(x_S)$ ;
- (ii)  $\int F_S(x) dx_i \equiv 0$  for every  $S \subseteq [n]$  and every  $i \in S$ .



Note that it follows from Definition 3.1(i) and (ii) that for every  $T \subseteq [n]$ , we have  $\int f dx_{[n]\setminus T} = \sum_{S \subseteq T} F_S$ . Consequently for every  $y \in X^n$ ,

$$(6) \quad F_S(y) = \sum_{T \subseteq S} (-1)^{|S \setminus T|} \int f(y_T, x_{[n]\setminus T}) dx_{[n]\setminus T},$$

which shows that the generalized Walsh expansion is unique. It follows from (6) that for every  $S \subseteq [n]$ , we have

$$(7) \quad \|F_S\|_\infty \leq 2^{|S|} \|f\|_\infty.$$

Consider two subsets  $S_1, S_2 \subseteq [n]$ . If  $S_1 \neq S_2$ , then Definition 3.1(i) and (ii) guarantee that  $\int F_{S_1} \overline{F_{S_2}} = 0$ , or in other words the functions  $\{F_S : S \subseteq [n]\}$  are pairwise orthogonal. As a consequence we have Parseval's identity:

$$(8) \quad \|f\|_2^2 = \sum_{S \subseteq [n]} \|F_S\|_2^2.$$

The influences of variables have simple descriptions in terms of the generalized Walsh expansion. For a measurable function  $f : X^n \rightarrow \{0, 1\}$ , and every  $i \in [n]$ , define

$$f^{(i)} = f - \int f dx_i = \sum_{S:i \in S} F_S.$$

It is easy to see that  $I_f(i) = 2\|f^{(i)}\|_2^2$ , which by Parseval's identity implies

$$I_f(i) = 2 \sum_{S:i \in S} \|F_S\|_2^2.$$

Thus the total influence of  $f$  is given by the formula

$$(9) \quad I_f = 2 \sum_{S \subseteq [n]} |S| \|F_S\|_2^2.$$

#### 4. The $p$ -biased case: Proof of Theorem 2.8

Without loss of generality we assume  $p \leq \frac{1}{2}$ . We start from the generalized Walsh expansion  $f = \sum_{S \subseteq [n]} F_S$ . In the first two steps we simplify this expansion by removing some insignificant terms from it. We will arrive at a set  $\mathcal{S} \subseteq \mathcal{P}([n])$  such that  $\|f - \sum_{S \in \mathcal{S}} F_S\|_2$  is small and meanwhile the functions  $F_S$  with  $S \in \mathcal{S}$  satisfy certain properties. Then in the main step of the proof we define the collection  $\mathcal{J} = \{J_S\}_{S \subseteq [n]}$  and show that  $\|g - \mathbb{E}[g|\mathcal{F}_{\mathcal{J}}]\|_2$  is small for  $g := \sum_{S \in \mathcal{S}} F_S$ . Once this is established, it is straightforward to finish the proof. In the case of the  $p$ -biased distribution, the functions  $F_S$  have simple descriptions: There exist real constants  $\{\hat{f}(S)\}_{S \subseteq [n]}$  such that  $F_S(x) = \hat{f}(S) \prod_{i \in S} r(x_i)$  where  $r(0) = -\sqrt{\frac{p}{1-p}}$  and  $r(1) = \sqrt{\frac{1-p}{p}}$ . Note that  $\hat{f}(S) = \int f(x) \prod_{i \in S} r(x_i) dx$ , which implies  $|\hat{f}(S)| \leq p^{|S|/2}$ . These properties of  $F_S$  are crucial in this proof.

We abbreviate  $\mu_p^n$  to  $\mu$ . We will define various constants, which for the convenience of the reader are listed here:

$$\begin{aligned} C &:= \lceil I_f \rceil, & \epsilon_0 &:= 10^{-3}\epsilon, & k &:= C\epsilon_0^{-1} = 10^3\epsilon^{-1}C, \\ \delta &:= 2^{-10^2k^2}, & \epsilon_1 &:= 3^{-10k^2}\epsilon_0^{10k}. \end{aligned}$$

Set  $C, k, \epsilon_0$  as above, and let  $\epsilon_1 > 0$  be a constant to be determined later. Let

$$\mathcal{S} := \{S \subseteq [n] : |S| \leq k, \|F_S\|_\infty > \epsilon_1\}.$$

We wish to show that the total contribution of  $F_S$  with  $S \notin \mathcal{S}$  to  $\|f\|_2^2$  is insignificant. First we deal with  $S$  with  $|S| > k$ .

4.1. *Step I: High frequencies.* By (8) and (9), we have

$$(10) \quad \sum_{S:|S|>k} \|F_S\|_2^2 \leq \frac{I_f}{k} \leq \epsilon_0.$$

4.2. *Step II: Bourgain’s argument.* Next we deal with  $S \notin \mathcal{S}$  with  $|S| \leq k$ . This step is not self-contained and is based on Bourgain’s argument in [Bou99]. Following his proof (but substituting  $k$  in place of  $10C$ ), one can replace [Bou99, inequality (2.17)] with the following:

$$\begin{aligned} \int \sum_{S:|S|\leq k} F_S^2 1_{\|F_S\|\leq\epsilon_1} &\leq M^k \int \max_{S:|S|\leq k} F_S^2 1_{\|F_S\|\leq\epsilon_1} + 2C3^{k/2}\epsilon^{2/3} + C^{1/2} \frac{3^k}{M^{1/2}\epsilon} \\ &\leq M^k \epsilon_1^2 + 2C3^{k/2}\epsilon^{2/3} + C^{1/2} \frac{3^k}{M^{1/2}\epsilon}, \end{aligned}$$

where  $\epsilon_1, \epsilon, M > 0$  are arbitrary constants. Then by setting  $\epsilon := 3^{-3k}\epsilon_0^3$ ,  $M := 3^{10k}\epsilon_0^{-10}$ , and  $\epsilon_1 := 3^{-10k^2}\epsilon_0^{10k}$ , we obtain

$$(11) \quad \sum_{S:|S|\leq k, S \notin \mathcal{S}} \|F_S\|_2^2 \leq \int \sum_{S:|S|\leq k} F_S^2 1_{\|F_S\|\leq\epsilon_1} \leq \frac{\epsilon_0}{k}.$$

4.3. *Step III: Main Step.* Next we define the collection of functions  $\mathcal{J} = \{J_T\}_{T \subseteq [n]}$ . Set  $\delta := 2^{-10^2k^2}$ , and for every  $T \subseteq [n]$  with  $|T| \leq k$ , define  $J_T : \{0, 1\}^T \rightarrow \{0, 1\}$  as

$$J_T : y \mapsto \begin{cases} 1 & \sum_{S \in \mathcal{S}, S \supseteq T} \int 1_{\|F_S\| \geq \epsilon_1} \geq \delta \mu(y), \\ 0 & \text{otherwise.} \end{cases}$$

For  $T \subseteq [n]$  with  $|T| > k$ , define  $J_T \equiv 0$ . Note that since  $p \leq \frac{1}{2}$ , the functions  $J_T$  are increasing. The required bound on  $\int |J_{\mathcal{J}}(x)|$  can be verified easily:

$$\begin{aligned} \int |J_{\mathcal{J}}| &\leq \sum_{T \subseteq [n], |T| \leq k} |T| \int J_T \leq k \sum_{T:|T| \leq k} \int J_T \\ &\leq \frac{k}{\delta} \sum_{T:|T| \leq k} \sum_{S \in \mathcal{S}, S \supseteq T} \int 1_{\|F_S\| \geq \epsilon_1} \int \frac{1}{\mu(x_T)} \end{aligned}$$

$$\begin{aligned}
 &= \frac{k}{\delta} \sum_{T:|T|\leq k} \sum_{S\in\mathcal{S}, S\supseteq T} 2^{|T|} \int 1_{\|F_S\|\geq\epsilon_1} \\
 &\leq \frac{k2^{2k}}{\delta} \sum_{S\in\mathcal{S}} \frac{\|F_S\|_2^2}{\epsilon_1^2} \leq \frac{k2^{2k}}{\delta\epsilon_1^2} \leq 2^{10^3k^2} \leq e^{10^{10}\epsilon^{-2}\lceil I_f \rceil^2}.
 \end{aligned}$$

By (10) and (11) for  $g := \sum_{S\in\mathcal{S}} F_S$ , we have  $\|f - g\|_2^2 \leq 2\epsilon_0$ . Our goal is now to show that  $\|g - \mathbb{E}[g|\mathcal{F}_{\mathcal{J}}]\|_2^2$  is small. Note that  $\mathbb{E}[g|\mathcal{F}_{\mathcal{J}}] = \sum_{S\in\mathcal{S}} \mathbb{E}[F_S|\mathcal{F}_{\mathcal{J}}]$ . However since  $\mathcal{F}_{\mathcal{J}}$  depends on all coordinates, it is difficult to analyze  $\mathbb{E}[F_S|\mathcal{F}_{\mathcal{J}}]$  directly. To remedy this we define some auxiliary  $\sigma$ -algebras. For every  $S \subseteq [n]$ , define the collection  $\mathcal{J}_S := \{\tilde{J}_T\}_{T\subseteq[n]}$  in the following way. For every  $T \subseteq [n]$ , we set  $\tilde{J}_T := J_T$  if  $T \subseteq S$ , and  $\tilde{J}_T \equiv 0$  otherwise. Define  $\tilde{g} := \sum_{S\in\mathcal{S}} \tilde{F}_S$ , where  $\tilde{F}_S := \mathbb{E}[F_S|\mathcal{F}_{\mathcal{J}_S}]$ . For every  $S \subseteq [n]$ , the function  $\tilde{F}_S$  depends only on the coordinates in  $S$  and furthermore the  $\sigma$ -algebra  $\mathcal{F}_{\mathcal{J}_S}$  is coarser than the  $\sigma$ -algebra  $\mathcal{F}_{\mathcal{J}}$ . It follows from the latter that  $\tilde{g}$  is measurable with respect to  $\mathcal{F}_{\mathcal{J}}$ , and hence

$$(12) \quad \|g - \mathbb{E}[g|\mathcal{F}_{\mathcal{J}}]\|_2 \leq \|g - \tilde{g}\|_2.$$

So in order to bound  $\|g - \mathbb{E}[g|\mathcal{F}_{\mathcal{J}}]\|_2$ , it suffices to bound  $\|g - \tilde{g}\|_2$ . Trivially for every  $S \in \mathcal{S}$ ,

$$\int F_S \tilde{F}_S = \int \tilde{F}_S^2.$$

For  $S_1, S_2 \in \mathcal{S}$ , since  $\tilde{F}_{S_1}$  and  $\tilde{F}_{S_2}$  depend respectively only on the coordinates in  $S_1$  and  $S_2$ , if  $S_1 \not\subseteq S_2$  and  $S_2 \not\subseteq S_1$ , then by Definition 3.1(ii),

$$\int F_{S_1} F_{S_2} = \int F_{S_1} \tilde{F}_{S_2} = \int \tilde{F}_{S_1} F_{S_2} = 0.$$

Similarly if  $S_1 \cap S_2 = \emptyset$ , then

$$\int \tilde{F}_{S_1} \tilde{F}_{S_2} = \left(\int \tilde{F}_{S_1}\right) \left(\int \tilde{F}_{S_2}\right) = \left(\int F_{S_1}\right) \left(\int F_{S_2}\right) = 0.$$

Thus

$$\|g - \tilde{g}\|_2^2 = \int \left(\sum_{S\in\mathcal{S}} F_S - \tilde{F}_S\right)^2 \leq (k+1) \sum_{r=0}^k \int \left(\sum_{S\in\mathcal{S}, |S|=r} F_S - \tilde{F}_S\right)^2$$

$$= (k+1) \int \sum_{S\in\mathcal{S}} F_S^2 - \tilde{F}_S^2 + \sum_{\substack{S_1, S_2 \in \mathcal{S}, S_1 \cap S_2 \neq \emptyset \\ S_1 \neq S_2, |S_1|=|S_2|}} \tilde{F}_{S_1} \tilde{F}_{S_2}$$

$$(13) \quad \leq 2k \int \sum_{S\in\mathcal{S}} F_S^2 - \tilde{F}_S^2$$

$$(14) \quad + 2k \sum_{\substack{S_1, S_2 \in \mathcal{S}, S_1 \neq S_2 \\ S_1 \cap S_2 \neq \emptyset}} \left| \int \tilde{F}_{S_1} \tilde{F}_{S_2} \right|.$$

*Bounding (13):* Note that if  $|F_S(x)| \geq \epsilon_1$ , then  $J_S(x) = 1$  and hence  $F_S(x) = \tilde{F}_S(x)$ . Consequently by (11), we have

$$(13) = 2k \int \sum_{S \in \mathcal{S}} |F_S - \tilde{F}_S|^2 \leq 2k \int \sum_{S \in \mathcal{S}} |F_S|^2 1_{[|F_S| \leq \epsilon_1]} \leq 2k \frac{\epsilon_0}{k} \leq 2\epsilon_0.$$

*Bounding (14):* First let us prove a simple inequality. Consider a set  $S \subseteq [n]$ , and note that for every  $x \in \{0, 1\}^S$ ,

$$\mu(x)|F_S(x)| = (p(1-p))^{|S|/2} |\hat{f}(S)|$$

is a constant that does not depend on  $x$ . This shows that  $|\tilde{F}_S(x)| \leq 2^{|S|} |F_S(x)|$  for every  $x \in \{0, 1\}^S$ . It follows that for every subset  $T \subseteq S$ , and every  $y \in \{0, 1\}^T$ , we have

$$(15) \quad \left| \int \tilde{F}_S(y, x_{S \setminus T}) dx_{S \setminus T} \right| \leq 2^{|S|} \int |F_S(y, x_{S \setminus T})| dx_{S \setminus T} \\ = 2^{|S|+|S \setminus T|} \frac{(p(1-p))^{|S|/2} |\hat{f}(S)|}{\mu(y)} \leq 2^{2|S|} \frac{p^{|S|}}{\mu(y)}.$$

Moreover if  $T \subsetneq S$  and  $J_T(y) = 1$ , then

$$(16) \quad \int \tilde{F}_S(y, x_{S \setminus T}) dx_{S \setminus T} = \int F_S(y, x_{S \setminus T}) dx_{S \setminus T} = 0.$$

Consider *distinct*  $S_1, S_2 \in \mathcal{S}$  and  $y \in \{0, 1\}^T$  where  $T := S_1 \cap S_2$ . If  $J_T(y) = 1$ , then by (16),

$$\int \tilde{F}_{S_1}(y, x_{[n] \setminus T}) \tilde{F}_{S_2}(y, x_{[n] \setminus T}) dx_{[n] \setminus T} = 0.$$

Hence by (15), we have

$$\left| \int \tilde{F}_{S_1} \tilde{F}_{S_2} \right| = \left| \int \tilde{F}_{S_1}(x_T, x_{S_1 \setminus T}) \tilde{F}_{S_2}(x_T, x_{S_2 \setminus T}) 1_{[J_T(x_T)=0]} \right| \\ = \left| \int \left( \int \tilde{F}_{S_1}(x_T, x_{S_1 \setminus T}) dx_{S_1 \setminus T} \right) \right. \\ \left. \times \left( \int \tilde{F}_{S_2}(x_T, x_{S_2 \setminus T}) dx_{S_2 \setminus T} \right) 1_{[J_T(x_T)=0]} dx_T \right| \\ \leq 2^{2|S_1|+2|S_2|} p^{|S_1|+|S_2|} \int \frac{1_{[J_T(x_T)=0]}}{\mu(x_T)^2} dx_T.$$

Since for  $S \in \mathcal{S}$  we have  $\int 1_{[|F_S| \geq \epsilon_1]} \geq p^{|S|}$ , it follows from the definition of  $J_T$  that for every  $T \subseteq [n]$ ,

$$\sum_{S \in \mathcal{S}, S \supseteq T} p^{|S|} \int \frac{1_{[J_T(x_T)=0]}}{\mu(x_T)^2} dx_T = \sum_{S \in \mathcal{S}, S \supseteq T} p^{|S|} \sum_{x_T \in \{0,1\}^T} \frac{1_{[J_T(x_T)=0]}}{\mu(x_T)} \\ \leq \sum_{x_T \in \{0,1\}^T} \delta \leq 2^{|T|} \delta,$$

and thus

$$\begin{aligned}
 (14) &\leq 2k \sum_{T \subseteq [n]} \sum_{\substack{S_1, S_2 \in \mathcal{S} \\ S_1 \neq S_2, S_1 \cap S_2 = T}} 2^{4k} p^{|S_1|+|S_2|} \int \frac{1_{[J_T(x_T)=0]}}{\mu(x_T)^2} \\
 &\leq 2^{5k} \sum_{T \subseteq [n]} \sum_{S \in \mathcal{S}, S \supseteq T} p^{|S|} 2^{|T|} \delta \\
 &\leq 2^{6k} \delta \sum_{T \subseteq [n]} \sum_{S \in \mathcal{S}, S \supseteq T} p^{|S|} \leq 2^{7k} \delta \sum_{S \in \mathcal{S}} p^{|S|} \leq 2^{7k} \delta \sum_{S \in \mathcal{S}} \frac{\|F_S\|_2^2}{\epsilon_1^2} \leq 2^{7k} \delta \epsilon_1^{-2} \leq \epsilon_0.
 \end{aligned}$$

Now from our bounds on (13) and (14), we conclude that  $\|g - \tilde{g}\|_2^2 \leq 3\epsilon_0$ .

4.4. *Step IV: Finishing the proof.* In the previous steps we have shown that both  $\|f - g\|_2^2$  and  $\|g - \tilde{g}\|_2^2$  are small. It follows that

$$\begin{aligned}
 \|f - \mathbb{E}[f|\mathcal{F}_{\mathcal{J}}]\|_2^2 &\leq \|f - \mathbb{E}[g|\mathcal{F}_{\mathcal{J}}]\|_2^2 \\
 &\leq 2\|f - g\|_2^2 + 2\|g - \mathbb{E}[g|\mathcal{F}_{\mathcal{J}}]\|_2^2 \leq 4\epsilon_0 + 2\|g - \tilde{g}\|_2^2 \leq 10\epsilon_0.
 \end{aligned}$$

Define  $h : \{0, 1\}^n \rightarrow \{0, 1\}$  as

$$h(x) := \begin{cases} 0 & \mathbb{E}[f|\mathcal{F}_{\mathcal{J}}] \leq \frac{1}{2} \\ 1 & \mathbb{E}[f|\mathcal{F}_{\mathcal{J}}] > \frac{1}{2}. \end{cases}$$

Note that  $h$  is a Boolean function and it is measurable with respect to  $\mathcal{F}_{\mathcal{J}}$ . Since  $f$  is a Boolean function, we have

$$\|f - h\|_1 = \int |f - h|^2 \leq \int 4|f - \mathbb{E}[f|\mathcal{F}_{\mathcal{J}}]|^2 \leq 40\epsilon_0 \leq \epsilon.$$

### 5. The general case: Proof of Theorem 2.6

First note that by Remark 1.1, we can assume that  $X$  is a finite probability space. We start from the generalized Walsh expansion  $f = \sum_{S \subseteq [n]} F_S$ . The first two steps are similar to the proof of Theorem 2.8 where we simplify this expansion by removing some insignificant terms from it. We will arrive at a set  $\mathcal{S} \subseteq \mathcal{P}([n])$  such that  $\|f - \sum_{S \in \mathcal{S}} F_S\|_2$  is small and meanwhile the functions  $F_S$  with  $S \in \mathcal{S}$  satisfy certain properties. In general the functions  $F_S$  are not necessarily as well behaved as in the case of the  $p$ -biased distribution. So in the third step we approximate  $\sum_{S \in \mathcal{S}} F_S$  with  $g := \sum_{S \in \mathcal{S}} G_S$  where for every  $S \in \mathcal{S}$ , the function  $G_S$  (similar to  $F_S$ ) satisfies the conditions in Definition 3.1(i) and (ii) and furthermore it has some other desirable properties. In the fourth and fifth steps we define the collection  $\mathcal{J} = \{J_S\}_{S \subseteq [n]}$  and show that  $\|g - \mathbb{E}[g|\mathcal{F}_{\mathcal{J}}]\|_2$  is small. Then in the last step, which is similar to the last step of the proof of Theorem 2.8, we conclude the theorem.

Set  $C := \lceil I_f \rceil$ . In the proof we will define various constants, which for the convenience of the reader are listed here:

$$\begin{aligned} \epsilon_0 &= 10^{-3}\epsilon, & k &= C\epsilon_0^{-1} = 10^3\epsilon^{-1}C, \\ \delta_0 &= 2^{-2k}, & \epsilon_1 &= 3^{-10k^2}\epsilon_0^{10k} \geq e^{-10^8 C^2 \epsilon^{-2}}, \\ \delta &= 2^{-10^3 k^2} \epsilon_1^{10} \geq e^{-10^{10} C^2 \epsilon^{-2}}, & \epsilon_2 &= \delta^{10k} \geq e^{-10^{14} C^3 \epsilon^{-3}}. \end{aligned}$$

5.1. *Step I: High frequencies.* Set  $\epsilon_0 := 10^{-3}\epsilon$  and  $k := C\epsilon_0^{-1}$ , and notice that by (8) and (9), we have

$$(17) \quad \sum_{S:|S|\geq k} \|F_S\|_2^2 \leq \frac{I_f}{k} \leq \epsilon_0.$$

5.2. *Step II: Bourgain's argument.* As we mentioned in the proof of Theorem 2.8, following Bourgain's proof in [Bou99] (but substituting  $k$  in place of  $10C$ ), one can replace [Bou99, inequality (2.17)] with the following:

$$\begin{aligned} \int \sum_{S:|S|\leq k} F_S^2 1_{\{|F_S|\leq\epsilon_1\}} &\leq M^k \int \max_{S:|S|\leq k} F_S^2 1_{\{|F_S|\leq\epsilon_1\}} + 2C3^{k/2}\epsilon^{2/3} + C^{1/2} \frac{3^k}{M^{1/2}\epsilon} \\ &\leq M^k \epsilon_1^2 + 2C3^{k/2}\epsilon^{2/3} + C^{1/2} \frac{3^k}{M^{1/2}\epsilon}, \end{aligned}$$

where  $\epsilon_1, \epsilon, M > 0$  are arbitrary constants. Then by setting  $\epsilon := 3^{-3k}\epsilon_0^3$ ,  $M := 3^{10k}\epsilon_0^{-10}$ , and  $\epsilon_1 := 3^{-10k^2}\epsilon_0^{10k}$ , we obtain

$$(18) \quad \int \sum_{S:|S|\leq k} F_S^2 1_{\{|F_S|\leq\epsilon_1\}} \leq \epsilon_0^2 k^{-1}.$$

Define

$$\mathcal{S} := \left\{ S \subseteq [n] : |S| \leq k, \int F_S^2 1_{\{|F_S|\leq\epsilon_1\}} \leq \epsilon_0 k^{-1} \int F_S^2 \right\},$$

and notice that by (18), we have

$$(19) \quad \sum_{S:|S|\leq k, S \notin \mathcal{S}} \int F_S^2 \leq \sum_{S:|S|\leq k} \frac{k}{\epsilon_0} \int F_S^2 1_{\{|F_S|\leq\epsilon_1\}} \leq \epsilon_0.$$

5.3. *Step III: Modifying the generalized Walsh functions.* Now we will focus our attention to  $S \in \mathcal{S}$  as (17) and (19) show that the generalized Walsh functions  $F_S$  with  $S \notin \mathcal{S}$  have a negligible contribution to the  $L_2$  norm of  $f$ .

Define  $\delta := 2^{-10^3 k^2} \epsilon_1^{10}$ , and for every  $S \in \mathcal{S}$ , let  $\psi_S : X^S \rightarrow \{0, 1\}$  be defined as

$$\psi_S(y) := \begin{cases} 1 & \max_{T \subseteq S} \delta^{-2|S \setminus T|} \int 1_{\{|F_S(y_T, x_{S \setminus T})| > \epsilon_1\}} dx_{S \setminus T} \geq 1, \\ 0 & \text{otherwise.} \end{cases}$$

The next lemma provides some information about the generalized Walsh expansion of  $F_S \psi_S$ .

LEMMA 5.1. *Consider an  $S \in \mathcal{S}$ . Let  $F_S\psi_S = \sum_{T \subseteq S} H_T$  be the generalized Walsh expansion of  $F_S\psi_S$ . Then  $\|H_T\|_\infty \leq \delta$  for every  $T \subsetneq S$ .*

*Proof.* We claim that for every  $R \subsetneq S$ ,

$$(20) \quad \left\| \int F_S\psi_S dx_{S \setminus R} \right\|_\infty \leq 2^{3k} \delta^2.$$

Suppose to the contrary that there exists  $y \in X^R$  such that

$$\left| \int F_S(y, x_{S \setminus R}) \psi_S(y, x_{S \setminus R}) dx_{S \setminus R} \right| > 2^{3k} \delta^2.$$

Then for every  $T \subseteq R$ , we have

$$(21) \quad \delta^{-2|S \setminus T|} \int 1_{[|F_S(y_T, x_{S \setminus T})| > \epsilon_1]} dx_{S \setminus T} < 1,$$

as otherwise  $\psi_S(z) = 1$ , for every  $z \in X^S$  with  $z_R = y$ , and in this case, since  $R \neq S$ , we would have

$$\int F_S(y, x_{S \setminus R}) \psi_S(y, x_{S \setminus R}) dx_{S \setminus R} = \int F_S(y, x_{S \setminus R}) dx_{S \setminus R} = 0.$$

Now by (7), (21), and the definition of  $\psi_S$ , we have

$$\begin{aligned} & \left| \int F_S(y, x_{S \setminus R}) \psi_S(y, x_{S \setminus R}) dx_{S \setminus R} \right| \\ & \leq 2^k \int \psi_S(y, x_{S \setminus R}) dx_{S \setminus R} \\ & \leq 2^k \int \sum_{T \subseteq S, T \not\subseteq R} \delta^{-2|S \setminus T|} \left( \int 1_{[|F_S| > \epsilon_1]} dx_{S \setminus T} \right) dx_{S \setminus R} \\ & = 2^k \sum_{T \subseteq S, T \not\subseteq R} \delta^{-2|S \setminus T|} \int 1_{[|F_S| > \epsilon_1]} dx_{S \setminus (T \cap R)} \\ & \leq 2^{2k} \delta^2 \sum_{T \subseteq R} \delta^{-2|S \setminus T|} \int 1_{[|F_S| > \epsilon_1]} dx_{S \setminus T} \leq 2^{3k} \delta^2, \end{aligned}$$

which establishes our claim (20). Now by (6) and (20), for every  $T \subsetneq S$ , we have

$$\begin{aligned} \|H_T\|_\infty &= \left\| \sum_{R \subseteq T} (-1)^{|T \setminus R|} \int F_S\psi_S dx_{[n] \setminus R} \right\|_\infty \\ &\leq \sum_{R \subseteq T} \left\| \int F_S\psi_S dx_{[n] \setminus R} \right\|_\infty \leq 2^{4k} \delta^2 \leq \delta. \quad \square \end{aligned}$$

For every  $S \in \mathcal{S}$ , define  $G_S := H_S$ , where  $F_S\psi_S = \sum_{T \subseteq S} H_T$  is the generalized Walsh expansion of  $F_S\psi_S$ . The function  $G_S$  satisfies the conditions in Definition 3.1(i) and (ii). Lemma 5.2 below lists some other properties of  $G_S$ . Part (a) shows that for  $S \in \mathcal{S}$ , similar to  $F_S$ , most of the  $L_2$  weight of  $G_S$  is concentrated on  $\{x : |F_S(x)| \geq \epsilon_1\}$ . Part (b) shows that  $G_S$  is a good

approximation of  $F_S$  in the  $L_2$  norm. Parts (c) and (d) in particular show that it is possible to bound the sum of the  $L_1$  norms of  $G_S$  for  $S \in \mathcal{S}$ . The functions  $F_S$  do not necessarily satisfy this latter property, and it is for this reason that we replace them with the functions  $G_S$ . For every  $S \in \mathcal{S}$ , define  $a_S : X^S \rightarrow \mathbb{R}$  as

$$a_S : y \mapsto 2^{3k} \delta^{-2k} \sum_{T \subseteq S} \int 1_{\{|F_S(y_{S \setminus T}, x_T)| > \epsilon_1\}} dx_T.$$

LEMMA 5.2. *For every  $S \in \mathcal{S}$ , we have*

- (a):  $\int G_S^2 1_{\{|F_S| \leq \epsilon_1\}} \leq 2\epsilon_0 k^{-1} \|F_S\|_2^2$ .
- (b):  $\|F_S - G_S\|_2^2 \leq 10\epsilon_0 k^{-1} \|F_S\|_2^2$ .
- (c): *For every  $y \in X^S$ , we have  $|G_S(y)| \leq a_S(y)$ .*
- (d):  $\sum_{S \in \mathcal{S}} \int a_S dx_{[n]} \leq \delta^{-3k}$ .

*Proof.* Consider  $S \in \mathcal{S}$ , and let  $F_S \psi_S = \sum_{T \subseteq S} H_T$  be the generalized Walsh expansion of  $F_S \psi_S$ . By Lemma 5.1, we have

$$(22) \quad \|F_S - G_S\|_\infty \leq \sum_{T \subseteq S} \|H_T\|_\infty \leq 2^k \delta.$$

Hence by (7), (8), and the assumptions that  $S \in \mathcal{S}$  and  $\delta = 2^{-10^3 k^2} \epsilon_1^{10}$ , we have

$$\begin{aligned} \int F_S^2 &\geq \int F_S^2 \psi_S^2 \geq \int G_S^2 \geq \int G_S^2 1_{\{|F_S| > \epsilon_1\}} \\ &\geq \int (|F_S| - 2^k \delta)^2 1_{\{|F_S| > \epsilon_1\}} \\ &\geq \int F_S^2 1_{\{|F_S| > \epsilon_1\}} - 2^{k+1} \delta \int |F_S| 1_{\{|F_S| > \epsilon_1\}} \\ &\geq \int F_S^2 1_{\{|F_S| > \epsilon_1\}} - 2^{3k} \delta \int 1_{\{|F_S| > \epsilon_1\}} \\ &\geq (1 - \epsilon_0 k^{-1}) \int F_S^2 - 2^{3k} \delta \epsilon_1^{-2} \int F_S^2 \geq (1 - 2\epsilon_0 k^{-1}) \int F_S^2. \end{aligned}$$

Then

$$\begin{aligned} \int G_S^2 1_{\{|F_S| \leq \epsilon_1\}} &= \int G_S^2 - \int G_S^2 1_{\{|F_S| > \epsilon_1\}} \\ &\leq \int F_S^2 - (1 - 2\epsilon_0 k^{-1}) \int F_S^2 \leq 2\epsilon_0 k^{-1} \|F_S\|_2^2, \end{aligned}$$

which verifies Part (a). In order to prove Part (b) note that by Part (a), (22), and the assumptions that  $S \in \mathcal{S}$  and  $\delta = 2^{-10^3 k^2} \epsilon_1^{10}$ , we have

$$\begin{aligned} \|F_S - G_S\|_2^2 &= \int |F_S - G_S|^2 1_{\{|F_S| \geq \epsilon_1\}} + \int |F_S - G_S|^2 1_{\{|F_S| < \epsilon_1\}} \\ &= 2^{2k} \delta^2 \int 1_{\{|F_S| \geq \epsilon_1\}} + 2 \int (F_S^2 + G_S^2) 1_{\{|F_S| < \epsilon_1\}} \end{aligned}$$



$$\begin{aligned} &\leq 2^{2k} \delta^2 \epsilon_1^{-2} \int F_S^2 + 2 \int F_S^2 1_{[|F_S| < \epsilon_1]} + 2 \int G_S^2 1_{[|F_S| < \epsilon_1]} \\ &\leq 2^{2k} \delta^2 \epsilon_1^{-2} \|F_S\|_2^2 + 2\epsilon_0 k^{-1} \|F_S\|_2^2 \\ &\quad + 4\epsilon_0 k^{-1} \|F_S\|_2^2 \leq 10\epsilon_0 k^{-1} \|F_S\|_2^2. \end{aligned}$$

To prove Part (c) notice that by (6), (7), and the definition of  $\psi_S$ ,

$$\begin{aligned} |G_S| &= \left| \sum_{R \subseteq S} (-1)^{|S \setminus R|} \int F_S \psi_S dx_{S \setminus R} \right| \leq \sum_{R \subseteq S} 2^k \int \psi_S dx_{S \setminus R} \\ &\leq 2^k \sum_{R \subseteq S} \int \delta^{-2k} \sum_{T \subseteq S} \left( \int 1_{[|F_S| > \epsilon_1]} dx_{S \setminus T} \right) dx_{S \setminus R} \\ &\leq 2^k \delta^{-2k} \sum_{R \subseteq S} \sum_{T \subseteq S} \int 1_{[|F_S| > \epsilon_1]} dx_{S \setminus (T \cap R)} \leq 2^{3k} \delta^{-2k} \sum_{T \subseteq S} \int 1_{[|F_S| > \epsilon_1]} dx_T. \end{aligned}$$

It remains to prove Part (d). We have

$$\begin{aligned} \sum_{S \in \mathcal{S}} \int a_S dx_{[n]} &= 2^{3k} \delta^{-2k} \sum_{S \in \mathcal{S}} \sum_{T \subseteq S} \int 1_{[|F_S| > \epsilon_1]} dx_T dx_{[n]} \\ &\leq 2^{4k} \delta^{-2k} \sum_{S \in \mathcal{S}} \int 1_{[|F_S| > \epsilon_1]} \\ &= 2^{4k} \delta^{-2k} \epsilon_1^{-2} \sum_{S \in \mathcal{S}} \|F_S\|_2^2 \leq 2^{4k} \delta^{-2k} \epsilon_1^{-2} \leq \delta^{-3k}. \quad \square \end{aligned}$$

5.4. *Step IV: The sigma algebra.* In Steps 1–3 of the proof we approximated  $f$  in the  $L_2$  norm with  $g := \sum_{S \in \mathcal{S}} G_S$ , where the functions  $G_S$  satisfy certain properties. Next we will define the collection  $\mathcal{J} = \{J_S\}_{S \subseteq [n]}$  so that  $\|g - \mathbb{E}[g | \mathcal{F}_{\mathcal{J}}]\|_2$  is small and  $\int |J_{\mathcal{J}}|$  is bounded by a constant that does not depend on  $n$ . In order to define  $\mathcal{J}$ , we need to introduce the auxiliary functions  $\xi_T : X^T \rightarrow \{0, 1\}$  for every  $T \subseteq [n]$  with  $|T| \leq k$ . Set  $\epsilon_2 := \delta^{10k}$ , and for every  $T \subseteq [n]$  with  $|T| \leq k$ , define  $\xi_T : X^T \rightarrow \{0, 1\}$  as

$$\xi_T : y \mapsto \begin{cases} 1 & \sum_{R \subseteq T} \sum_{S \in \mathcal{S} : S \supseteq T} \int a_S(y_R, x_{S \setminus R}) dx_{S \setminus R} > \epsilon_2, \\ 0 & \text{otherwise.} \end{cases}$$

Then set  $\delta_0 := 2^{-2k}$ , and for every  $T \subseteq [n]$  with  $|T| \leq k$ , define  $J_T : X^T \rightarrow \{0, 1\}$  as

$$J_T : y \mapsto \begin{cases} 1 & \max_{R \subseteq T} \delta_0^{-2|T \setminus R|} \int \xi_T(y_R, x_{T \setminus R}) dx_{T \setminus R} \geq 1, \\ 0 & \text{otherwise.} \end{cases}$$

For  $T \subseteq [n]$  with  $|T| > k$ , define  $J_T \equiv 0$ . In the sequel it will be useful to bear in mind that for  $S \in \mathcal{S}$ ,

$$(23) \quad 1_{[|F_S| > \epsilon_1]} \leq \xi_S \leq J_S,$$

and in general for every  $T \subseteq [n]$  with  $|T| \leq k$ , it holds that

$$(24) \quad \xi_T \leq J_T.$$

In order to be able to bound  $\|g - \mathbb{E}[g|\mathcal{F}_{\mathcal{J}}]\|_2$  and  $\int |J_{\mathcal{J}}|$  we need to prove a few lemmas.

LEMMA 5.3. *Consider a set  $T \subseteq [n]$  with  $|T| \leq k$ . For every  $R \subseteq T$  and every  $y \in X^R$ , either  $\int J_T(y, x_{T \setminus R}) dx_{T \setminus R} \leq \delta_0$ , or  $J_T(z) = 1$  for every  $z \in X^T$  with  $z_R = y$ .*

*Proof.* Suppose that  $J_T(z) = 0$  for at least one  $z \in X^T$  with  $z_R = y$ . Then by the definition of  $J_T$ , for every  $A \subseteq R$ , we have

$$(25) \quad \int \xi_T(y_A, x_{T \setminus A}) dx_{T \setminus A} < \delta_0^{2|T \setminus A|}.$$

Hence for every  $B \subseteq T$  with  $B \not\subseteq R$ , setting  $A := B \cap R$ , we have

$$\begin{aligned} & \int 1_{\left[\int \xi_T(y_A, x_{T \setminus A}) dx_{T \setminus A} \geq \delta_0^{2|T \setminus B|}\right]} dx_{T \setminus R} \\ &= \int \left( \int 1_{\left[\int \xi_T(y_A, x_{T \setminus A}) dx_{T \setminus A} \geq \delta_0^{2|T \setminus B|}\right]} dx_{T \setminus B} \right) dx_{T \setminus R} \\ &\leq \int \left( \int \xi_T(y_A, x_{T \setminus A}) \delta_0^{-2|T \setminus B|} dx_{T \setminus B} \right) dx_{T \setminus R} \\ &= \delta_0^{-2|T \setminus B|} \int \xi_T(y_A, x_{T \setminus A}) dx_{T \setminus A} \\ &\leq \delta_0^{-2|T \setminus B|} \delta_0^{2|T \setminus A|} \leq \delta_0^2. \end{aligned}$$

Then using (25) we conclude that

$$\begin{aligned} \int J_T(y_R, x_{T \setminus R}) dx_{T \setminus R} &\leq \int \sum_{A \subseteq R} 1_{\left[\int \xi_T dx_{T \setminus A} \geq \delta_0^{2|T \setminus A|}\right]} dx_{T \setminus R} \\ &\quad + \int \sum_{\substack{B: B \subseteq T, \\ B \not\subseteq R}} 1_{\left[\int \xi_T dx_{T \setminus B} \geq \delta_0^{2|T \setminus B|}\right]} dx_{T \setminus R} \\ &= \int \sum_{B: B \subseteq T, B \not\subseteq R} 1_{\left[\int \xi_T dx_{T \setminus B} \geq \delta_0^{2|T \setminus B|}\right]} dx_{T \setminus R} \\ &= \sum_{B: B \subseteq T, B \not\subseteq R} \delta_0^2 \leq 2^k \delta_0^2 \leq \delta_0. \quad \square \end{aligned}$$

The following lemma shows that for a random  $x \in X^T$ , the expected value of  $\sum_{T: |T| \leq k} J_T(x)$  is bounded from above by the constant  $\epsilon_2^{-2}$ .

LEMMA 5.4. *We have*

$$\sum_{T: |T| \leq k} \int J_T \leq \epsilon_2^{-2}.$$

*Proof.* Note that

$$\begin{aligned}
 (26) \quad \sum_{T:|T|\leq k} \int J_T dx_T &\leq \sum_{T:|T|\leq k} \int \delta_0^{-2k} \sum_{R\subseteq T} \int \xi_T dx_{T\setminus R} dx_R \\
 &\leq \delta_0^{-2k} \sum_{T:|T|\leq k} \sum_{R\subseteq T} \int \xi_T dx_T \\
 &\leq 2^k \delta_0^{-2k} \sum_{T:|T|\leq k} \int \xi_T dx_T.
 \end{aligned}$$

Furthermore by the definition of  $\xi_T$  and Part (d) of Lemma 5.2,

$$\begin{aligned}
 (27) \quad \sum_{T:|T|\leq k} \int \xi_T dx_{[n]} &\leq \epsilon_2^{-1} \sum_{T:|T|\leq k} \int \left( \sum_{R\subseteq T} \sum_{S\in\mathcal{S}, S\supseteq T} \int a_S dx_{[n]\setminus R} \right) dx_{[n]} \\
 &\leq 2^k \epsilon_2^{-1} \sum_{T:|T|\leq k} \sum_{S\in\mathcal{S}, S\supseteq T} \int a_S dx_{[n]} \leq 2^{2k} \epsilon_2^{-1} \int \sum_{S\in\mathcal{S}} a_S dx_{[n]} \\
 &\leq 2^{2k} \epsilon_2^{-1} \delta^{-3k}.
 \end{aligned}$$

From (26), (27), and the assumptions  $\epsilon_2 = \delta^{10k}$ , we conclude that

$$\int \sum_{T:|T|\leq k} J_T dx_T \leq 2^{3k} \delta_0^{-2k} \delta^{-3k} \epsilon_2^{-1} \leq \epsilon_2^{-2}. \quad \square$$

It follows immediately from Lemma 5.4 that

$$(28) \quad \int |J_{\mathcal{J}}| \leq \int \sum_{T\subseteq[n], |T|\leq k} |T| J_T \leq k \sum_{T:|T|\leq k} \int J_T \leq k \epsilon_2^{-2}.$$

5.5. *Step V: Bounding the error.* Here we show that  $\|g - \mathbb{E}[g|\mathcal{F}_{\mathcal{J}}]\|_2$  is small. By linearity of expectation,  $\mathbb{E}[g|\mathcal{F}_{\mathcal{J}}] = \sum_{S\in\mathcal{S}} \mathbb{E}[G_S|\mathcal{F}_{\mathcal{J}}]$ . However since  $\mathcal{F}_{\mathcal{J}}$  depends on all coordinates, it is difficult to analyze  $\mathbb{E}[G_S|\mathcal{F}_{\mathcal{J}}]$  directly. To remedy this we define some auxiliary  $\sigma$ -algebras. For every  $S \subseteq [n]$ , define the collection  $\mathcal{J}_S := \{\tilde{J}_T\}_{T\subseteq[n]}$  in the following way. For every  $T \subseteq [n]$ , we set  $\tilde{J}_T := J_T$  if  $T \subseteq S$ , and  $\tilde{J}_T \equiv 0$  otherwise. Note that  $\mathcal{J} = \mathcal{J}_{[n]}$ . Define  $\tilde{g} := \sum_{S\in\mathcal{S}} \tilde{G}_S$ , where  $\tilde{G}_S := \mathbb{E}[G_S|\mathcal{F}_{\mathcal{J}_S}]$ . For every  $S \subseteq [n]$ , the function  $\tilde{G}_S$  depends only on the coordinates in  $S$ , and furthermore the  $\sigma$ -algebra  $\mathcal{F}_{\mathcal{J}_S}$  is coarser than the  $\sigma$ -algebra  $\mathcal{F}_{\mathcal{J}}$ . It follows from the latter that  $\tilde{g}$  is measurable with respect to  $\mathcal{F}_{\mathcal{J}}$ , and hence

$$(29) \quad \|g - \mathbb{E}[g|\mathcal{F}_{\mathcal{J}}]\|_2 \leq \|g - \tilde{g}\|_2.$$

So in order to bound  $\|g - \mathbb{E}[g|\mathcal{F}_{\mathcal{J}}]\|_2$ , it suffices to bound the right-hand side of (29). Trivially for every  $S \in \mathcal{S}$ ,

$$\int G_S \tilde{G}_S = \int \tilde{G}_S^2.$$

For  $S_1, S_2 \in \mathcal{S}$ , since  $\tilde{G}_{S_1}$  and  $\tilde{G}_{S_2}$  depend respectively only on the coordinates in  $S_1$  and  $S_2$ , if  $S_1 \not\subseteq S_2$  and  $S_2 \not\subseteq S_1$ , then

$$\int G_{S_1} G_{S_2} = \int G_{S_1} \tilde{G}_{S_2} = \int \tilde{G}_{S_1} G_{S_2} = 0.$$

Similarly for  $S_1, S_2 \in \mathcal{S}$ , if  $S_1 \cap S_2 = \emptyset$ , then

$$\int \tilde{G}_{S_1} \tilde{G}_{S_2} = \left( \int \tilde{G}_{S_1} \right) \left( \int \tilde{G}_{S_2} \right) = \left( \int G_{S_1} \right) \left( \int G_{S_2} \right) = 0.$$

Thus

$$\begin{aligned} \|g - \tilde{g}\|_2^2 &= \int \left( \sum_{S \in \mathcal{S}} G_S - \tilde{G}_S \right)^2 \leq (k+1) \sum_{r=0}^k \int \left( \sum_{S \in \mathcal{S}, |S|=r} G_S - \tilde{G}_S \right)^2 \\ &= (k+1) \int \sum_{S \in \mathcal{S}} G_S^2 - \tilde{G}_S^2 + \sum_{\substack{S_1, S_2 \in \mathcal{S}, S_1 \cap S_2 \neq \emptyset \\ S_1 \neq S_2, |S_1|=|S_2|}} \tilde{G}_{S_1} \tilde{G}_{S_2} \\ (30) \quad &\leq 2k \int \sum_{S \in \mathcal{S}} |G_S^2 - \tilde{G}_S^2| \end{aligned}$$

$$(31) \quad + 2k \sum_{\substack{S_1, S_2 \in \mathcal{S} \\ S_1 \cap S_2 \neq \emptyset, S_1 \neq S_2}} \left| \int \tilde{G}_{S_1} \tilde{G}_{S_2} \right|.$$

*Bounding (30):* Consider an  $S \in \mathcal{S}$  and an  $x \in X^S$ . By (23), if  $|F_S(x)| > \epsilon_1$ , then  $J_S(x) = 1$ , which implies that  $J_{\mathcal{J}_S}(x) = S$ , and consequently  $G_S(x) = \tilde{G}_S(x)$ . Hence it follows from Part (a) of Lemma 5.2 and Parseval's identity that

$$\begin{aligned} (32) \quad (30) &= 2k \int \sum_{S \in \mathcal{S}} |G_S^2 - \tilde{G}_S^2| \\ &= 2k \int \sum_{S \in \mathcal{S}} |G_S - \tilde{G}_S|^2 = 2k \int \sum_{S \in \mathcal{S}} |G_S - \tilde{G}_S|^2 1_{\{|F_S| \leq \epsilon_1\}} \\ &\leq 2k \int \sum_{S \in \mathcal{S}} G_S^2 1_{\{|F_S| \leq \epsilon_1\}} \leq 4\epsilon_0 \sum_{S \in \mathcal{S}} \|F_S\|_2^2 \leq 4\epsilon_0. \end{aligned}$$

*Bounding (31):* We start with a lemma about the size of the atoms of the  $\sigma$ -algebra  $\mathcal{F}_{\mathcal{J}_S}$ .

**LEMMA 5.5.** *Consider  $S \in \mathcal{S}$ , a subset  $A \subseteq S$ , and an element  $y \in X^S$  with  $J_{\mathcal{J}_S}(y) = A$ . Then  $\int 1_{[J_{\mathcal{J}_S}(y_A, x_{S \setminus A}) = A]} dx_{S \setminus A} \geq \frac{1}{2}$ .*

*Proof.* Since  $J_{\mathcal{J}_S}(y) = A$ , for every  $R \subseteq S$  with  $R \not\subseteq A$ , we have  $J_R(y) = 0$ . Then it follows from Lemma 5.3 that for every such  $R$ , we have

$$\int J_R(y_{A \cap R}, x_{R \setminus A}) dx_{R \setminus A} \leq \delta_0.$$

Hence

$$\begin{aligned} \int 1_{[J_{\mathcal{J}_S}(y_A, x_{S \setminus A})=A]} dx_{S \setminus A} &= \int \prod_{R \subseteq S, R \not\subseteq A} 1_{[J_R(y_{A \cap R}, x_{R \setminus A})=0]} dx_{S \setminus A} \\ &\geq 1 - \sum_{R \subseteq S, R \not\subseteq A} \int J_R(y_{A \cap R}, x_{R \setminus A}) dx_{S \setminus A} \\ &\geq 1 - 2^k \delta_0 \geq \frac{1}{2}. \end{aligned} \quad \square$$

LEMMA 5.6. Consider  $S \in \mathcal{S}$  and a subset  $T \subseteq S$ . For every  $y \in X^T$ , we have

$$\left| \int \tilde{G}_S(y, x_{S \setminus T}) dx_{S \setminus T} \right| \leq 2^{k+1} \sum_{R \subseteq T} \int a_S(y_R, x_{S \setminus R}) dx_{S \setminus R}.$$

Furthermore if  $T \subsetneq S$  and  $J_T(y) = 1$ , then  $\int \tilde{G}_S(y, x_{S \setminus T}) dx_{S \setminus T} = 0$ .

*Proof.* Consider  $A \subseteq S$ , and  $z \in X^S$  with  $J_{\mathcal{J}_S}(z) = A$ . By Lemma 5.5 and Part (c) of Lemma 5.2, we have

$$\begin{aligned} |\tilde{G}_S(z)| &= \frac{\left| \int G_S(z_A, x_{S \setminus A}) 1_{[J_{\mathcal{J}_S}(z_A, x_{S \setminus A})=A]} dx_{S \setminus A} \right|}{\int 1_{[J_{\mathcal{J}_S}(z_A, x_{S \setminus A})=A]} dx_{S \setminus A}} \\ &\leq 2 \left| \int G_S(z_A, x_{S \setminus A}) 1_{[J_{\mathcal{J}_S}(z_A, x_{S \setminus A})=A]} dx_{S \setminus A} \right| \\ &\leq 2 \int a_S(z_A, x_{S \setminus A}) 1_{[J_{\mathcal{J}_S}(z_A, x_{S \setminus A})=A]} dx_{S \setminus A} \leq 2 \int a_S(z_A, x_{S \setminus A}) dx_{S \setminus A}. \end{aligned}$$

Hence for every  $y \in X^T$ ,

$$\begin{aligned} \left| \int \tilde{G}_S(y, x_{S \setminus T}) dx_{S \setminus T} \right| &\leq \sum_{A \subseteq S} \left| \int \tilde{G}_S(y, x_{S \setminus T}) 1_{[J_{\mathcal{J}_S}(y, x_{S \setminus T})=A]} dx_{S \setminus T} \right| \\ &\leq 2 \sum_{A \subseteq S} \int a_S(y_{A \cap T}, x_{S \setminus (A \cap T)}) dx_{S \setminus (A \cap T)} \\ &\leq 2^{k+1} \sum_{R \subseteq T} \int a_S(y_R, x_{S \setminus R}) dx_{S \setminus R}. \end{aligned}$$

This verifies the first assertion of the lemma. Now suppose that  $T \subsetneq S$  and  $J_T(y) = 1$ . Then for every  $z \in X^S$  with  $z_T = y$ , we have  $J_{\mathcal{J}_S}(z) \supseteq T$ . Hence

$$\begin{aligned} \int \tilde{G}_S(y, x_{S \setminus T}) dx_{S \setminus T} &= \sum_{A: T \subseteq A \subseteq S} \int \tilde{G}_S 1_{[J_{\mathcal{J}_S}=A]} dx_{S \setminus T} \\ &= \sum_{A: T \subseteq A \subseteq S} \int \int \tilde{G}_S 1_{[J_{\mathcal{J}_S}=A]} dx_{S \setminus A} dx_{S \setminus T} \\ &= \sum_{A: T \subseteq A \subseteq S} \int \int G_S 1_{[J_{\mathcal{J}_S}=A]} dx_{S \setminus A} dx_{S \setminus T} \\ &= \int G_S(y, x_{S \setminus T}) dx_{S \setminus T} = 0. \end{aligned} \quad \square$$

Now consider two distinct  $S_1, S_2 \in \mathcal{S}$  with  $T := S_1 \cap S_2 \neq \emptyset$ . By the second assertion of Lemma 5.6, for  $y \in X^T$ , if  $J_T(y) = 1$ , then

$$\int \tilde{G}_{S_1}(y, x_{[n] \setminus T}) \tilde{G}_{S_2}(y, x_{[n] \setminus T}) dx_{[n] \setminus T} = 0.$$

It follows that

$$\begin{aligned} (31) &= 2k \sum_{\substack{S_1, S_2 \in \mathcal{S} \\ S_1 \neq S_2, S_1 \cap S_2 \neq \emptyset}} \left| \int \tilde{G}_{S_1} \tilde{G}_{S_2} \right| \\ &= 2k \sum_{T \subseteq [n]} \sum_{\substack{S_1, S_2 \in \mathcal{S} \\ S_1 \neq S_2, S_1 \cap S_2 = T}} \left| \int \tilde{G}_{S_1} \tilde{G}_{S_2} 1_{[J_T=0]} \right| \\ &= 2k \sum_{T \subseteq [n]} \sum_{\substack{S_1, S_2 \in \mathcal{S} \\ S_1 \neq S_2, S_1 \cap S_2 = T}} \left| \int \left( \int \tilde{G}_{S_1} dx_{[n] \setminus T} \right) \left( \int \tilde{G}_{S_2} dx_{[n] \setminus T} \right) 1_{[J_T=0]} dx_T \right| \\ &\leq 2k \sum_{T \subseteq [n]} \int \left( \sum_{S \in \mathcal{S}, S \supseteq T} \left| \int \tilde{G}_S dx_{[n] \setminus T} \right| \right)^2 1_{[J_T=0]} dx_T. \end{aligned}$$

Now applying the first assertion of Lemma 5.6, we have

$$(33) \quad (31) \leq 2k \sum_{T \subseteq [n]} \int \left( 2^{k+1} \sum_{S \in \mathcal{S}, S \supseteq T} \sum_{R \subseteq T} \int a_S dx_{[n] \setminus R} \right)^2 1_{[J_T=0]} dx_T.$$

Recall from (24) that  $\xi_T \leq J_T$ . Hence if  $J_T(y) = 0$  for  $y \in X^T$ , then

$$\sum_{R \subseteq T} \sum_{S \in \mathcal{S}: S \supseteq T} \int a_S(y_R, x_{[n] \setminus R}) dx_{[n] \setminus R} \leq \epsilon_2.$$

So (33), together with Part (d) of Lemma 5.2, implies that

$$\begin{aligned} (34) \quad (31) &\leq 2k \sum_{T \subseteq [n]} 2^{3k} \epsilon_2 \int \left( \sum_{S \in \mathcal{S}, S \supseteq T} \sum_{R \subseteq T} \int a_S dx_{[n] \setminus R} \right) dx_T \\ &\leq 2k 2^{3k} \epsilon_2 \sum_{T \subseteq [n]} \sum_{R \subseteq T} \sum_{\substack{S \in \mathcal{S} \\ S \supseteq T}} \int a_S \leq 2k 2^{5k} \epsilon_2 \sum_{S \in \mathcal{S}} \int a_S dx_{[n]} \\ &\leq 2k 2^{5k} \epsilon_2 \delta^{-3k} \leq \epsilon_0. \end{aligned}$$

Combining (29), (32), and (34), we conclude that

$$(35) \quad \|g - \mathbb{E}[g | \mathcal{F}_{\mathcal{J}}]\|_2^2 \leq \|g - \tilde{g}\|_2^2 \leq 5\epsilon_0.$$

5.6. *Step VI: Finishing the proof.* By (17) and (19), we have

$$\int \left| f - \sum_{S \in \mathcal{S}} F_S \right|^2 \leq 2\epsilon_0.$$

Recall that  $g = \sum_{S \in \mathcal{S}} G_S$ . By Part (b) of Lemma 5.2,

$$\begin{aligned} \int \left| \sum_{S \in \mathcal{S}} F_S - \sum_{S \in \mathcal{S}} G_S \right|^2 &= \sum_{S \in \mathcal{S}} \int |F_S - G_S|^2 \\ &\leq \sum_{S \in \mathcal{S}} 10\epsilon_0 k^{-1} \|F_S\|_2^2 \leq 10k^{-1}\epsilon_0 \leq \epsilon_0. \end{aligned}$$

Hence

$$\int |f - g|^2 \leq 2 \int \left| f - \sum_{S \in \mathcal{S}} F_S \right|^2 + \left| \sum_{S \in \mathcal{S}} F_S - \sum_{S \in \mathcal{S}} G_S \right|^2 \leq 4\epsilon_0 + 2\epsilon_0 \leq 6\epsilon_0.$$

Then by (35),

$$\begin{aligned} \int |f - \mathbb{E}[f|\mathcal{F}_{\mathcal{J}}]|^2 &\leq \int |f - \mathbb{E}[g|\mathcal{F}_{\mathcal{J}}]|^2 \\ &\leq 2 \int |f - g|^2 + |g - \mathbb{E}[g|\mathcal{F}_{\mathcal{J}}]|^2 \leq 12\epsilon_0 + 10\epsilon_0 \leq 22\epsilon_0. \end{aligned}$$

Define  $h : X^n \rightarrow \{0, 1\}$  as

$$h(x) := \begin{cases} 0 & \mathbb{E}[f|\mathcal{F}_{\mathcal{J}}] \leq \frac{1}{2} \\ 1 & \mathbb{E}[f|\mathcal{F}_{\mathcal{J}}] > \frac{1}{2}. \end{cases}$$

Note that  $h$  is a Boolean function and it is measurable with respect to  $\mathcal{F}_{\mathcal{J}}$ . Since  $f$  is a Boolean function, we have

$$\|f - h\|_1 = \int |f - h|^2 \leq \int 4|f - \mathbb{E}[f|\mathcal{F}_{\mathcal{J}}]|^2 \leq 90\epsilon_0 \leq \epsilon.$$

We also verified in (28) that  $\int |J_{\mathcal{J}}(x)| dx \leq k\epsilon_2^{-2} \leq e^{10^{15}\epsilon^{-3}\lceil I_f \rceil^3}$ .

### 6. Concluding remarks

- For graph properties which are also *increasing* (as is the case when one studies the threshold phenomenon), what is shown by Friedgut [Fri99] is slightly stronger than the assertion of Theorem 2.8. He showed that in this case one can deduce from the structure provided in Theorem 2.8 that the property is essentially generated by small minimal elements. He then conjectured this statement (Conjecture 1.5 in [Fri99]) for general increasing set properties. Friedgut’s conjecture remains unsolved as we do not know how to refine the structure provided in Theorem 2.8 for increasing functions without appealing to any symmetry assumptions.

- In [BKK<sup>+</sup>92] a different notion of the influence is defined, and in [Hat09] the structure of Boolean functions that have small total influences with the definition of [BKK<sup>+</sup>92] is determined.

### Acknowledgements

The author is grateful to Serguei Norine for pointing the connection to the FKG inequality that led to the proof of Corollary 2.10. He also wishes to thank him and Pooya Hatami for many valuable discussions, and Elchanan Mossel and Mokshay Madiman for pointing out the references [ES81], [Hoe48].

### References

- [AS00] N. ALON and J. H. SPENCER, *The Probabilistic Method*, second ed., Wiley-Intersci. Ser. Discrete Math. Optim., Wiley-Interscience [John Wiley & Sons], New York, 2000. MR 1885388. Zbl 0996.05001. <http://dx.doi.org/10.1002/0471722154>.
- [Bou99] J. BOURGAIN, Appendix to “Sharp thresholds of graph properties, and the  $k$ -sat problem”, *J. Amer. Math. Soc.* **12** (1999), 1046–1053. MR 1678031. Zbl 0932.05084. <http://dx.doi.org/10.1090/S0894-0347-99-00305-7>.
- [Bou00] ———, Harmonic analysis and combinatorics: how much may they contribute to each other?, in *Mathematics: Frontiers and Perspectives*, Amer. Math. Soc., Providence, RI, 2000, pp. 13–32. MR 1754764. Zbl 0963.42001.
- [BKK<sup>+</sup>92] J. BOURGAIN, J. KAHN, G. KALAI, Y. KATZNELSON, and N. LINIAL, The influence of variables in product spaces, *Israel J. Math.* **77** (1992), 55–64. MR 1194785. Zbl 0771.60002. <http://dx.doi.org/10.1007/BF02808010>.
- [BK97] J. BOURGAIN and G. KALAI, Influences of variables and threshold intervals under group symmetries, *Geom. Funct. Anal.* **7** (1997), 438–461. MR 1466334. Zbl 0982.20004. <http://dx.doi.org/10.1007/s000390050015>.
- [ES81] B. EFRON and C. STEIN, The jackknife estimate of variance, *Ann. Statist.* **9** (1981), 586–596. MR 0615434. Zbl 0481.62035. <http://dx.doi.org/10.1214/aos/1176345462>.
- [Fri98] E. FRIEDGUT, Boolean functions with low average sensitivity depend on few coordinates, *Combinatorica* **18** (1998), 27–35. MR 1645642. Zbl 0909.06008. <http://dx.doi.org/10.1007/PL00009809>.
- [Fri99] ———, Sharp thresholds of graph properties, and the  $k$ -sat problem, *J. Amer. Math. Soc.* **12** (1999), 1017–1054. MR 1678031. Zbl 0932.05084. <http://dx.doi.org/10.1090/S0894-0347-99-00305-7>.
- [Fri05] ———, Hunting for sharp thresholds, *Random Structures Algorithms* **26** (2005), 37–51. MR 2116574. Zbl 1059.05092. <http://dx.doi.org/10.1002/rsa.20042>.
- [FK96] E. FRIEDGUT and G. KALAI, Every monotone graph property has a sharp threshold, *Proc. Amer. Math. Soc.* **124** (1996), 2993–3002. MR 1371123. Zbl 0864.05078. <http://dx.doi.org/10.1090/S0002-9939-96-03732-X>.



- [Hat09] H. HATAMI, Decision trees and influences of variables over product probability spaces, *Combin. Probab. Comput.* **18** (2009), 357–369. MR 2501432. Zbl 1193.60007. <http://dx.doi.org/10.1017/S0963548309009833>.
- [Hoe48] W. HOEFFDING, A class of statistics with asymptotically normal distribution, *Ann. Math. Statistics* **19** (1948), 293–325. MR 0026294. Zbl 0032.04101.
- [KKL88] J. KAHN, G. KALAI, and N. LINIAL, The influence of variables on boolean functions, in *29-th Annual Symposium on Foundations of Computer Science*, 1988, pp. 68–80.
- [Mar74] G. A. MARGULIS, Probabilistic characteristics of graphs with large connectivity, *Problemy Peredači Informacii* **10** (1974), 101–108. MR 0472604.
- [MOO10] E. MOSSEL, R. O’DONNELL, and K. OLESZKIEWICZ, Noise stability of functions with low influences: invariance and optimality, *Ann. of Math.* **171** (2010), 295–341. MR 2630040. Zbl 1201.60031. <http://dx.doi.org/10.4007/annals.2010.171.295>.
- [Rus81] L. RUSSO, On the critical percolation probabilities, *Z. Wahrsch. Verw. Gebiete* **56** (1981), 229–237. MR 0618273. Zbl 0457.60084. <http://dx.doi.org/10.1007/BF00535742>.
- [Rus82] ———, An approximate zero-one law, *Z. Wahrsch. Verw. Gebiete* **61** (1982), 129–139. MR 0671248. Zbl 0501.60043. <http://dx.doi.org/10.1007/BF00537230>.
- [Tal94] M. TALAGRAND, On Russo’s approximate zero-one law, *Ann. Probab.* **22** (1994), 1576–1587. MR 1303654. Zbl 0819.28002. Available at [http://links.jstor.org/sici?sici=0091-1798\(199407\)22:3<1576:ORAZL>2.0.CO;2-S&origin=MSN](http://links.jstor.org/sici?sici=0091-1798(199407)22:3<1576:ORAZL>2.0.CO;2-S&origin=MSN).

(Received: September 17, 2010)

(Revised: November 12, 2011)

SCHOOL OF COMPUTER SCIENCE, MCGILL UNIVERSITY, MONTRÉAL, CANADA  
E-mail: hatami@cs.mcgill.ca