Noise sensitivity

Ilias Zadik

Part III Essay

Essay setter: Professor Bela Bollobas

May 2014, Cambridge University, Cambridge, UK

Contents

Acknowledgement

I would like to thank Holden Lee for helpful discussions during the creation of this essay and Jason Miller and Sam Watson for letting me use their figures in subsection 7.3.

1 Introduction

In this essay we will study the notion of noise sensitivity. Intuitively for a sequence of Boolean functions f_n we say that it is (asymptotically) noise-sensitive if even a small perturbation on the input $x \in \{-1,1\}^n$ (asymptotically) causes all information we had about the value of f_n to be lost. More rigorously, suppose that for $\varepsilon \in (0, \frac{1}{2})$ we flip independently each coordinate of $x \in \{-1,1\}^n$ with probability $\varepsilon > 0$ and we call the outcome x_ε. Then we say that the sequence of Boolean functions f_n is noise sensitive if $f_n(x)$ and $f_n(x_\varepsilon)$ are asymptotically independent in the sense that if x is chosen randomly and uniformly from $\{-1,1\}^n$ then: $\mathrm{Cov}(f_n(x), f_n(x_\varepsilon)) \to 0$.

From a historical point of view, ideas relevant to noise sensitivity like the noise operator and the Bonami-Gross-Beckner inequality (see Section 5) were firstly studied from Kahn, Kalai and Linial in [KKL88]. What is really important about this paper is that they showed the power that Fourier Analysis on the Boolean hypercube has to solve combinatorial problems on the Boolean hypercube. It was ten years after, that noise sensitivity was rigorously defined from Benjamini,Kalai and Schramm in [BKS99]. In this paper they defined noise sensitivity and stated and proved some equivalent definitions and results about the notion.

After that, in a relevant short period of time many results about noise sensitivity became known. From a mathematical point of view, a lot of big results emerged from the field of Percolation theory. A famous example is the fact that the crossing events for both the bond model for the square lattice and for the site model for the triangular lattice models are noise sensitive. Many results about noise sensitivity are related with the study of Complexity theory. One of these results is the (finally proven) conjecture called "Majority is stablest" proven in [KKMO07]. Some years after, Gil Kalai made another connection between noise sensitive and the field of Social Choice Theory in [Kal10]. He proved that noise sensitivity is translated equivalently to a notion in Social Choice Theory, called social chaos. Finally, many other applications of noise sensitivity became known as for example from the field of Learning theory (for details see [BJT99]). All of the above results have rich combinatorial and probabilistic aspects. Moreover they are more or less characterized by the use of Fourier analysis on the hypercube which became a standard and very useful tool for the subject.

Now let's present the structure we will follow in this essay. We start by devoting Section 2 into explaining the basics about Boolean functions and introducing Fourier analysis on the hypercube. Section 3 is also introductory and states the basic facts we will need about Percolation Theory.

In Section 4 we introduce noise sensitivity. We give some examples and some equivalent characterizations by using Fourier analysis on the hypercube. Then we state and prove an equivalent definition of noise sensitivity given from Benjamini,Kalai and Schramm in [BKS99]. Finally, we state a criterion for a sequence of Boolean function to be noise sensitive, firstly stated and proved also in [BKS99]. We devote the next two sections into proving this criterion with two different ways. In Section 5 we prove the criterion under an extra assumption by using a very important technique coming from Harmonic analysis, called Hypercontractivity method. This method was introduced in [KKL88]. In Section 6, we are presenting the proof of Keller and Kindler gave in [KK13] for the above criterion that actually makes the relation behind this criterion quantitative. We continue with Section 7 where we prove that the percolation crossing events are noise sensitive. We do this in two ways also. The first proof uses an argument based on arm-events theory (we explain the basic facts of this theory in subsection 3.3) and the criterion from [BKS99] we mentioned above. The second proof we refer to is independent and it is based on a connection between randomized algorithms and Boolean functions. We finish the section by stating some important and recent results about the Fourier spectrum of the crossing events. In Section 8 we focus on the special role that Majority Boolean functions has concerning noise sensitivity. We end this section by stating an important proven conjecture, mentioned also above, which is known as "Majority is stablest". We end the essay with Section 9, where we refer to the notion of social chaos in Social Choice Theory which is an equivalent notion to noise sensitivity. This was an equivalence discovered by Kalai in [Kal10].

2 Boolean functions and Fourier analysis

In this section we start by defining what a Boolean function is and we present some key concepts relevant to this definition. Moreover we analyse one basic tool for studying them, which is the use of Fourier analysis on the hypercube.

2.1 Boolean functions

Let $\Omega_n = \{-1,1\}^n$, $n \in \mathbb{N}$.

Definition 2.1. *A **Boolean function** f is a function from the hypercube Ω_n to $\{0,1\}$ or $\{-1,1\}$. For the purpose of this essay we will suppose that the range of a Boolean function is always $\{-1,1\}$ unless something different is said.*

Remark 2.2. *We could have defined Ω_n as the Hamming cube $\{0,1\}^n$ and nothing in this essay would have changed essentially. The reason we have chosen $\{-1,1\}^n$ instead is because it seems a more convenient choice if someone wishes to use Fourier Analysis on the Boolean hypercube as we will do later.*

The space Ω_n will be endowed with the uniform product measure $(\frac{\delta_{-1}}{2} + \frac{\delta_1}{2})^n$, which will be denoted by $\mathbb{P}$ from now on. We will also denote by $\mathbb{E}[\cdot]$ the expectation with respect to this measure.

To illustrate now our definition of a Boolean function we will give some examples:

Example 1 (Dictatorship): $\mathbf{Dict}(x_1,..,x_n) = x_1$. In dictatorship the first coordinate totally decides the outcome.

Example 2 (Parity): $\mathbf{Par}(x_1,..,x_n) = \prod_{i=1}^n x_i$.

But of course a Boolean function need not be defined in a closed form;

Example 3 (Majority): $\mathbf{Maj}(x_1, .., x_n) = \mathbf{sign}(\sum_{i=1}^{n} x_i)$,assuming n is odd.

Example 4 (Graph Property): Take the set of undirected graphs on n vertices and let $P = P_n$ be any graph property (for example P can be the property that a graph G with $|V(G)| = n$ has a clique of size r). Then, whether G has the property P or not, can be characterized by a Boolean function from $\Omega^{\binom{n}{2}}$ to $\{0, 1\}$, firstly identifying G with its edges and after that sending G to 1 if and only if G has the property P.

Before presenting our last example we should mention an easy but important way of thinking about Boolean functions (especially for Section 9 purposes). A Boolean function can also be treated as a voting scheme. Let's try to explain more what we mean by this. Imagine that we have a society consisting of n members and that the society wants to vote between two candidates A and B. Then the way that they will decide after the voting which candidate will be elected can be represented by a Boolean function from Ω_n to $\{-1, +1\}$ (identifying naturally A with 1 and B with -1). Of course the converse is also true: every Boolean function can be interpreted as a voting scheme. Examples 1,3 are cases of known voting schemes but there are more possible ways of creating a reasonable voting method. We now present a similar voting method with Majority but which ,as we will see in the last section, has a fundamental difference with it:

Example 5 (Ternary Majority): Let S be a society with 3^r members who vote between two candidates. To find the result of the election we act as following : Before they vote we order them. Then after they have voted we apply simple majority rule at the 3^{r-1} triples $(3k-2, 3k-1, 3k)$ where $k = 1, .., 3^{r-1}$. Then we take the elected president from each of the triples and we apply the same method again for them but in this case as we had a society of 3^{r-1} members. After r steps we end up with one candidate who is the one that is elected.

Remark 2.3. *Another useful way of thinking about Boolean functions comes after observing that a Boolean function can be identified with the subset of the hypercube* $A_f = \{x \in \Omega_n | f(x) = 1\}$. *Even if we will not really need this interpretation for the rest of the essay it will may end up being very helpful for the understanding of some theorems in later sections.*

2.2 Important concepts

Suppose now that we have a Boolean function $f : \Omega_n \to \{-1, 1\}$, $n \in \mathbb{N}$.

We denote $\{1, 2, .., n\}$ by $[n]$ and for x in Ω^n we write x^i for the vector of the hypercube Ω^n such that $(x^i)_j = x_j$ for $j \neq i$ and $(x^i)_i = -x_i$. It is useful to think that x^i is x with the i-th coordinate flipped. Finally we should remind ourselves that we have endowed Ω_n with the uniform product measure.

Definition 2.4. *We say that* $i \in [n]$ *is pivotal for* f *at* x *if* $f(x) \neq f(x^i)$.

Definition 2.5. *The pivotal set for* f *at* x *is defined to be the random set:*

$$\mathcal{P}(x) = \mathcal{P}_f(x) := \{i \in [n] |\ i \text{ pivotal for } f \text{ at } x\}.$$

We should mention now that under the canonical identification of the Boolean function f with the subset A_f of the hypercube that we mentioned at Remark 2.3, the pairs $\{\{x, x^i\} | i \in \mathcal{P}(x), x \in \Omega_n\}$ define a very important quantity in combinatorics which is called the edge boundary of A_f.

Definition 2.6. *The influence of* $i \in [n]$ *at* f *is defined by* $I_i(f) := \mathbb{E}\big[|f(x) - f(x^i)|\big] = 2\mathbb{P}(i \in \mathcal{P})$.

Denote by $\mathrm{Inf}(f)$ the vector with coordinates $I_i(f)$.

Definition 2.7. *The total influence of f is denoted by $I(f)$ and it is given by:*

$$I(f) = \sum_{i=1}^{n} I_i(f) = ||\mathrm{Inf}(f)||_1.$$

2.3 Fourier analysis on the hypercube

It turns out that to understand and analyse Boolean functions with respect to the uniform measure on the hypercube Ω_n a key tool is to use Fourier Analysis on the hypercube. For this purpose we will introduce now the basic tool we will need from this field, which is called the Fourier-Walsh expansion.

It is more convenient for us to work on the bigger (Hilbert) space $L^2(\Omega_n)$. This is the space of all real valued functions starting from Ω_n with inner product $< f, g >= \sum_{x\in\Omega_n} 2^{-n} f(x)g(x) = \mathbb{E}[fg]$. Consider now for every $S \subset [n]$, the function $\chi_S : \Omega_n \to \mathbb{R}$, given by $\chi_S(x) = \prod_{i\in S} x_i$.

It is an easy check that the functions χ_S where $S \subset [n]$ constitute an orthonormal basis of $L^2(\Omega^n)$. Therefore, every $f : \Omega_n \to \mathbb{R}$ can be written as $f = \sum_{S\subset[n]} \widehat{f}(S)\chi_S$, where $\widehat{f}(S) = \mathbb{E}[f\chi_S]$. This is called the **Fourier-Walsh** expansion of f. Parseval's identity yields that for every function $f \in L^2(\Omega^n)$:

$$\sum_{S\subset[n]} \widehat{f}(S)^2 = ||f||_2^2.$$

If in particular we had that f is a Boolean function then $\mathbb{E}\left[f^2\right] = \mathbb{E}[1] = 1$ which means,

$$\sum_{S\subset[n]} \widehat{f}(S)^2 = 1.$$

Finally, we need one last definition:

Definition 2.8. *For every function $f : \Omega^n \to \mathbb{R}$, the energy spectrum of f at k , $E_f(k)$, is given by:*

$$E_f(k) = \sum_{S\subset[n],|S|=k} \widehat{f}^2(S).$$

3 Basics of Percolation

Since Percolation Theory is an interesting area of Mathematics where noise sensitivity found important applications during the past years, we devote this section on describing some of the area's basic results that we will need for later chapters.

3.1 The model

The two models of Percolation theory that we will be concerned in this essay are 2-dimensional and are based on the **square** lattice $\mathbb{Z}^2$ and one the **triangular** lattice $\mathbb{T}$.

We will first describe the bond percolation model for $\mathbb{Z}^2$. We should first remind ourselves that the lattice $\mathbb{Z}^2$ can be defined as the subset of $\mathbb{C}$ with vertices the complex numbers of the form

$\mathbb{Z}+i\mathbb{Z}$ and two vertices are connected with an edge if and only if the corresponding complex points have distance equal to 1. Now to define bond percolation on $\mathbb{Z}^2$ we act as following. For every $p \in [0,1]$ we consider independently every edge of the graph open with probability p and closed with probability $1-p$. Equivalently for every edge $e \in E(\mathbb{Z}^2)$ we consider a random variable $w(e)$ that follows the Bernoulli(p) distribution. Moreover we demand that the random variables from the family $\{w(e)|e \in E(\mathbb{Z}^2)\}$ are independent. From basic probability theory we know that a big sample space Ω that contains a family of random variables with these properties always exist.

Now let's introduce the site percolation model for the lattice $\mathbb{T}$. The triangular lattice $\mathbb{T}$ can be defined as the subset of $\mathbb{C}$ with vertices the complex numbers of the form $\mathbb{Z}+e^{\frac{\pi i}{3}}\mathbb{Z}$ and we again connect two vertices with an edge if and only if the corresponding complex points have distance equal to 1. To define the site percolation now on $\mathbb{T}$ we act in the exact similar way as with bond on $\mathbb{Z}^2$ with the only difference that in this case we consider the vertices to be closed and open and not the edges.

We should mention now that for convenience we will usually refer to the first model on $\mathbb{Z}^2$ as **bond** percolation and to the second on $\mathbb{T}$ as **site** percolation.

The law of both of these models is denoted by $\mathbb{P}_p(\cdot)$ but every time it will be clear on which model we are referring at. If $p=\frac{1}{2}$ the law will be denoted by $\mathbb{P}$ for convenience.

3.2 Interesting questions and RSW theorem

Let us start with an interesting question that arises naturally from the construction of the two models (and is possibly the reason that these models were firstly introduced). Imagine that we have a liquid that can travel through neighbouring open edges in $\mathbb{Z}^2$ (or through neighbouring open sites in $\mathbb{T}$). Assuming the liquid starts its journey from 0 what is the probability that the liquid will be able to find its way to infinity? For which values of $p \in [0,1]$ this probability is positive? If we consider the subgraph of open edges (vertices) of our graph, the last question can take the following more precise from: For which values of $p \in [0,1]$, the probability that the connected component of 0 in this random subgraph is infinite? We denote this event by $\{0 \leftrightarrow \infty\}$.

The answer to this last question for both of this models is known. It is a celebrated theorem of Kesten in [Kes80] for bond percolation on $\mathbb{Z}^2$ that proves that this probability is zero if and only if $p \leq \frac{1}{2}$. The same is true for site percolation on $\mathbb{T}$. We call $\frac{1}{2}$ the critical value of $\mathbb{Z}^2$ and $\mathbb{T}$ and we denote this by writing $p_c(\mathbb{Z}^2)=p_c(\mathbb{T})=\frac{1}{2}$.

After that it is natural to study the behaviour of the probability of $\{0 \leftrightarrow \infty\}$ as a function of p near $p=\frac{1}{2}$. It is worth mentioning that many basic questions for this behaviour (for example whether this function for bond percolation is right-continuous at $\frac{1}{2}$ or not) are still open.

Another interesting topic that arise quite naturally when one studies percolation models is the study of the "crossing events". Suppose that we have a rectangle. What is the probability that there exist a left-right crossing with open edges (vertices) as a function of p ? The study of such questions are becoming of special interest when $p=p_c(\mathbb{Z}^2)=p_c(\mathbb{T})=\frac{1}{2}$.

About the crossing events there is a very important theorem that we will use a lot in the following chapters.

Theorem 3.1. ***(R.S.W. Theorem*** *) For bond percolation on $\mathbb{Z}^2$ at $p=\frac{1}{2}$, one has the following property for crossing events.Let $a,b>0$.There is a constant $c=c(a,b)>0$,such that, for every $n \geq 1$, if A_n denotes the event that there exist a left right crossing in the rectangle $([0,a\cdot n]\times[0,b\cdot n])\cap\mathbb{Z}^2$, then $c<\mathbb{P}_{\frac{1}{2}}(A_n)<1-c$.*

The same result hold also in the case of site-percolation on $\mathbb{T}$, at $p = \frac{1}{2}$ with the only modification that in this case the part of the rectangle we care about is $([0, a \cdot n] \times [0, b \cdot n]) \cap \mathbb{T}$.

3.3 Arm-events

We will state in this subsection some basic results without proof about a particular type of percolation events, called arm-events. We will not prove the results we will state but we will give appropriate references for the big results we will use. In everything that follows we assume that the percolation model we are treating each time is for $p = \frac{1}{2}$.

Let's focus firstly on the **site** percolation on $\mathbb{T}$.

Let $A(r, R) \subset \mathbb{C}$ be an annulus with radii $0 \leq r < R$. Let j be an odd natural number. Then the **j-th arm event** is denoted by $A_j(r, R)$ and it is the event that there exist j disjoint paths where each connect the inner boundary of the annulus with the outer boundary and every path's vertices are either all open or all closed. We ask also that the status of the paths alternate in circular order between consisting only of open and only of closed vertices. In the case j is odd the definition is similar except that in this case we ask for $j-1$ of the paths to have alternating status in circular order and for the extra path we ask just to consist of open vertices. From now on when we refer to those paths we will call them arms.

Let's illustrate more the case where $r = 0$ and $j = 1, 2, 4, 5$ to get a better feeling on what is happening. For any radius $R > 1$, we set $A_R^1 := A_1(0, R)$ the event that the site 0 is connected to the circle $C(0, R)$ by some open arm. By saying this, we mean that there exists a path with open vertices (except possibly 0) which starts from site 0 and crosses the circle with radius R and centre 0. Similarly we set A_R^2 the event that the site 0 is connected with one open arm and one closed arm to distance R which are disjoint, A_R^4 the event that the site 0 is connected to distance R with 4 disjoint arms of alternating status (open-closed-open-closed) and finally A_R^5 the event that there are 5 arms from 0 to distance R where 4 of them alternate between being open and closed and 5th arm is just open. We note that in all of the above events, as we mentioned after defining A_R^1, the site 0 can be either open or closed. It should be clear by now that when we say that the arm has distance R we do not mean that the graph-theoritic length of the path that defines the arm is R but that the arm starts from 0 and crosses the circle with center 0 and distance R. Of course because of the translation invariance of our percolation models the fact that we picked 0 as the start of our arms is not important. We could have chosen any other site instead.

Now back to the more general annulus case we denote $a_i(r, R) = \mathbb{P}(A_i(r, R))$ for $i = 1, 2, 4, 5$.

It was proved in [LSW02] that $a_1(R) = R^{-\frac{5}{48}+o(1)}$,in [SW01] that $a_2(R) = R^{-\frac{1}{4}+o(1)}$ and that $a_4(R) = R^{-\frac{5}{4}+o(1)}$. Finally in [Wer09] it is proven that $a_5(R) = \Omega(R^{-2})$. For the 5-arm event we know also that in the general annulus case: $a_5(r, R) = \Omega((\frac{r}{R})^2)$.

We need now to define some other types of arm events also. Define the upper hyperplane in $\mathbb{C}$ by $\mathbb{H}$. The **2-arm event in the half-plane** $\mathbb{H}$ for distance R is defined to be the event that there exist one open and one closed arm from site 0 to distance R that stay inside $\mathbb{H}$ and are disjoint. The **3-arm event in the half plane** $\mathbb{H}$ for distance R demands that there are three disjoint arms from site 0 to distance R that stay inside $\mathbb{H}$ and have alternating status in circular order (two open arms and one closed arm in the middle). We name the corresponding probabilities by $a_2^+(R), a_3^+(R)$. As one can find in [Wer09] it holds: $a_2^+(r, R) = \Omega(\frac{r}{R})$ and $a_3^+(r, R) = \Omega((\frac{r}{R})^2)$.

Finally we are interested in one more type of arm-event, the **2-arm event in quarter plane** for distance R which corresponds to the event of having one closed and one open arm starting from site 0 to distance R which are disjoint and we demand also that both lie inside the quarter plane. We denote the probability of this event as $a_2^{++}(R)$ which can be proved that it equals to $R^{-2+o(1)}$. Furthermore with the previous notation it can be shown that $a_2^{++}(r,R) = (\frac{r}{R})^{2+o(1)}$. Both results were proved in [SW01].

Now we will state the corresponding but weaker known results about the arm-event for the model of **bond** percolation on $\mathbb{Z}^2$. The definition for arm-events is exactly the same. We take $A(r,R) \subset \mathbb{C}$ an annulus and we define the j-th arm-event as following: If j is even we demand the existence of j (vertex)-disjoint "one-status arms" that connect the inner boundary of the annulus to the outer and we ask also for the status of the arms to alternate between being open and closed. If j is odd we ask for the existence of $j-1$ arms with the above properties and moreover we ask for one extra disjoint arm with open edges also. Of course now by saying open (or closed) arm we refer to the status of the edges.

We should mention now that one reason there is a gap between understanding the arm events for site percolation on $\mathbb{Z}^2$ and bond percolation on $\mathbb{T}$ is that Stanislav Smirnov proved at 2001 an important result about conformal invariance of the site triangular model which is still a big open problem for the bond percolation on $\mathbb{Z}^2$.

Now what do we know for $\mathbb{Z}^2$: By using R.S.W. theorem one can deduce that $a_1(r,R) \leq (\frac{r}{R})^{\alpha}$, for some $\alpha > 0$. It can also be proven that $c(\frac{r}{R})^{2-e} \leq a_4(r,R) \leq C(\frac{r}{R})^{1+e'}$, for some constants $c, e, e', C > 0$ and that $a_5(r,R) = \Omega((\frac{r}{R})^2)$.

For the $\mathbb{H}$-arm events we have the exact same results as with site percolation: $a_2^+(r,R) = \Omega(\frac{r}{R})$ and $a_3^+(r,R) = \Omega((\frac{r}{R})^2)$.

Finally for both our models we will use a lot a property of the 4-arm event called quasi-multiplicity:

Theorem 3.2. *(Quasi-multiplicity)*
For $r_1 < r_2 < r_3$ both for site percolation on $\mathbb{T}$ and bond percolation on $\mathbb{Z}^2$ percolation it holds:

$$a_4(r_1,r_2) \cdot a_4(r_2,r_3) = \Omega(a_4(r_1,r_3))$$

4 Noise sensitivity

In this section we start by introducing the fundamental concept of noise sensitivity. We will illustrate it with some examples and relevant computations and we will state and prove an equivalent definition. We are ending the section with mentioning a very important theorem about this concept.

4.1 Definition of noise sensitivity and stability

Let Ω_n be, as in the first section, the set $\{-1,1\}^n$, $n \in \mathbb{N}$ endowed with the uniform product measure. Suppose now that $\varepsilon \in (0,1)$ and x is uniformly chosen from Ω_n. Set x_ε to be the perturbed x in the sense that each coordinate of x is **remains unchanged** with probability $1-\varepsilon$ and it is **"rerandomized"** otherwise. By "rerandomized" we mean that independently it chooses its new value from $\{-1,1\}$ with probability $\frac{1}{2}$. Of course the interesting cases are when ε is close to zero. One can easily check that x_ε and x has the same (uniform) distribution.

Remark 4.1. *In the first paper on noise sensitivity , [BKS99], the perturbed x denoted again by x_ε, was defined a bit differently; according to this paper each coordinate of x_ε remains the same with probability $1-\varepsilon$ and it changes its value with probability ε. The important thing is that there* *is no essential difference between the two models; the model we describe now is our model for noise equal to $2 \cdot \varepsilon$. To see this observe that in our model each coordinate of x is flipped with probability exactly $\frac{\varepsilon}{2}$.*

Now we are ready to introduce the fundamental notion of this essay:

Definition 4.2. *Let $f_n : \Omega_{m_n} \to \{-1,1\}$ be a sequence of Boolean function with $m_n \to \infty$. Then we say that the sequence f_n is (asymptotically)* ***noise sensitive*** *if for every $\varepsilon \in (0,1)$ it holds that:*

$$\mathbb{E}[f_n(x_\varepsilon)f_n(x)] - \mathbb{E}[f_n(x)]^2 = \mathrm{Cov}(f_n(x), f_n(x_\varepsilon)) \to 0, \quad \text{as} \quad n \to \infty.$$

Intuitively, noise sensitivity says that (asymptotically) the value of $f_n(x)$ becomes independent of the value of $f_n(x_\varepsilon)$, for every fixed $\varepsilon > 0$.

One should observe that,

$$\mathrm{Cov}(f_n(x), f_n(x_\varepsilon)) = 4 \cdot \mathrm{Cov}(\frac{f_n(x)+1}{2}, \frac{f_n(x_\varepsilon)+1}{2})$$

and therefore the definition of noise sensitivity is valid equivalently if we consider the corresponding Boolean functions that take values in $\{0,1\}$ by changing from f to $2f-1$ and vice versa.

An opposite notion to noise sensitivity is noise stability.

Definition 4.3. *We say that a sequence $f_n : \Omega_{m_n} \to \{-1,1\}$ of Boolean functions with $m_n \to \infty$ is* ***noise stable*** *if*

$$\sup_n \mathbb{P}(f_n(x) \neq f_n(x_\varepsilon)) \to 0,$$

as $\varepsilon \to 0$.

Intuitively noise stability implies that with high probability $f_n(x)$ remains unchanged under small random perturbations of x.

A natural question that arises is whether it is possible for some sequence of Boolean function to be also noise sensitive and noise stable or not. It ends up that this family of sequences is degenerate (as it should) in the following sense,

Proposition 4.4. *If a sequence of Boolean functions $f_n : \Omega_{m_n} \to \{-1,1\}$ with $m_n \to \infty$ is noise sensitive and noise stable then* $\mathrm{Var}(f_n) \to 0$.

Proof. Since f_n are Boolean functions for every $\varepsilon > 0$:

$$4 \cdot \mathbb{P}(f_n(x) \neq f_n(x_\varepsilon)) = \mathbb{E}\big[|f_n(x_\varepsilon) - f_n(x)|^2\big] = 2 \cdot (\mathbb{E}\big[f_n(x)^2\big] - \mathbb{E}[f_n(x)f_n(x_\varepsilon)])$$

and therefore:

$$4 \cdot \mathbb{P}(f_n(x) \neq f_n(x_\varepsilon)) = 2 \cdot (\mathrm{Var}(f_n(x)) - \mathrm{Cov}(f(x), f(x_\varepsilon)),$$

or,

$$2 \cdot \mathbb{P}(f_n(x) \neq f_n(x_\varepsilon)) + \mathrm{Cov}(f_n(x), f_n(x_\varepsilon)) = \mathrm{Var}(f_n(x)) \text{ for every } \varepsilon > 0.$$

Now from this point we first choose $\varepsilon > 0$ small enough to make the first term sufficiently small using the noise stability of f_n. Then we send n to infinity and noise sensitivity of f_n gives the desired result. □

The above proof yields the following very interesting formula which we should keep in our mind for later use:

Proposition 4.5. *For every Boolean function $f : \Omega_n \to \{-1, 1\}$, it is true that*

$$2 \cdot \mathbb{P}(f(x) \neq f(x_\varepsilon)) + \mathrm{Cov}(f(x), f(x_\varepsilon)) = \mathrm{Var}(f(x))$$

Now let's try to see some detailed examples of functions with these properties to gain some better understanding on what is happening.

Example 1 (Dictatorship)
For every $n \in \mathbb{N}$, let $\mathbf{Dict}_n : \Omega_n \to \{-1, 1\}$ be the function given by $\mathbf{Dict}_n(x_1, .., x_n) = x_1$.

We claim that this function is noise stable and not noise sensitive. This can be intuitively explained based on the fact that we have dependence only in the first variable and therefore the probability that the reuslt will change as n increases remains the same positive number. Let's make this more precise.

The first part of the claim follows since $\mathbb{P}(\mathbf{Dict}_n(x) \neq \mathbf{Dict}_n(x_\varepsilon)) = \frac{\varepsilon}{2}$, for every $n \in \mathbb{N}$.

For the second part of the claim we observe that $\mathrm{Cov}(\mathbf{Dict}_n(x), \mathbf{Dict}_n(x_\varepsilon)) = \mathbb{E}[x_1(x_\varepsilon)_1] = 1-\varepsilon > 0$, for every $n \in \mathbb{N}$ and we are done. We could have argued based on Proposition 4.1 instead, by observing that $\mathrm{Var}(\mathbf{Dict}_n(x)) = 1$ for every $n \in \mathbb{N}$ and therefore noise sensitivity and noise stability can not happen simultaneously.

Example 2 (Parity)
For every $n \in \mathbb{N}$, let $\mathbf{Par}_n : \Omega_n \to \{-1, 1\}$, be the function given by $\mathbf{Par}_n(x) = \prod_{i=1}^n x_i$.

We claim that this sequence is noise sensitive and not noise stable. This can be intuitively explained based on the fact that any odd numbers of flipped coordinates will change the result.

For the first claim we observe that: $\mathrm{Cov}(\mathbf{Par}_n(x), \mathbf{Par}_n(x_\varepsilon)) = \mathbb{E}[\prod_{i=1}^n x_i(x_\varepsilon)_i] = (1-\varepsilon)^n \to 0$, for every $\varepsilon \in (0, 1)$.

For the second claim we see that $\mathrm{Var}(\mathbf{Par}_n(x)) = 1$ and a call to Proposition 4.1 gives the result.

Majority (Majority)
For every odd $n \in \mathbb{N}$, let $\mathbf{Maj}_n : \Omega_n \to \{-1, 1\}$ be the function given by $\mathbf{Maj}_n(x_1, .., x_n) = \mathbf{sign}(\sum_{i=1}^n x_i)$.

The claim is that this sequence, like dictatorship, it is noise stable and it is not noise sensitive.

Since $\mathrm{Var}(\mathbf{Maj}_n(x)) = 1$ by Proposition 4.1 it is enough to prove it is noise stable. It can actually be proved something more quantitative that goes back to Sheppard:

For every $n \in \mathbb{N}$ the quantity $\mathbb{P}(\mathbf{Maj}_n(x) \neq \mathbf{Maj}_n(x_\varepsilon))$ behaves like $\frac{\sqrt{2\varepsilon}}{2\pi}$ for big values of n with an error term progressing like $O\left(\frac{1}{\sqrt{1-\varepsilon^2}\sqrt{n}}\right)$.

The above theorem implies indeed the noise stability of the sequence of Majority Boolean functions. We will not give a proof of this theorem but we will try to give a convincing heuristic explanation of the fact that it behaves like $\sqrt{\varepsilon}$ (from [Kal10]):

From central limit theorem we know that the median gap between the number of coordinates of x that have the value 1 and of the coordinates that take the vaue 0 is close to $\sqrt{n}$. Now observe that around $\varepsilon \cdot n$ will be flipped after the "re-randomization" and the median gap in this case between the 0's that become 1's and the 1's that become 0's, is also based on central limit theorem around $\sqrt{\varepsilon \cdot n}$. Therefore for the "re-randomization" to change the first result, the first median gap should be also around $\sqrt{\varepsilon \cdot n}$. Now by making a last call at central limit theorem, this probability is close to $\mathbb{P}(Z \text{ is near } \sqrt{\varepsilon})$, where Z follows the $N(0,1)$ distribution. But for small values of ε is indeed close to $\sqrt{\varepsilon}$ as we wanted.

As we will mention also later there is a very big Theorem concerning the "stability" of Majority functions, stating that among all Boolean functions with uniformly low influences the majority function is asymptotically the "stablest". (For the rigorous statement of this result see Section 8).

Example 4 (Crossing events)

One of the most early and important application of noise sensitivity in Percolation theory is the fact that the crossing events of the rectangles $[0, a \cdot n] \times [0, b \cdot n]$ for bond percolation on $\mathbb{Z}^2$ and for site percolation on $\mathbb{T}$ are noise sensitive. The proofs of these facts are non trivial and we will focus on them in chapter 7.

4.2 The use of Fourier analysis

As one may have noticed, even when we were dealing with easily defined Boolean function as Majority or indicators of "crossing" events we could not easily say whether they are noise sensitive or not. This proposes that we may need some new machinery to work with and indeed the use of Fourier analysis is the key we need.

Take a Boolean function $f : \Omega_n \to \{-1, 1\}$ and it's Fourier-Walsh expansion $f = \sum_{S \subset [n]} \widehat{f}(S)\chi_S$.

Let's try now to express $\text{Cov}(f(x), f(x_\varepsilon))$ in term of the Fourier coefficients of f.

By direct computations we get $\mathbb{E}[f(x)f(x_\varepsilon)] = \sum_{S,T \subset [n]} \widehat{f}(S)\widehat{f}(T)\mathbb{E}[\chi_S(x)\chi_T(x_\varepsilon)]$.

Therefore we need to calculate $\mathbb{E}[\chi_S(x)\chi_T(x_\varepsilon)]$, for $S, T \subset [n]$. It easy to see that as the coordinates take values independently of each other and have mean zero, if $S \neq T$ then $\mathbb{E}[\chi_S(x)\chi_T(x_\varepsilon)]$ contains as a factor, either $\mathbb{E}[x_i] = 0$ or $\mathbb{E}[(x_\varepsilon)_i] = 0$.

Therefore in the case $S \neq T$ it holds $\mathbb{E}[\chi_S(x)\chi_T(x_\varepsilon)] = 0$.

That means that we need only to consider the terms of the form $\mathbb{E}[\chi_S(x)\chi_S(x_\varepsilon)]$. But for every $i \in [n]$, we can easily compute $\mathbb{E}[x_i(x_\varepsilon)_i] = 1 - \varepsilon$,hence: $\mathbb{E}[\chi_S(x)\chi_S(x_\varepsilon)] = (1 - \varepsilon)^{|S|}$.

It follows that:

$$\mathbb{E}[f(x)f(x_\varepsilon)] = \sum_{S \subset [n]} \widehat{f}(S)^2(1 - \varepsilon)^{|S|}$$

or

$$\text{Cov}(f(x), f(x_\varepsilon)) = \mathbb{E}[f(x)f(x_\varepsilon)] - \mathbb{E}[f(x)]^2 = \sum_{S \subset [n], S \neq \varnothing} \widehat{f}(S)^2(1 - \varepsilon)^{|S|}, \tag{4.1}$$

since $\widehat{f}(\varnothing) = \mathbb{E}[f(x)]$.

An easy but very important corollary of the above formula and Parseval's identity is the following one:

Proposition 4.6. *([BKS99])*

For a sequence of Boolean functions $f_n : \Omega_{m_n} \to \{-1, 1\}$ with $m_n \to \infty$, f_n is noise sensitive if and only if, for every $k \geq 1$, $\sum_{S \subset [m_n], |S| = k} \widehat{f_n}(S)^2 \to 0$.

Proof. For the one direction, fix $k \geq 1$. Then for fixed $\varepsilon \in (0,1)$:

$$\sum_{S \subset [m_n], |S|=k} \widehat{f_n}(S)^2 (1-\varepsilon)^k \leq \sum_{S \subset [n], S \neq \varnothing} \widehat{f}(S)^2 (1-\varepsilon)^{|S|} = \mathrm{Cov}(f(x), f(x_\varepsilon)) \to 0$$

and we are done.

For the other direction, fix any $\varepsilon \in (0,1)$. Then:

For every $N \in \mathbb{N}$,

$$\mathrm{Cov}(f(x), f(x_\varepsilon)) = \sum_{S \subset [n], S \neq \varnothing} \widehat{f}(S)^2 (1-\varepsilon)^{|S|}$$

is at most,

$$\leq (1-\varepsilon)^N \sum_{S \subset [m_n], |S| \geq N} \widehat{f_n}(S)^2 + \sum_{S \subset [n], S \neq \varnothing, |S| < N} \widehat{f}(S)^2 (1-\varepsilon)^{|S|}.$$

which by Parseval's identity, since f_n are Boolean functions, is bounded above by:

$$(1-\varepsilon)^N + \sum_{S \subset [n], S \neq \varnothing, |S| < N} \widehat{f}(S)^2 (1-\varepsilon)^{|S|}.$$

Now choose first N big enough to make the first term small and use our hypothesis after to make the second term also small. We are done.

□

But a better look at the previous statement and it's proof gives the following result:

Proposition 4.7. *For a sequence of Boolean function $f_n : \Omega_{m_n} \to \{-1,1\}$ with $m_n \to \infty$ if* $\mathrm{Cov}(f_n(x), f_n(x_\varepsilon)) \to 0$ *for some $\varepsilon \in (0,1)$, then* $\mathrm{Cov}(f_n(x), f_n(x_\varepsilon)) \to 0$ *for every $\varepsilon \in (0,1)$.*

Hence, if for one $\varepsilon \in (0,1)$, $\mathrm{Cov}(f(x), f(x_\varepsilon)) \to 0$ *then the sequence $(f_n)_{n \in \mathbb{N}}$ is noise-sensitive.*

Proof. Indeed if for one $\varepsilon \in (0,1)$,

$$\sum_{S \subset [n], S \neq \varnothing} \widehat{f}(S)^2 (1-\varepsilon)^{|S|} = \mathrm{Cov}(f(x), f(x_\varepsilon)) \to 0$$

then as in the above proof,

$$\sum_{S \subset [m_n], |S|=k} \widehat{f_n}(S)^2 \to 0 \text{ for every } k \geq 1.$$

But then by the converse in Proposition 4.6. f_n is noise sensitive which by definition means

$$\mathrm{Cov}(f(x), f(x_\varepsilon)) \to 0 \text{ for every } \varepsilon \in (0,1).$$

□

Now about the notion of noise stability we observe:

$$2 \cdot \mathbb{P}(f(x) \neq f(x_\varepsilon)) = \mathrm{Var}(f(x)) - \mathrm{Cov}(f(x), f(x_\varepsilon)) = \mathbb{E}\big[f^2(x)\big] - \mathbb{E}[f(x) f(x_\varepsilon)] = \sum_{S \subset [n]} \widehat{f}(S)^2 \left(1 - (1-\varepsilon)^{|S|}\right),$$

where for the last equality we have used Parseval's identity and (4.1).

Hence,

$$2 \cdot \mathbb{P}(f(x) \neq f(x_\varepsilon)) = \sum_{S \subset [n]} \widehat{f}(S)^2 \left(1 - (1-\varepsilon)^{|S|}\right), \tag{4.2}$$

which it is easy based again on Parseval's identity as in the proof of Proposition 4.6 to give:

Proposition 4.8. *([BKS99])*

For a sequence of Boolean functions $f_n : \Omega_{m_n} \to \{-1,1\}$ with $m_n \to \infty$,
f_n is noise stable if and only if for every $\varepsilon > 0$ there exists $k \in \mathbb{N}$ such that for all $n \in \mathbb{N}$,

$$\sum_{|S| \geq k} \widehat{f_n}(S)^2 < \varepsilon.$$

4.3 An alternative definition

In the very first paper on noise sensitivity, [BKS99], from Benjamini, Kalai and Schramm noise sensitivity was defined in the following way:

Definition 4.9. *A sequence of Boolean functions $f_n : \Omega_{m_n} \to \{-1,1\}$ wtih $m_n \to \infty$ is (asymptotically)* BKS-*noise sensitive if for every $\varepsilon \in (0,1)$, $\mathbb{E}[f_n(x_\varepsilon)|x] \to \mathbb{E}[f_n(x)]$, in probability as $n \to \infty$.*
***Caution**:The first expectation is taken only with respect to the noise.*

We will prove in this subsection that the definition we gave and this one are actually equivalent, i.e. that a sequence of Boolean function is (asymptotically) noise sensitive if and only if it is (asymptotically) B.K.S.-noise sensitive.

Definition 4.10. *For a Boolean function $f : \Omega_n \to \{-1,1\}$, define for $\delta, \varepsilon \in (0,1)$,*

$$\gamma(f,\delta,\varepsilon) = \mathbb{P}(|\mathbb{E}[f(x_\varepsilon)|x)] - \mathbb{E}[f(x)]\,| > \delta)$$

*We say that the **sensitivity gauge** of f is the quantity*

$$\varphi(f,\varepsilon) = \inf\{\delta > 0 | \gamma(f,\delta,\varepsilon) < \delta\}$$

It is easy to prove that :

Proposition 4.11. *A sequence of Boolean functions $f_n : \Omega_{m_n} \to \{-1,1\}$ is (asymptotically)* BKS-*noise sensitive if and only if for every $\varepsilon \in (0,1)$, $\varphi(f_n,\varepsilon) \to 0$.*

Now we are ready to state the key proposition for the equivalence.

Proposition 4.12. *For a Boolean function $f : \Omega_n \to \{-1,1\}$ and $\varepsilon \in (0,1)$ set $g(x) = \mathbb{E}[f(x_\varepsilon)|x]$. Then it holds,*

$$\frac{\mathrm{Var}(g)}{5} \leq \varphi(f,\varepsilon) \leq \mathrm{Var}(g)^{\frac{1}{3}}.$$

Hence, $\varphi(f_n,\varepsilon) \to 0$ if and only if $\mathrm{Var}(\mathbb{E}[f_n(x_\varepsilon)|x]) \to 0$, and therefore: a sequence f_n is B.K.S.-noise sensitive if and only if $\mathrm{Var}(\mathbb{E}[f_n(x_\varepsilon)|x]) \to 0$.

Proof. Set $\delta := \varphi(f,\varepsilon) > 0$ and $Y = \{y \in \Omega_n \Big| |g(y) - \mathbb{E}[f(x)]| \geq \delta\}$.

Then by the definition of δ, it holds: $\mathbb{P}(Y) \geq \delta$.

Therefore,

$$Var(g) \geq \delta^2 \cdot \mathbb{P}(\delta) \geq \delta^3.$$

For the opposite inequality, set $Y' = \{y \in \Omega_n \Big| |g(y) - \mathbb{E}[f(x)]| > \delta\}$.

Then again by the definition of δ, $\mathbb{P}(Y') \leq \delta$.

Hence,

$$Var(g) = \mathbb{E}\big[|g(y) - \mathbb{E}[f(x)]|^2 \cdot 1(Y'^c)\big] + \mathbb{E}\big[|g(y) - \mathbb{E}[f(x)]|^2 \cdot 1(Y')\big] \leq \delta + 4\mathbb{P}(Y') \leq 5\delta.$$

,where we have used the fact that $|g(y) - \mathbb{E}[f(x)]| \leq 2$.
We are done. □

Having now Proposition 4.11 in our disposal it is reasonable to try to compute the variance of $\mathbb{E}[f(x_\varepsilon)|x]$ with some Fourier analysis tools.

For a Boolan function $f : \Omega_n \to \{-1, 1\}$ and $i \in [n]$ it holds: $\mathbb{E}[(x_\varepsilon)_i|x] = (1-\varepsilon)x_i$ and because of independence $\mathbb{E}[\chi_S(x_\varepsilon)|x] = (1-\varepsilon)^{|S|}\chi_S(x)$.

Therefore by taking the Fourier-Walsh expansion of f,

$$\mathbb{E}[f(x_\varepsilon)|x)] = \sum_{S\subset[n]} \widehat{f}(S)(1-\varepsilon)^{|S|}\chi_S(x)$$

which gives that,

$$\mathrm{Var}(\mathbb{E}[f(x_\varepsilon)|x]) = \sum_{S\subset[n],S\neq\varnothing} \widehat{f}(S)^2(1-\varepsilon)^{2|S|}.$$

Now we are ready to prove the equivalence:

Theorem 4.13. *A sequence of Boolean functions $f_n : \Omega_{m_n}$ with $m_n \to \infty$ is noise sensitive if and only if it is B.K.S.-noise sensitive.*

Proof. We have proved that f_n is B.K.S.- noise sensitive if and only if $\mathrm{Var}(\mathbb{E}[f_n(x_\varepsilon)|x]) \to 0$.

But as we just proved:

$$\mathrm{Var}(\mathbb{E}[f_n(x_\varepsilon)|x]) = \sum_{S\subset[m_n],S\neq\varnothing} \widehat{f}(S)^2(1-\varepsilon)^{2|S|}.$$

Now since for every $n \in \mathbb{N}$,

$$\sum_{S\subset[m_n]} \widehat{f_n}(S)^2 = 1$$

it is easy to prove that:

$$\sum_{S\subset[n],S\neq\varnothing} \widehat{f_n}(S)^2(1-\varepsilon)^{2|S|} \to 0 \text{ if and only if } \sum_{S\subset[n],S\neq\varnothing} \widehat{f_n}(S)^2(1-\varepsilon)^{|S|} \to 0.$$

Therefore from relation 4.1 we get our desired result. □

4.4 Influences and noise sensitivity

Having defined noise sensitivity and analysed it through Fourier analysis we have translated the notion in a different language but still we did not get an efficient criterion to be able to judge whether a sequence of Boolean function is noise sensitive or not. Computing the squares the Fourier-Walsh coefficients requires a lot of computations which in some situations it is quite hard. It turns out that there is a beautiful B.K.S. theorem again proved in [BKS99] that gives a criterion which involves influences. Influences, as noise sensitivity, measure how likely it is for the value to change when the "input" is perturbed. It turns out that, maybe counter-intuitively, functions with low influences are noise sensitive.

Theorem 4.14. *(B.K.S. theorem,[BKS99]) Let $f_n : \Omega_{m_n} \to \{-1,1\}$ be a sequence of Boolean function with $m_n \to \infty$. If $\mathbb{I}(f_n) := \sum_{i=1}^{m_n} I_i(f_n)^2 \to 0$, then the sequence f_n is noise sensitive.*

What about the converse? It is easy to see that it is **not** generally true.

For example, if f_n is the sequence of parity functions then it is noise sensitive (example 2 in subsection 4.1) but for every i,

$$I_i(f_n) = \mathbb{E}\left[|x_j \prod_{j\neq i} x_i - (-x_j)\prod_{j\neq i} x_i\right] = \mathbb{E}\left[2|\prod_{j\neq i} x_j|\right] = 2.$$

Despite that it turns out that the converse is true for sequences of **monotone** Boolean functions: We say that a Boolean function $f : \Omega_n \to \{-1,1\}$ is **monotone** increasing (decreasing) if for every pair $x, y \in \Omega_n$ with $x \leq y$ (meaning that $x_i \leq y_i$ for every $i \in [n]$) it holds $f(x) \leq f(y)$ $(f(x) \geq f(y))$.

Theorem 4.15. *If a sequence $f_n : \Omega_{m_n} \to \{-1,1\}$ of monotone Boolean functions is noise sensitive then $\mathbb{I}(f_n) \to 0$.*

Proof. By switching from f_n to $-f_n$ if necessary we may assume that f_n is a sequence of monotone increasing Boolean functions. Since the sequence is noise sensitive from Proposition 4.6 and $k = 1$ we have

$$\sum_{i=1}^{m_n} \widehat{f_n}(\{i\})^2 \to 0.$$

Now we need to observe that if $f : \Omega_n \to \{-1,1\}$ is an increasing Boolean function then for every $i \in [n]$,

$$\widehat{f}(\{i\}) = \mathbb{E}[f(x)x_i] = \frac{1}{2}\mathbb{E}_{x_{-i}\in\{-1,1\}^{[n]-\{i\}}}[f(x_{-i}|1) - f(x_{-i}|-1)] = \frac{I_i(f)}{2}.$$

where the last equality follows by the increasing property of f and we denote by $(x_{-i}|j)$ the vector x which equals x_{-i} for the coordinates different from i and satisfies $x_i = j$.

Therefore for f increasing,

$$I_i(f) = 2\widehat{f}(\{i\}).$$

Hence,

$$\mathbb{I}(f_n) = \sum_{i=1}^{m_n} I_i(f_n)^2 = 4\sum_{i=1}^{m_n} \widehat{f_n}(\{i\})^2 \to 0$$

as we wanted. □

A next important question is whether the relation between $\mathbb{I}(f)$ and $\mathrm{Cov}(f(x), f(x_\varepsilon))$ which is hidden behind B.K.S. theorem can be made quantitative. Keller and Kindler in [KK13] answered affirmative to this question:

Theorem 4.16. *([KK13])*

There exists $C > 0.234$ such that the following holds: Let $f : \Omega_n \to \{-1, 1\}$ be a Boolean function. Then for every $\varepsilon \in (0, 1)$:

$$\mathrm{Cov}(f(x), f(x_\varepsilon)) \leq (6e + 1)(\frac{\mathbb{I}(f)}{4})^{C\cdot\varepsilon}$$

Theorem 4.14 under an extra assumption will be proved in the next section. Theorem 4.16 will be proved in the section after.

5 Use of Hypercontractivity to prove B.K.S. theorem under polynomial decay

At this section we present the method of Hypercontractivity in the hypercube (firstly introduced in [KKL88]) and as an application of this method we present a proof of B.K.S. theorem (theorem 4.14) under the extra assumption that $\mathbb{I}(f_n) \leq m_n^{-\delta}$ for some $\delta > 0$. We follow the path of the proof given in [GS12]. The theorem even under this assumption remains important because many interesting sequences of Boolean functions, like the indicators of "crossing" events in percolation theory (see subsection 7.1) satisfy this extra hypothesis.

5.1 The method

Let's first state what Hypercontractivity is in the continuous case.

Let K_t be the heat kernel in $\mathbb{R}^n$.

Theorem 5.1. *(Hypercontractivity)*
Consider $\mathbb{R}^n$ with the standard Gaussian measure. If $1 < q < 2$, then there is some $t = t(q) > 0$ such that for every $f \in L^q(\mathbb{R}^n)$ then $||K_t * f||_2 \leq ||f||_q$.

This means the following regularization result. If someone starts with a function $f \in L^q$ and wait some time $t = t(q) > 0$, under the heat kernel, then this function will end up being in L^2 and we will have a very good control on it's L^2-norm. In one sentence: hypercontractivity is a result for the existence of a contraction from L^q to L^2.

Now let's see the corresponding result we need for the discrete case on the hypercube.

We need first to define the **noise operator**.

Definition 5.2. *For $p \in [0, 1]$ define T_p to be the noise operator on the set of functions from Ω_n to $\mathbb{R}$ in the following way: Every function $f = \sum \widehat{f}(S)\chi_S$ is mapped into $T_p(f) := \sum \widehat{f}(S)p^{|S|}\chi_S$.*

It can be easily checked from the work we did in subsection 4.3 that for every f,

$$\mathcal{T}_p(f) = \mathbb{E}[f(x_{1-p})|x]\,,$$

where the expectation is taken with respect to the noise. It is this observation that justifies the name of our operator.

Now we are ready to state our result.

Theorem 5.3. *(Bonami-Gross-Beckner)*
For every $f:\Omega_n \to \mathbb{R}$, it is true that if $1 \leq s \leq t$ and $0 \leq p \leq \sqrt{\frac{s-1}{t-1}}$ then

$$||T_p(f)||_t \leq ||f||_s.$$

In particular for $t = 2, s = 1 + p^2, p \in [0, 1]$ we get (B.G.B. inequality):

$$||T_p(f)||_2 \leq ||f||_{1+p^2}$$

The correspondence between Theorem 5.3 and B.G.B. inequality is transparent: T_p is a contraction from L^{1+p^2} to L^2.

Now let's try to explain what actually theorem 5.1 implies about the Fourier coefficients of f that interest us.

Remark 5.4. *(from [GS12]) B.G.B. inequality tells us, non-rigorously speaking, that every function $f : \Omega_n \to \{-1, 0, 1\}$ with small support has it's frequencies $\{\widehat{f}(S)\}$ supported on large sets.*
Let's try to support this claim. Assume that f has small support in the sense that $||f||_2$ is small. Then for every $p \in (0, 1)$, B.G.B. inequality and the low support together imply that:

$$||T_p(f)||_2 \leq ||f||_{1+p^2} = ||f||_2^{\frac{2}{1+p^2}} \ll ||f||_2,$$

where we have used the fact that $|f| = 0, 1$ implies $||f||_{1+p^2}^{1+p^2} = ||f||_2^2$.
Now squaring and using Parseval's identity we get:

$$\sum p^{|S|}\widehat{f}(S)^2 \ll \sum \widehat{f}(S)^2$$

The last relation indeed implies intuitively that $\widehat{f}(S) \neq 0$ for many large sets S as we wanted.

It is worth mentioning that this is a result that comes in accordance with the intuition coming from Weyl-Heisenberg uncertainty principle according to which small support of a function implies that it's Fourier spectrum must be quite spread out and therefore supported on high frequencies.

5.2 Proof of BKS under polynomial decay

Now let's try to work on the proof of theorem 4.14 **under the extra assumption** that $\mathbb{I}(f_n) \leq m_n^{-\delta}$ for some $\delta > 0$.

Summarizing, we will prove that if a sequence of Boolean functions $f_n : \Omega_{m_n} \to \{-1, 1\}$ with $m_n \to \infty$ satisfies $\mathbb{I}(f_n) \leq m_n^{-\delta}$ for some $\delta > 0$ then it is noise sensitive.

Based on Remark 5.4., it is reasonable to try to apply B.G.B. inequality efficiently. We need some definitions to achieve that.

Let $f : \Omega_n \to \{-1, 1\}$ be a Boolean function.

For every $i \in [n]$ define the function $\nabla_i(f) : \Omega_n \to \{-1, 0, 1\}$ given by: $\nabla_i(f)(x) = \frac{1}{2}(f(x) - f(x^i))$, for every $x \in \Omega_n$.
Then $2||\nabla_i(f)||_q^q = 2||\nabla_i(f)||_2^2 = 2\mathbb{E}[|\nabla_i(f)|] = I_i(f)$, for every $q > 0$ since $|\nabla_i(f)(x)| = 0$ or 1, for every $x \in \Omega_n$.

Finally it is straightforward to check that $\widehat{\nabla_i(f)}(S) = \widehat{f}(S)$ if $i \in S$ and zero otherwise. Therefore by Parseval's identity it holds:

$$I_i(f) = 2||\nabla_i(f)||_2^2 = 2\sum_{i\in S} \widehat{f}(S)^2.$$

Now given the last identity let's try to understand why Theorem 4.14 is true:

If we have a sequence of Boolean function f_n with $\mathbb{I}(f_n) \to 0$, then for every $i \in [m_n]$ the quantity $I_i(f_n) = 2||\nabla_i(f_n)||_2^2$ is small. Hence from remark 5.4,, $\widehat{\nabla_i(f_n)}(S)$ is supported on large sets S. But $\widehat{\nabla_i(f_n)}(S)$ equals to $\widehat{f_n}(S)$ if $i \in S$ and we have that this holds for every $i \in [m_n]$. Therefore given $S \neq \varnothing$, $\widehat{f_n}(S)$ is non-zero for many large S. But according to proposition 4.16 this is very close to what noise sensitivity is about.

Let's try to make our efforts precise and prove theorem 4.14 under polynomial decay.

Proof. (Theorem 4.14 under polynomial decay)

We will prove that $\sum_{S\subset[m_n],|S|\leq M\log(m_n)} \widehat{f_n}(S)^2 \to 0$ for some $M > 0$, which is enough from Proposition 4.6 .

Fix $M > 0, p \in (0,1)$ which we will be specified later.

Then we have :

$$\sum_{\varnothing\neq S\subset[m_n],|S|\leq M\log(m_n)} \widehat{f_n}(S)^2 \leq p^{-2M\log(m_n)} \sum_{S\subset[m_n]} p^{2|S|}|S|\widehat{f_n}(S)^2$$

$$= p^{-2M\log(m_n)} \sum_{S\subset[m_n]}\sum_{i=1}^{m_n} p^{2|S|}\widehat{\nabla_i}(f_n)(S)^2 = p^{-2M\log(m_n)}\sum_{i=1}^{m_n} ||T_p(\nabla_i(f_n))||_2^2$$

$$\leq p^{-2M\log(m_n)}\sum_{i=1}^{m_n} ||\nabla_i(f_n)||_{1+p^2}^2 = p^{-2M\log(m_n)}\sum_{i=1}^{m_n} ||\nabla_i(f_n)||_2^{\frac{4}{1+p^2}}$$

$$= 2^{\frac{-2}{1+p^2}} p^{-2M\log(m_n)}\sum_{i=1}^{m_n} I_i(f_n)^{\frac{2}{1+p^2}} \leq 2^{\frac{-2}{1+p^2}} p^{-2M\log(m_n)}(m_n^{\frac{p^2}{1+p^2}})(\sum_{i=1}^{n} I_i(f_n)^2)^{\frac{1}{1+p^2}}$$

,where the last inequality follows from Holder's inequality and the one before by B.G.B inequality.

Therefore, using our hypothesis we get

$$\sum_{\varnothing\neq S\subset[m_n],|S|\leq M\log(m_n)} \widehat{f_n}(S)^2 \leq 2^{\frac{-2}{1+p^2}} p^{-2M\log(m_n)}(m_n)^{\frac{p^2-\delta}{1+p^2}}$$

Now we choose $p, M > 0$ small enough so that the exponent of m_n is negative in the right hand side and we are done.

□

Now one may ask what happens when we assume that the norm $||\mathrm{Inf}(f_n)||_q$ for $q > 2$ decays polynomially to zero. Does this imply also that f_n is noise sensitive? The answer is yes if the polynomial decay is strong enough. Let's generalise a bit Theorem 4.14:

Proposition 5.5. *Let $f_n : \Omega_{m_n} \to \{-1,1\}$ be a sequences of Boolean functions with $\sum_{i=1}^{m_n} I_i(f_n)^q \leq m_n^{-\delta}$, for $\delta > \frac{q}{2} - 1, q \geq 2$. Then f_n is noise sensitive.*

Proof. For $M, p > 0$ that will be specified later following the same path as in the last proof and using Holder inequality we have:

$$\sum_{\varnothing \neq S \subset [m_n], |S| \leq M \log(m_n)} \widehat{f_n}(S)^2 \leq 2^{\frac{-2}{1+p^2}} p^{-2M \log(m_n)} \sum_{i=1}^{m_n} I_i(f_n)^{\frac{2}{1+p^2}}$$

$$\leq 2^{\frac{-2}{1+p^2}} p^{-2M \log(m_n)} (m_n)^{1 - \frac{2}{q(1+p^2)}} (\sum_{i=1}^{m_n} I_i(f_n)^q)^{\frac{2}{q(1+p^2)}}$$

Now using our hypothesis we get

$$\sum_{S \subset [m_n], |S| \leq M \log(m_n)} \widehat{f_n}(S)^2 \leq 2^{\frac{-2}{1+p^2}} p^{-2M \log(m_n)} (m_n)^{1 - \frac{2}{q(1+p^2)}} (m_n)^{-\delta \frac{2}{q(1+p^2)}}$$

Now since $\delta > \frac{q}{2} - 1$ there exists $p > 0$ such that $1 < (\delta + 1)\frac{2}{q(1+p^2)}$. Now choose $M > 0$ small enough so that the exponent of m_n in the right hand side remains negative. We are done. □

6 Quantitative relation between noise sensitivity and influences

In this section we will illustrate the proof of Keller and Kindler given in [KK13] for theorem 4.16 which of course imply Theorem 4.14 on it's whole generality. We will try to explain the reason behind most of the steps of the proof trying to understand why the have chosen this line of thinking.

The structure of this section will be the following one. In the first subsection we focus on some basic results we need, in the second we will work on the proof of a very important lemma and in the last one we will combine our tools to reach the desired result.

Finally for notation reasons:
We set $I_i'(f) = \frac{I_i(f)}{2}$, where $I_i(f)$ is the i-th influence of the Boolean function $f : \Omega_n \to \{-1,1\}$, and $\mathbb{W}(f) = \frac{\mathbb{I}(f)}{4}$. Therefore we have to prove that there exists a constant $C > 0$ such that for every $n \in \mathbb{N}$ and for every Boolean function $f : \Omega_n \to \{-1,1\}$ it holds:

$$\text{Cov}(f(x), f(x_\varepsilon)) \leq (6e + 1)(\mathbb{W}(f))^{C \cdot \varepsilon}$$

6.1 Preliminary results

A crucial component of the proof relies on a bound on the large deviation of low Fourier degree multivariate polynomials. Formally, for $d \geq 2$ we will bound the probability $\mathbb{P}(|f| \geq t), t \in \mathbb{R}_+$ for every function with Fourier degree at most d, i.e. with $\widehat{f}(S) = 0$ when $|S| \geq d+1$. It turns out the the general Bonami-Groos-Beckner inequality (Theorem 5.3) will help us out here.

Lemma 6.1. *If $r \geq 2$ then for every function $f : \Omega_n \to \mathbb{R}$ with Fourier degree at most d it holds:*

$$||f||_r \leq (r-1)^{\frac{d}{2}} ||f||_2.$$

Proof. We apply Theorem 5.3 for the function $g = \sum_{S\subset[n]} p^{-|S|}\widehat{f}(S)\chi_S$ with $t = r$, $s = 2$, $p = \sqrt{\frac{1}{r-1}}$ and we get

$$||f||_r^2 = ||T_p(g)||_r^2 \leq ||g||_2^2 = \sum_{S\subset[n]} p^{-2|S|}\widehat{f}(S)^2 \leq p^{-2d}||f||_2^2 = (r-1)^d||f||_2.$$

which is what we wanted to prove.

□

Now we are ready for our large deviation result:

Lemma 6.2. *([DFKO07])*

Let $f : \Omega_n \to \mathbb{R}$ be a function with Fourier degree at most d and $||f||_2 = 1$. Then for any $t \geq (2e)^{\frac{d}{2}}$,

$$\mathbb{P}(|f| \geq t) \leq \exp(-\frac{d}{2e}t^{\frac{2}{d}})$$

Proof. For any $r \geq 2$,

$$\mathbb{P}(|f| \geq t) = \mathbb{P}(|f|^r \geq t^r) \leq \frac{||f||_r^r}{t^r} \leq (r-1)^{\frac{rd}{2}}\frac{||f||_2^r}{t^r} < r^{\frac{rd}{2}}\frac{1}{t^r}$$

,where we have used Markov's inequality and Lemma 6.1 .
Now we minimize the right-hand side with respect to r and we get that the minimum is for $r = \frac{t^{\frac{2}{d}}}{e}$. One can check that it gives the desired bound. □

A quite technical lemma we will need is the following:

Lemma 6.3. *Let $d \geq 2$ be a positive integer, and let $t_0 > (4e)^{\frac{d-1}{2}}$.*
Then,

$$\int_{t=t_0}^{\infty} t^2 \exp(-\frac{d-1}{2e}\cdot t^{\frac{2}{d-1}})dt \leq 5\cdot e\cdot t_0^{3-\frac{2}{d-1}}\cdot\exp(-\frac{d-1}{2e}\cdot t_0^{\frac{2}{d-1}}).$$

We sketch the proof:

Proof. We first change variables to see what the statement is really about : We set $s = -\frac{d-1}{2e}\cdot t^{\frac{2}{d-1}}$ and the left hand side becomes:

$$(\frac{2e}{d-1})^{\frac{3(d-1)}{2}}\cdot\frac{d-1}{2}\cdot\int_{s=s_0}^{\infty} s^{\frac{3d-5}{2}}\cdot\exp(s)ds,$$

where $s_0 = \frac{d-1}{2e}\cdot t_0^{\frac{2}{d-1}}$.

Set $g(s) = s^{\frac{3d-5}{2}}\cdot\exp(-s)$, the integrand function of the last integral.

One can check that $s \geq s_0$ and $t_0 > (4e)^{\frac{d-1}{2}}$ implies $\frac{g(s+1)}{g(s)} \leq \exp(-\frac{1}{4})$.

Therefore our integral is bounded above from

$$(s_0)^{\frac{3d-5}{2}} \cdot \exp(s_0) \cdot \frac{1}{1-\exp(-\frac{1}{4})} \leq 5(s_0)^{\frac{3d-5}{2}} \cdot \exp(s_0).$$

Changing now from s_0 to t_0 we get the desired bound. □

Finally we need an immediate application of Fubini's theorem:

Lemma 6.4. *Let Ω be a probability space and let $f, g : \Omega \to \mathbb{R}$ where g is non-negative. Now for any $t \in \mathbb{R}$ let $L(t) = \{x \in \Omega | g(x) > t\}$. Then*

$$\mathbb{E}[(f(x)g(x)] = \int_{t=0}^{\infty} \mathbb{E}[f(x)1(L(t))(x)]\, dt,$$

where $1(A)$ is the indicator of the event A and the expectations is with respect to Ω.

Proof.

$$\mathbb{E}[f(x)g(x)] = \mathbb{E}\left[f(x)\int_{t=0}^{g(x)} 1dt\right] = \mathbb{E}\left[\int_{t=0}^{\infty} f(x)1(L(t)(x)\right] dt = \int_{t=0}^{\infty} \mathbb{E}[f(x)1(L(t))(x)]\, dt$$

□

6.2 The key lemma

Now we will state and prove the key lemma that will help us prove Theorem 4.16 .

Lemma 6.5. *For all $d \geq 2$ and for every function $f : \Omega_n \to \{-1, 1\}$ such that $\mathbb{W}(f) \leq \exp(-2(d-1))$ we have:*

$$\sum_{|S|=d} \widehat{f}(S)^2 \leq \frac{5 \cdot e}{d}(\frac{2e}{d-1})^{d-1}\mathbb{W}(f)(\log(\frac{d}{\mathbb{W}(f)}))^{d-1}.$$

One should be able to observe that the lemma 6.5. by itself implies Theorem 4.14.

Before we try to see why such a result is true let's discuss a bit the history behind.

Talagrand (in [Tal96]) was the first one that proved a result similar to lemma 6.5. but only for the case $d = 2$. He did not use it for noise sensitivity results but for estimations about the correlation of monotone events. Benjamini, Kalai and Schramm in [BKS99] proved Theorem 4.14 by generalising Talagrand result to a statement very similar to Lemma 6.5. The differences is that their result did not compute the constant and also it was referring only to monotone Boolean functions. Keller and Kindler finally in [KK13] stated and proved Lemma 6.5. as it appears above.

Let's start discussing the proof they gave:

Proof. (Lemma 6.5)

We need to bound the energy spectrum of f at level d from a function of the influences of f. A brute force attempt would be to consider the function $g_d = \sum_{|S|=d} \widehat{f}(S)\chi_S$ and try to bound $||g_d||_2^2$ which is exactly, by Parseval's identity, what we want to bound. This by itself does not seem a promising idea as there is no apparent way of bringing up the n different influences. What we should probably do it to find which summands are related to the j influence for every $j = 1, 2, .., n$. To do this we have to find out how the influences can arise naturally from bounding Fourier coefficients. An important observation is that for every $j = 1, 2, .., n$:

$$\mathbb{E}_{x\in\{-1,1\}^{[n]-j}}[|\mathbb{E}_{x_j\in\{-1,1\}}[\chi_j(x,x_j)\cdot f(x,x_j)]|] = \frac{I_j(f)}{2}.$$

This can be verified by a straightforward computation:

$$\mathbb{E}_{x\in\{-1,1\}^{[n]-\{j\}}}[|\mathbb{E}_{x_j\in\{-1,1\}}[\chi_j(x,x_j)f(x,x_j)]|] = \frac{1}{2}\mathbb{E}_{x\in\{-1,1\}^{[n]-\{j\}}}[|f(x,1)-f(x,-1)|] = \frac{I_j(f)}{2} = I'_j(f).$$

Hence, it seems that the influence will appear by bounding the Fourier coefficients of the form

$$\widehat{f}(T\cup\{j\}) = \mathbb{E}[f\cdot\chi_j\cdot\chi_T] = \mathbb{E}_{x\in\{-1,1\}^{[n]-j}}[\chi_T\cdot\mathbb{E}_{x_j\in\{-1,1\}}[\chi_j(x,x_j)\cdot f(x,x_j)]].$$

This suggests that $\sum_{T\subset[n]-j,|T|=d-1}\widehat{f}(T\cup\{j\})^2$ could be bounded by a function of $I_j(f)$, or equivalently that for $J\subset[n]$ the $\sum_{j\in J}\sum_{T\subset[n]-J,|T|=d-1}\widehat{f}(T\cup\{j\})^2$ could be bounded by a function of the influences $I_j(f), j\in J$.

And yes actually this is the case:

Claim 6.6. *For every partition $\{I,J\}$ of $[n]$, it is true that:*

$$\sum_{T\subset I,|T|=d-1}\sum_{j\in J}\widehat{f}(T\cup\{j\})^2 \le 5\cdot(\frac{2e}{d-1})^{d-1}\cdot(\sum_{j\in J}(I'_j(f)^2)\cdot(\log(\frac{1}{(\sum_{j\in J}I'_j(f)^2)}))^{d-1}$$

Now the question changes to given such inequalities whether we can find appropriate weights to multiply each of them and sum them up to derive a good bound for $\sum_{|S|=d}\widehat{f}(S)^2$? And the answer is again affirmative:

Claim 6.7. *If for every partition $\{I,J\}$ of $[n]$,*

$$\sum_{T\subset I,|T|=d-1}\sum_{j\in J}\widehat{f}(T\cup\{j\})^2 \le 5\cdot(\frac{2e}{d-1})^{d-1}\cdot(\sum_{j\in J}I'^2_j(f))\cdot(\log(\frac{1}{(\sum_{j\in J}I'^2_j(f))}))^{d-1}$$

then Lemma 6.5. holds.

Therefore we now have to prove Claims 6.6 and 6.7. We start with the proof of Claim 6.7. to clear the picture:

Proof. (of Claim 6.7) Keller and Kindler proved this claim with an elegant but mysterious argument. We will try to present their proof in some small steps that will make more understandable why Keller and Kindler argument have chosen such an argument.

Suppose that for every partition the corresponding inequality is true.

As we said above we want to adjust appropriate weights to every such inequality to get a bound for $\sum_{|S|=d}\widehat{f}(S)^2$.

Let's try to see what weights are appropriate. Assume that we multiply the inequality for the partition $\{[n]-J,J\}$ by $a_J\ge 0$ and without loss of generality we may assume that the weights are normalised, i.e. we may assume that $\sum_{J\subset[n]} a_J = 1$. Before summing them up we need some notation. For every $S \subset [n]$ with d elements set

$$L_S = \{J \subset [n] \Big| \text{ there exist } T \subset [n] - J \text{ and } j \in J \text{ such that } S = T \cup \{j\}\}$$

or equivalently

$$L_S = \{J \subset [n] \Big| \ |J \cap S| = 1 \ \}.$$

Also define the function g given by $g(x) = x \log{(\frac{1}{x})}^{d-1}$. Now summing up the weighted inequalities over all $J \subset [n]$ we get:

$$\sum_{S\subset[n],|S|=d} (\sum_{J\in L_S} a_J)\cdot \widehat{f}(S)^2 = \sum_{J\subset[n]} a_J \cdot \left(\sum_{T\subset I,|T|=d-1} \sum_{j\in J} \widehat{f}(T\cup\{j\})^2 \right) \leq 5\cdot(\frac{2e}{d-1})^{d-1} \sum_{J\subset[n]} a_J g(\sum_{j\in J} I'_j(f)^2).$$

Now one can check that g is concave in the interval $[0, \exp{(-(d-1))}]$. Therefore since for every $J \subset [n]$:

$$\sum_{j\in J} I'_j(f)^2 \leq \sum_{j=1}^{n} I'_j(f)^2 \leq \exp{(-(d-1))},$$

we have:

$$\sum_{J\subset[n]} a_J \cdot g(\sum_{j\in J} I'_j(f)^2) \leq g\left(\sum_{J\subset[n],j\in J} a_J \cdot I'_j(f)^2 \right) = g(\sum_{j\in[n]} (\sum_{j\in J\subset[n]} a_J) \cdot I'_j(f)^2).$$

Thus we have proved that for any choice of normalised weights:

$$\sum_{S\subset[n],|S|=d} (\sum_{J\in L_S} a_J) \cdot \widehat{f}(S)^2 \leq 5 \cdot (\frac{2e}{d-1})^{d-1} g(\sum_{j\in[n]} (\sum_{J\subset[n] \text{ s.t } j\in J} a_J) \cdot I'_j(f)^2).$$

Since we want the quantity $\mathbb{W}(f) = \sum_{j=1}^{n} I'_j(f)^2$ to appear on the right hand side we need for every $j \in [n]$ the quantities $\sum_{J\subset[n] \text{ s.t } j\in J} a_J$ to be equal for $j = 1, 2, 3, .., n$.

Therefore if we interpret the weights $\{a_J, J \subset [n]\}$ as a probability measure on the subsets of $[n]$, the quantity $\sum_{J\subset[n] \text{ s.t } j\in J} a_J$ is the probability that j is a member of the random set. So we ask that this probability equals $\alpha > 0$ for every $j \in [n]$.

Now for the left hand side what we care about to get a bound is that the quantities $\sum_{J\in L_S} a_J$ are uniformly bounded from below. But again using our probabilistic interpretation of the weights, the quantity $\sum_{J\in L_S} a_J$ is the probability that our random set has exactly a single point as intersection with S.

Therefore we want these probabilities to be uniformly bounded from below for every set $S \subset [n]$ with d elements. If we make now the extra assumption that we choose the elements of our random set independently of each other this probability equals exactly to $d \cdot \alpha(1-\alpha)^{d-1}$. Now choosing $\alpha = \frac{1}{d}$ seems a correct choice as then

$$d \cdot \alpha(1-\alpha)^{d-1} = (1 - \frac{1}{d})^{d-1} > e^{-1}.$$

Therefore a way we can choose our weights can be based on the construction of a random set J which independently contains $i \in [n]$ with probability d^{-1}. That means that for every $J \subset [n]$, we can set

$$a_J = d^{-|J|}(1 - \frac{1}{d})^{n-|J|}$$

and this is exactly the choice for weights Keller and Kindler made.

Thus, based on the discussion above for this choice of weights we have:

$$e^{-1} \sum_{S \subset [n], |S|=d} \widehat{f}(S)^2 \leq \cdot \sum_{S \subset [n], |S|=d} (\sum_{J \in L_S} a_J) \cdot \widehat{f}(S)^2$$

$$\leq 5 \cdot (\frac{2e}{d-1})^{d-1} g(\sum_{j \in [n]} (\sum_{J \subset [n] \text{ s.t } j \in J} a_J) \cdot I'_j(f)^2) = 5 \cdot (\frac{2e}{d-1})^{d-1} g(\frac{\mathbb{W}(f)}{d}),$$

which means:

$$\sum_{S \subset [n], |S|=d} \widehat{f}(S)^2 \leq \frac{5 \cdot e}{d} (\frac{2e}{d-1})^{d-1} \mathbb{W}(f) (\log(\frac{d}{\mathbb{W}(f)}))^{d-1}.$$

But this is exactly the statement of Lemma 6.5. We are done.

□

Now we have to prove the inequality at Claim 6.6 which finds an upper bound for

$$\sum_{T \subset I, |T|=d-1} \sum_{j \in J} \widehat{f}(T \cup \{j\})^2.$$

Proof. (of Claim 6.6)
We have to bound from above

$$\sum_{T \subset I, |T|=d-1} \sum_{j \in J} \widehat{f}(T \cup \{j\})^2.$$

Fix $j \in J$. We know,

$$\mathbb{E}_{x \in \{-1,1\}^{[n]-j}}[|\mathbb{E}_{x_j \in \{-1,1\}}[\chi_j(x, x_j) \cdot f(x, x_j)]|] = \frac{I_j(f)}{2} = I'_j(f).$$

and

$$\widehat{f}(T \cup \{j\}) = \mathbb{E}[f \cdot \chi_j \cdot \chi_T] = \mathbb{E}_{x \in \{-1,1\}^{[n]-j}}[\chi_T(x) \cdot \mathbb{E}_{x_j \in \{-1,1\}}[\chi_j(x, x_j) \cdot f(x, x_j)]] \tag{6.1}$$

Now an immediate bound based on the above relations is:

$$\widehat{f}(T \cup \{j\})^2 \leq \Big(\mathbb{E}_{x \in \{-1,1\}^{[n]-j}}[|\chi_T(x)| \cdot |\mathbb{E}_{x_j \in \{-1,1\}}[\chi_j(x, x_j) \cdot f(x, x_j)]|]\Big) = I'_j(f)^2.$$

which gives that

$$\sum_{j \in J} \sum_{T \subset I, |T|=d-1} \widehat{f}(T \cup \{j\})^2$$

is bounded above by

$$\binom{|I|}{d-1} \cdot \sum_{j\in J} I_j(f)^2$$

which is not the type of bound we need as it depends on $|I|$.

Therefore we should proceed in a different way.

Firstly since $T \subset I$ we can change a bit (6.1) to:

$$\widehat{f}(T \cup \{j\}) = \mathbb{E}_{x\in\{-1,1\}^I}[\chi_T(x) \cdot \mathbb{E}_{y\in\{-1,1\}^J}[\chi_j(x,y) \cdot f(x,y)]]. \tag{6.2}$$

so that we will have a uniform treat over all $j \in J$. This minor change will end up being very useful in what follows.

Now let $f'_j = \sum_{T\subset I,|T|=d-1} \widehat{f}(T \cup \{j\})\chi_T$. Then from relation (6.2) we get:

$$\sum_{T\subset I,|T|=d-1} \widehat{f}(T \cup \{j\})^2 = \mathbb{E}\big[f \cdot \chi_j \cdot f'_j\big] = \mathbb{E}_{x\in\{-1,1\}^I}[f'_j(x) \cdot \mathbb{E}_{y\in\{-1,1\}^J}[\chi_j(x,y) \cdot f(x,y)]].$$

Now since $||f'_j||_2^2 = \sum_{T\subset I,|T|=d-1} \widehat{f}(T \cup \{j\})^2$ setting $f_j = \frac{f'_j}{||f_j||_2}$ we derive:

$$\sum_{T\subset I,|T|=d-1} \widehat{f}(T \cup \{j\})^2 = (\mathbb{E}[f_j \cdot \chi_j \cdot f])^2.$$

This is much progress. If for some reason f_j was bounded by $M > 0$ then triangle inequality yields:

$$\sum_{T\subset I,|T|=d-1} \widehat{f}(T \cup \{j\})^2 \leq M^2 \mathbb{E}_{x\in\{-1,1\}^I}[|\mathbb{E}_{y\in\{-1,1\}^J}[\chi_j(x,y) f(x,y)]|]$$

$$\leq M^2 \left(\mathbb{E}_{x\in\{-1,1\}^{[n]-j}}[|\mathbb{E}_{x_j\in\{-1,1\}}[\chi_j(x,x_j) \cdot f(x,x_j)]|]\right)^2 = M^2 \frac{I_j^2(f)}{4} = M^2 I'_j(f)^2.$$

Unfortunately, there is not an obvious reason for f_j to be bounded but since it has fixed Fourier degree $d \geq 2$ and it's 2-norm is equal to 1 based on results from subsection 6.1. we know that the quantities $\mathbb{P}(|f_j| > t)$ decay exponentially with t. Therefore what we should do is on the event $\{|f_j| \leq t\}$ bound the expectation by the square of the influence multiplied by the factor t^2 as if it were bounded on the first place and on the complement of the event use our estimate for $\mathbb{P}(|f_j| > t)$ we proved in Lemma 6.2.

Let's make this work. First we observe that from triangle inequality and the fact that f_j does not depend on J by construction:

$$\sum_{T\subset I,|T|=d-1} \widehat{f}(T \cup \{j\})^2 = (\mathbb{E}[f_j \cdot \chi_j \cdot f])^2 = (\mathbb{E}_{x\in\{-1,1\}^I} f_j(x) \cdot [\mathbb{E}_{y\in\{-1,1\}^J} \chi_j(x,y) \cdot f(x,y)]])^2$$

$$\leq (\mathbb{E}_{x\in\{-1,1\}^I}[|f_j(x)| \cdot |\mathbb{E}_{y\in\{-1,1\}^J}[\chi_j(x,y) f(x,y)]|])^2.$$

Now using Lemma 6.4. this can be rewritten as:

$$= (\int_{t=0}^{\infty} (\mathbb{E}_{x\in\{-1,1\}^I}[1(|f_j(x)| > t)|\mathbb{E}_{y\in\{-1,1\}^J}[\chi_j(x,y) f(x,y)]|]dt)^2$$

Now following the line of thinking we said we write the integral $\int_0^\infty$ as a sum of $\int_0^{t_0}$ and $\int_{t_0}^\infty$ and we will use the bound with the influence we mentioned above for the first and the tail estimate for the second. To do this we use the basic $(a+b)^2 \leq 2(a^2+b^2)$ and we get that for every $t_0 > 0$:

$$\sum_{T \subset I, |T|=d-1} \widehat{f}(T \cup \{j\})^2 \leq 2 \cdot (A(t_0, j) + B(t_0, j)) \tag{6.3}$$

where

$$A(t_0, j) = \int_{t=0}^{t_0} (\mathbb{E}_{x \in \{-1,1\}^I}[[1(|f_j(x)| > t)|\mathbb{E}_{y \in \{-1,1\}^J}[\chi_j(x,y)f(x,y)]|]dt)^2$$

and,

$$B(t_0, j) = \int_{t=t_0}^{\infty} (\mathbb{E}_{x \in \{-1,1\}^I}[[1(|f_j(x)| > t)|\mathbb{E}_{y \in \{-1,1\}^J}[\chi_j(x,y)f(x,y)]|]dt)^2$$

Now let's bound $A(t_0, j), B(t_0, j)$ separately.

To bound $A(t_0, j)$ we firstly bound the indicator function by 1 and based on triangle inequality we get as expected that:

$$A(t_0, j) \leq \int_{t=0}^{t_0} (\mathbb{E}_{x \in \{-1,1\}^I}[|\mathbb{E}_{y \in \{-1,1\}^J}[\chi_j(x,y)f(x,y)]|]dt)^2$$

$$\leq \int_{t=0}^{t_0} (\mathbb{E}_{x \in \{-1,1\}^{[n]-\{j\}}}[|\mathbb{E}_{x_j \in \{-1,1\}}[\chi_j(x,x_j)f(x,x_j)]|]dt)^2 = t_0^2 \cdot I_j'(f)^2$$

Concluding,

$$A(t_0, j) \leq t_0^2 \cdot I_j'(f)^2, \quad \text{for every } j \in J \tag{6.4}$$

Now for $B(t_0, j)$:

We may assume $t_0 > (4e)^{\frac{d-1}{2}}$ for the moment. Then starting with Cauchy Scwharz we get:

$$B(t_0, j) = (\int_{t=t_0}^{\infty} \frac{1}{t} \cdot t \cdot \mathbb{E}_{x \in \{-1,1\}^I}[1(|f_j(x)| > t) \cdot |\mathbb{E}_{y \in \{-1,1\}^J}[\chi_j(x,y)f(x,y)]|]dt)^2$$

$$\leq \int_{t=t_0}^{\infty} \frac{1}{t^2} dt \cdot \int_{t=t_0}^{\infty} t^2 (\mathbb{E}_{x \in \{-1,1\}^I}[[1(|f_j(x)| > t) \cdot |\mathbb{E}_{y \in \{-1,1\}^J}[\chi_j(x,y)f(x,y)]|])^2 dt \tag{6.5}$$

But since we want to use our estimate in Lemma 6.2. we use again Cauchy-Scwharz on $\{-1,1\}^I$ at the expectation inside the integral and hence the quantity in (6.4) is bounded above by:

$$\frac{1}{t_0} \int_{t=t_0}^{\infty} t^2 \cdot \mathbb{P}_{x \in \{-1,1\}^I}(|f_j(x)| > t) \cdot \mathbb{E}_{x \in \{-1,1\}^I}[(\mathbb{E}_{y \in \{-1,1\}^J}[\chi_j(x,y)f(x,y)])^2]dt,$$

and of course now using Lemma 6.2 we derive:

$$B(t_0,j) \leq \frac{1}{t_0} \cdot \int_{t=t_0}^{\infty} t^2 \exp\left(-\frac{(d-1)}{2e} t^{\frac{2}{d-1}}\right) \cdot \mathbb{E}_{x\in\{-1,1\}^I}[(\mathbb{E}_{y\in\{-1,1\}^J}[\chi_j(x,y)f(x,y)])^2]dt.$$

Now we need to bound $\int_{t=t_0}^{\infty} t^2 \exp\left(-\frac{(d-1)}{2e} t^{\frac{2}{d-1}}\right)dt$ and this is exactly what Lemma 6.3 was about. Hence, using Lemma 6.3:

$$B(t_0,j) \leq 5 \cdot e \cdot t_0^{2-\frac{2}{d-1}} \exp\left(-\frac{d-1}{2e} \cdot t_0^{\frac{2}{d-1}}\right) \mathbb{E}_{x\in\{-1,1\}^I}[(\mathbb{E}_{y\in\{-1,1\}^J}[\chi_j(x,y)f(x,y)])^2] \tag{6.6}$$

Therefore from (6.2),(6.3) and (6.5) under the condition that $t_0 > (4e)^{\frac{d-1}{2}}$ we proved that for every $j \in J$:

$$\sum_{T\subset I,|T|=d-1} \widehat{f}(T\cup\{j\})^2 \leq 2\cdot t_0^2 \cdot I_j'(f)^2 + 10\cdot e\cdot t_0^{2-\frac{2}{d-1}} \exp\left(-\frac{d-1}{2e} \cdot t_0^{\frac{2}{d-1}}\right) \mathbb{E}_{x\in\{-1,1\}^I}[(\mathbb{E}_{y\in\{-1,1\}^J}[\chi_j(x,y)f(x,y)])^2].$$

Summing now over all $j \in J$ it follows that :

$$\sum_{j\in J} \sum_{T\subset I,|T|=d-1} \widehat{f}(T\cup\{j\})^2$$

is controlled above by:

$$\leq 2\cdot t_0^2 \cdot \left(\sum_{j\in J} I_j'(f)^2\right) + 10\cdot e\cdot t_0^{2-\frac{2}{d-1}} \exp\left(-\frac{d-1}{2e} \cdot t_0^{\frac{2}{d-1}}\right) \mathbb{E}_{x\in\{-1,1\}^I}\left[\sum_{j\in J} (\mathbb{E}_{y\in\{-1,1\}^J}[\chi_j(x,y)\cdot f(x,y)])^2\right]$$

The first term is what we wanted. Now we have to focus on the second one.

The important observation is that

$$\sum_{j\in J} (\mathbb{E}_{y\in\{-1,1\}^J}[\chi_j(x,y)\cdot f(x,x_j)])^2 \leq 1.$$

To see this, set for every $x \in \{-1,1\}^I$, $f_x : \{-1,1\}^J \to \{-1,1\}$, the Boolean function given by $f_x(y) = f(x,y)$. Then :

$$\widehat{f_x}(\{j\})^2 = (\mathbb{E}_{y\in\{-1,1\}^J}[\chi_j(x,y)f(x,y)])^2,$$

and using Parseval identity it follows:

$$\sum_{j\in J} (\mathbb{E}_{y\in\{-1,1\}^J}[\chi_j(x,y)f(x,y)])^2 = \sum_{j\in J} \widehat{f_x}(\{j\})^2 \leq ||f_x||_2^2 = 1.$$

It is worth mentioning now that if someone works from the beginning based on relation (6.1) and not relation (6.2), i.e. for every $j \in J$ with the partition $([n]-\{j\},\{j\})$ for every $j \in J$ and not with the same partition (I,J) for every $j \in J$ he will end up trying to bound

$$\sum_{j\in J} \mathbb{E}_{x\in\{-1,1\}^{[n]-\{j\}}}\left[\Big(\mathbb{E}_{x_j\in\{-1,1\}}[\chi_j(x,x_j)f(x,x_j)]\Big)^2\right].$$

But for example if f is the parity function then this is equal to $|J|$ and therefore there was no hope to bound this uniformly from a constant as we did in the other case.

Now back to our proof we have:

$$\sum_{j\in J}\sum_{T\subset I,|T|=d-1}\widehat{f}(T\cup\{j\})^2 \leq 2\cdot t_0^2\cdot(\sum_{j\in J}I'_j(f)^2)+10\cdot e\cdot t_0^{2-\frac{2}{d-1}}\exp\big(-\frac{d-1}{2e}\cdot t_0^{\frac{2}{d-1}}\big).$$

To improve our bound we choose t_0 such that

$$\exp\big(-\frac{d-1}{2e}\cdot t_0^{\frac{2}{d-1}}\big)=\sum_{j\in J}I'_j(f)^2,$$

which means

$$t_0^2=(\frac{2e}{d-1})^{d-1}\cdot(\log(\frac{1}{\sum_{j\in J}I'_j(f)^2}))^{d-1}.$$

One can see that $\mathbb{W}(f)\leq\exp(-(\frac{d-1}{2}))$ implies $t_0=(\frac{2e}{d-1})^{\frac{d-1}{2}}\cdot(\log(\frac{1}{\sum_{j\in J}I'_j(f)^2}))^{\frac{d-1}{2}}>(4e)^{\frac{d-1}{2}}$ for every possible $J\subset[n]$ as we needed.

Moreover substituting t_0 in our bound we get:

$$\sum_{j\in J}\sum_{T\subset I,|T|=d-1}\widehat{f}(T\cup\{j\})^2\leq 5\cdot(\frac{2e}{d-1})^{d-1}(\sum_{j\in J}I'_j(f)^2)\cdot\log\big(\frac{1}{\sum_{j\in J}I'_j(f)^2}\big),$$

as we wanted. □

Claims 6.6 and 6.7 complete the proof of Lemma 6.5. The proof is complete. □

6.3 Proof of the theorem

Now we are ready to combine our results and prove Theorem 4.16.

What we need to prove is that if $f:\Omega_n\to\{-1,1\}$ is a Boolean function then

$$\mathrm{Cov}(f(x),f(x_\varepsilon))\leq(6e+1)\mathbb{W}(f)^{C\cdot\varepsilon}\text{ for every }\varepsilon\in(0,1).$$

Kellen and Kindler compute $C>0$ exactly but we will not be bothered with this during our proof.

We know that

$$\mathrm{Cov}(f(x),f(x_\varepsilon))=\sum_{S\neq\varnothing}\widehat{f}(S)^2(1-\varepsilon)^{|S|}.$$

The strategy now will be the following:

We will bound the terms coming from big sets by a big power of $(1-\varepsilon)$ and for the small sets S we will use Lemma 6.4 bound. Of course we sould be careful as Lemma 6.4 is for the the sets with $d\geq 2$ elements so we need to find another bound for $d=1$.

Let's begin the proof. Take a positive $L>0$.

Let's start now bounding the quantity $\mathrm{Cov}(f(x),f(x_\varepsilon))$:

As we said for the high degree terms we act as following:

$$\sum_{|S|>L}(1-\varepsilon)^{|S|}\widehat{f}(S)^2\leq(1-\varepsilon)^{L+1}(\sum_{|S|>L}\widehat{f}(S)^2)\leq(1-\varepsilon)^{L+1}$$

,where we have used Parseval's identity.

For the low-degree terms now:

As we said for the sets with the singletons we need a new estimate: Observe that for every $i \in [n]$,

$$|\widehat{f}(\{i\})| = |\mathbb{E}[\chi_i f]\,| \le \mathbb{E}_{x\in\{-1,1\}^{[n]-\{j\}}}|[\mathbb{E}_{x_j\in\{-1,1\}}[\chi_i(x,x_j)f(x,x_j)]|] = \frac{1}{2}I_i(f) = I_i'(f)$$

Therefore

$$\sum_i \widehat{f}(\{i\})^2 \le \mathbb{W}(f).$$

Now since for the bigger level sets we want to apply Lemma 6.5 we want for $2 \le d \le L$ to have $\mathbb{W}(f) \le \exp(-2(d-1))$ or equivalently $\mathbb{W}(f) \le \exp(-2(L-1))$ or $2(L-1) \le \log(\frac{1}{\mathbb{W}(f)})$. Therefore we choose $L = C\log(\frac{1}{\mathbb{W}(f)})$ with $C < \frac{1}{2}$ which will fulfil this demand.

For this choice of L we can, by using Lemma 6.5 and the above estimate for the singletons, to get the following bound:

$$\sum_{0<|S|\le L}(1-\varepsilon)^{|S|}\widehat{f}(S)^2 \le \sum_{d=1}^{L}\sum_{|S|=d}\widehat{f}(S)^2 \le \mathbb{W}(f) + \sum_{d=2}^{L}\frac{5\cdot e}{d}(\frac{2e}{d-1})^{d-1}\mathbb{W}(f)(\log(\frac{d}{\mathbb{W}(f)}))^{d-1}.$$

We observe that if $d < L$ then the ratio between the $d+1$ term and the d term in the right hand side of the last inequality is bounded below from 2. This is a calculation we avoid but it can be checked straightforward based on the fact that $L = C\log(\frac{1}{\mathbb{W}(f))}) < \frac{1}{2}\cdot\log(\frac{1}{\mathbb{W}(f))})$.

Therefore this ratio result implies that the sum of the terms is at most twice the last term which means:

$$\sum_{0<|S|\le L}(1-\varepsilon)^{|S|}\widehat{f}(S)^2 \le \mathbb{W}(f) + \frac{10\cdot e}{L}(\frac{2e}{L-1})^{L-1}\mathbb{W}(f)(\log(\frac{L}{\mathbb{W}(f)}))^{L-1} = K(L)$$

Hence we have proven that for any $L, C > 0$ such that $C < \frac{1}{2}$ and $L = C\log(\frac{1}{\mathbb{W}(f))})$,

$$\mathrm{Cov}(f(x), f(x_\varepsilon)) \le (1-\varepsilon)^{L+1} + K(L) \tag{6.7}$$

Now what we have to do is to optimize with respect to $L > 0$ or equivalently $C > 0$ to get the best bound.

Claim 6.8. *For $C > 0$ sufficiently small:*

$$(1-\varepsilon)^{L+1} \le \mathbb{W}(f)^{C\cdot\varepsilon}$$

and

$$K(L) \le 6e\cdot\mathbb{W}(f)^{C\cdot\varepsilon}.$$

Moreover proving the above two inequalities and using relation (6.7) the proof of our theorem is complete.

Proof. The fact that these two inequalities suffices the get the desired result is immediate.

Let's try to prove them.

The first comes from the basic $1 + x \leq e^x$:

$$(1-\varepsilon)^{L+1} < (1-\varepsilon)^L \leq \exp(-L\cdot\varepsilon) = \exp\left(-\varepsilon \cdot C \cdot \log\left(\frac{1}{\mathbb{W}(f)}\right)\right) = \mathbb{W}(f)^{C\cdot\varepsilon},$$

for every $C > 0$.

For the second inequality now.

We can write $K(L)$ as,

$$K(L) = \mathbb{W}(f) + \frac{10\cdot e}{L}\cdot\left(\frac{2e}{C}\right)^{L-1}\cdot\left(\frac{C}{L}\right)^{L-1}\left(\frac{L}{L-1}\right)^{L-1}\cdot\mathbb{W}(f)\cdot\left(\log\left(\frac{L}{\mathbb{W}(f))}\right)\right)^{L-1}$$

then since $(\frac{L}{L-1})^{L-1} < e$ and $L = C\log\left(\frac{1}{\log\left(\frac{1}{W}\right)}\right)$ we get:

$$\leq \mathbb{W}(f) + \frac{10e^2}{L}\cdot\left(\frac{1}{\log\frac{1}{\mathbb{W}(f)}}\right)^{L-1}\cdot\left(\log\left(\frac{L}{\mathbb{W}(f)}\right)\right)^{L-1}\cdot\left[\mathbb{W}(f)\cdot\left(\frac{2e}{C}\right)^{L-1}\right]$$

which can be written as:

$$= \mathbb{W}(f) + \frac{10e^2}{L}\cdot\left(1 + \frac{\log(L)}{\log\frac{1}{\mathbb{W}(f)})}\right)^{L-1}\cdot\left[\mathbb{W}(f)\cdot\left(\frac{2\varepsilon}{C}\right)^{L-1}\right]$$

and since $C < \frac{1}{2} < 1$ and therefore $L < \log(\frac{1}{\mathbb{W}(f)})$ this is at most:

$$\leq \mathbb{W}(f) + \frac{10e^2}{L}\cdot\left(1 + \frac{\log(L)}{\log\left(\frac{1}{\mathbb{W}(f)}\right)}\right)^{\log\left(\frac{1}{\mathbb{W}(f)}\right)}\cdot\left[\mathbb{W}(f)\cdot\left(\frac{2\varepsilon}{C}\right)^{L-1}\right].$$

From this a call again to the basic $1 + x \leq e^x$ shows that our last quantity can be bounded by:

$$\leq \mathbb{W}(f) + 10e^2\cdot\left[\mathbb{W}(f)\cdot\left(\frac{2\varepsilon}{C}\right)^{L-1}\right].$$

Concluding so far we have:

$$K(L) \leq \mathbb{W}(f) + \cdot\left[\mathbb{W}(f)\cdot 10e^2\cdot\left(\frac{2\varepsilon}{C}\right)^{L-1}\right].$$

We remind here again that we assume $\mathbb{W}(f) < 1$ as we care only for small values of it.

What we just have to understand is the factor:

$$\left(\frac{2\varepsilon}{C}\right)^{L-1}.$$

But,

$$\left(\frac{2\varepsilon}{C}\right)^{L-1} = \frac{C}{2\varepsilon}\cdot\left(\frac{2\varepsilon}{C}\right)^{C\log\left(\frac{1}{\mathbb{W}(f)}\right)} = \frac{C}{2e}\cdot\mathbb{W}(f))^{-C\log\left(\frac{2e}{C}\right)}.$$

Therefore

$$K(L) \leq \mathbb{W}(f) + 5e \cdot C \cdot \mathbb{W}(f)^{1 - C\log\left(\frac{2e}{C}\right)}.$$

Now we just have to choose $C < \frac{1}{2}$ small enough such that $1 > 1 - C\log\left(\frac{2e}{C}\right) > C > C \cdot \varepsilon$ and then since $\mathbb{W}(f) < 1$,

$$K(L) \leq \mathbb{W}(f)^{C\cdot\varepsilon} + 5e \cdot \mathbb{W}(f))^{C\cdot\varepsilon} \leq 6e\mathbb{W}(f)^{C\cdot\varepsilon}.$$

The proof of the claim and therefore of the theorem is complete.

□

Remark 6.9. *Continuing the last line in our proof Kellen and Kinder proved that for every* $\varepsilon \in (0,1)$*, we can take as constant* $C(\varepsilon)$ *the quantity:*

$$\frac{1}{\varepsilon + \log(2e) + 3\log\log\{(2e)\}}$$

Moreover they proved that this choice of $C = C(\varepsilon)$ *is tight in the sense that there is a sequence of Boolean function that approximately turn the inequality to an equality up to a constant factor.*

To be more precise they proved the following result:

Suppose that for every $n \in \mathbb{N}$ *we have a society with* n *people and that the people are divided into* r *tribes. Suppose that they vote between two tasks "*-1*" and "*1*". Define the tribes sequence of functions ,* t_n*, as following. For every* n*,* $t_n = 1$ *if and only if there is at least one tribe that all of it's members voted for* 1*, and* $t_n = -1$ *otherwise.*

Then they proved that Theorem 4.16 becomes tight in the sense that

$$\log\left(\mathrm{Cov}(t_n(x), t_n(x_\varepsilon)\right) \sim C(\varepsilon) \cdot \varepsilon \log\left(\mathbb{W}(t_n)\right),$$

for n *sufficiently big and* ε *sufficiently small.*

7 Noise sensitivity of crossing events and beyond

In this section we start by proving that the percolation crossing events for site percolation for $\mathbb{T}$ and for bond percolation on $\mathbb{Z}^2$ satisfy the B.K.S. criterion (Theorem 4.14) and therefore are noise sensitive. After that we will discuss another approach toward this problem using the notion of randomized algorithms and the connection it has with Boolean functions. The last line of thinking will lead us naturally to study the Fourier spectrum of the percolation crossing events and we close the section by mentioning some very important results towards this last direction. The path we will follow is based on material from [GS12].

We will assume that everything that it is mentioned at the introductory section 2 about percolation theory is known.

Moreover we remind here, to avoid possible confusion, that we defined site percolation only for the triangular lattice $\mathbb{T}$ and bond percolation only for the square lattice $\mathbb{Z}^2$. Therefore from now on when we say **site percolation** we mean on the triangular lattice $\mathbb{T}$ and when we say **bond percolation** we refer to the square lattice $\mathbb{Z}^2$.

7.1 Crossing events are noise sensitive (via Arm-events theory)

Assume firstly that we work with the site percolation model on the triangular lattice $\mathbb{T}$.

Fix $a, b > 0$ and for every $n \in \mathbb{N}$ set $R_n = ([0, a \cdot n] \times [0, b \cdot n]) \cap \mathbb{T}$ and $m_n = |R_n|$.

Using B.K.S. theorem we want to prove that if for every $n \in \mathbb{N}$, $f_n : \Omega_{m_n} \to \{-1, 1\}$ is the Boolean function which is defined on all the possible configurations of R_n and is 1 if and only if there is a left to the right crossing in R_n with open vertices, then:

$$\mathbb{I}(f_n) = \sum_{x \in R_n} I_x(f_n)^2 \to 0.$$

For convenience from now on when we will say left-right crossing we will mean of course left-right right crossing with open vertices.

Let's study this sum:

For $x \in R_n$, we know that $I_x(f_n) = 2 \cdot \mathbb{P}(x \text{ is pivotal for } f_n)$.

Therefore to continue we need to gain some intuition towards what being pivotal really means for a site inside the rectangle.

An important observation is that for x to be pivotal there must exist two paths that pass through x with the following properties: One of them will consist only of open vertices (except possibly x) and will connect the left side of R_n with the right one and the other path will consist only of closed vertices (except possibly x) and will connect the upper side of R_n with the lower. Therefore if x becomes open a left-right crossing will rise up and if x becomes closed all the possible left-right crossings will be blocked from the top-bottom closed path.

The proof that this situation is necessary comes from the observation that the connected component with respect to the open vertices that contains the left side of the rectangle actually decides whether there exist a left-right crossing or not. A left-right crossing exists if and only if this component touches the right boundary of the rectangle. This has as a corollary that if a left-right crossing does not exist then the "right boundary" of the connected component will be an upper-lower crossing of the rectangle with only closed vertices.

The above observation yields that a pivotal $x \in R_n$ of distance $d > 0$ from the boundary satisfies the 4-arm event with distance d, which by section 2 is an event with probability of order $d^{-\frac{5}{4}+o(1)}$.

Therefore if we focus on the set of vertices in R_n with distance for the boundary at least $c \cdot n$,$c > 0$, call this set C_n, we can estimate from above the probability that a vertex in C_n is pivotal. Indeed, for every $x \in C_n$, then $I_x(f_n) \leq 2 \cdot a_4(c \cdot n) \leq 2 \cdot (c \cdot n)^{-\frac{5}{4}+o(1)}$ which gives

$$\sum_{x \in C_n} I_x(f_n)^2 \leq 4 \cdot n^2 \cdot (c \cdot n)^{-\frac{5}{2}+o(1)} = 4 \cdot c^{-\frac{5}{2}+o(1)} \cdot n^{-\frac{1}{2}+o(1)} \to 0.$$

This means that we are ok for the vertices that are "far" from the boundary.

Unfortunately for the general case we need something more since the distance from the boundary can become negligible.

What we actually need is a better geometrical understanding of what it is happening if a vertex near to the boundary of the rectangle is pivotal.

The reader should be comfortable with the subsection 2.3 of arm-events before he continues with the rest of the proof.

Recall two important facts for site percolation on $\mathbb{T}$:
-The 3-arms events in the half plane $\mathbb{H}$ at the annulus $A(r,R)$ is the event that there exist three disjoint arms of alternating status (open-closed-open) who are connecting the inner boundary of the annulus with the outer boundary and we ask also that they stay inside the half plane $\mathbb{H}$. The probability of such an event is denoted by $a_3^+(r,R)$ and it can be proved that it behaves like $(\frac{r}{R})^2$ both for site and bond percolation.

-The 2-arm in the quarter plane at the annulus $A(r,R)$ is the event that there exist two arms of different status (open and closed) connecting the inner boundary of the annulus with the outer that stay inside the quarter plane. The probability of such an event is denoted by $a_2^{++}(r,R)$ and it can be proved that it behaves like $(\frac{r}{R})^{2+o(1)}$ for the site percolation.

Now we are ready to state a key geometrical observation.

Take $x \in R_n$ which is pivotal. Let n_0 the distance of x from the boundary of R_n and x_0 a point in the boundary of R_n with $|x-x_0| = n_0$. Now let n_1 to be the smallest distance between x_0 and a vertex of R_n. Call the vertex with this minimal distance x_1.

Then we claim the following:

Claim 7.1. *If $x \in R_n$ is pivotal then (the notation is defined above):*

1)The 4-arm event is satisfied in the ball with center x and radius n_0.
2)The 3-arm event in $\mathbb{H}$ is satisfied in the annulus with center x_0 and radii $2n_0$ and n_1, assuming $2n_0 \leq n_1$,
and
3)The 2-arm event in the quarter plane is satisfied in the annulus with center x_1 and radii $2n_1$ and $a \cdot n$, assuming $2n_1 \leq a \cdot n$.

From the above we get that for every $x \in R_n$:

$$I_x(f_n) \leq 2a_4(n_0) \cdot a_3^+(2n_0, n_1) \cdot a_2^{++}(2n_1, a \cdot n),$$

since the ball and the annuli are mutually disjoint.

We set $a_3^+(2n_0, n_1) = 1$ or $a_2^{++}(2n_1, n) = 1$ if $n_1 \geq 2n_0$ or $2n_1 \leq n$ respectively for reasons of generality. Under this assumption the upper bound estimations we gave for these probabilities hold trivially.

The proof of the claim is easy if someone draws a picture, so we omit it.

Now using the claim and the facts we mentioned for site percolation we deduce the following:

For every $x \in R_n$,

$$I_x(f_n) \leq O(1) \cdot n_0^{-\frac{5}{4}} \cdot (\frac{n_0}{n_1})^2 \cdot (\frac{n_1}{n})^{2+o(1)} \leq O(1) n^{-\frac{5}{4}+\varepsilon}$$

and therefore

$$\mathbb{I}(f_n) \leq O(1) n^2 \cdot (n^{-\frac{5}{4}+\varepsilon})^2 = O(1) n^{-\frac{1}{2}+2\varepsilon} \to 0,$$

which means we are done for the case of site percolation on the triangular lattice.

Now let's see what can we say for the case of bond percolation on $\mathbb{Z}^2$.

In this case we set $R_n = ([0, a \cdot n] \times [0, b \cdot n]) \cap \mathbb{Z}^2$ and f_n the corresponding Boolean functions for the crossing events.

Then following the above proof everything follows with the exact same way using some minor modifications and the correspondence between vertices and edges, until the first line after the claim. Indeed the inequality that for every $x \in R_n$:

$$I_x(f_n) \leq 2a_4(n_0)a_3^+(2n_0, n_1)a_2^{++}(2n_1, a \cdot n).$$

holds again. The problem is just after that we can not use all the estimates for the arm-events we need since some of them are known only for the triangular lattice.

Let's see what do we know. We know that for the half plane the critical exponent is the same for bond and site percolation, $a_3^+(r,R) = \Omega((\frac{r}{R})^2)$, for the 4-arm event we know $a_4(r,R) \leq O(1)(\frac{r}{R})^{1+c}$ for some $c \in (0,1)$ and finally we do not have a result for the critical exponent for quarter-plane arm event $a_2^{++}(r,R)$.

Thus we need to bound the arm-event for the quarter plane. We need to recall F.K.G. inequality. F.K.G. inequality says that for up-sets $A, B \subset \Omega_n$ it holds: $\mathbb{P}(A \cap B) \geq \mathbb{P}(A)\mathbb{P}(B)$.

With this inequality we can prove that:

$$a_2^{++}(r,R)^2 \leq a_3^+(r,R) = \Omega((\frac{r}{R})^2).$$

Indeed, divide the $\mathbb{H}$ into two quarter planes and denote the event of having two quarter-plane arms in the left annulus as $A_1 = A_1(r,R)$ and the event of having two quarter-plane arms in the right annulus as $A_2 = A_2(r,R)$. Finally set B the 3-arm event in the half plane $\mathbb{H}$. Then $A_1 \cap A_2 \subset B$ and the F.K.G. inequality yields:

$$\mathbb{P}(A_1)\mathbb{P}(A_2) \leq \mathbb{P}(A_1 \cap A_2) \leq \mathbb{P}(B),$$

which is what we wanted to prove. Hence,

$$a_2^{++}(r,R) \leq O(1)\frac{r}{R}.$$

For bond percolation the critical exponent for the 5-arm event is known. It has actually the same behaviour with the 3-arm event in the half plane since both are behaving like $(\frac{r}{R})^2$, therefore $a_5(r,R) = \Omega(a_3(r,R)) = \Omega((\frac{r}{R})^2)$. Now observe that if the 5-arm event is happening then the 4 arm is happening also and therefore using the above two observations:

$$a_4(r,R) \geq a_5(r,R) = \Omega(a_3^+(r,R)).$$

This is actually really helpful because of the quasi-multiplicity property the 4-arm event has (see Theorem 3.2). Indeed based on the above facts we get:

$$a_4(n_0)a_3^+(2n_0, n_1) \leq O(1)a_4(n_0)a_4(2n_0, n_1) \leq O(1)a_4(n_1).$$

Thus for every edge $x \in R_n$ using also our quarter-plane bound:

$$I_x(f_n) \leq 2a_4(n_0)a_3^+(2n_0, n_1)a_2^{++}(2n_1, a \cdot n) \leq O(1)a_4(n_0)a_4(2n_0, n_1)\frac{n_1}{n} \leq O(1)a_4(n_1)\frac{n_1}{n} \quad (7.1)$$

Now using that $a_4(n_1) \leq O(1)n_1^{-(1+c)}$, for fixed $c \in (0,1)$, we get that for every $x \in R_n$:

$$I_x(f_n) \leq O(1)\frac{1}{n} \cdot n_1^{-c}$$

Therefore :

$$\sum_{x \in R_n} I_x(f_n)^2 \leq O(1)\frac{1}{n^2} \cdot \sum_{x \in R_n} n_1^{-2c}$$

which means that to prove noise sensitivity it is enough to show:

$$\frac{1}{n^2} \cdot \sum_{x \in R_n} n_1^{-2c} \to 0.$$

But $n_1 = n_1(x)$ is defined above as the distance between the projection of x to the boundary and the projection's nearest vertex. Therefore for each k, $1 \leq k \leq O(n)$, $n_1(x) = k$ for $O(k)$ edges. This means,

$$\sum_{x \in R_n} n_1^{-c} \leq O(1) \sum_{k=1}^{O(1)n} k^{1-2c}$$

which can be bounded above by:

$$\leq \int_0^{O(1)n} x^{1-2c} dx = O(1)n^{2-2c}.$$

Therefore:

$$\frac{1}{n^2} \cdot \sum_{x \in R_n} n_1^{-2c} \leq O(1)n^{-2c} \to 0,$$

and we are done.

Combining all the above we have our Theorem,

Theorem 7.2. *The percolation crossing events for site percolation on $\mathbb{T}$ and for bond percolation on $\mathbb{Z}^2$ are (asymptotically) noise sensitive*

Remark 7.3. *For the sequence of Boolean functions f_n of left-right crossing events either for $\mathbb{T}$ or $\mathbb{Z}^2$ we proved that $\mathbb{I}(f_n)$ decay polynomially and therefore we can deduce noise sensitivity even from the first proof we gave for B.K.S. theorem using Hypercontractivity at Section 5.*

7.2 Randomized algorithms

It was actually in the very first paper on noise sensitivity that Benjamini, Kalai and Schramm proved that percolation crossing events on $\mathbb{Z}^2$ are noise sensitive and the way they did it was by using a "correlation with majority" argument combined with a randomised algorithm. More about arguments of the first type will be discusses in Section 8. Here we would like to reveal the connection between randomised algorithms and noise sensitivity on general and in the next subsection we sketch a way that we can derive from that the noise sensitivity of crossing events.

Let us firstly introduce the notion of a randomised algorithm for a function $f : \Omega_n \to \mathbb{R}$.

An **algorithm** for f, A, is an algorithm that queries the coordinates of an element $x \in \Omega_n$ and determines $f(x)$ completely after it terminates. The way that it chooses the next coordinate to be queried can be based on the values of the coordinates A learnt before.

A **randomised algorithm** for f, A, is an algorithm for f as above with the additional property that randomness can also be used to choose which bit is the next one to be queried.

An important quantity for a function $f : \Omega_n \to \mathbb{R}$ and its' randomised algorithms, is the notion of **revealment** of f, called δ_f. Given a randomised algorithm A for f let J_A to be the random subset of the coordinates that are queried.

Definition 7.4. *For every randomised algorithm A define the revealment of A to be the quantity,*

$$\delta_A = \max_{i=1,..,n} \mathbb{P}(i \in J_A).$$

For a function $f : \Omega_n \to \mathbb{R}$, the revealment of f is defined to be the quantity:

$$\delta_f = \inf_A \delta_A,$$

where the infimum it is taken over all randomised algorithms A for f.

A very important theorem, proved in [SS11], that showed a connection between the revealment of a a function $f : \Omega_n \to \mathbb{R}$ and it's Fourier spectrum is the following:

Theorem 7.5. *[SS11] For every function $f : \Omega_n \to \mathbb{R}$ and $k \in [n]$,*

$$\sum_{S \subset [n], |S|=k} \widehat{f}(S)^2 \leq k \cdot \delta_f \cdot ||f||_2^2.$$

But given proposition 4.6, which characterises the property of noise sensitivity with respect to the Fourier Spectrum, combined with the above theorem we get a criterion for noise sensitivity via randomised algorithms:

Theorem 7.6. *([SS11]) For a sequence of Boolean function $f_n : \Omega_{m_n} \to \{-1, 1\}$ with $m_n \to \infty$ we have that if $\delta_{f_n} \to 0$ then the sequence f_n is noise sensitive.*

The proof of the Theorem 7.4. is elegant but it is technical therefore we omit it but we will try to present a non-rigorous argument why such a result is true.

Let's state first a word of caution given by Schramm and Steif made in [SS11] for the above theorem.

Assume that we have a randomised algorithm A that reveals $f = \sum \widehat{f}(S)\chi_S$. It seems rational to believe that every non-zero coefficient of the above representation of f should become known after running A to be able to decide f. Therefore at least one element of S for every $\varnothing \neq S \subset [n]$ with $\widehat{f}(S) \neq 0$ should be revealed , i.e. $J_A \cap S \neq \varnothing$. But that means that:

$$\sum_{S \subset [n], |S|=k} \widehat{f}(S)^2$$

should be near to

$$\sum_{S \subset [n], |S|=k} \widehat{f}(S)^2 \cdot \mathbb{P}(S \cap J_A \neq \varnothing)$$

which is less than:

$$k \cdot \delta_A \cdot \sum_{S \subset [n], |S|=k} \widehat{f}(S)^2.$$

Unfortunately such an argument is **not** true as Schramm and Steif noticed; there is another case that we can influence and even erase a coefficient of χ_S without revealing any element of S.

Indeed denote, for the t-th step of the algorithm, by $f_t(x)$ the function which equals $f(x)$ but with the known coordinates, until the t step, of x substituted with their exact values. Now assume that $S \subset [n]$ and $i \in [n] - S$ is revealed by A at step $t+1$. Then by substituting it's value on the part of the Fourier expansion of f_t:

$$\widehat{f}_t(S \cup \{i\}) \cdot \chi_{S\cup\{i\}} + \widehat{f}_t(S) \cdot \chi_S$$

the quantity becomes:

$$(x_i \widehat{f}_t(S \cup \{i\}) + \widehat{f}_t(S))\chi_S = \widehat{f}_{t+1}(S) \cdot \chi_S.$$

But it could be the case that

$$x_i \widehat{f}_t(S \cup \{i\}) + \widehat{f}_t(S) = 0$$

which means that we can actually get rid of the coefficient of χ_S without revealing any element of S as we stated. We call this phenomenon **"collapsing from above"**.

The proof of the theorem uses a decomposition theorem to handle exactly these collapses.

Let's ask ourselves now something different. How many collapses can happen in an algorithm with small revealment?

Assuming for the sake of the heuristic argument that the algorithm A is choosing the bits queried independently. Then a set with m elements collapses to a set of k elements if $m \geq k$ with probability at most $\binom{m}{k} \cdot \delta_A^{m-k}(1-\delta_A)^k$. Therefore the sum of the squares of the coefficients that will collapse from to above to level k is at most

$$(\sum_{m \geq k+1} \binom{m}{k} \cdot \delta_A^{m-k}) \cdot (1-\delta_A)^k \cdot (\sum_{S \subset [n]} \widehat{f}(S)^2)$$

which can be bounded from above by

$$\delta_A \cdot ||f||_2^2.$$

Therefore if δ_A is much smaller than $\sum_{S\subset[n],|S|=k} \widehat{f}(S)^2$ many of the sets in level k will not collapse and therefore at least one element from each of them should be revealed making the first mistaken heuristic we presented valid.

Details of a rigorous proof can be found in [SS11] and in [GS12].

Now one can ask what happens with the **converse** of Theorem 7.5: is the condition $\delta_{f_n} \to 0$ **necessary** for a sequence f_n if it is noise sensitive?

Unfortunately one can easily find a counterexample which is the sequence of Parity Boolean functions (example 2,subsection 2.1), call them f_n. This sequence of Parity Boolean functions as we proved at section 4 is noise sensitive but also it can be checked that a randomised algorithm that decides the n-th parity function needs to query every $i \in [n]$ to decide the outcome. The reason that this happens is that every $i \in [n]$ is pivotal for every input $x \in \Omega_n$. Therefore $\delta_{f_n} = 1$ for every $n \in \mathbb{N}$ and therefore $\delta_{f_n} \not\to 0$ as we wanted.

It is worth noticing that it is the same sequence that provides a counterexample for the converse of B.K.S. theorem (Theorem 4.14). In that case we wanted a noise sensitive sequence of Boolean functions f_n such that $\mathbb{I}(f_n) \not\to 0$ and the sequence of Parity Boolean functions satisfied $I_i(f_n) = 1$ for every $i \in [n]$ for the same reason again that i is pivotal for every input $x \in \Omega_n$. Despite that we were able to show that the converse for the B.K.S. theorem holds for monotone Boolean functions. Unfortunately such a result does not continue to hold here as well:

Proposition 7.7. *Whether a sequence of graphs of n vertices contains a clique or not of k_n vertices (where k_n is appropriately chosen) provides a counterexample of a sequence of monotone Boolean functions which is noise sensitive but their revealements are not going to zero.*

The details and proof of the above proposition can be found in Garban and Steif lecture notes ([GS12]).

7.3 Noise sensitivity of crossing events (via randomised algorithms)

Let's illustrate how we can derive the noise sensitivity of crossing events via the revealment theorem we mentioned.

From the previous subsection it is enough to show that if f_n, g_n are the sequences of Boolean functions deciding whether there is a left-right crossing with open vertices in the rectangle $R_n = ([0, a \cdot n] \times [0, b \cdot n]) \cap (\mathbb{T})$ and $R_n = ([0, a \cdot n] \times [0, b \cdot n]) \cap (\mathbb{Z}^2)$ respectively, then $\delta_{f_n} \to 0$ and $\delta_{g_n} \to 0$.

The theorem is the following:

Theorem 7.8. *([SS11])*

With the notation defined above it holds that:

$$\delta_{f_n} \leq n^{-\frac{1}{4}+o(1)} \text{ and } \delta_{g_n} \leq n^{-c},$$

as $n \to \infty$ for some $c > 0$.

Proof. (Sketch) We will focus on the site percolation for the triangular lattice $\mathbb{T}$. The case of bond percolation for the square lattice $\mathbb{Z}^2$ can be handled with some minor modifications.

We can find naturally an algorithm that decides whether there is a left-right crossing with open vertices or not, which is deterministic given a configuration of open and closed sites for the rectangle.

First of all, let's define the hexagonal lattice which can be constructed on top of the triangular lattice in the following way: put an hexagon on top of every site with the same orientation such that two hexagons will share an edge if and only if the corresponding sites are connected with an edge in $\mathbb{T}$. About the percolation model of $\mathbb{T}$ we colour red every hexagon that contains an open site and otherwise we colour it black.(See Figure 1.)

Therefore what we look for is a left right crossing with red hexagons.

Now using the hexagonal vertices and edges we can define the exploration path which will be able to check for us whether there is a left-right crossing with red hexagons or not. We start with the bottom-right hexagonal vertex in R_n. Then we construct the exploration path on the vertices of the hexagons following the following rule: if the hexagon the last vertex of our path encounters (in the sense that if it continued straight then it will enter the interior of this hexagon - with the exception of the initial vertex where we can assume that the hexagon encountered is the bottom right hexagon) is red it will turn left, otherwise it will turn right. By turning left or correspondingly right we mean that it goes through the hexagonal edge on it's right (left) to move to another hexagonal vertex. To make this work we assume that it treats the hexagons outside the rectangle in the following way: if the hexagon touches the bottom side we treat it as if it was black, if it touches the right side as if it was red and if our path reaches the left or the top side of the rectangle it stops. This keeps all the time red hexagons to the right of our path and black hexagons to the left. (See Figure 2.)

Firstly observe that there is a left-right crossing of red hexagons if and only if the exploration path is reaching the left side of the rectangle.

Now given the probabilistic structure that the colours are given to the hexagons one can make some comments about the probability that an hexagon in the rectangle will be queried. Observe that for an hexagon to be queried from the algorithm, from next to this hexagon there must be starting a black path of hexagons and a red path of other hexagons such that both of these paths are reaching the boundary. This is true based on the way the exploration path is constructed since from the right it has neighbours only red hexagons and from the left only black. Therefore if the hexagon has distance R from the boundary and it is queried then it must satisfy the 2-arm event of radius R and therefore the probability that it is revealed is $R^{-\frac{1}{4}+o(1)}$. That means that the probability of revealment for hexagons far from the boundary is small but of course the revealment on general is quite big as the right bottom hexagon is queried with probability 1.

What Schramm and Steif did in [SS11] to produce a randomised algorithm with better revealment is to pick randomly an hexagonal edge in the middle third of the right side of R_n, call it e and start from the midpoint of it, call it p, two paths. The first will be an exploration path that will go up as for the first step and will be able therefore to detect whether there is a left-right crossing with starting hexagon from the right side above p. The second now will go down and it will create a path treating the colours in the opposite way of what an exploration path is doing, in the sense that if it encounters a red hexagon it will turn right, if it encounters a black hexagon it will turn left and it will treat the hexagons outside the rectangle from the top side as they were black, from the right side as if they were red and if it reaches the left or the bottom side of the rectangle it stops. This path will be able to detect whether there is a left-right crossing with red hexagons with starting hexagon from the right side below p. Then indeed there exist a left-right crossing with red hexagons if and only if at least one of the paths will reach the left boundary. Moreover it can be shown that the revealment in this case is bounded from above by $n^{-\frac{1}{4}+o(1)}$ as we wanted. The exact computation of the revealment requires a deeper geometrical observation similar to Claim 7.1 that uses also the 1-arm half-plane critical exponent for site percolation which it is known that it equals $\frac{1}{3}$.

For bond percolation on $\mathbb{Z}^2$ we act similarly by defining an exploration path in a similar construction like the hexagonal lattice. A drawback is that for bond percolation the critical exponents for the 2-arm event and the 1-arm half-plane event are not known. Therefore to overcome this, we bound them both from above, using R.S.W. theorem, by R^{-c} for some constant $c > 0$. It is this exponent that gives the unknown positive exponent in the statement of the theorem.

For details see [SS11]. $\square$

Using Theorem 7.8 and Theorem 7.4 we deduce that for every k and n we get:

$$\sum_{S\subset[m_n],|S|=k} \widehat{f_n}(S)^2 \leq k \cdot n^{-\frac{1}{4}+o(1)}.$$

This yields of course that :

$$\sum_{S\subset[m_n],|S|\leq M_n} \widehat{f_n}(S)^2 \leq O(M_n^2) \cdot n^{-\frac{1}{4}+o(1)},$$

and therefore if $M_n = o(n^{\frac{1}{8}-o(1)})$ then:

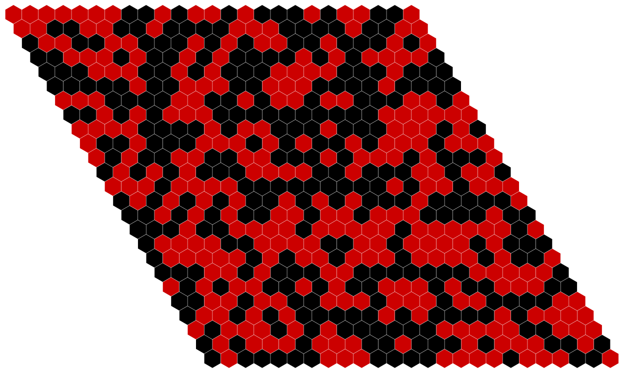

Figure 1: The hexagonal lattice with red and black colours

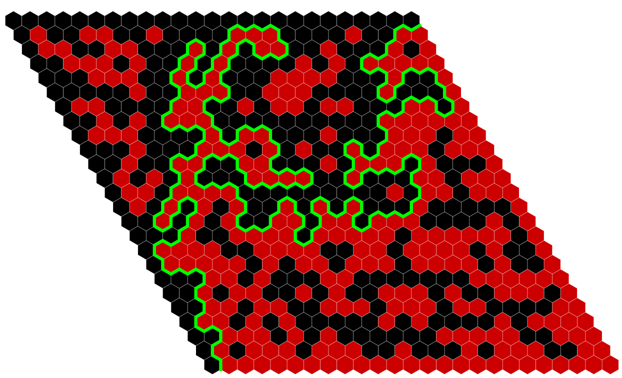

Figure 2: Exploration path for the hexagonal lattice

$$\sum_{S\subset[m_n],|S|\leq M_n} \widehat{f_n}(S)^2 \to 0$$

Similarly for bond percolation and the sequence g_n one gets that if $M_n' = o(n^{\frac{c}{2}})$ where $c > 0$ is given by Theorem 7.8 then :

$$\sum_{S\subset[m_n],|S|\leq M_n'} \widehat{g_n}(S)^2 \to 0.$$

That means firstly that sequences f_n, g_n are noise sensitive but it actually tells more than this. Noise sensitivity from Proposition 4.6. is a property that demands the Fourier Spectrum to do not be accumulated in low ($O(1)$) frequencies but the above theorem shows that their Fourier spectrum is not only not consecrated on $O(1)$-frequencies but it is consecrated only on polynomially high frequencies. This is an advantage of the proof given via randomised algorithms comparing to using the B.K.S. criterion and arm events. Indeed the polynomial decay for the L^2 norm of the influences gives, based on the work we did at section 5.2, only that the Fourier spectrum is consecrated on logarithmically high frequencies.

Can we actually find more information about the consecration of this spectral mass for the crossing events? The answer is yes and this will be the topic of our next subsection.

7.4 Fourier spectrum of crossing events

To study the Fourier spectrum we need an observation. Take a sequence of Boolean function $f_n : \Omega_{m_n} \to \{-1, 1\}$ where $m_n \to \infty$. Imagine that for every $n \in \mathbb{N}$ we don't study the case where we have the same noise $\varepsilon > 0$ for every $n \in \mathbb{N}$ but instead we introduce a different noise $\varepsilon_n > 0$ for every $n \in \mathbb{N}$. In that case applying the known results from section 4 we get:

$$\mathrm{Cov}(f_n(x), f_n(x_{\varepsilon_n})) = \sum_{S\subset[m_n],S\neq\varnothing} \widehat{f_n}(S)^2(1-\varepsilon_n)^{|S|}$$

and

$$2 \cdot \mathbb{P}(f_n(x) \neq f_n(x_{\varepsilon_n})) = \sum_{S\subset[n]} \widehat{f_n}(S)^2(1-(1-\varepsilon_n)^{|S|})$$

Then,

Proposition 7.9. *For a sequence $\varepsilon_n \to 0$ it holds that:*

$\mathrm{Cov}(f_n(x), f_n(x_{\varepsilon_n})) \to 0$ *,implies* $\sum_{S\subset[m_n],|S|<\varepsilon_n^{-1}} \widehat{f_n}(S)^2 \to 0$

and,

$\mathbb{P}(f_n(x) \neq f_n(x_{\varepsilon_n})) \to 0$ *implies* $\sum_{S\subset[m_n],|S|>\varepsilon_n^{-1}} \widehat{f_n}(S)^2 \to 0.$

Proof. For the first implication:
Our hypothesis implies that

$$\sum_{S\subset[m_n],|S|<\varepsilon_n^{-1}} \widehat{f_n}(S)^2(1-\varepsilon_n)^{|S|} \to 0.$$

But for $S \subset [m_n]$ with $|S| < \varepsilon_n^{-1}$ we have

$$(1-\varepsilon_n)^{|S|} > (1-\varepsilon_n)^{\varepsilon_n^{-1}}$$

Therefore

$$(1-\varepsilon_n)^{\varepsilon_n^{-1}} \cdot \sum_{S\subset[m_n],|S|<\varepsilon_n^{-1}} \widehat{f_n}(S)^2 \to 0$$

and as the first term tends to e^{-1} we are done.
The second implication follows similarly. □

The above theorem means that by looking how a sequence of Boolean function reacts under a sequence of different noises can give us important information about their Fourier Spectrum and it's consecration.

What we have proved so far for the percolation events is that for the triangular lattice the spectrum is consecrated on frequencies of sets with cardinality a lot bigger than $n^{\frac{1}{8}-o(1)}$ and similarly for the square lattice the consecration happens for sets with cardinality a lot bigger than $n^{\frac{c}{2}-o(1)}$. A natural question is now to ask for an upper bound. At this point we remind ourselves that we work on Ω_{m_n} with $m_n = O(n^2)$. The following theorem gives a bound on this type.

Theorem 7.10. *([GS12]) Both for site percolation on $\mathbb{T}$ and for bond percolation on $\mathbb{Z}^2$ and the corresponding sequences of Boolean functions f_n and g_n of the crossing events it holds:*

$$\sum_{S\subset[m_n],|S|>>n^2\cdot a_4(n)} \widehat{f_n}(S)^2 \to 0,$$

and similarly for g_n:

$$\sum_{S\subset[m_n],|S|>>n^2\cdot a_4(n)} \widehat{g_n}(S)^2 \to 0,$$

where $a_4(n)$ is the probability of the 4-arm event of radius n, in the first inequality for site percolation and in the second for bond percolation.

Proof. Using Proposition 7.9 it is enough to show that if $\varepsilon_n = o((n^2 \cdot a_4(n))^{-1})$ then

$$\mathbb{P}(f_n(x_{\varepsilon_n}) \neq f_n(x)) \to 0.$$

We need a lemma:

Lemma 7.11. *The total influence of f_n, $I(f_n) := \sum_{x\in R_n} I_x(f_n)$, is $O(n^2 \cdot a_4(n))$.*

Proof. We will prove it simultaneously for $\mathbb{T}$ and for $\mathbb{Z}^2$.

We remind ourselves from relation (7.1) that for every $x \in R_n$, $I_x(f_n) \leq O(1)a_4(n_1)\frac{n_1}{n}$, where $n_1 = n_1(x)$ was defined as the distance between the projection of x to the boundary of R_n and the projection's nearest vertex.

Therefore $I(f_n) \leq O(1)\sum_{x\in R_n} a_4(n_1)\frac{n_1}{n}$. But as we argued also at section 7.1 $n_1(x) = k$ for $O(k)$ vertices.

That gives:

$$I(f_n) \leq \frac{1}{n} \cdot \sum_{k=1}^{C \cdot n} a_4(k) \cdot k^2,$$

where $C > 0$ is an appropriate constant.

Now from the quasi-multiplicity property (Theorem 3.2), the fact that the 5-arm event implies the 4-arm event and the fact that $a_5(r, R)$ behaves like $(\frac{r}{R})^2$ (all of these hold either for site or bond percolation) we get for every $k < C \cdot n$ that

$$a_4(k) \leq O(1)\frac{a_4(n)}{a_4(k, C \cdot n)} \leq O(1)\frac{a_4(n)}{a_5(k, C \cdot n)} \leq O(1)\frac{n^2 a_4(n)}{k^2}$$

Combining the above relations we get:

$$I(f_n) \leq O(1) n^2 a_4(n),$$

as we wanted. □

Now back to our original proof let: $\varepsilon_n = o((n^2 \cdot a_4(n))^{-1})$. Remember that the noise acts on every vertex of R_n independently. Order them as $\{x_1, .., x_{n^2}\}$ and assume that the noise acts with respect to that order. Moreover suppose that $y = x_{\varepsilon_n}$ with: $\{y_1, .., y_{n^2}\}$, i.e. we keep the same order as before for the vertices but now they have the status after the noise. (Note that the vertices are $O(n^2)$ but we treat them as they were n^2 for convenience as the constants does not matter for our purposes).

Now for $f(x) \neq f(x_{\varepsilon_n})$ to happen there exist at least one i such that x_i flipped and $f(y_1, .., y_{i-1}, x_i, .., x_n) \neq f(y_1, .., y_{i-1}, y_i, x_{i+1}, .., x_n)$.

The probability of the last to happen, as y follows the same (uniform) distribution with x, is at most $\frac{\varepsilon_n}{2} \cdot \frac{I_i(f_n)}{2}$.

Therefore

$$\mathbb{P}(f(x) \neq f(x_{\varepsilon_n})) \leq O(1)\varepsilon_n \cdot I(f_n) \leq O(1)\varepsilon_n \cdot n^2 a_4(n),$$

where the last inequality follows from Lemma 7.11.
Given our hypothesis for ε_n we are done and the proof is complete. □

Therefore we can say that for site percolation the spectrum is mostly consecrated on sets with cardinality in the interval $[n^{\frac{1}{8}-o(1)}, n^2 a_4(n)]$, ignoring constants.

Since $a_4(n) = \Omega(n^{-\frac{5}{4}+o(1)})$ the interval becomes: $[n^{\frac{1}{8}-o(1)}, n^{\frac{3}{4}+o(1)}]$.

Similarly for bond percolation the interval where the spectrum is consecrated as $a_4(n) = O(n^{-(1+c_2)})$ is $[n^{c_1}, n^{1-c_2}]$ for some positive constants $c_1, c_2 > 0$.

Can we say even more?

Actually yes. A lot of research is done and more results became known about the spectrum of the crossing events. The most surprising is that it was finally proved from Garban, Pete and Schramm in [GPS11] that for both site and bond percolation almost all the mass of their Fourier spectrum is consecrated on sets with cardinality, up to constants, equal to the upper bound we gave, i.e. equal to $n^2 a_4(n)$. The proof is quite technical and requires a lot of new theory to be established. (For details see [GPS11], [GS12]).

8 The role of Majority in noise sensitivity

We devote this section on trying to present a leading role that Majority Boolean functions plays in the context of noise sensitivity. We remind ourselves that for every $n \in \mathbb{N}$ we define the n-th majority funtion to be $M_n : \Omega_n \to \{-1, 1\}$ given by

$$M_n(x) = \mathbf{sign}(x_1 + .. + x_n)$$

if $x_1 + .. + x_n \neq 0$ and 1 otherwise. This convention we do by setting it equal to 1 on the middle level set it is just for including the case of all possible hypercubes Ω_n and not just those with odd dimension. It is a convention with not actual contribution to the asymptotic behaviour of the Majority functions which is exactly what we care about.

8.1 Correlation with majority

It was in the very first paper on noise sensitivity from Benjamini,Kalai and Schramm, [BKS99], that it was mentioned an elegant criterion for a sequence of monotone Boolean function to satisfy $\mathbb{I}(f_n) \to 0$, and therefore to be noise sensitive. The criterion non-rigorously said that if a sequence of monotone Boolean function is much (asymptomatically) uncorrelated with the majority functions then the sequence is noise sensitive. We will present and prove this criterion now.

For every $K \subset [n]$ set $M_K : \Omega_n \to \{-1, 1\}$ be the Boolean function given by

$$M_K(x) = \mathbf{sign}(\sum_{j \in K} x_j),$$

given that $\sum_{j \in K} x_j \neq 0$ and set it equal to 1 otherwise. Notice that if $K = [n]$, $M_K = M_{[n]}$ is the known majority function on the n-coordinates.

Now let $f : \Omega_n \to \{-1, 1\}$ be a Boolean function. We define "correlation with majority" the quantity given by :

$$\mathbb{L}(f) = \max_{K \subset [n]} \{|\mathbb{E}[f \cdot M_K|]\}.$$

Then the criterion is the following one:

Theorem 8.1. *([BKS99]) If $f : \Omega_n \to \{-1, 1\}$ is a monotone Boolean function with $\mathbb{L}(f) \leq e^{-\frac{1}{2}}$ then:*

$$\mathbb{I}(f) \leq C \cdot \mathbb{L}(f)^2 \cdot (1 - \log(\mathbb{L}(f))) \cdot \log(n),$$

where $C > 0$ is a universal constant.

Therefore if a sequence of monotone Boolean functions $f_n : \Omega_{m_n} \to \{-1, 1\}$ satisfies

$$\mathbb{L}(f_n)^2 \cdot (1 - \log(\mathbb{L}(f_n))) \cdot \log(m_n) \to 0,$$

then

$$\mathbb{I}(f_n) \to 0$$

and therefore the sequence $\{f_n\}$ is noise sensitive.

Let's present now the proof of Theorem 8.1. divided into some steps. We will first prove some intermediate results and then explain the final proof.

We prove first a result about a quantitative relation between the total influence of a monotone Boolean function and it's correlation with majority.

Lemma 8.2. *For a monotone Boolean function* $f : \Omega_n \to \{-1, 1\}$ *, it holds:*

$$I(f) \leq C\sqrt{n} \cdot \mathbb{E}\big[fM_{[n]}\big] \cdot (1 + \sqrt{-\log\left(\mathbb{E}\big[fM_{[n]}\big]\right)}),$$

where $C > 0$ *is some universal constant.*

Proof. (Sketch)
Set for every k with $1 \leq k \leq n$:

$$F(k) = \frac{1}{\binom{n}{k}} \cdot \left(\sum_{\{x : |\{i | x_i = 1\}| = k\}} f(x) \right).$$

Now since f is monotone one can show with direct computations that:

$$I(f) = 2 \cdot 2^{-n} (\sum_{k > \frac{n}{2}} \binom{n}{k} \cdot (F(k) - F(n-k)) \cdot (2k - n))$$

and,

$$\mathbb{E}\big[fM_{[n]}\big] = 2^{-n} (\sum_{k > \frac{n}{2}} \binom{n}{k} (F(k) - F(n-k))).$$

Therefore for $L > 0$ (using that $F(k) \leq 1$ for every k) we get:

$$I(f) \leq O(1) \cdot (2^{-n} (\sum_{L > k > \frac{n}{2}} \binom{n}{k} \cdot (F(k) - F(n-k))) \cdot (2L - n) + \sum_{L < k} 2^{-n} \cdot \binom{n}{k} \cdot (2k - n)$$

which is less than,

$$C_1 \mathbb{E}\big[fM_{[n]}\big] (2L - n) + \sum_{L < k} 2^{-n} \cdot \binom{n}{k} \cdot (2k - n),$$

for some constant $C_1 > 0$.

Now for convenience we set $L = \frac{n}{2} + S$ and therefore our bound equals:

$$C_1 \cdot \mathbb{E}\big[fM_{[n]}\big] \cdot S + \sum_{\frac{n}{2} + S < k} 2^{-n} \cdot \binom{n}{k} \cdot (2k - n).$$

Now it holds:

$$2^{-n} \cdot \binom{n}{k} \cdot (2k - n) \leq C_2 \exp(-\frac{(2k - n)^2}{C_3 n})$$

for constants $C_2, C_3 > 0$. Therefore applying this inequality and bounding the sum of the exponentials by it's integral, we have as an upper bound the:

$$C_1 \cdot \mathbb{E}\big[fM_{[n]}\big] \cdot S + C_4 \sqrt{n} \exp(-\frac{S^2}{n}).$$

Now we choose $S = C_5 \sqrt{n} \sqrt{-\log(\mathbb{E}\big[fM_{[n]}\big])}$ and thus get the desired inequality.

□

Now given a monotone Boolean function f and $K \subset [n]$ we set $f_K(x) = 2^{n-|K|} \sum_{y \in \{-1,1\}^{[n]-K}} f(x,y)$ for $x \in \{-1,1\}^K$. Then f_K is monotone and moreover one can check that

$$I_K(f) := \sum_{i \in K} I_i(f) = I(f_K).$$

Given that observation, Lemma 8.2 applied to f_K yields that:

Corollary 8.3. *For a monotone Boolean function $f : \Omega_n \to \{-1,1\}$ and $K \subset [n]$ it holds that:*

$$I_K(f) \leq C\sqrt{|K|}\mathbb{E}[fM_K]\,(1 + \sqrt{-\log\left(\mathbb{E}[fM_K]\right)}),$$

where $C > 0$ is some universal constant.

Now we are ready for the proof of our theorem:

Proof. (of Theorem 8.1)

Order the influences without loss of generality into a decreasing order,
i.e. assume that: $I_{i+1}(f) \leq I_i(f)$, for $i = 1, .., n-1$.

Now one can check that for $x \leq e^{-\frac{1}{2}}$ the function

$$g(x) := x(1 + \sqrt{-\log(x)})$$

is increasing.

Therefore Theorem 8.3 applied for $\{1, 2, .., k\}$ for $k = 1, .., n$ yields :

$$\sum_{i=1}^{k} I_i(f) \leq C \cdot \mathbb{L}(f) \cdot (1 + \sqrt{-\log\left(\mathbb{L}(f)\right)}) \cdot \sqrt{k} \text{ for } k = 1, .., n$$

Now we need to observe that since $I_i(f)$ is a decreasing sequence of non-negative numbers, to maximize

$$\mathbb{I}(f) = \sum_i I_i(f)^2$$

we must have equality in all of the above constraints.

Let's illustrate the reason behind. For the sake of this argument assume $I_i(f), i = 1, .., n$ is just any decreasing sequence of non-negative numbers satisfying the above inequality constraints and maximizing $\mathbb{I}(f)$.

Assume that we don't have equality for the first time at the k-th constraint. We distinguish two cases.

If $k = n$ we can increase $I_n(f)$ without a problem as more as possible since in that case every other influence will be equal to $C \cdot \mathbb{L}(f)\left(1 + \sqrt{-\log\left(\mathbb{L}(f)\right)}\right)(\sqrt{i} - \sqrt{i-1})$ for $i = 1, .., n-1$, and the n-th constraint will become

$$I_n(f) \leq C \cdot \mathbb{L}(f)\left(1 + \sqrt{-\log\left(\mathbb{L}(f)\right)}\right)(\sqrt{n} - \sqrt{n-1}).$$

Therefore given that $\sqrt{i} - \sqrt{i-1}$ decreases with i , we can set $I_n(f)$ equal to it's upper bound. But that will increase $\mathbb{I}(f)$ which is a contradiction.

In the other case now where $k < n$: Firstly we argue that we can not have $I_k(f) = I_{k-1}(f)$. This is true since the linear system of equalities for the first $k-1$ constraints gives

$$I_{k-1}(f) = C \cdot \mathbb{L}(f) \left(1 + \sqrt{-\log\left(\mathbb{L}(f)\right)}\right) (\sqrt{k-1} - \sqrt{k-2})$$

and if we had

$$I_k(f) = I_{k-1}(f) = C \cdot \mathbb{L}(f) \left(1 + \sqrt{-\log\left(\mathbb{L}(f)\right)}\right) (\sqrt{k-1} - \sqrt{k-2})$$

the k-th constraint will be violated since as before $\sqrt{i} - \sqrt{i-1}$ decreases with i. Therefore indeed $I_k(f) < I_{k-1}(f)$.

Now we increase $I_k(f)$ by $\alpha > 0$ and decrease $I_{k+s}(f)$ for the amount $\frac{\alpha}{n-k} > 0$,for every $s = 1,..,n-k$. Finally we keep all the other influences the same. Then if $\alpha > 0$ is small enough all the constraints are valid plus the decreasing property. Finally the sum of squares changes by something bigger than

$$2 \cdot \alpha \cdot \left(I_k(f) - \frac{1}{n-k} \cdot \sum_{s=1}^{n-k} I_{k+s}(f) \right) \geq 0$$

and we are done.

Therefore indeed we have that in the extreme case that every constraint holds with equality and therefore every influence equals to

$$C \cdot \mathbb{L}(f) \left(\left(1 + \sqrt{-\log\left(\mathbb{L}(f)\right)}\right) (\sqrt{i} - \sqrt{i-1} \right), \text{ for } i = 1,..,n.$$

Given the above argument we have:

$$\mathbb{I}(f) \leq O(1) \sum_{i=1}^{n} \left(\mathbb{L}(f) \left(1 + \sqrt{-\log\left(\mathbb{L}(f)\right)}\right) (\sqrt{i} - \sqrt{i-1}) \right)^2$$

$$\leq O(1)\mathbb{L}(f)^2 \left(1 - \log\left(\mathbb{L}(f)\right)\right) \cdot \sum_{i=1}^{n} \frac{1}{i} \leq O(1) \cdot \mathbb{L}(f)^2 \cdot \left(1 - \log\left(\mathbb{L}(f)\right) \cdot \log n.\right.$$

The proof of Theorem 8.1 is complete. □

An elementary but basic remark about the sequence of Majority Boolean functions that one can deduce directly from the computations inside the proof of Theorem 8.2 is the following one:

Proposition 8.4. *Let $n \in \mathbb{N}$. Then for a monotone boolean function $f : \Omega_n \to \{-1, 1\}$ it holds:*

$$I(f) \leq I(M_{[n]}) = \Omega(\sqrt{n})$$

Proof. It follows from the following observations for a monotone Boolean function f: Firstly as in the proof of Theorem 8.2 if

$$F(k) = \frac{1}{\binom{n}{k}} \cdot \left(\sum_{\{x:|\{i|x_i=1\}|=k\}} f(x) \right)$$

then:

$$I(f) = 2^{-n}(\sum_{k>\frac{n}{2}} \binom{n}{k} \cdot (F(k) - F(n-k)) \cdot (2k-n)).$$

Secondly for every $k > \frac{n}{2}$ it holds : $F(k) - F(n-k) \leq 2$ with equality to all of them if and only if $f = M_{[n]}$ up to a change in the elements with $\frac{n}{2}$,if n is even, 1's since they have no contribution in the sum.

Finally some final computations indeed give

$$I(M_{[n]}) = 2^{-n}(\sum_{k>\frac{n}{2}} \binom{n}{k} \cdot 2 \cdot (2k-n)) = \Omega(\sqrt{n}).$$

□

Remark 8.5. *Benjamini,Kalai and Schramm proved the noise sensitivity of the crossing events using Theorem 8.1 and a randomised algorithm. For more details see [BKS99].*

8.2 Majority is stablest

In this section we will try to present one of the biggest proven conjectures in the area of noise sensitivity known as "Majority is stablest".

As we mentioned in Example 3 of subsection 4.1 based on central limit theorem we can make an heuristic argument about the fact that denoting $M_{[n]}$ the sequence of majority Boolean functions then the quantity $\mathbb{P}(M_{[n]}(x) \neq M_{[n]}(x_\varepsilon))$ behaves like $\sqrt{\varepsilon}$ for small values of ε and big values of $n \in \mathbb{N}$. Once more we mention that fact that majority function is setted to 1 on the middle level set if n is even does not affect the limit behaviour. It can actually be proven the following:

Proposition 8.6. *(Sheppard 1890) Asymptoticaly*

$$\mathbb{P}(M_{[n]}(x) \neq M_{[n]}(x_\varepsilon)) \to \frac{1}{\pi} \cdot \arccos{(1 - \frac{\varepsilon}{2})} \text{ ,}as\ n \to \infty.$$

We also mentioned that the sequence of Majority Boolean functions play an extreme role between all Boolean functions with low influences in the sense that it is asymptotically the stablest. Let's try to work more on this thing.

Based on our last subsection results we proved that if a sequence of monotone Boolean functions has sufficient low correlation with the sequence of Majority Boolean functions then it is noise sensitive. This seems to imply that Majority Boolean functions are very far from being noise sensitive in the sense that even if a monotone Boolean function is correlated with it it becomes not noise sensitive.

Therefore it is naturally to ask whether we can make more precise this relation. For example on can ask whether generally for Boolean functions the quantity that should go to zero if a sequence is noise stable, $\mathbb{P}(f(x) \neq f(x_\varepsilon))$, is minimized asymptotically for the sequence of Majority Boolean functions. The answer to this question is unfortunately negative.

Using Fourier analysis we have that

$$2 \cdot \mathbb{P}(f(x) \neq f(x_\varepsilon)) = \sum_{\varnothing \neq S \subset [n]} \widehat{f}(S)^2(1 - (1-\varepsilon)^{|S|}) \geq (||f||_2^2 - \mathbb{E}[f]^2) \cdot \varepsilon = \mathrm{Var}(f) \cdot \varepsilon,$$

using Parseval's identity. Therefore it behaves at least like ε and actually this behaviour is realised for the Dictatorship Boolean functions(example 1 in subsection 4.1) , i.e. for the Boolean functions given by $\mathbf{Dict}(x_1, .., x_n) = x_1$. Indeed the noise in that case changes the result if and only if it flips the first coordinate which happens with probability $\frac{1}{2}\varepsilon$. Of course as we mentioned above the noise stability of the majority Boolean function behaves like $\sqrt{\varepsilon}$ which is strictly bigger than ε.

The question then should change to something more reasonable. A crucial difference between dictatorship and majority is the behaviour of their influences. In the example of dictatorship the first coordinate has the maximal possible influence and on the contrary the influences for $M_{[n]}$ uniformly go to 0. Indeed a computation made at the proof of Proposition 8.4 gave that $I(M_{[n]}) = \Omega(\sqrt{n})$ and therefore because of the symmetry between the coordinates every influence behaves like $\frac{1}{\sqrt{n}}$.

Therefore one can ask what happens if we restrict ourselves to Boolean functions with uniformly low influences? Is now Majority Boolean functions asymptotically the stablest? It turns out that there is in that case an affirmative answer under the extra symmetry assumption that our Boolean function has mean zero:

Theorem 8.7. *([MOO10])(**"Majority is stablest"**)*

Fix a noise $\varepsilon > 0$. Then for every $p > 0$ there is a $\delta = \delta(\varepsilon, p) > 0$ such that for every Boolean function $f : \Omega_n \to \{-1, 1\}$ with $\max_{i=1,..n} I_i(f) \leq \delta$ and $\mathbb{E}[f] = 0$ it holds:

$$\mathbb{P}(f(x) \neq f(x_\varepsilon)) \geq \frac{1}{\pi} \cdot \arccos{(1 - \frac{\varepsilon}{2})} - p.$$

Notice now that theorems 8.6 , 8.7 indeed reason the title that Majority Boolean functions are asymptotically the stablest.

We will not present a proof of this theorem 8.6 but we will make some comments on the interesting background of this theorem. The "Majority is stablest" as a conjecture was firstly formally stated in [KKMO07] from Mossel,Khot,Kindler and O'Donnell. Despite that, the belief that Hamming balls will minimize noise stability quantities goes back at least to the [KKL88] paper of 1988. Then B.K.S. paper [BKS99] by showing "correlation with majority results" made the belief stronger. After that a result came from Bourgain that showed that the quantity $\sum_{|S| \geq k} \widehat{f}(S)^2$ among all Boolean functions with mean zero and low influences is asymptotically minimized for the sequence of Majority Boolean functions (for more details see [Bou02]). This result, based on our results in section 4 is really close to the conjecture. Finally strong motivation on proving sharp lower bounds for the noise stability of low influences Boolean functions came firstly from Theoretical computer science and more specifically Hardness of Approximation results (see Khot's paper [Kho02]) and secondly from Social Choice theory (see Kalai's paper [Kal02]). Finally it was proved by Mossel, O'Donnell, and Oleszkiewicz in [MOO10].

Hence one can see that noise sensitivity has connections with Theoretical computer science and Social Choice theory. For the first we will not mention something more in the rest of the essay (though one can find a lot informations in [Kho02],[KKMO07],[MOO10] and in the very informative thesis of O'Donnel, [O'D03]), but we will devote the next section on presenting a connection that noise sensitivity has with Social Choice theory.

9 Noise sensitivity and social chaos in Social Choice Theory

Copying Gil Kalai's word from his paper [Kal10]:
How likely is it that small random mistakes in counting the votes in an election between two

candidates will reverse the election's outcome? And if there are three alternatives and the society prefers alternative a to alternative b and alternative b to alternative c what is the probability that it will prefer alternative a to alternative c ?

The first question seems to be closely related with the notion of noise sensitivity and things we have studied so far. The second is a question in social choice theory that goes back to Condorcet at the late 18th century. It turns out that these two questions are closely related and this is exactly what we will try to illustrate at this section. It is a connection made from Gil Kalai in [Kal10], which uses some results he had already proven in [Kal02].

We start by presenting the basic notions of social choice theory and after that we present the theorem that reveals this connection, proving some parts of it and sketching some others.

9.1 Some basic notions

Let's start with the notion of a simple game.

A **simple game** (or voting game) G on a set of $N = \{1, 2, .., n\}$ players (voters) is described by a Boolean function $f : \Omega_n \to \{-1, 1\}$ satisfying $f(-1, -1, .., -1) = -1$ and $f(1, 1, .., 1) = 1$. We say that a subset S of N forms a **winning coalition** if the element $x_S \in \Omega_n$ that corresponds to the set S (i.e. the element $x_S \in \Omega_n$ such that $(x_S)_i = 1 \leftrightarrow i \in S$) satisfies $f(x_S) = 1$.

We say that:

- the game is **monotone** if f is a monotone increasing Boolean function

and

- the game is **strong** if f satisfies $f(-x) = -f(x)$ for every $x \in \Omega_n$.

From now on, will care mostly for strong simple games.

A **neutral social welfare function** , (neutral S.W.F.) $F(A)$ for a set A of m alternatives based on a strong simple game G is defined as follows:

Assume that we have a society with n voters and that they want to vote between m alternatives. Each of the voters is giving to the society a linear order with his preferences on the set of the alternatives. All together these linear orders form a **profile**. Given this profile of linear orders $R_1, .., R_n$, F assigns on the alternatives an asymmetric relation R in the way that aRb if and only if the set $S_{a,b} = \{i \in [n] | aR_ib\}$ forms a winning coalition for the strong simple game G. That means that the society prefers a to b if the set of the voters that prefer a to b form a winning coalition for the strong simple game G.

Note that F is completely determined by G and the number of the alternatives and moreover that the relation it gives is well defined and asymmetric exactly because G is assumed to be strong.

We say that an asymmetric relation R is **rational** if it is an order relation in the alternatives. What we mean by this is that if the alternatives are given without loss of generality from $A = \{1, 2, ..., m\}$ then they can be renumbered such that iRj for every $i < j$. Equivalently an asymmetric relation is rational if and only if there is no circle in the preferences between the alternatives, i.e. there are no alternatives $i_1, .., i_k \in A$ such that i_jRi_{j+1} for every $j \in [k]$ where the indices are going modulo k. Naturally if our relation is not rational we call it **non-rational**. In the case we have three alternatives the equivalent **cyclic asymmetric relation** will be preferred.

A well-known fact is that a neutral S.W.F. does not certainly results into a rational asymmetric relation. It is something observed by Condorcet at the late 18th century. He gave a fairly easy example where the Boolean function for the simple game G was the Majority and there were 3 voters and 3 alternatives. Let's present it: Imagine that our voters profile is: aR_1bR_1c, cR_2aR_2b and bR_3cR_3a where for example aR_1bR_1c means that voter 1 prefers a to b, a to c and b to c. In this case it turns out that the S.W.F. based on the majority Boolean function gives that society prefers alternative a to alternative b , alternative b to alternative c and alternative c to alternative a. After that example it was conjectured that this was a problem rising up from the nature of the Majority Boolean functions. Unfortunately after more than one century and a half, Arrow generalised the result for any non-dictatorship Boolean function :

Theorem 9.1. ***(Special case of Arrows impossibility theorem)***

Let F be a neutral social welfare function based on a strong simple game G.

Then, unless the boolean function f that describes the game G is the ***Dictatorship*** *Boolean function, there always exist a profile of linear orders such that F results in a cyclic asymmetric relation.*

Now to study the social welfare functions we endow a uniform probability measure on the possible profiles. More specifically to every voter we assign the same probability $\frac{1}{m!}$ for each one of the $m!$ order preference relation on the alternatives. Furthermore we assume that the individual preference relations are independent.

Now we are ready to introduce the notion of social chaos. Suppose first that there are three alternatives and that G is a strong simple game with n voters. Let F also be the neutral S.W.F. based on G for three alternatives. Then given the uniform measure on the individual profiles we write $p_{cycl}(G)$ for the probability that the random uniform profile will result in a cyclic asymmetric relation. In that context, Theorem 9.1 states that, unless G is a dictatorship, $p_{cycl}(G) > 0$.

Definition 9.2. *Let a sequence (G_n) of strong simple games on m_n voters with $m_n \to \infty$. Then we say that G_n leads to* ***social chaos*** *on 3 alternatives if $p_{cycl}(G_n) \to \frac{1}{4}$, as $n \to \infty$.*

Now let's work a bit with this definition. There are 8 possible asymmetric relations and exactly 6 of them are rational. Symmetry considerations imply that the probabilities that each of the 6 rational relations will happen is equal and the probabilities the each of the 2 cyclic relations will happen is equal also. Therefore if the sequence (G_n) leads to social chaos then $p_{cycl}(G_n) \to \frac{1}{4}$ and from the above comments for any asymmetric relation R the probability $p_R(G_n)$ that the neutral S.W.F. based on G_n will lead to R tends to $\frac{1}{8}$, i.e. $p_R(G_n) \to \frac{1}{8}$ for every asymmetric relation R on 3 alternatives. Of course it goes also the other way around. Therefore:

Proposition 9.3. *For a sequence (G_n) on m_n voters with $m_n \to \infty$ of strong simple games, (G_n) leads to social chaos on 3 alternatives if and only if for every asymmetric relation R on 3 alternatives, $p_R(G_n) \to \frac{1}{8}$.*

We use Proposition 9.3 to generalise on m alternatives. On m alternatives there are $2^{\binom{m}{2}}$ asymmetric relations.

Definition 9.4. *We say that a sequence (G_n) of strong simple games on m_n voters with $m_n \to \infty$ leads to social chaos on m alternatives if and only if for every asymmetric relation R on m alternatives, the probability that the neutral welfare function based on (G_n) on m alternatives is equal to R tends to $2^{-\binom{m}{2}}$, i.e. $p_R(G_n) \to 2^{-\binom{m}{2}}$.*

Quite surprisingly it turns out that for a sequence of strong simple games (G_n), (G_n) is leading to social chaos on 3 alternatives if and only if it is leading to social chaos on m alternatives, for every $m \geq 3$. We will come back to this later.

9.2 The introduction of Fourier analysis and an important result

Fourier analysis on the hypercube is actually the key tool we need to understand the quantity $p_{cycl}(G)$ for a strong game G.

Let remind ourselves how F the neutral S.W.F. based on G acts on 3 alternatives $\{a, b, c\}$. Assume that our society has n voters. F receives n asymmetric relations $R_1, .., R_n$ that we assume are rational and leads to an asymmetric relation R such that aRb if and only if the set $\{i \in [n] | aR_ib\}$ is a winning coalition for G. Now to use Fourier analysis we make the following observation: We can assume that the asymmetric relations $R_1, .., R_n$ are described by $x, y, z \in \Omega_n$ such that $x_i = 1$ if and only if aR_ib , $y_i = 1$ if and only if bR_ic and $z_i = 1$ if and only if cR_ia. Moreover we set T to be the subset of the 3n-dimensional hypercube of vertices that arises from a rational profile, i.e. from x, y, z such that for every $i = 1, 2, .., n$:

$$(x_i, y_i, z_i) \notin \{(-1, -1, -1), (1, 1, 1)\}.$$

Now we can see that to define $p_{cycl}(G)$ we assumed a uniformly random profile. By giving the uniform measure on the 3n-dimensional hypercube that is equivalent with assuming that (x, y, z) is uniformly chosen from T. Note that under this measure $\mathbb{P}(T) = \left(\frac{6}{8}\right)^n$.

Therefore we have defined our probability in the context of the uniform random measure on the hypercube and hence we can use Fourier analysis.

Theorem 9.5. *([Kal02])*

For a strong simple game G based on Boolean function f,

$$p_{cycl}(G) = \frac{1}{4} - \frac{1}{4} \cdot \sum_{\varnothing \neq S \subset [n]} \widehat{f}(S)^2 (-\frac{1}{3})^{|S|-1}$$

Proof. We start with a claim:

Claim 9.6. *With the set T defined above:*

$$p_{cycl}(G) = \mathbb{P}(T)^{-1} \cdot \sum_{(x,y,z) \in T} 2^{-3(n+1)} \left((1 - f(x))(1 - f(y))(1 - f(z)) + (f(x) + 1)(f(y) + 1)(f(z) + 1)\right).$$

Proof. Denote by A the set of all (x, y, z) in the 3n-dimensional cube that lead to a cyclic asymmetric relation under the neutral S.W.F. based on G. Note that we assumed that S.W.F. receives as inputs only rational profiles but we may assume for purposes of this proof that that it receives also non-rational individual profiles. The way that it chooses its' outcome remains the same.

Then:

$$\mathbb{P}(T) \cdot p_{cycl}(G) = \mathbb{P}(T \cap A) = 2^{-3n} \cdot \sum_{(x,y,z) \in T} 1((x, y, z) \in A)$$

,where $1(A)$ is the indicator function for the set A.
Since the game G is based on f one can check that:
$(x, y, z) \in T$ leads to a cyclic asymmetric relation if and only if $f(x) = f(y) = f(z)$.

Therefore since f is a Boolean function it holds:

$$8 \cdot 1((x, y, z) \in A) = ((1 - f(x))(1 - f(y))(1 - f(z)) + (f(x) + 1)(f(y) + 1)(f(z) + 1))$$

and the proof is complete. □

Since $(x, y, z) \in T$ if and only if for every $i = 1, .., n$: (x_i, y_i, z_i) takes values different from $(1, 1, 1)$ and $(-1, -1, -1)$ we have that $\mathbb{P}(T) = (\frac{6}{8})^n$.

Therefore using Claim 9.6:

$$p_{cycl}(G) = \mathbb{P}(T)^{-1} \cdot \sum_{(x,y,z) \in T} 2^{-3(n+1)} \left((1 - f(x))(1 - f(y))(1 - f(z)) + (f(x) + 1)(f(y) + 1)(f(z) + 1)\right)$$

$$= \mathbb{P}(T)^{-1} \cdot \sum_{(x,y,z) \in T} 2^{-3(n+1)} 2 \cdot (1 + f(x)f(y) + f(y)f(z) + f(z)f(x)) = \frac{1}{4} + 6 \cdot (\frac{8}{6})^n \cdot \sum_{(x,y,z) \in T} 2^{-3(n+1)} f(x)f(y),$$

and therefore:

$$p_{cycl}(G) = \frac{1}{4} + (\frac{8}{6})^{n-1} \cdot \sum_{(x,y,z) \in T} 2^{-3n} f(x)f(y). \tag{9.1}$$

So we just need to compute the quantity:

$$\sum_{(x,y,z) \in T} 2^{-3n} f(x)f(y).$$

Set $g = 1(T)$ the indicator function for the event T and $h(x, y, z) = f(x)f(y)$. Then,

$$\sum_{(x,y,z) \in T} 2^{-3n} f(x)f(y) = \mathbb{E}_{(x,y,z)}[h(x, y, z) \cdot g(x, y, z)] = \sum_{U \subset [3n]} \widehat{h}(U) \cdot \widehat{g}(U)$$

(remember that Fourier-Walsh expansion was defined for every real function defined on Ω_n , not just Boolean functions.)

Given the above we only need to find the Fourier coefficients of g, h. Note that every $U \subset [3n]$ can be partitioned uniquely into 3 subsets of $[n]$,S_1, S_2, S_3 corresponding to the variables x, y, z respectively.

-For h: It is a straightforward calculation that $\widehat{h}(U) = \widehat{h}(S_1, S_2, S_3) = \widehat{f}(S_1)\widehat{f}(S_2)$ if $S_3 = \varnothing$ and zero otherwise.

-For g: Observe that g has also a multiplicative structure: asking for $(x, y, z) \in T$ is something that depends independently on each of the n triples (x_i, y_i, z_i). Therefore $T = T_n$ is the Cartesian product of n-copies of the set $T_1 = \{(x, y, z) \in \Omega_3 \Big| (x, y, z) \neq (1, 1, 1) \text{ or } -(1, 1, 1)\}$.

But for $g_1 = 1(T_1)$ we can compute easily the Fourier coefficients: $\widehat{g_1}(\varnothing) = \frac{3}{4}, \widehat{g_1}(U) = -\frac{1}{4}$ for $|U| = 2$ and zero otherwise.

It follows that the $\widehat{g}(S_1, S_2, S_3)$ is a product of n expressions, one for each $i = 1, 2, .., n$,as from independence $\widehat{s(x)t(y)} = \widehat{s}(x)\widehat{t}(y)$ for function s, t with different ranges.

Therefore the contribution of i in the coefficient is $\frac{3}{4}$ if $i \notin S_1 \cup S_2 \cup S_3$ and $-\frac{1}{4}$ if it belongs to exactly two of the S_i.

Otherwise it contributes zero.

In summary, $\widehat{g}(S_1, S_2, S_3) = 0$ unless every index i belongs to zero or two of the sets S_i.

In that special case: $\widehat{g}(S_1, S_2, S_3) = (-\frac{1}{4})^{|S_1 \cup S_2 \cup S_3|} \cdot (\frac{3}{4})^{n-|S_1 \cup S_2 \cup S_3|}$.

Combining the above for the product $\widehat{h}(U) \cdot \widehat{g}(U)$ to be non-zero, if $U = (S_1, S_2, S_3)$, we must have that $S_3 = \varnothing$ and $S_1 = S_2 = S$. It follows that:

$$\sum_{(x,y,z)\in T} 2^{-3n} f(x)f(y) = \sum_{S\subset[n]} \widehat{f}(S)^2(-\frac{1}{4})^{|S|} \cdot (\frac{3}{4})^{n-|S|}$$

and by (9.1) :

$$p_{cycl}(G) = \frac{1}{4} - \frac{1}{4} \cdot \sum_{S\subset[n]} \widehat{f}(S)^2(-\frac{1}{3})^{|S|-1},$$

as we wanted. □

Remark 9.7. *Kalai actually used Theorem 9.5 to give a* ***new proof for Theorem 9.1*** *in [Kal02]. The proof is not difficult after this step. Using Cauchy Schwartz one can show that $p_{cycl}(G) = 0$ implies that f should have it's spectral mass only on sets with 0 or 1 elements. But one can prove that this implies that f is a dictatorship ($f(x) = x_1$) , anti-dictatorship ($f(x) = -x_1$) or constant. Now since f satisfies $f(1, 1, .., 1) = 1$ and $f(-(1, 1, ., 1)) = -1$ by assumptions for a Boolean function that characterizes a strong simple game, we get the desired result.*

9.3 Noise sensitivity and social chaos

We are finally in the position to discuss the relation between noise sensitivity of a sequence of Boolean functions and the property of social chaos. Let (G_n) be a sequence of strong simple games. Name P_3 the property for a sequence (G_n) to bring social chaos on 3 alternatives. This property as we said in the first subsection means that every asymmetric relation on 3 alternatives is asymptotically equi-probable for the neutral Social Welfare Functions arising from (G_n) given a random uniform profile. Equivalently, based on Proposition 9.3, we ask for the probability that G_n will lead to a cyclic asymmetric relation to tend to $\frac{1}{4}$.

We will prove that social chaos on 3 alternatives and noise sensitivity are actually equivalent.

Theorem 9.8. *Let (G_n) be a sequence of strong simple games with m_n voters with $m_n \to \infty$. Let $f_n, n \in \mathbb{N}$ be (G_n)'s corresponding Boolean function. Then (G_n) leads to social chaos on 3 alternatives if and only if the sequence f_n is noise-sensitive.*

Proof. We need a lemma first:

Lemma 9.9. *If $f : \Omega_n \to \{-1, 1\}$ is a Boolean function describing a strong simple game then $\widehat{f}(S) = 0$ whenever $|S|$ is an even integer.*

Proof. The fact that the game is strong yields $f(-x) = -f(x)$ for every $x \in \Omega_n$. Therefore if $S \subset [n]$ has even number of elements the contribution from x and $-x$ to $f(x)\chi_S$ cancels out and we get the desired result.

□

Back to our original proof: We know that (G_n) leads to social chaos on 3 alternatives if and only if

$$p_{cycl}(G_n) \to \frac{1}{4}.$$

But by applying Theorem 9.5 and Lemma 9.11 we get that this is equivalent to:

$$\sum_{\varnothing \neq S \subset [m_n]} \widehat{f_n}(S)^2 (\frac{1}{3})^{|S|-1} \to 0.$$

But relation (4.1) yields:

$$\mathrm{Cov}(f_n(x), f_n(x_\varepsilon)) = \sum_{\varnothing \neq S \subset [m_n]} \widehat{f}(S)^2 (1-\varepsilon)^{|S|}$$

for every $\varepsilon \in (0,1)$.

Therefore for the sequence f_n and $\varepsilon = \frac{2}{3}$ we know that

$$\mathrm{Cov}(f_n(x), f_n(x_{\frac{2}{3}})) = \sum_{\varnothing \neq S \subset [m_n]} \widehat{f_n}(S)^2 (\frac{1}{3})^{|S|-1} \to 0..$$

Hence G_n leads to social chaos on 3 alternatives if and only if

$$\mathrm{Cov}(f_n(x), f_n(x_{\frac{2}{3}})) \to 0.$$

But Proposition 4.7 says that a sequence is noise sensitive if and only if $\mathrm{Cov}(f_n(x), f_n(x_\varepsilon)) \to 0$ for one $\varepsilon \in (0,1)$ and therefore our proof is complete.

□

Now one can ask what happens with the social chaos on m alternatives. Is that relevant to noise sensitivity as well? Kalai answered affirmative to this. Name P_m the property for a sequence (G_n) of strong simple games to bring social chaos on m alternatives.

He proved that:

Theorem 9.10. *(Noise sensitive implies P_m)*

Let (G_n) be a sequence of strong simple games on m_n voters with $m_n \to \infty$ and f_n their corresponding Boolean functions. Fix also a natural number $m \geq 3$. Then the following holds:

if the sequence f_n is noise sensitive then (G_n) leads to social chaos on m alternatives, i.e. (G_n) satisfies P_m.

and also that,

Theorem 9.11. *(P_m implies P_3)*
Let (G_n) be a sequence of strong simple games on m_n voters with $m_n \to \infty$. If (G_n) leads to social chaos on m alternatives for some $m \geq 3$ then it leads also to social chaos on 3 alternatives.

The proof of Theorem 9.11 is not difficult and it can be proved based on the fact that every random uniform profile on m alternatives implies a random uniform profile on any subset of the alternatives as well.

The proof of Theorem 9.10 though is more involved. The proof firstly observes that it is enough to prove that if two asymmetric relations differ on exactly one pair of alternatives then they are asymptotically equi-probable and then uses a coupling method to deduce this from noise sensitivity. We will not mention more about it here but the details can be found in [Kal10].

Combining now Theorem 9.9, Theorem 9.10 and Theorem 9.11 we get the desired equivalence result:

Theorem 9.12. *([Kal10]) Let (G_n) be a sequence of strong simple games with m_n players and $m_n \to \infty$ and let f_n be the Boolean function describing G_n for every $n \in \mathbb{N}$. Then for any $m \geq 3$ the following are equivalent:*

1)The sequence f_n is noise sensitive.

2)The sequence (G_n) satisfies the property P_3, i.e. leads to social chaos on 3 alternatives.

2)The sequence (G_n) satisfies the property P_m, i.e it leads to social chaos on m alternatives.

Observe that as a corollary we get that for any $m \geq 3$, P_m is equivalent with P_3.

9.4 The example of ternary majority

It is an interesting question whether there exists a reasonable sequence of realistic voting games that leads to social chaos or equivalently from Theorem 9.12 a reasonable sequence of Boolean function corresponding to realistic voting games that is noise sensitive.

It turns out that there is a natural example that leads to social chaos and this is the example of Ternary Majority (example 5 at subsection 2.1). As we said at subsection 2.1 Simple Majority rule besides it's superficial similarity with Ternary Majority rule it has fundamental difference. This is the fact that the former produces a noise stable sequence of Boolean function while the latter ,as we will prove here, produces a noise sensitive sequence of Boolean functions and therefore leads to social chaos.

Let's remind ourselves how the sequence of Boolean functions $t_n : \Omega_{3^n} \to \{-1, 1\}$ of Ternary Majority with 3^n players is defined. We will define them inductively.

For $n = 1$ it is just the simply majority rule with 3 voters. Suppose we have defined it for $n - 1$ and let's define it for n. Assume we have a society S with 3^n members who vote between two candidates. To find the result of the election we act as following :

Before they vote we order them. Then they vote and we receive an element $x \in \Omega_{3^n}$. Then we apply simple majority rule at the 3^{n-1} triples $(3k-2, 3k-1, 3k)$ where $k = 1, .., 3^{n-1}$. In this way we are getting an element $y \in \Omega_{3^{n-1}}$. Then we set $t_n(x) = t_{n-1}(y)$ meaning that we apply the same ternary majority method again but in this case for votes given by the coordinates of y like we had a society of 3^{n-1} members. Note that t_n satisfies $t_n(1, 1, .., 1) = 1$ and $t_n(-x) = -t_n(x)$ for every $x \in \Omega_n$. Therefore for every n we can define the G_n the strong simple game described by the Boolean function t_n. Then the following theorem holds:

Theorem 9.13. *The sequence t_n of Boolean functions on ternary majority is noise sensitive and therefore their corresponding sequence (G_n) of strong simple games leads to social chaos.*

Proof. Since t_n satisfies $t_n(-x) = -t_n(x)$ for every $x \in \Omega_{3^n}$, it holds $\mathbb{E}[t_n] = 0$ for every $n \in \mathbb{N}$. Therefore as t_n are Boolean functions $\mathrm{Var}(t_n) = 1$ for every $n \in \mathbb{N}$. That means from Proposition 4.5 that to prove noise sensitivity of t_n we need to prove that for a fixed $\varepsilon \in (0, 1)$,

$$\mathbb{P}(t_n(x_\varepsilon) \neq t_n(x)) \to \frac{1}{2}.$$

Claim 9.14. *Let the function $v : \mathbb{R} \to \mathbb{R}$ given by $v(z) = z^3 - \frac{3}{2}z^2 + \frac{3}{2}z$ and fix $\varepsilon \in (0, 1)$. Then for every $n \in \mathbb{N}$,*

$$\mathbb{P}(t_n(x_\varepsilon) \neq t_n(x)) = v^{(n)}(\frac{\varepsilon}{2}),$$

where by $v^{(n)}$ we mean n compositions of v.

Proof. We apply induction on n. For $n = 1$ we compute the Fourier coefficients for simple majority with 3 voters and using relation (4.2) from subsection 4.1 we get :

$$\mathbb{P}(t_n(x_\varepsilon) \neq t_n(x)) = \frac{\varepsilon^3}{8} - \frac{3}{8}\varepsilon^2 + \frac{3}{4}\varepsilon = v(\frac{\varepsilon}{2}),$$

as we wanted. For the inductive step:

Assume that the claim is true for all natural numbers less than n.

Take a random input $x \in \Omega_{3^n}$. Then $f(x)$ is the simply majority for the triple (z_1, z_2, z_3) that arise after we apply the ternary majority rule for the voters in the three families,

$$F_1 = \{i \in [3^n] \Big| 1 \leq i \leq 3^{n-1}\},$$

$$F_2 = \{i \in [3^n] \Big| 3^{n-1} + 1 \leq i \leq 2 \cdot 3^{n-1}\},$$

and

$$F_3 = \{i \in [3^n] \Big| 2 \cdot 3^{n-1} \leq i \leq 3^n\}.$$

Then after applying noise $\varepsilon \in (0, 1)$ to the coordinates of x, the probability that z_i will be flipped is by induction equal to $v^{(n-1)}(\frac{\varepsilon}{2})$. Therefore since it is easy to prove that v sends $(0, 1)$ to $(0, 1)$ the situation we get is that we are applying simple majority rule to the 3 votes z_1, z_2, z_3 whom choice is flipped with probability $v^{(n-1)}(\varepsilon)$. Therefore the final outcome will be reversed based on the case $n = 1$ with probability exactly equal to $v(v^{(n-1)}(\frac{\varepsilon}{2})) = v^{(n)}(\frac{\varepsilon}{2})$. The claim is proven. □

Based on claim 9.45 we need to prove that for $\varepsilon \in (0, 1)$, $v^{(n)}(\frac{\varepsilon}{2}) \to \frac{1}{2}$ where $v(z) = z^3 - \frac{3}{2}z^2 + \frac{3}{2}z$. Fix $\varepsilon \in (0, 1)$. Observe that for $z \in (0, \frac{1}{2})$, it is true that $v(z) > z$ and that v is increasing on the whole $\mathbb{R}$. Therefore $v\left((0, \frac{1}{2})\right) \subset (0, \frac{1}{2})$ since $v(0) = 0$ and $v(\frac{1}{2}) = \frac{1}{2}$ and therefore based on $v(z) > z$ for $z \in (0, \frac{1}{2})$, we derive that $v^{(n)}(\frac{\varepsilon}{2})$ is an increasing sequence. Assume the limit of the sequence equals to $\alpha \in (0, \frac{1}{2}]$. But then since v is continuous it must hold $v(\alpha) = \lim_n v^{(n+1)}(\frac{\varepsilon}{2}) = \alpha$ which gives $\alpha \in \{0, \frac{1}{2}, 1\}$. As $\alpha \in (0, \frac{1}{2}]$ we get $\alpha = \frac{1}{2}$ as we wanted.

The proof is complete.

□

References

[BJT99] Nader H. Bshouty, Jeffrey C. Jackson, and Christino Tamon. Uniform-distribution attribute noise learnability. In *Proceedings of the Twelfth Annual Conference on Computational Learning Theory (Santa Cruz, CA, 1999)*, pages 75–80 (electronic). ACM, New York, 1999.

[BKS99] Itai Benjamini, Gil Kalai, and Oded Schramm. Noise sensitivity of Boolean functions and applications to percolation. *Inst. Hautes Études Sci. Publ. Math.*, (90):5–43 (2001), 1999.

[Bou02] J. Bourgain. On the distributions of the Fourier spectrum of Boolean functions. *Israel J. Math.*, 131:269–276, 2002.

[DFKO07] Irit Dinur, Ehud Friedgut, Guy Kindler, and Ryan O'Donnell. On the Fourier tails of bounded functions over the discrete cube. *Israel J. Math.*, 160:389–412, 2007.

[GPS11] Christophe Garban, Gábor Pete, and Oded Schramm. The Fourier spectrum of critical percolation. In *Selected works of Oded Schramm. Volume 1, 2*, Sel. Works Probab. Stat., pages 445–530. Springer, New York, 2011.

[GS12] Christophe Garban and Jeffrey E. Steif. Noise sensitivity and percolation. In *Probability and statistical physics in two and more dimensions*, volume 15 of *Clay Math. Proc.*, pages 49–154. Amer. Math. Soc., Providence, RI, 2012.

[Kal02] Gil Kalai. A Fourier-theoretic perspective on the Condorcet paradox and Arrow's theorem. *Adv. in Appl. Math.*, 29(3):412–426, 2002.

[Kal10] Gil Kalai. Noise sensitivity and chaos in social choice theory. In *Fete of combinatorics and computer science*, volume 20 of *Bolyai Soc. Math. Stud.*, pages 173–212. Jánosa Bolyai Math. Soc., Budapest, 2010.

[Kes80] Harry Kesten. The critical probability of bond percolation on the square lattice equals $\frac{1}{2}$. *Comm. Math. Phys.*, 74(1):41–59, 1980.

[Kho02] Subhash Khot. On the power of unique 2-prover 1-round games. In *Proceedings of the Thirty-Fourth Annual ACM Symposium on Theory of Computing*, pages 767–775. ACM, New York, 2002.

[KK13] Nathan Keller and Guy Kindler. Quantitative relation between noise sensitivity and influences. *Combinatorica*, 33(1):45–71, 2013.

[KKL88] Jeff Kahn, Gil Kalai, and Nathan Linial. The influence of variables on boolean functions (extended abstract). pages 68–80, 1988.

[KKMO07] Subhash Khot, Guy Kindler, Elchanan Mossel, and Ryan O'Donnell. Optimal inapproximability results for MAX-CUT and other 2-variable CSPs? *SIAM J. Comput.*, 37(1):319–357, 2007.

[LSW02] Gregory F. Lawler, Oded Schramm, and Wendelin Werner. One-arm exponent for critical 2D percolation. *Electron. J. Probab.*, 7:no. 2, 13 pp. (electronic), 2002.

[MOO10] Elchanan Mossel, Ryan O'Donnell, and Krzysztof Oleszkiewicz. Noise stability of functions with low influences: invariance and optimality. *Ann. of Math. (2)*, 171(1):295–341, 2010.

[O'D03] Ryan William O'Donnell. *Computational applications of noise sensitivity*. ProQuest LLC, Ann Arbor, MI, 2003. Thesis (Ph.D.)–Massachusetts Institute of Technology.

[SS11] Oded Schramm and Jeffrey E. Steif. Quantitative noise sensitivity and exceptional times for percolation [mr2630053]. In *Selected works of Oded Schramm. Volume 1, 2*, Sel. Works Probab. Stat., pages 391–444. Springer, New York, 2011.

[SW01] Stanislav Smirnov and Wendelin Werner. Critical exponents for two-dimensional percolation. *Math. Res. Lett.*, 8(5-6):729–744, 2001.

[Tal96] Michel Talagrand. How much are increasing sets positively correlated? *Combinatorica*, 16(2):243–258, 1996.

[Wer09] Wendelin Werner. Lectures on two-dimensional critical percolation. In *Statistical mechanics*, volume 16 of *IAS/Park City Math. Ser.*, pages 297–360. Amer. Math. Soc., Providence, RI, 2009.