Noise Stability versus Pivotals and Sparse Reconstruction of Functions in Spins Systems

Pál Galicza

A thesis presented for the degree of
Doctor of Philosophy in Mathematics

Department of Mathematics and its Applications
Central European University
Budapest, Hungary
September 30, 2019

Contents

Abstract

In Chapter 1 we give a concise introduction into the analysis of Boolean functions. Several equivalent definitions of noise sensitivity are discussed. We highlight the complex relation between noise sensitivity/stability and the pivotal set. In particular we construct a noise stable sequence of monotone, transitive Boolean functions which has many pivotals with high probability.

In Chapter 2 we introduce the central concept of our thesis: For a sequence of functions $f_n : \{-1,1\}^{V_n} \longrightarrow \mathbb{R}$ defined on increasing configuration spaces we talk about sparse reconstruction if there is a sequence of subsets $U_n \subseteq V_n$ of coordinates satisfying $U_n = o(V_n)$ such that knowing the coordinates of U_n gives us some information about the value of f_n.

We first show that if the underlying measure is a product measure, then for transitive functions no sparse reconstruction is possible. We discuss the question with an L_2-type and an information theoretic concept of information. Furthermore we show that the left-right crossing event for critical planar percolation on the square lattice does not admit sparse reconstruction either.

Chapter 3 extends the question of sparse reconstruction to some larger classes of sequences of measures. We find that if the average correlation of spins in a sequence of spin systems decays slower then $1/m_n$, where m_n is the size of the coordinate set, then sparse reconstruction is possible. We also investigate the question for sequences converging to a finitary factor of IID system and we find that the expected coding volume plays a crucial role in determining whether there is sparse reconstruction or not.

Finally, we apply our results and methods to investigate Ising models on sequences of locally convergent graphs. We show that there is sparse reconstruction for low temperature and critical Ising models, and that there is no sparse reconstruction on the high temperature Curie-Weiss model.

Chapter 1

Noise Sensitivity and the Pivotal Set

1.1 Introduction to Noise Sensitivity and Noise Stability

Noise Sensitivity for Boolean functions was introduced in the seminal work of Benjamini, Kalai and Schramm [BKS99]. It was originally created to understand the behavior of crossing events for critical Bernoulli percolation, but it turned out to be of interest on its own right.

A sequence of Boolean functions is called noise sensitive, if distorting each input bit with any fixed small probability asymptotically destroys all information on the original value of f_n.

Definition 1.1.1 (Non-degenerated sequences)**.** A sequence of Boolean functions $f_n : \{-1,1\}^{V_n} \longrightarrow \{-1,1\}$ is called non-degenerated if there exists an $\epsilon > 0$ such that for all n

$$-(1-\epsilon) < \mathbb{E}[f_n] < 1-\epsilon.$$

Definition 1.1.2 (Noise Sensitivity)**.** Let ϵ be a positive real number. For a uniform random vector $\omega \in \{-1,1\}^{k_n}$ denote $N_\epsilon(\omega)$ the random vector which we obtain from ω by resampling each of its bits independently with probability ϵ. A sequence of non-degenerated functions $f_n : \{-1,1\}^{k_n} \longrightarrow \{-1,1\}$ is noise sensitive if and only if for every $\epsilon > 0$

$$\lim_{n\to\infty} \frac{\mathrm{Var}(\mathbb{E}[f_n \mid N_\epsilon(\omega)])}{\mathrm{Var}(f_n)} = 0 \tag{1.1.1}$$

The notion of noise sensitivity has applications in complexity and social choice theory. A Boolean function $f : \{-1,1\}^n \longrightarrow \{-1,1\}$ can be interpreted as a voting scheme or an aggregation rule. Each coordinate stands for a voter and the values -1 or 1 represents a choice between two alternatives. f may be seen as rule telling how the individual votes aggregate to a group decision. In this set up noise sensitivity of a voting system means that even a small ϵ ratio of counting mistakes in the voting has reasonable chance to turn the outcome. See [K05] for applications of noise sensitivity in the Social Choice Theory setting.

Remark 1.1.1. The usual definition of noise sensitivity is slightly different. It states that the expected correlation between f_n and f_n applied to the noisy input decays to 0 as n approaches to infinity. It is easy to see that the two definitions are equivalent see Theorem 1.1.4. We have a preference for this form since it features the notion of clue (Definition 2.1.1). (1.1.1) states that $\lim_n \mathsf{clue}(f_n \mid N_\epsilon(\omega)) = 0$

Note that the above definition naturally extends to $\mathbb{R}$-valued functions requiring asymptotic independence. (For binary-valued function decorrelation implies independence.)

A sort of opposite case of noise sensitivity is a sequence where the value of the function is, under a small amount of noise, highly correlated with the value of the noisy version. This phenomenon is expressed by the notion of noise stability:

Definition 1.1.3 (Noise Stability)**.** A sequence of functions $f_n : \{-1,1\}^{k_n} \longrightarrow \{-1,1\}$ is noise stable if and only if

$$\lim_{\epsilon\to 0} \sup_n \mathbb{P}[f_n(\omega) \neq f_n(N_\epsilon(\omega))] = 0 \tag{1.1.2}$$

1.1.1 The Fourier-Walsh expansion

We introduce a function transform for functions on the hypercube which is widely used in the analysis of Boolean functions.

Definition 1.1.4 (Fourier-Walsh expansion)**.** For any $f \in L_2(\{-1,1\}^V)$ and $\omega \in \{-1,1\}^V$

$$f(\omega) = \sum_{S\subset V} \widehat{f}(S), \chi_S(\omega) \quad \chi_S(\omega) := \quad ()i \in S\omega_i \text{ (and } \chi_S(\emptyset) := 1). \tag{1.1.3}$$

This is in fact the Fourier transform, the event space naturally identified with the group $\mathbb{Z}_2^V$ by assigning a generator g_x to every $x \in V$. The functions χ_S are the characters of $\mathbb{Z}_2^V$

Let us introduce the natural inner product $(f,g) = \mathbb{E}[fg]$ on the space of real functions on the hypercube. It is straightforward to check that the functions χ_S form an orthonormal basis with respect to this inner product. We will now state a few important consequences of this fact.

Since the Fourier-Walsh transform, as it is straightforward to check is an orthonormal transformation with respect to the standard inner product, Parseval's formula applies and therefore

$$\sum_{S\subseteq V} \widehat{f}(S)^2 = \|f\|^2 .$$

Noting that $\widehat{f}(\emptyset) = \mathbb{E}[f]$, we also have

$$\mathrm{Var}(f) = \sum_{\emptyset\neq S\subseteq V} \widehat{f}(S)^2. \tag{1.1.4}$$

For a subset $T \subseteq V$ let us denote by $\mathcal{F}_T$ the σ-algebra generated by the bits belonging to T. So $\mathcal{F}_T$ expresses knowing the coordinates in T. It turns out that the conditional expectation of any function $f : \{-1,1\}^n \longrightarrow \mathbb{R} \quad S \subset E$ with respect to $\mathcal{F}_T$ can be expressed in terms of the squared Fourier-Walsh expansion coefficients see [GS15]:

$$\mathbb{E}[f \mid \mathcal{F}_T] = \sum_{S\subseteq T} \widehat{f}(S)\chi_S$$

The proof is fairly simple: we only need to observe that if $S \subseteq U$ then $\mathbb{E}[\chi_S \mid \mathcal{F}_U] = \chi_S$ in any other case $\mathbb{E}[\chi_S \mid \mathcal{F}_U] = 0$.

Using (1.1.4) we get a concise spectral expression for the variance of the conditional expectation.

$$\mathrm{Var}(\mathbb{E}[f \mid \mathcal{F}_T]) = \sum_{\emptyset\neq S\subseteq T} \widehat{f}(S)^2 \tag{1.1.5}$$

Fourier-Walsh transformation as eigenbasis of the noise operator

Consider a continuous time simple random walk $\{\omega^t \ : t \in [0,\infty)\}$ on the hypercube. More precisely we have a rate 1 Poisson clock for every $i \in V$, and each time the clock of i rings the bit ω_i switches to $-\omega_i$. It is easy to see that after time t the joint distributions of (ω^0, ω^t) and $(\omega, N_\epsilon(\omega))$ are the same with the conversion $\epsilon = 1 - e^{-t}$.

So on the one hand, we have an interpretation of information theoretic flavour, that we try to compute a piece of information (represented by f) but the input is corrupted with noise. The question is: can we recover the original information?

On the other hand, we have a geometric interpretation: We perform a simple random walk on the discrete hypercube and we have a subset of vertices $\mathcal{A}$ (again represented by f). The question is whether after a fixed amount of time do we remember if we started the walk from $\mathcal{A}$ or not.

One can now think about the Fourier-Walsh expansion in a more probabilistic way: the functions χ_S are the eigenfunctions of the simple random walk on the hypercube or, which is the same, the operator N_ϵ.

Indeed, observe that for any $i \in V$ $\mathbb{E}[N_\epsilon(\omega_i)|\omega_i] = (1-\epsilon)\omega_i$, and using that for any coordinates $i \neq j$ $\mathbb{E}[N_\epsilon(\omega_i)|\omega_i]$ and $\mathbb{E}[N_\epsilon(\omega_j)|\omega_j]$ are independent, we have for any $S \subseteq V$:

$$\mathbb{E}[\chi_S(N_\epsilon(\omega))|\omega] = (1-\epsilon)^{|S|}\chi_S$$

Let us introduce the operator

$$T_\epsilon[f] := \mathbb{E}[f(N_\epsilon(\omega)|\omega]$$

for any $f : \{-1,1\}^n \longrightarrow \mathbb{R}$. Based on the above and the linearity of conditional expectation:

$$T_\epsilon[f] = \sum_{S \subset V} (1-\epsilon)^{|S|}\widehat{f}(S)\chi_S(\omega) \tag{1.1.6}$$

Using now Parseval's formula it is an easy calculation to establish the following spectral description of Noise Sensitivity and Noise Stability.

Theorem 1.1.2. *A sequence of functions $f_n : \{-1,1\}^{V_n} \longrightarrow \mathbb{R}$ is noise sensitive if and only if for any $k \in \mathbb{N}$*

$$\lim_{n\to\infty} \frac{1}{\mathrm{Var}(f_n)} \sum_{0<|S|<k} \widehat{f_n}(S)^2 = 0.$$

and noise stable if and only if for every ϵ there is a large enough k such that

$$\lim_{n\to\infty} \frac{1}{\mathrm{Var}(f_n)} \sum_{|S|>k} \widehat{f_n}(S)^2 < \epsilon.$$

The Spectral Sample

It turns out to be useful to think about the squared Fourier coefficients $\widehat{f}(S)^2$ as a random subset of the spins called the Spectral Sample It is convenient to normalize this measure to get a probability measure. The random subset according to this measure $\mathscr{S}_f$ is called the spectral sample.

Definition 1.1.5 (Spectral Sample)**.** Let $f \in \mathcal{L}_2(\{-1,1\}^V$. The Spectral Sample $\mathscr{S}_f$ of f is a random subset of V chosen according to the distribution

$$\mathbb{P}[\mathscr{S}_f = S] = \frac{\widehat{f}(S)^2}{\|f\|^2} \quad S \subseteq V.$$

The advantage of this concept that it introduces a new and rather compact language, where the concepts that we introduced so far has a straightforward translation. Indeed, noise sensitivity of a sequence of functions is equivalent to the fact that the respective Spectral Measure - the measure corresponding to the Spectral Sample is concentrated on large subsets, while noise stability means that the Spectral Sample is concentrated on bounded subsets.

Theorem 1.1.3 (Noise sensitivity and stability via Spectral Sample)**.** *A sequence of functions* $f_n : \{-1,1\}^{k_n} \longrightarrow \mathbb{R}$ *is*
(1) noise sensitive if conditioned on the events $|\mathscr{S}_{f_n}| \neq 0, \quad |\mathscr{S}_{f_n}| \to \infty$ *in probability,*
(2) noise stable if the sequence $|\mathscr{S}_{f_n}|$ *is tight.*

Another concept that translates very well to the Spectral Sample language is notion of clue. The clue of a function f with respect to a subset of coordinates U, defined as $\mathsf{clue}(f \mid U) = \frac{\mathrm{Var}(\mathbb{E}[f \mid \mathcal{F}_U])}{\mathrm{Var}(f)}$ (see Definition 2.1.1) is one of the central concept of this work. Using (1.1.4) and (1.1.5) we get that

$$\mathsf{clue}(f \mid U) = \mathbb{P}[\mathscr{S}_f \subseteq U \mid \mathscr{S}_f \neq \emptyset]. \tag{1.1.7}$$

This observation is in fact the key step in proving Theorem 2.1.1.

1.1.2 Equivalent Characterisation of Noise Sensitivity

Here we collect a few statements that are equivalent to being Noise Sensitive. These equivalences are fairly easy and implicitly known to the community, but (some of them) have not been explicitly spelled out and it seems to be of some use to include them here.

For better readability, we introduce the shorthand notation $\omega^\epsilon = N_\epsilon(\omega)$.

Theorem 1.1.4. *Let* $f_n : \{-1,1\}^{V_n} \longrightarrow \{-1,1\}$ *be sequence of non-degenerate Boolean functions. The following statements are equivalent*

1. f_n *is noise sensitive.*

2. *For every* $\epsilon > 0$

$$\lim_{n\to\infty} \mathbb{E}[f_n(\omega^\epsilon) f_n(\omega))] - \mathbb{E}[f_n]^2 = 0$$

3. *For every* $p \in (0,1)$ *denoting by* $\mathcal{H}^p$ *the Bernoulli random subset at level* p *(i.e. each* $i \in V$ *is in* $\mathcal{H}^p$ *with probability* p*, independently from what happens to the other elements). Let* $\mathbb{E}[\mathsf{clue}(f_n \mid \mathcal{H}^p)] := \mathbb{E}\left[\frac{\mathrm{Var}(\mathbb{E}[f_n \mid \mathcal{H}^p])}{\mathrm{Var}(f_n)}\right]$ *(See also Definition 2.1.1). With this notation*

$$\lim_{n\to\infty} \mathbb{E}[\mathsf{clue}(f_n \mid \mathcal{H}^p)] = 0.$$

4. *Let $\mathbb{P}_f$ be the uniform measure on $\{f_n = 1\}$ (we suppress the n from the subfix to simplify the notation) and let consider a simple random walk $\{X^t \ : t \in [0,\infty)\}$ on the hypercube with initial distribution $\mathbb{P}_f$. Let us denote by $\mathbb{P}^t_f[\omega]$ the measure according to the distribution of X^t. Then for every $t > 0$*
$$\lim_{n\to\infty} \|\mathbb{P} - \mathbb{P}^t_f\|_1 = 0,$$
where $\| \quad \|_1$ is the total variation distance of measures.

5. *For every $\epsilon \in (0,1)$*
$$\lim_{n\to\infty} \mathbb{E}[f_n(\omega^\epsilon) \mid f_n(\omega) = 1] = \mathbb{E}[f].$$

6. *For every $\epsilon \in (0,1)$*
$$\mathbb{E}[f_n(\omega_\epsilon) \mid \omega] \xrightarrow{p} \mathbb{E}[f].$$

Proof. ($1 \Leftrightarrow 2$) Note that $\mathbb{E}[\chi_S(\omega)\chi_S(\omega^\epsilon)] = \prod_{i\in S}\mathbb{E}[\omega(i)\omega^\epsilon_i] = (1-\epsilon)^{|S|}$. Consequently, for general functions, using that $\mathbb{E}[\chi_{S_1}(\omega)\chi_{S_2}(\omega^\epsilon)] = 0$ whenever $S_1 \neq S_2$ we have the following formula
$$\mathbb{E}[f(\omega)f(\omega^\epsilon)] - \mathbb{E}[f_n]^2 = \sum_{\emptyset\neq S\subset V} \widehat{f}(S)^2\mathbb{E}[\chi_S(\omega)\chi_S(\omega^\epsilon)] = \sum_{S\subset V} \widehat{f}(S)^2(1-\epsilon)^{|S|}.$$
At the same time (1.1.6) shows that
$$\mathrm{Var}(\mathbb{E}[f_n(\omega^\epsilon) \mid \omega]) = \sum_{\emptyset\neq S\subset V} \widehat{f}(S)^2(1-\epsilon)^{2|S|}.$$
Since (ω^ϵ,ω) has the same distribution as (ω,ω^ϵ) we have $\mathrm{Var}(\mathbb{E}[f_n(\omega^\epsilon) \mid \omega]) = \mathrm{Var}(\mathbb{E}[f_n(\omega) \mid \omega^\epsilon])$. By assumption f_n is non-degenerated, therefore $1/\mathrm{Var}(f_n)$ is just a constant factor and the equivalence follows.

($1 \Rightarrow 3$)
$$\mathbb{E}[\mathrm{Var}(\mathbb{E}[f \mid \mathcal{H}^p])] = \mathbb{E}[\sum_{\emptyset\neq S\subset V} \widehat{f}(S)^2\mathbb{1}_{S\subseteq\mathcal{H}^p}] =$$
$$\sum_{\emptyset\neq S\subset V} \widehat{f}(S)^2\mathbb{P}[S \subseteq \mathcal{H}^p] = \sum_{\emptyset\neq S\subset V} \widehat{f}(S)^2 p^{|S|}$$
Using $p = 1-\epsilon$ and noting again that the variance of f_n is of constant order, we get the desired equivalence.

($1 \Rightarrow 4$)

Observe that the Radon-Nykodim derivative $\frac{d\mathbb{P}_f}{d\mathbb{P}}$ is $\frac{2}{\mathbb{E}[f_n]+1}$ if $f_n(\omega) = 1$ and 0 otherwise so for any $\omega \in \{-1,1\}^{V_n}$
$$\frac{d\mathbb{P}_f}{d\mathbb{P}}(\omega) = \frac{f_n(\omega)+1}{\mathbb{E}[f_n]+1}.$$
Similarly,
$$\frac{d\mathbb{P}^t_f}{d\mathbb{P}}(\omega) = \frac{\mathbb{E}[f_n(\omega^t)+1 \mid \omega]}{\mathbb{E}[f_n]+1}.$$

So we can write

$$\|\mathbb{P}[\omega]-\mathbb{P}_f^t\|_1=\sum_{\omega\in\{-1,1\}^{V_n}}|\mathbb{P}[\omega]-\mathbb{P}_f^t[\omega]|=\frac{1}{2^{|V_n|}}\sum_{\omega\in\{-1,1\}^{V_n}}\left|1-\frac{\mathbb{E}[f_n(\omega^t)\mid\omega]+1}{\mathbb{E}[f_n]+1}\right|.$$

Using the Cauchy-Schwarz inequality we get

$$\|\mathbb{P}[\omega]-\mathbb{P}_f^t\|_1\le\left\|1-\frac{\mathbb{E}[f_n(\omega^t)\mid\omega]+1}{\mathbb{E}[f_n]+1}\right\|_2.$$

Now we can use the Fourier-Walsh transform to conclude that

$$\|\mathbb{P}[\omega]-\mathbb{P}_f^t\|_1\le\sqrt{\frac{1}{\mathbb{E}[f_n]+1}\sum_{S\subset V}\widehat{f}(S)^2e^{-2|S|t}}.$$

($4\Rightarrow 5$) Let X_t as before the simple random walk with initial distribution $\mathbb{P}_f$ and t such that $1-e^{-t}=\epsilon$.

$$\mathbb{E}[f_n(\omega^\epsilon)\mid f_n(\omega)=1]=\mathbb{E}[f_n(X_t)\mid f_n(X_0)=1]=\mathbb{P}_f^t[f_n=1]-\mathbb{P}_f^t[f_n=-1].$$

By assumption, for large enough n the total variation distance between $\mathbb{P}_f^t$ and the uniform measure is smaller then δ and therefore

$$|\mathbb{E}[f_n(\omega^\epsilon)\mid f_n(\omega)=1]-\mathbb{E}[f_n]|<2\delta.$$

($5\Rightarrow 6$) Immediately follows as the arguments ($4\Rightarrow 5$) and ($5\Rightarrow 6$) can be repeated with $f_n(\omega)=-1$.

($6\Rightarrow 1$) Since

□

Note that in the usual definition of noise sensitivity, see for example Property 2. ([GS15]) or Property 6. ([BKS99]) from Theorem **??** it is not asymptotic decorrelation, but only asymptotic independence is required. (Ironically, in this special case decorrelation seems stronger then independence).

As the difference is whether we scale down by the variance, this question only effects degenerate sequences. According to the usual definition degenerated sequences are automatically noise sensitive, since asymptotic independence is guaranteed. The intuition behind is that with high probability we know everything about a degenerated sequences of functions and what else a probabilist can ask for? Our point is that it might be meaningful to differentiate between degenerated sequences as well, depending on the speed of decorrelation. Also, defining noise sensitive via covariance or conditional variance is semantically vague, as these notions - in contrast with clue or correlation - are not dimensionless concepts expressing information content.

1.2 Noise Sensitivity versus Pivotals

1.2.1 Pivotal set, Influence and the Spectrum

The influence of Boolean functions, a very natural discrete partial derivative concept had been studied way before noise sensitivity, most notably in [KKL88]. We introduce the main concept and investigate its relationship with the previously introduced concepts.

We are going to use the following notation: For a configuration $\omega \in \{-1,1\}^V$ we denote by ω^j the configuration which is the same as ω except its jth coordinate which is flipped.

Definition 1.2.1 (Pivotal Set)**.** Let $f : \{-1,1\}^V \longrightarrow \{-1,1\}$ and $\omega \in \{-1,1\}^V$. We call a coordinate j pivotal for f with respect to ω if $f(\omega) \neq f(\omega^j)$. The pivotal set $\mathscr{P}_f$ is the (ω-measurable) random set of pivotal coordinates.

Influence of a variable is the probability that it is pivotal.

Definition 1.2.2 (Influence)**.** Let $f : \{-1,1\}^V \longrightarrow \{-1,1\}$ than for an $j \in V$ the influence of the coordinate j is

$$I_j(f) := \mathbb{P}[f(\omega) \neq f(\omega^j)].$$

The total influence is defined as $I(f) := \sum_{j \in V} I_j(f)$

Unsurprisingly the influence also admits a concise formulation in terms of the Fourier-Walsh transform:

$$I_j(f) = \sum_{S:j\in S} \hat{f}^2(S) \quad \text{and} \quad I(f) = \sum_{S:\subseteq V} |S|\hat{f}^2(S) \tag{1.2.1}$$

This is easy to derive using the fact that the pivotal set of χ_S is constant S. A remarkable consequence is the following link between the Spectral Sample and the Pivotal Set $I(f) = \mathbb{E}[|\mathscr{S}_f|]$. On the other hand, by definition $I(f) = \mathbb{E}[|\mathscr{P}_f|]$, so the expected size of these random sets is the same. In fact even more is true.

Proposition 1.2.1. *Let $f : \{-1,1\}^V \longrightarrow \{-1,1\}$ then for every $i,\ j\ \in V$*

$$\Pr[i \in \mathscr{P}_f] = \Pr[i \in \mathscr{S}_f] \quad \textit{and} \quad \Pr[i,j \in \mathscr{P}_f] = \Pr[i,j \in \mathscr{S}_f]$$

For a proof see [GS15] Corollary IX.7. The fact that the one dimensional marginals are equal comes directly from (1.2.1). The equality of the two dimensional marginals is a consequence of a generalization of (1.1.7), the so-called Random Restriction Lemma.

The sudden idea that the two random sets might be the same can be easily discarded as already on three bits one can find counterexamples. Still these observations raises the possibility of characterising Noise Sensitivity and Stability with the help of influences or the pivotal set. According to Theorem 1.1.3, a sequence of function is noise sensitive when the Spectral Sample is typically large, and Proposition 1.2.1 suggests that there might be some connection between the sizes of the Spectral Sample and the Pivotal Set.

Indeed in some well-studied cases the two distributions show similar behavior. For example, this is the case for the crossing event in critical planar percolation. In [GPS10] a thorough analysis of the Fourier expansion shows that most of the Fourier spectrum is typically on sets larger then n^ϵ for some ϵ, while it is known that the pivotal set of the critical crossing event is also typically of polynomial size. See [GPS10] or Chapter X in [GS15] for further details on the similarities and differences between the size of the spectrum and the pivotal set.

Another important example is the Majority on $2n+1$ bits defined as

$$\mathsf{Maj}_{2n+1} = \begin{cases} 1 & \text{if } \sum_i \omega(i) > 0 \\ -1 & \text{if } \sum_i \omega(i) < 0. \end{cases}$$

This function is noise stable, that is, most of its Spectral Sample is concentrated on small (bounded) subsets although the average size of the spectrum is going to infinity in the order of $\sqrt{n}$ (See [OD14]). Similarly, the pivotal set is typically empty since in order to have a pivotal bit one must have $\sum_{i\in n}\omega(i)=\pm 1$.

At the same time this example shows that Spectral Sample does not need to be concentrated, i.e. the knowledge of their expected size is irrelevant for noise sensitivity or stability.

It would be quite useful to infer noise sensitivity via influences. While the Fourier-Walsh transform is a strong theoretical tool, it is very challenging to calculate or even estimate the spectrum of a sequence of Boolean Function (see for example [GPS10], a highly technical paper that estimates the typical size of the Spectral Sample for the crossing event of planar percolation). On the other hand, the influences is usually easier to calculate and in particular, the pivotal set is easy to simulate via a uniformly random string of bits, while there is no known way to sample the spectral sample.

There is, in fact one very important and slightly mysterious result that links noise sensitivity to influences:

Theorem 1.2.2 ([BKS99]). *Let* $f_n : \{-1,1\}^{V_n} \longrightarrow \{-1,1\}$ *if*

$$\lim_{n\to\infty}\sum_{j\in V_n} I_j(f_n)^2 = 0$$

then f_n *is noise sensitive.*

The proof is not long, but rather technical. It uses the method of hypercontractivity, an analytic tool introduced already in [KKL88]. The converse is not true in general, exemplified by χ_{V_n}, but it is for monotone Boolean functions. We note that $\sum_{j\in V_n} I_j(f_n)^2$ in the pivotal set language means the expected size of the intersection of two independent samples of the pivotal set. So we can rephrase Theorem 1.2.2 as follows: if two independent copies of the pivotal set are asymptotically disjoint then the sequence is noise sensitive.

1.2.2 A paradoxical sequence

Apart from Proposition 1.2.1 and Theorem 1.2.2 there is no further known general connection between the behaviour of $\mathscr{S}_f$ and $\mathscr{P}_f$.

Indeed [GS15] Section XII.2 features a number of 'paradoxical' sequences . Among others, a sequence of a noise sensitive sequence of monotone, non-degenerate functions has been constructed for which the pivotal set is empty with high probability or a noise stable sequence which has many pivotals with high probability.

Along these lines the following question was posed by Gil Kalai: Is there a sequence of Boolean functions $f_n : \{-1,1\}^{k_n} \longrightarrow \{-1,1\}$ such that f_n is transitive and noise stable, but at the same time $\mathbb{P}[\mathscr{P}_n(\omega)\neq\emptyset] > c$ for some constant $c>0$ for all $n\in\mathbb{N}$?

Definition 1.2.3 (Transitive function)**.** A function $f : \{-1,1\}^V \longrightarrow \mathbb{R}$ is transitive if there is a transitive group action G on V such that for every $g\in G$ $f^g = f$

In the language of Social Choice Theory the transitivity of a voting scheme can be interpreted as no voter is privileged or treated differently from teh others.

We are going to show that the answer is positive.

Theorem 1.2.3. *There exists a sequence of transitive monotone functions* $f_n : \{-1,1\}^{k_n} \longrightarrow \{-1,1\}$ *such that* f_n *is transitive, noise stable and* $\lim_n \mathbb{P}[\mathscr{P}_n > a_n] = 1$ *(here* $\mathscr{P}_n$ *is the pivotal set, see below) for some sequence of integers* $a_n \to \infty$.

Our result is another indication that, apart from the known connections in general the Spectral Sample and the Pivotal set can show very different behavior.

Remark 1.2.4. One can relax the stability condition to the lack of noise sensitivity. In this case the answer is almost trivial. We now sketch an example of a sequence of monotone functions which is transitive, not noise sensitive and the pivotal set is nonempty with a uniformly positive probability.

Let $A_n, B_n \subseteq \{-1,1\}^{k_n}$ be two sequences of monotone transitive events satisfying the conditions of Theorem 1.2.3 except for noise stability, with the property that there exists $c > 0$ such that for all (large) n, $\mathbb{P}[A_n \cup B_n] - \mathbb{P}[A_n] > c$ (We can, for example, choose A_n and B_n to be two tribes events defined on different tribe partitions). Let Maj_{k_n} be the Majority function on the same k_n bits. Now let

$$f_n = \begin{cases} \mathbb{1}_{A_n} & \text{if } \mathsf{Maj}_{k_n} = -1 \\ \mathbb{1}_{A_n \cup B_n} & \text{if } \mathsf{Maj}_{k_n} = 1. \end{cases}$$

It is clear that f_n is monotone, transitive and admits pivotals with a positive probability. At the same time it is positively correlated with Maj_{k_n} and therefore cannot be noise sensitive.

In the sequel, we shall construct a sequence of functions $f_n : \{-1,1\}^{k_n} \longrightarrow \{-1,0,1\}$ with the following properties:

1. f_n is transitive
2. $\lim_n \mathbb{P}[f_n = 0] = 1$
3. $\lim_n \mathbb{P}[\exists\, i,j \in [k_n] \;:\; f_n(\omega^i) = 1 \text{ and } f_n(\omega^j) = -1] = 1.$

where ω^i denotes ω with its ith coordinate flipped. We will call a sequence of functions bribable if it satisfies the above conditions. The name is coming from the Social Choice Theory interpretation. It is an impartial (transitive) monotone voting scheme (this time with three possible results) with the property that although in most of the times the result is the same, with high probability we can buy some people who can turn the result in a particular direction.

It might be also interesting to think of the bribable sequence geometrically. This is a sequence of invariant and monotone subsets of the hypercube with a density going to 0, but with the property that almost any vertex of hypercube is a neighbour, meaning that it can be reached from the set by changing the value of a single coordinate.

Using a bribable sequence f_n one can easily construct a transitive noise stable Boolean function which admits a pivotal bit with high probability. Namely, let Maj_n denote the majority function on the corresponding bit set. Let

$$g_n = \begin{cases} \mathsf{Maj}_n & \text{if } f_n = 0 \\ f_n & \text{if } f_n \neq 0. \end{cases}$$

Obviously g_n is noise stable because of property 2 of f_n. On the other hand, conditioned on $\{f_n = 0\}$ there is a pivotal bit with high probability because of property 3 of the sequence f_n.

It is also straightforward to verify that if we choose a bribable sequence f_n which is monotone then the resulting g_n sequence will be monotone as well.

Again looking at this from the social choice perspective, this is an impartial, transitive voting scheme, which is noise stable - that is, small random perturbations, such as miscounting or a few (random) people changing there mind in the last moment are not likely to effect the results. However, with high probability there are some powerful voters who can change the result of the voting, if they change their mind.

Construction of a monotone bribable sequence

Now we turn to the construction of a monotone bribable sequence. Define the Boolean function $\mathsf{Tribes}(l,k) : \{-1,1\}^{lk} \longrightarrow \{0,1\}$ as follows: we group the bits in k l-element subsets, these are the so called tribes. The function takes on 1 if there is a tribe T such that for every $i \in T \ : \ \omega(i) = 1$, and 0 otherwise. The Tribes function is standard example, when k_n and l_n are defined in such a way that the function is non-degenerate. It is well know that such a sequence testifies that the Kahn-Kalai-Linial theorem about the maximal influence of sequences of Boolean functions (Theorem 1.14 in [GS15]) is sharp.

We are going to show that in case the two sequences l_n, k_n are properly chosen, a slight modification of $\mathsf{Tribes}(l_n, k_n)$ is bribable.

Proposition 1.2.5. *Suppose that l_n and k_n are sequences such that*

$$\lim_{n\to\infty}\left(1-\frac{1}{2^{l_n}}\right)^{k_n} = 1 \tag{1.2.2}$$

and

$$\lim_{n\to\infty} k_n l_n \frac{1}{2^{l_n}} = \infty \tag{1.2.3}$$

then the sequence of functions $f_n(\omega) := \mathsf{Tribes}(l_n,k_n)(\omega) - \mathsf{Tribes}(l_n,k_n)(-\omega)$ is bribable. Moreover, there is a sequence of positive integers $a_n \to \infty$ such that $\mathbb{P}[|\mathscr{P}_n| > a_n] \to 1$

Proof. Let us call a tribe T pivotal if there is exactly one $j \in T$ such that $\omega(j) = -1$. Define the random variable X_n as the number of pivotal tribes in a configuration. Note that $\mathbb{E}[X_n] = k_n l_n \frac{1}{2^{l_n}}$.

It is clear that conditioned on the event $\{\mathsf{Tribes}(l_n,k_n) = 0\}$ we have $|\mathscr{P}_n| = X_n$, where $|\mathscr{P}_n|$ denotes the pivotal set of $\mathsf{Tribes}(l_n,k_n)$. Consequently, for the respective conditional expected values:

$$\mathbb{E}[\mathscr{P}_n|\mathsf{Tribes}(l_n,k_n) = 0] = \mathbb{E}[X_n|\mathsf{Tribes}(l_n,k_n) = 0].$$

We can write $X_n = \sum_j^{k_n} Y_j$ where Y_j is the indicator of the event that the jth tribe is pivotal. For any $j \in [k_n]$ we have

$$\mathbb{P}[Y_j = 1|\mathsf{Tribes}(l_n,k_n) = 1] = \frac{\mathbb{P}[Y_j = 1]\mathbb{P}[\mathsf{Tribes}(l_n,k_n-1) = 1]}{\mathbb{P}[\mathsf{Tribes}(l_n,k_n) = 1]} \le \mathbb{P}[Y_j = 1],$$

using that if the jth tribe is pivotal and there is a full 1 tribe then the latter is among the remaining $k_n - 1$ tribes. This implies

$$\mathbb{E}[X_n|\mathsf{Tribes}(l_n,k_n) = 1] \le \mathbb{E}[X_n] \le \mathbb{E}[X_n|\mathsf{Tribes}(l_n,k_n) = 0]$$

and therefore

$$\mathbb{E}[\mathscr{P}_n|\mathsf{Tribes}(l_n,k_n)=0] \geq \mathbb{E}[X_n] = k_n l_n \frac{1}{2^{l_n}} \to \infty.$$

As X_n is binomially distributed with $\mathbb{E}[X_n] \to \infty$, being the sum of i.i.d $0-1$-valued random variables, there is a $a_n \to \infty$ such that

$$\lim_{n\to\infty} \mathbb{P}[X_n > a_n] = 1.$$

Note that

$$\mathbb{P}[\mathsf{Tribes}(l_n,k_n)=0] = \left(1-\frac{1}{2^{l_n}}\right)^{k_n}$$

and this probability tends to 1 as n approaches ∞ by our assumption. So clearly

$$\mathbb{P}[X_n > a_n \text{ and } \mathsf{Tribes}(l_n,k_n)=0] = \mathbb{P}[|\mathscr{P}_n| > a_n, \text{ and } \mathsf{Tribes}(l_n,k_n)=0] \to 1$$

and therefore also

$$\lim_{n\to\infty} \mathbb{P}[|\mathscr{P}_n| > a_n \mid \mathsf{Tribes}(l_n,k_n)=0] = 1.$$

The same argument can be repeated for $-\mathsf{Tribes}(l_n,k_n)(-\omega)$. The event that neither $\mathsf{Tribes}(l_n,k_n)(\omega)$ nor $\mathsf{Tribes}(l_n,k_n)(-\omega)$ happens while the pivotal set of both is larger than a_n still holds with high probability. That is, we find pivotal bits for both $\mathsf{Tribes}(l_n,k_n)(\omega)$ and $\mathsf{Tribes}(l_n,k_n)(-\omega)$ with high probability and thus push $f_n = \mathsf{Tribes}(l_n,k_n)(\omega) - \mathsf{Tribes}(l_n,k_n)(-\omega)$ to 1 or -1, respectively.

Furthermore $\mathsf{Tribes}(l_n,k_n)(\omega) - \mathsf{Tribes}(l_n,k_n)(-\omega)$ is monotone increasing as the sum of monotone increasing functions. □

Now it only remains to show that with an appropriate choice of the sequences k_n and l_n (1.2.2) and (1.2.3) are satisfied.

First, note that

$$\left(1-\frac{1}{2^{l_n}}\right)^{k_n} \to 1 \text{ if and only if } \frac{k_n}{2^{l_n}} \to 0,$$

or equivalently

$$\log k_n - l_n \to -\infty, \tag{1.2.4}$$

while after taking the logarithm in both sides (1.2.3) becomes

$$\log k_n + \log l_n - l_n \to \infty. \tag{1.2.5}$$

If we now choose $l_n = \log k_n + \frac{1}{2}\log\log k_n$ then clearly (1.2.4) is satisfied. As for (1.2.5), using that $\log l_n \geq \log\log k_n$

$$\log k_n + \log l_n - l_n \leq \log k_n + \log\log k_n - (\log k_n + \frac{1}{2}\log\log k_n) = \frac{1}{2}\log\log k_n \to \infty.$$

Finally, we note that the argument remains valid with some elementary modifications in case if, instead of the uniform measure we endow the hypercube with the product measure $\mathbb{P}_p = (1-p\delta_{-1}+p\delta_1)^{\otimes k_n}$ for some $p \in (0,1)$.

The case of a general sequence of p_n

So far we assumed that the hypercube is endowed with the uniform measure. Now we consider a sequence of measures where the hypercube $\{-1,1\}^{m_n}$ is endowed with the measure $\mathbb{P}_{p_n} = (1 - p_n\delta_{-1} + p_n\delta_1)^{\otimes k_n}$.

In the argument used to prove Proposition 1.2.5 we only made use of the uniform measure in explicit calculations. Therefore in the general case we simply have to replace (1.2.2) with

$$\lim_{n\to\infty} \left(1 - p_n^{l_n}\right)^{k_n} = 1 \tag{1.2.6}$$

and (1.2.3) with

$$\lim_{n\to\infty} k_n l_n (1 - p_n) p_n^{l_n - 1} = \infty, \tag{1.2.7}$$

respectively. Furthermore, since we also want to use simultaneously the function $\mathsf{Tribes}(l_n, k_n)(-\omega)$, the above asymptotics should hold if we replace p_n with $q_n = 1 - p_n$ as well.

We will replace k_n by m_n/l_n in the sequel, The question we would like to answer is for what pair of m_n and p_n can we find an appropriate sequence l_n that satisfy (1.2.6) and (1.2.7).

First, observe, in case $0 < \inf_n p_n \leq \sup_n p_n < 1$ basically the argument used in the uniform case continues working. So we will investigate two cases: when $\lim_n p_n = 0$ and when $\lim_n p_n = 1$. Case 1 : $\lim_n p_n = 0$ Using that $p_n^{l_n} \to 0$ and taking logarithm from (1.2.6) we get

$$\frac{m_n}{l_n} \log\left(1 - p_n^{l_n}\right) \asymp -\frac{m_n}{l_n} p_n^{l_n} \to 0,$$

which after taking logarithm again, becomes

$$\log m_n - \log l_n - \log \frac{1}{p_n} l_n \to -\infty. \tag{1.2.8}$$

As for (1.2.7), we get

$$\log(a_n) := \log m_n - \log \frac{1}{p_n}(l_n - 1) \to \infty \tag{1.2.9}$$

ignoring the term $\log q_n \to 0$. So (1.2.8) can be written as

$$d_n - \log p_n - \log l_n = d_n - \log(p_n l_n) \to -\infty.$$

Case 2 : $\lim_n p_n = 1$ In this case $p_n^{l_n} = (1 - q_n)^{l_n} \asymp e^{-q_n l_n}$ using that q_n tends to 0. Now let us suppose that we can choose l_n such that $p_n^{l_n} \asymp e^{-q_n l_n} \to 0$. That is we have the condition $q_n l_n \to \infty$

Under these assumptions taking logarithm from (1.2.6) gives:

$$\frac{m_n}{l_n} \log\left(1 - p_n^{l_n}\right) = -\frac{m_n}{l_n} p_n^{l_n} = -\frac{m_n}{l_n} e^{-q_n l_n} \to 0$$

After taking logarithm one more time we get the equivalent

$$\log m_n - \log l_n - q_n l_n \to -\infty. \tag{1.2.10}$$

While taking logarithm from (1.2.7) we get

$$d_n := \log m_n - \log \frac{1}{q_n} - q_n l_n \to \infty \tag{1.2.11}$$

where we ignored the term $\log \frac{1}{p_n} \to 0$. We can express l_n in terms of m_n, q_n and d_n and plug it into (1.2.8):

$$\log m_n - \left(\log(\log m_n - \log \frac{1}{q_n} - d_n) + \log \frac{1}{q_n} \right) - (\log m_n - \log \frac{1}{q_n} - d_n) =$$
$$d_n - \log(\log (q_n m_n) - d_n) \to -\infty.$$

If we choose d_n as for example $\frac{1}{2} \log\log(q_n m_n)$, then obviously (1.2.10) and (1.2.11) are satisfied. It is also straightforward to verify that whenever $q_n m_n \to \infty$ then $l_n < m_n$ moreover $q_n l_n \to \infty$ is consistent with this choice of d_n, and it also guarantees that, in particular $l_n \to \infty$.

1.2.3 Volatility

Another dynamical property of Boolean functions, which may look, at first glance, almost the same as noise sensitivity, is volatility, studied in [JS16]. It roughly says that if we are updating the input bits in continuous time, then the output changes very often.

Our construction also implies, see Corollary 1.2.8 below, that every (monotone) Boolean function is close to a (monotone) Boolean function that has many pivotals with high probability. As functions with these properties are also volatile, this is a strengthening of Theorem 1.4 in [F18].

Let $X_n(t)$ be the continuous time random walk on the k_n hypercube (where $X_n(0)$ is sampled according to the stationary measure) with rate 1 clocks on the edges. For a sequence of Boolean functions f_n let C_n denote the (random) number of times $f_n(X_n(t))$ changes value in the interval $[0, 1]$. The following concepts where introduced in [JS16].

Definition 1.2.4 (Volatility, tameness)**.** A sequence of functions $f_n : \{-1, 1\}^{k_n} \longrightarrow \{-1, 1\}$ is called volatile if the sequence C_n tends to ∞ in distribution and tame, if the sequence C_n is tight.

It is a (rather intuitive) fact that a non-degenerate noise sensitive sequence is volatile (Proposition 1.17 in [JS16]) and all tame sequences are noise stable (Proposition 1.13 in [JS16]). The Maj function, for example, is noise stable, but not tame and not volatile either.

Now we are going to relate our conditions to volatility.

Lemma 1.2.6. *Let $f_n : \{-1, 1\}^{k_n} \longrightarrow \{-1, 1\}$ be a sequence of Boolean functions with the property that there is a sequence of positive integers $a_n \to \infty$ such that $\mathbb{P}[|\mathscr{P}_n| > a_n] \to 1$ (where $\mathscr{P}_n$ denotes the pivotal set of of f_n). Then f_n is volatile.*

Proof. Let $A_n := \{|\mathscr{P}_n| \leq a_n\}$. It is clear that $\mathbb{E}[\int_0^1 \mathbb{1}_{X_n(t) \in A_n} dt] = \mathbb{P}[|\mathscr{P}_n| \leq a_n] \to 0$ so for every ϵ for large enough n it holds that

$$\mathbb{E}[\int_0^1 \mathbb{1}_{X_n(t) \in A_n} dt] < \epsilon^2$$

and therefore, using Markov's inequality

$$\mathbb{P}[\int_0^1 \mathbb{1}_{X_n(t) \in A_n} dt > \epsilon] < \epsilon.$$

By Lemma 1.5 in [JS16] volatility is equivalent with the condition

$$\lim_n \mathbb{P}[C_n = 0] = 0.$$

Now we show that $\mathbb{P}[C_n = 0]$ can be arbitrary small. If we choose n large enough so that $e^{-(1-\epsilon)a_n} < \epsilon$

$$\mathbb{P}[C_n = 0] \leq \mathbb{P}[\int_0^1 \mathbb{1}_{X_n(t)\in A_n} dt > \epsilon] + \mathbb{P}[\int_0^1 \mathbb{1}_{X_n(t)\in A_n} dt \leq \epsilon \text{ and } C_n = 0] \leq \epsilon + e^{-(1-\epsilon)a_n} < 2\epsilon,$$

where we used that $C_n = 0$ can only hold as long as no pivotal bit is switched during the time we are outside of A_n. □

Hence we obtain the following

Corollary 1.2.7. *There exists a noise stable and volatile sequence of transitive monotone functions.*

We say that the sequences f_n and g_n $o(1)$-close to each other if $\lim_n \mathbb{P}[f_n \neq g_n] = 0$. In [F18] it is proved (Theorem 1.4) that for every sequence of Boolean functions there is a volatile sequence $o(1)$-close to it and in this sense volatile sequences are dense among all sequences of Boolean functions. Our construction has a similar conclusion. Using the fact that any sequence of Boolean functions can be slightly modified with a bribable sequence in the same way as we did with Maj, we obtain the following strengthening of Theorem 1.4 from [F18]:

Corollary 1.2.8. *Any sequence of (monotone) Boolean functions is $o(1)$-close to a (monotone) volatile sequence with the property that $\mathbb{P}[\mathscr{P}_n > a_n] \to 1$ for some sequence of integers $a_n \to \infty$.*

Although here we consider the uniform measure on the hypercube the same type of questions are meaningful when the uniform measure is replaced by the sequence of product measures $\mathbb{P}_{p_n} = (1 - p_n\delta_{-1} + p_n\delta_1)^{\otimes k_n}$. It has to be noted that Theorem 1.4 in [F18] is valid for basically all possible sequences p_n under which the question is meaningful, while our construction works in a more restricted range of sequences p_n. Most importantly, our results extend to all sequences p_n that satisfy $0 < \liminf p_n \leq \limsup p_n < 1$.

Furthermore, in [F18] a sequence of Boolean functions is constructed which is noise stable and volatile, but at the same time it is not $o(1)$-close to any non-volatile sequence. Such a sequence, of course cannot be obtained with a small modification from some non-volatile stable sequence.

This naturally lead to the following questions:

Question 1.2.9. *Is there a transitive, noise stable (volatile?) sequence f_n such that $\mathbb{P}[\mathscr{P}_n(\omega) \neq \emptyset] \to 1$ and f_n is not $o(1)$-close to any sequence which does not have these properties?*

We think that the answer is positive to this question.

Question 1.2.10. *Is there a transitive, monotone and noise stable (volatile?) sequence f_n such that $\mathbb{P}[\mathscr{P}_n(\omega) \neq \emptyset] \to 1$ and f_n is not $o(1)$-close to any sequence which does not have these properties?*

This looks more difficult and it might be the case that the answer is negative.

An alternative bribable sequence

Here we sketch a completely different way of constructing a bribable sequence. Its disadvantage is that it is not-monotone therefore it only implies a weaker result (without monotonicity). We shall include here because the ideas in it might be of interest.

Let $L, k \in \mathbb{N}$ be such that $L > 6k$ and $n = \binom{L}{2k+1}$. In fact, we are going to identify the set of bits $[n]$ with the $2k+1$ element subsets of $[L]$.

Define for $j = 1, 2, \ldots L$ the subset of bits $H_j \subset [n]$ by

$$H_j := \left\{ t \in \binom{L}{2k+1} : j \in t \right\}.$$

We introduce a spin system indexed by $j = 1, 2, \ldots L$ which is a factor of the ω:

$$\sigma_j := \chi_{H_j}(\omega) = \prod_{t \in H_j} \omega_t$$

That is, we multiply all the ω_t bits corresponding to subsets that contain j.

The crucial property of this spin system with respect to the original bits is the following simple observation:

Lemma 1.2.11. *For every $2k+1$ element subset $t \subseteq [L]$ the corresponding bit ω_t has the following property: If one flips the value of ω_t then the values of all the spins $\sigma_j : j \in t$ are flipped, while all the other spins $\sigma_k : k \notin t$ are kept unchanged.*

Proof. Flipping the value of any bit t that is in H_j will change the value of $\sigma_j = \chi_{H_j}(\omega)$ and it will obviously not change the value of any $\sigma_l = \chi_{H_l}(\omega)$ which does not contain t. By definition, t is contained in H_j if and only if j is contained in t. Since every for every $2k+1$ element subset there is a corresponding bit t the statement follows. □

Lemma 1.2.12. *If $\omega_t,\ t \in \binom{L}{2k+1}$ is a uniform i.i.d spin system, then so is $\{\sigma_j,\ j \in [L]\}$.*

Proof. First we observe that if for any $\emptyset \neq S \subseteq [L]$ it holds that $\mathbb{E}[\sigma_S] = \mathbb{E}[\prod_{j \in s} \sigma_j] = 0$ then the random variables $\{\sigma_j,\ j \in [L]\}$ are independent, unbiased coin flips.

We show this by induction with respect to L. For $L = 1$ the statement is trivially true. Now suppose we have a system of $L-1$ spins which satisfies the condition above. By the induction hypothesis this is a uniform i.i.d spin system.

It is easy to see that any event A which is measurable with respect to $\{\sigma_j,\ j \in [L-1]\}$ is independent from σ_L. Indeed, $\mathbf{1}_A$ can be written as a linear combination of functions $\sigma_S :\ S \subseteq [L-1]$ (i.e. the Fourier-Walsh transform of $\mathbf{1}_A$), but $\mathbb{E}[\sigma_S \sigma_L] = \mathbb{E}[\sigma_{S \cup \{L\}}] = 0$ and consequently $\mathrm{Cov}(\mathbf{1}_A, \sigma_L) = 0$. This shows that σ_L is independent from the σ-algebra generated by $\{\sigma_j,\ j \in [L-1]\}$. Together with $\mathbb{E}[\sigma_L] = 0$, this shows that $\{\sigma_j,\ j \in [L]\}$ is a uniform i.i.d spin system. Now we are going to show that $\mathbb{E}[\sigma_S] = 0$ for any $\emptyset \neq S \subseteq [L]$.

Indeed,

$$\sigma_S = \prod_{j \in s} \sigma_j = \prod_{j \in S} \prod_{t \in H_j} \omega_t = \prod_{j \in S} \prod_{t : j \in t} \omega_t.$$

Notice that for any particular subset $t \in \binom{L}{2k+1}$, the bit ω_t appears in this product for all $j \in S$ for which $j \in t$ also holds, that is $|S \cap t|$ times. So we get that

$$\sigma_S = \prod_{t \in \binom{L}{2k+1}} \omega_t^{|S \cap t|}.$$

Consequently, σ_S is uniform on ± 1 whenever there exist some $2k+1$ element subset t for which $|S \cap t|$ is odd. But this is always true. Indeed, in case $|S| \geq 2k+1$ then there exists a $t \in \binom{L}{2k+1}$ such that $t \subseteq S$, and therefore $|S \cap t| = |t| = 2k+1$. If $|S| < 2k+1$ we can choose one element from S and another $2k$ elements from $L \setminus S$ (which is possible since $L > 4k$) and then $|S \cap t| = 1$. □

Now we can define the sequence f_n.

Let $k = \lceil L^{\frac{1}{2}+\epsilon} \rceil$, and define the following two events:

$$A_n := \left\{ \sum_{j \in [L]} \sigma_j \geq 2k \right\}$$

and

$$B_n := \left\{ \sum_{j \in [L]} \sigma_j \leq -2k \right\}.$$

Now let $f_n := \mathbb{1}_{A_n} - \mathbb{1}_{B_n}$.

It is clear that $\lim_n \Pr[f_n = 0] = \lim_n (1 - \Pr[A_n] - \Pr[B_n]) = 1$ because of the Central Limit Theorem. At the same time, conditioned on the event $\{f_n = 0\}$ (which happens with high probability) we can always find (many) bits that change f_n to 1 or -1, respectively.

Indeed, define $M^+ = \{j \in [L] : \sigma_j = 1\}$ and M^- in a similar way. Obviously, $|M^+| + |M^-| = L$ and $-2k < |M^+| - |M^-| < 2k$ on $\{f_n = 0\}$. So $|M^+| > L/2 - k > 2k$ using that $L > 6k$. (In fact, $|M^+| = L/2 - o(L)$ while $2k+1 = o(L)$.) By symmetry, the same lower bound holds for M^-. So we can always choose a $t \subseteq M^+$ and a $t' \subseteq M^+$ with $t, t' \in \binom{L}{2k+1}$.

On the other hand $\sum_{j \in [L]} \sigma_j = |M^+| - |M^-|$ increases (decreases) by $4k+2$ whenever we change the value of any $t \subseteq M^+$ ($t' \subseteq M^-$). Therefore, as $-2k < |M^+| - |M^-| < 2k$ holds on $\{f_n = 0\}$, by changing the value of a bit t (or respectively t') as above one can achieve that $\sum_{j \in [L]} \sigma_j \geq 2k$ (respectively, $\sum_{j \in [L]} \sigma_j \leq -2k$).

1.2.4 Revealment

The fundamental paper [BKS99] used hypercontractivity estimates to prove that crossing events are noise sensitive. There is, however another tool coming from the theory of randomised algorithms that allows for more quantitative noise sensitivity results. The revealment of a randomized algorithm for a Boolean function f is the maximum probability that a particular bit is queried during the algorithm. The revealment of the Boolean function is the infimum of the revealments over all randomized algorithms.

Definition 1.2.5 (Revealment)**.** Let $J_{\mathcal{A}}$ denote the random set of edges queried by the algorithm $\mathcal{A}$ until it learns the value of f. Let R denote all possible random algorithms on $\{-1,1\}^E$ The **revealment** of f is

$$\delta_f = \inf_{\mathcal{A} \in R} \max_{e \in E} \mathbb{P}[e \in J_{\mathcal{A}}] \tag{1.2.12}$$

Now the important result that links noise sensitivity to this notion is the following ([SS10]):

Theorem 1.2.13. *Let $f : \{-1,1\}^E \longrightarrow \mathbb{R}$ then*

$$\frac{\sum_{|S|=k} \widehat{f}(S)^2}{\|f\|_2^2} \leq \delta_f k \tag{1.2.13}$$

If the revealment δ_n of a sequence of Boolean functions f_n goes to 0 then it is noise sensitive. Moreover, δ_n gives a quantitative bound for noise sensitivity.

Chapter 2

Sparse Reconstruction in Product Measures

2.1 Sparse Reconstruction for Transitive Functions

Let G be a vertex transitive graph with vertex set V and let us put uniformly random bits (we will think about them as ± 1) on the vertices of the graph. Now take an event which is invariant under graph automorphisms. The question we are going to investigate is the following: Is it possible that knowing the bits of a small subset of vertices specified in advance (independently from the value of the bits) will give enough information to decide whether the event has occurred or not?

In this section we will answer this question and some of its generalizations. In order to make this question precise we need to measure the amount of information we gain about an event by learning a subset of the coordinate values of a configuration. For a subset of vertices $U \subseteq V$ let $\mathcal{F}_U$ denote the σ-algebra generated by the random bits belonging to vertices in U.

Definition 2.1.1 (Clue)**.** Let $f : \{-1, 1\}^V \longrightarrow \mathbb{R}$ and $U \subseteq V$.

$$\mathsf{clue}(f \mid U) = \frac{\mathrm{Var}(\mathbb{E}[f \mid \mathcal{F}_U])}{\mathrm{Var}(f)}$$

In the definition we allowed for any real function f, not only events (which may be represented by their indicator functions), as the definition extends naturally.

The notion of $\mathsf{clue}_f(U)$ quantifies the proportion of the total variance of f attributed to the variance of the function projected onto $\mathcal{F}_U$. The clue is always a number between 0 and 1, as projection can only decrease the variance.

It is worth noting that

$$\mathsf{clue}(f \mid U) = \frac{\mathrm{Cov}^2(f, \mathbb{E}[f \mid \mathcal{F}(U)])}{\mathrm{Var}(f)\mathrm{Var}(\mathbb{E}[f \mid \mathcal{F}(U)])} = \mathrm{Corr}^2(f, \mathbb{E}[f \mid \mathcal{F}(U)]). \qquad (2.1.1)$$

using that $\mathrm{Cov}(f, \mathbb{E}[f \mid \mathcal{F}(U)]) = \mathrm{Var}(\mathbb{E}[f \mid \mathcal{F}(U)])$, since conditional expectation is orthogonal projection.

Natural it may seem, clue is obviously not the only possible way to quantify the information content of a subset of coordinates about a function. Later on in this section we will consider a few alternatives.

Here we mention a similar concept introduced in [BL89]. For a subset $U \subseteq V$ the influence of U is defined as follows:

$$I(U) = \mathbb{P}[\text{ f is not determined by the bits on } U^c]$$

Influence is, however much weaker then clue (in the sense that it is much easier to have high influence then high clue). Like in Social Choice Theory, one may think about coordinates as individual agents trying to influence the value(outcome) of f by the values of the respective bits. In this framework the influence of a subset quantifies the probability that the set of agents in U can change the value of f by coordinating their values. While in this setting coordinates are allowed to cooperate, clue rather quantifies the average gain of information (measured in variance) for a uniformly random configuration of U.

Indeed, one can easily see that among n coordinates any set of size $n^{\frac{2}{3}}$ has influence close to 1 with respect to $\mathsf{Maj}(n)$. On the other hand, it is an easy exercise to verify that subsets of size $o(n)$ give asymptotically 0 clue (besides it follows immediately from Theorem 2.1.1).

We continue formalising the informal question posed at the beginning. Let G be a group acting transitively on the set of coordinates V. This action can be extended in the natural way to the configuration space $\{-1,1\}^V$, and in turn to any function $f : \{-1,1\}^V \longrightarrow \mathbb{R}$. For a Boolean function f, and for $g \in G$ we denote by $f^g(x) := f(x^g)$ the action of G on the function f.

Definition 2.1.2 (Transitive function)**.** A function $f : \{-1,1\}^V \longrightarrow \mathbb{R}$ is transitive if there is a transitive group action G on V such that for every $g \in G$ $f^g = f$

We introduce a concept that helps us to ask what we want in a concise way. Besides, this is one of the central notions of this thesis.

Definition 2.1.3 (Sparse Reconstruction)**.** Let $f_n : \{-1,1\}^{V_n} \longrightarrow \{-1,1\}$ be a sequence of Boolean functions and let $\mu_n =$. We say that there is Sparse Reconstruction for f_n if there is a sequence of subsets $U_n \subseteq V_n$ such that for some $c > 0$

$$\liminf_n \mathsf{clue}(f_n \mid U_n) > c$$

We are now ready to formulate our question. Is there a $f_n : \{-1,1\}^{V_n} \longrightarrow \{-1,1\}$ be a sequence of transitive Boolean functions for which there is Sparse Reconstruction?

One may guess that the answer is negative for transitive functions and this is indeed the case. The proof, however, is surprisingly short and it demonstrates the power of the notion of spectral sample in an impressive way. (For an introduction on the Fourier-Walsh transform on the hypercube and the spectral sample see Section 1.1.1).

Theorem 2.1.1 (Clue of Transitive Functions)**.** *If $f : \{-1,1\}^V \longrightarrow \{-1,1\}$ transitive, $U \subseteq V$ then*

$$\mathsf{clue}(f \mid U) \leq \frac{|U|}{|V|}$$

Proof. Let X be a uniformly random element from the spectral sample $\mathscr{S}_f$ of f conditioned on being non-empty. Because f is transitive X is uniform on V. Using (1.1.7) we get the following:

$$\mathsf{clue}(f \mid U) = \mathbb{P}[\mathscr{S} \subseteq U \,|\, \mathscr{S} \neq \emptyset] \leq \mathbb{P}[X \in U] = \sum_{u \in U} \mathbb{P}[X = u] = \frac{|U|}{|V|} \tag{2.1.2}$$

□

Remark 2.1.2. The bound in Theorem 2.1.1 is sharp, as it is testified by the function $\sum_{v\in V}\omega_v$.

Remark 2.1.3. There is no obvious way to relax the condition of transitivity. We now sketch an example of a sequence of Boolean functions where the individual influences $\mathcal{I}_v(f_n)$ (see Definition 1.2.2) are (almost) equal for every n, however there is a sparse subset of coordinates U_n (i. e. $\lim_n \frac{|U_n|}{|V_n|}=0$) such that $\lim_n \mathsf{clue}_{f_n}(U_n)=1$.

Let a_n be a sequence of integers such that $a_n\to\infty$. Let us define non-symmetrical majority functions

$$\mathsf{Maj}^{a_n}(n)=\begin{cases}1 & \text{if } \sum_i \omega(i)>a_n\sqrt{n}\\ -1 & \text{if } \sum_i \omega(i)<a_n\sqrt{n}.\end{cases}$$

We can choose a_n in such a way that for some small $\epsilon>0$

$$\mathcal{I}_i(\mathsf{Maj}_n^{a_n})=\frac{\binom{n}{n/2+a_n\sqrt{n}}}{2^n}\sim\frac{1}{n^{2/3}}$$

holds. The Tribes function $\mathsf{Tribes}(l_n,k_n)$ which has already been defined in Section **??** is known to be balanced if $l_n=\log n-\log\log n$ and $k_n=n/l_n$. Let us denote this balanced version of the tribes on n bits by $\mathsf{Tribes}(n)$. An easy calculation shows that $\mathcal{I}_i(\mathsf{Tribes}_n)\sim\frac{\log n}{n}$.

Let $V_n=M_n\cup T_n$. In the following definition, where the domain of $\mathsf{Maj}_{m_n}^{a_n}$ is M_n and the domain of $\mathsf{Tribes}(t_n)$ is T_n.

$$f_n:=\begin{cases}\mathsf{Maj}^{a_n}(m_n) & \text{if } \mathsf{Tribes}(t_n)=1\\ \mathsf{Maj}^{-a_n}(m_n) & \text{if } \mathsf{Tribes}(t_n)=-1.\end{cases}$$

We adjust the size of M_n and T_n in such a way that the influence of each coordinate is the same. So we have the equation $\frac{\log t_n}{t_n}=\frac{1}{n^{2/3}}$, or equivalently

$$m_n=\left(\frac{t_n}{\log t_n}\right)^{3/2}.$$

So the density of T_n goes to 0 compared to $|V_n|=t_n+m_n$. At the same time, from the Central Limit Theorem it is clear that $\lim_n\mathbb{P}[\mathsf{Maj}^{a_n}(m_n)=1]=0$ and $\lim_n\mathbb{P}[\mathsf{Maj}^{-a_n}(m_n)=1]=1$. Consequently, $\lim_n\mathsf{clue}_{f_n}(T_n)=1$.

It is worth noting that the result does not only apply for sequences of Boolean functions, but also for any sequences of real-valued functions, no matter bounded or not. One may ask whether a similar result can be derived in case we replace the $\{-1,1\}$ space in the domain with something more complicated. It turns out that there is an important generalization of the Fourier-Walsh transform and the spectral sample for general product spaces that enables us to extend the inequality in Theorem 2.1.1 to arbitrary product spaces. We will need the following simple observation, which turns out to be crucial for the orthogonality of the Efron-Stein decomposition

Lemma 2.1.4. *Let $f\in L^2(\Omega^n,\pi^{\otimes n})$ and let $K,L\subseteq[n]$. then*

$$\mathbb{E}[\mathbb{E}[f\,|\,\mathcal{F}_L]\,|\,\mathcal{F}_K]=\mathbb{E}[f\,|\,\mathcal{F}_{L\cap K}]$$

Proof. Rewriting the conditional expectations as integral and using Fubini's Theorem

$$\int_{X^{K^c}}\left(\int_{X^{L^c}}f(X_L,x_{L^c}dx_{L^c})\right)dx_{K^c}=\int_{X^{K^c\cup L^c}}f(X_{L\cap K},x_{K^c\cup L^c})dx_{K^c\cup L^c}.$$

□

Theorem 2.1.5 (Efron-Stein decomposition, 1981)**.** *For $f \in L^2(\Omega^n, \pi^{\otimes n})$, there is a unique decomposition*

$$f = \sum_{S \subseteq [n]} f^{=S},$$

where $f^{=S}$ is a function that depends only on the coordinates in S, and $(f^{=S}, f^{=T}) = 0$ whenever $S \neq T$.

Proof. Our proof follows the ideas from [OD14].

Notice first that assuming such a decomposition exists, then much like in the case of the hypercube,

$$\mathbb{E}[f \,|\, \mathcal{F}_T] = \sum_{S \subseteq T} f^{=S}.$$

Indeed, since $\mathbb{E}[f \,|\, \mathcal{F}_T]$ only depends on coordinates in T, for every $S \not\subseteq T$ we expect that $\mathbb{E}[f \,|\, \mathcal{F}_T]^{=S} = 0$. Therefore using the (assumed) orthogonality $(f, \mathbb{E}[f \,|\, \mathcal{F}_T]) = \sum_{L \subseteq T} (f^{=L}, \mathbb{E}[f \,|\, \mathcal{F}_T]^{=L})$ and since $\mathbb{E}[f \,|\, \mathcal{F}_T]$ maximizes (f, g) among all g $\mathcal{F}_T$-measurable functions, we have $f^{=L} = \mathbb{E}[f \,|\, \mathcal{F}_T]^{=L}$ for every $L \subseteq T$.

This means that we can reconstruct the functions $f^{=S}$ via a Moebius inversion (in this case, an exclusion-inclusion principle) from the conditional expectations:

$$f^{=S} = \sum_{L \subseteq S} (-1)^{S-L} \mathbb{E}[f \,|\, \mathcal{F}_L].$$

It is obvious from the construction that $f^{=T}$ only depends on coordinates in T. So what is left to show is that $f^{=T}$ and $f^{=S}$ are orthogonal, if they are not equal. First we show that if g is $\mathcal{F}_T$-measurable and $S \setminus T \neq \emptyset$ then $f^{=T}$ and g are orthogonal. We can pick an $i \in S \setminus T$ and write the above inner product as

$$\mathbb{E}[g f^{=S}] = \sum_{L \subseteq S \setminus \{i\}} (-1)^{S-L} \mathbb{E}[g \mathbb{E}[f \,|\, \mathcal{F}_L]] - \mathbb{E}[g \mathbb{E}[f \,|\, \mathcal{F}_{L \cup \{i\}}]]$$

using that $(-1)^{S-L}$ and $(-1)^{S-L\cup\{i\}}$ has opposite signs. Conditioning on T and after on L before taking the expectation and applying Lemma 2.1.4 twice gives that

$$\mathbb{E}[g \mathbb{E}[f \,|\, \mathcal{F}_L]] = \mathbb{E}[\mathbb{E}[g \,|\, \mathcal{F}_{T \cap L}] \mathbb{E}[f \,|\, \mathcal{F}_{T \cap L}]] =$$
$$\mathbb{E}[\mathbb{E}[g \,|\, \mathcal{F}_{T \cap (L \cup \{i\})}] \mathbb{E}[f \,|\, \mathcal{F}_{T \cap (L \cup \{i\})}]] = \mathbb{E}[g \mathbb{E}[f \,|\, \mathcal{F}_{L \cup \{i\}}]].$$

We used that $T \cap (L \cup \{i\}) = T \cap L$, since $i \notin L$ and $i \notin T$. This shows that $\mathbb{E}[g f^{=S}] = 0$. From this to $\mathbb{E}[f^{=T} f^{=S}]$ and switching the roles, it follows $\mathbb{E}[f^{=T} f^{=S}] = 0$ if either $S \setminus T \neq \emptyset$ or $T \setminus S \neq \emptyset$ which is equivalent to $T \neq S$. □

Observe that this is indeed a generalization of the Fourier-Walsh transform, with $f^{=S} = \widehat{f}(S) \chi_S$. What is important for our purpose is that we can again define a Spectral Sample $\mathbb{P}[\mathscr{S} = S] := \frac{\|f^{=S}\|^2}{\|f\|^2}$ for every square-integrable function, as in the case of the hypercube and thus Theorem 2.1.1 generalizes for product measures.

Theorem 2.1.6 (Small Clue Theorem for Product Spaces)**.** *Let $f \in L^2(\Omega^n, \pi^{\otimes n})$ and suppose that there is a $G \leq S_n$ acting on the n copies of Ω transitively. Suppose f is invariant under the action of G. If $U \subseteq [n]$ then*

$$\mathsf{clue}(f \mid U) \leq \frac{|U|}{n}$$

The proof is exactly the same as for Theorem 2.1.1, the only difference being that we need to use the Efron-Stein decomposition instead of the Fourier-Walsh transform to build the Spectral Sample.

2.2 Sparse Reconstruction and Mutual Information

Our setup remains the same, but we formulate it in a somewhat different way. Let $\{X_v : v \in V\}$ be a set of real-valued discrete random variables defined in a common probability space. Let G be a group acting on V transitively and we assume that the joint distribution of $\{X_v : v \in V\}$ is invariant under the group action. We introduce the following notation for a $S \subseteq V$ we have $X_S = \{X_j : j \in S\}$ and as before $\mathcal{F}_S$ denotes the σ–algebra generated by X_S. The variables $X_v : v \in V$ obviously playing the role of the coordinates. Let $f : \mathbb{R}^V \mapsto \mathbb{R}$ and let $Z = f(X_V)$. In this section we are going to discuss an alternative way of measuring the amount of information a subset $S \subseteq V$ of coordinates contains about the function f. In the sequel we use concepts from information theory and define an information-theoretic clue accordingly.

Our main interest is still the special case where the variables X_v and Z are ± 1-valued variables (spins) (the case $f : \{-1,1\}^V \to \{-1,1\}$), but all the argument we present here work in this slightly more general framework.

For a (possibly vector valued) random variable (or a probability distribution) entropy measures the amount of randomness or information.

Definition 2.2.1 (Entropy)**.** Let X be a random variable. Then the entropy of X is

$$H(X) = -\sum_{x \in \mathrm{Ran}(X)} \mathbb{P}[X = x] \log \mathbb{P}[X = x]$$

We will also need the concept of conditional entropy. The entropy of X conditioned on the random variable Y expresses how much randomness remains in X on average if we learn the value of Y .

Definition 2.2.2 (Conditional Entropy)**.** The conditional entropy of X given Y is

$$H(X|Y) = \mathbb{E}[H(X)|Y]$$

The mutual information quantifies the common information present in two variables. In a way it measures how far the joint distribution of the two variables is from being independent.

Definition 2.2.3 (Mutual Information)**.** Suppose that $H(X)$ and $H(Y)$ are both finite then the mutual information between X and Y is:

$$I(X : Y) = H(X) + H(Y) - H(X,Y) = H(X) - H(X|Y) \tag{2.2.1}$$

Now the definition of clue in this framework:

Definition 2.2.4 (I-Clue)**.** Let $f : \Omega^n \mapsto \mathbb{R}$ and $Z = f(X_V)$. The information theoretic clue (I-clue) of f with respect to $U \subseteq [n]$ is

$$\mathsf{clue}^I(f \mid U) = \frac{I(Z : X_U)}{H(Z)}$$

Note that if Z is X_U-measurable then $H(Z|X_U) = 0$ and therefore $I(Z : X_U) = H(Z)$, while if Z is independent from X_U then $I(Z : X) = 0$, in accordance with what we expect from a clue type notion. Now we are ready to prove an equivalent of Theorem 2.1.6 for the I-clue. The following Theorem, however, as well as the definition of I-Clue only works well in the discrete case, as the continuous counterpart of entropy, differential entropy has some drawbacks (for example, it can be negative).

Theorem 2.2.1. *Let $\{X_v : v \in V\}$ be discrete valued, i.i.d, random variables with finite entropy. Let $f : \Omega^n \mapsto \mathbb{R}$ be a transitive function and $Z = f(\{X_v : v \in V\})$. Then*

$$\mathsf{clue}^I(f \mid U) \le \frac{|U|}{n} \tag{2.2.2}$$

For the proof we will use the following well-known inequality which finds numerous applications in combinatorics. For a proof see [Ga12].

Theorem 2.2.2 (Shearer's inequality)**.** *Let $X_1, X_2, \dots X_n$ random variables defined on the same probability space. Let $S_1, S_2 \dots, S_L$ subsets of $[n]$ such that for every $i \in [n]$ there are at least k among $S_1, S_2 \dots, S_L$ containing i. Then*

$$kH(X_{[n]}) \le \sum_{l=1}^{L} H(X_{S_l}),$$

First we need the following consequence of Shearer's inequality.

Lemma 2.2.3. *Suppose $X_1, X_2, \dots X_n$ are independent. Let $S_1, \dots S_L$ be a system of subsets of $[n]$ such that each $i \in [n]$ appears in at most k sets. Then*

$$\sum_{j}^{L} I(Z : X_{S_j}) \le kI(Z : X_{[n]}) \tag{2.2.3}$$

Proof. Without loss of generality we can assume that each i appears in exactly k sets. Indeed the right hand side does not change and the left hand side can only increase by this.

Since the variables X_i are independent:

$$\sum_{j}^{L} H(X_{S_j}) = \sum_{j} \sum_{i \in S_j} H(X_i) = k \sum_{i \in [n]} H(X_i) = kH(X_{[n]}) \tag{2.2.4}$$

On the other hand, using Shearer's inequality

$$-\sum_{j}^{L} H(X_{S_j}|Z) \le -kH(X_{[n]}|Z) \tag{2.2.5}$$

Using that $I(Z : X_{S_j}) = H(X_{S_j}) - H(X_{S_j}|Z)$ and adding up 2.2.4 and 2.2.5 completes the proof. $\square$

Now the proof of the clue-theorem:

Proof. Recall that G acts transitively on V. We assume that both the product measure μ and the function f are G-invariant. Let $U \subseteq V$ arbitrary. then for each $g \in G$

$$I(Z : X_U) = I(Z : X_{U^g})$$

where $U^g = \{ug \ : \ g \in G\}$.
Observe that $v \in U^g \iff vg^{-1} \in U \iff u = vg^{-1}$. For each pair of $v \in V$ and $u \in U$ there are $|G_v|$ such g, where G_v is the stabilizer subgroup of G with respect to v. (Since the action is transitive such a g exists, moreover the cardinality of the stabilizer subgroup G_v is the same for every $v \in V$.) The conclusion is that each $v \in V$ appears in exactly $|U||G_i|$ translated version of U. Applying Lemma 2.2.3 gives

$$|G|I(Z : X_U) = \sum_{g \in G} I(Z : X_{U^g}) \leq |U||G_v|I(Z : X_V) = H(Z)$$

which is what we wanted since $|G| = n|G_v|$ by the orbit-stabilizer theorem. $\square$

The concept of clue and I-clue are close to each other as long as the variables Z_n are non-degenerate, in the sense that the variables Z^n are uniformly bounded and its variance is $\Omega(1)$. This follows from Proposition 3.1.7. So in the non-degenerate case, Theorem 2.1.6 and Theorem 2.2.1 are equivalent. In full generality, however we cannot say anything. In light of this, it is remarkable that we have the exact same bound (at least in the case product measures) for the clue and I-clue.

2.2.1 Measuring clue via Relative Entropy

The Kullback-Liebler(KL) divergence or relative entropy between probability measures μ and ν on the same probability space is a way of measuring distance between two measures.

Definition 2.2.5 (Relative entropy)**.** Let ν and μ measures on the same probability space, where $\nu \ll \mu$. the relative entropy between ν and μ is

$$D(\nu_U||\mu_U) = -\sum_x \mu(x) \log \frac{\mu(x)}{\nu(x)}$$

Observe that although it means to express a concept of distance between two distributions, the relative entropy is not a metric. In particular $D(\nu_U||\mu_U) \neq D(\mu_U||\nu_U)$

Let us consider again a product space generated by respective random variables $\{X_v \ : \ v \in V\}$. For convenience let us assume that the random variables are discrete. A measurable random variable with respect to X_V induces an absolutely continuous probability measure and a respective Radon-Nikodym derivative.

On the other hand every $\phi : \{-1, 1\}^n \mapsto [0, \infty]$ with $\mathbb{E}[\phi] > 0$ can be interpreted as a density, and can be used to define another measure on the same space by

$$\nu(\omega) := \frac{1}{\mathbb{E}[\phi]} \phi(\omega)\mu(\omega)$$

If X_V is distributed according to μ and a $U \subseteq V$ we denote by μ_U the projection of μ onto X_U.

If $D(\nu_U||\mu_U)$ expresses the total 'information distance' between μ and ν, we can interpret the quantity $D(\nu_U||\mu_U)$ as the 'information distance' restricted to the respective subset of coordinates. Like this we may introduce another possible clue measure as follows:

$$\mathsf{clue}^{KL}(f \mid U) := \frac{D(\nu_U||\mu_U)}{D(\nu_U||\mu_U)}$$

We can observe that

$$D(\nu||\mu) = \mathsf{Ent}(\phi)$$

where $\mathsf{Ent}(\phi) := \mathbb{E}[\phi \log \phi] - \mathbb{E}[\phi] \log \mathbb{E}[\phi]$.

The reason why it worth mentioning this concept is that the respective version of Theorem 2.1.6 and Theorem 2.2.1 is true for clue^{KL}. Moreover, in contrast with mutual information, relative entropy is a concept that remains meaningful for continuous random variables as well. Indeed, the following Shearer-type inequality holds:

Lemma 2.2.4. *Let μ be a product measure on the hypercube (in fact, any product measure will do) and and ν another measure on the same space satisfying $\nu \ll \mu$.*

Let $S_1, \dots S_L$ be a system of subsets of V such that each $i \in V$ appears in at most k sets. Then

$$\sum_{j}^{L} D(\nu_{S_i}||\mu_{S_i} \leq kD(\nu||\mu)$$

The proof of this Lemma is also a simple consequence of Shearer's inequality (Theorem 2.2.2), for a proof see [Ga12]. The corresponding clue theorem follows in the same way as Lemma 2.2.3 implies Theorem 2.2.1.

2.3 Sparse Reconstruction and Cooperative Game Theory

The field of cooperative game theory starts with the following setup: There is a set of players which we denote by V here (to be consistent) and the game is defined by assigning a positive real number $v(S)$ to every subset S of the players. Usually it is assumed that $v(\emptyset) = 0$. The function $v : 2^V \longrightarrow \mathbb{R}$ is referred to as the characteristic function. This aims to model a situation where individuals can gain profit, but the profit may change (typically increases) in case certain individuals cooperate and form a coalition. Thus $v(S)$ is the common payoff of the individuals in S provided that they cooperate.

Cooperative game theory is mostly concerned with finding some sort of fair distribution of the payoff given the characteristic function v. One of these concepts is the Shapley value, which aims to distribute the payoff based on the average marginal contribution of the individuals.

Definition 2.3.1 (Shapley value)**.**

$$\phi_i(v) = \frac{1}{|V|} \sum_{S \subseteq V \setminus \{i\}} \frac{v(S \cup \{i\}) - v(S)}{\binom{|V|-1}{|S|}} \tag{2.3.1}$$

Observe that for a given $f : \{-1,1\}^V \longrightarrow \{-1,1\}$ we can define a cooperative game via $v_f(U) := \mathrm{Var}[\mathbb{E}[f \mid \mathcal{F}_U]]$ for any $U \subseteq V$. Besides fitting the mathematical definition, it also fits into the interpretation of the Theory. It is a sort of information game, where the payoff (we can interpret it as an expected gain or profit) depends on how accurately we know a piece of information (represented by the value of the function). Each individual possesses one piece of information (the value of the corresponding coordinate) but only together they determine the valuable piece of information.

In the proof of Theorem 2.1.1 we introduced the random element X of the index set, which is a uniformly random element of the Spectral Sample. In fact, X is distributed according to the Shapley value.

Proposition 2.3.1. *Let $f : \{-1, 1\}^V \longrightarrow \mathbb{R}$. Then*

$$\frac{\phi_i(v_f)}{v_f(V)} = \mathbb{P}[X = i]$$

Proof. Without loss of generality we may assume that $\mathrm{Var}(f) = 1$. Let n = —V—. First, observe that

$$\mathbb{P}[X = u] = \sum_{u \in S} \widehat{f}(S)^2 \frac{1}{S}$$

Now we calculate $\phi_i(v_f)$ via Fourier-Walsh expansion and show that it equals to $\mathbb{P}[X = u]$. Using that $v_f(S) = \sum_{T \subseteq S} \widehat{f}(T)^2$ we get that

$$\phi_i(v) = \frac{1}{n} \sum_{S \subseteq V \setminus \{i\}} \frac{\sum_{T \subseteq S} \widehat{f}(T \cup \{i\})^2}{\binom{n-1}{|S|}} = \frac{1}{n} \sum_{T \subseteq V \setminus \{i\}} \widehat{f}(T \cup \{i\})^2 \sum_{S \subseteq [n] \setminus \{i\} : T \subseteq S} \frac{1}{\binom{n-1}{|S|}}$$

For a fixed T there is $\binom{n-1-|T|}{k-|T|}$ k-element subset S which contains T. Therefore we have

$$\phi_i(v) = \frac{1}{n} \sum_{T \subseteq V \setminus \{i\}} \widehat{f}(T \cup \{i\})^2 \sum_{k=|T|}^{n-1} \frac{\binom{n-1-|T|}{k-|T|}}{\binom{n-1}{k}}$$

With some elementary manipulation of the binomial coefficients we get that

$$\frac{\binom{n-1-|T|}{k-|T|}}{\binom{n-1}{k}} = \frac{\binom{k}{|T|}}{\binom{n-1}{|T|}}$$

and using that by the Hockey-stick identity ($\sum_{k=|T|}^{n-1} \binom{k}{|T|} = \binom{n}{|T|+1}$), we get the desired formula. $\square$

Remark 2.3.2. The Shapley value has a more analytic interpretation as well in the Boolean analysis framework.

The stability of f at level p is

$$\mathbf{Stab}_f(p) := \sum_{S \subseteq E_n} \widehat{f}(S) p^{|S|}$$

Stability has two interpretations. First, this quantity measures the noise stability of f. If f is defined on the p-correlated bit sets x and y than $\mathbf{Stab}_f(p) = \mathbb{E}[f(x)f(y)]$. On the other hand, it is also the expected clue of a Bernoulli random set of coordinates $\mathcal{B}^p$ of density p: $\mathbf{Stab}_f(p) = \mathbb{E}[\mathsf{clue}(f \mid \mathcal{B}^p)]$.

Stability can be generalized as a polynomial of $|V|$ variables. Than the quantity

$$\mathbf{Stab}_f(\overline{p}) = \sum_{S \subseteq V} \widehat{f}(S) \quad (_)i \in Sp_i$$

Can be interpreted as the expected clue of a random subset where the bit i is selected with probability p_i, independently from other bits.

Denote by $\overline{p}$ the vector with all of its coordinates is equal to p. An easy calculation shows that

$$\int_0^1 \frac{\partial \mathbf{Stab}_f(\overline{p})}{\partial p_u} dp = \sum_{u \in S} \widehat{f}(S)^2 \frac{1}{S} = \mathbb{P}[X = u]$$

This can be understood as the average increase in clue over all p values, induced by a small increase in the probability of selecting u into the random set.

Given how naturally the Shapley value shows up in the proof of Theorem 2.1.1 it is perhaps not surprising that there is proof that does not use Fourier-Walsh transform, only simple concepts from Cooperative Theory and Combinatorics. The advantage of this approach is that it makes it more clear what are the conditions under which a small clue Theorem can be true. It should be also noted that this approach entails both the L_2 and the entropy version of the Theorem.

We introduce another concept of fair distribution which is related to our topic. the core defines those distributions of the profit in which every coalition of players gets in total at least as much as they deserve (according to the characteristic function).

Definition 2.3.2 (Core)**.** The core of a Cooperative game V is $C(v) \subseteq \mathbb{R}^{|V|}$ in such a way that $x \in C(v)$ if and only if

$$\sum_{i \in V} x_v = v(V)$$

and for every $S \subset V$

$$\sum_{i \in S} x_i \geq v(S)$$

We have the following simple observation.

Proposition 2.3.3. *Let v be a transitive game. If the Shapley value vector $\phi(v)$ is in the core $C(v)$ then for every $S \subseteq V$*

$$v(S) \leq \frac{|S|}{|V|} v(V)$$

Proof. For transitive games, obviously $\phi_i(v) = \frac{v(V)}{|V|}$. Using that $\phi(v) \in C(v)$, we get that

$$v(S) \leq \sum_{i \in S} \phi_i(v) = \frac{|S|}{|V|} v(V)$$

□

We are going to show that a class of cooperative games, the so-called convex games satisfy the conditions of Proposition 2.3.3.

Definition 2.3.3 (Convex games)**.** A cooperative game v is convex if the characteristic function is supermodular. That is for every subset of players $S, T \subseteq [n]$

$$v(S) + v(T) \leq v(S \cup T) + v(S \cap T) \tag{2.3.2}$$

Recall that with any function f on a product space we can associate a game v_f by $v_f(U) := \mathrm{Var}[\mathbb{E}[f \mid \mathcal{F}_U]]$. We have another game if we define the information we gain via information theoretic concepts (see Definition 2.2.4) .

$$v_f^I(S) = I(Z : X_S)$$

It is not difficult to see that for product measures, both v_f and v_f^I are convex games. The entropy version is immediate from the submodularity of entropy, which can be written as:

$$-H(X_S|Z) - H(X_T|Z) \leq -H(X_{S\cap T}|Z) - H(X_{S\cup T}|Z)$$

Using that for independent variables the submodularity inequality is sharp we get

$$\begin{aligned} &H(X_S) - H(X_S|Z) + H(X_T) - H(X_T|Z) \leq \\ &H(X_{S\cap T}) - H(X_{S\cap T}|Z) + H(X_{S\cup T}) - H(X_{S\cup T}|Z). \end{aligned}$$

For the L_2 version the supermodularity of $\mathrm{Var}(\mathbb{E}[f \mid \mathcal{F}_U])$ follows easily from the spectral description. Here we show an argument that does not require Fourier-Walsh expansion or Efron-Stein decomposition.

Proposition 2.3.4. *Let $f : X^V \longrightarrow \{-1, 1\}$ endowed with a product measure. The set function (cooperative game) $v(S) = \mathrm{Var}(\mathbb{E}[f \mid \mathcal{F}_U])$ for ($S \subseteq V$) is supermodular (convex).*

Proof. First observe that whenever $S \subseteq T$ then $\mathbb{E}[\mathbb{E}[f \mid \mathcal{F}_T] \mid \mathcal{F}_S] = \mathbb{E}[f \mid \mathcal{F}_S]$ by the towering property, and using that conditional expectation is an orthogonal projection we get that

$$\mathrm{Var}(\mathbb{E}[f \mid \mathcal{F}_T]) - \mathrm{Var}(\mathbb{E}[f \mid \mathcal{F}_S]) = \mathrm{Var}(\mathbb{E}[f \mid \mathcal{F}_T] - \mathbb{E}[f \mid \mathcal{F}_S]),$$

and therefore (2.3.2) can be rewritten as

$$\mathrm{Var}(\mathbb{E}[f \mid \mathcal{F}_T] - \mathbb{E}[f \mid \mathcal{F}_{S\cap T}]) \leq \mathrm{Var}(\mathbb{E}[f \mid \mathcal{F}_{S\cup T}] - \mathbb{E}[f \mid \mathcal{F}_S]). \tag{2.3.3}$$

Fix $S, T \subseteq V$ such that $S \subseteq T$. Using Lemma 2.1.4 for $(T \setminus S)^c$ and T we get

$$\mathbb{E}[f \mid \mathcal{F}_S] = \mathbb{E}[\mathbb{E}[f \mid \mathcal{F}_{(T\setminus S)^c}] \mid \mathcal{F}_T].$$

Note that this is the only place in the argument where the fact that the underlying measure is a product measure is exploited.

This identity allows us to write $\mathbb{E}[f \mid \mathcal{F}_{S\cap T}] = \mathbb{E}[\mathbb{E}[f \mid \mathcal{F}_{(T\setminus S\cap T)^c}] \mid \mathcal{F}_T]$ and $\mathbb{E}[f \mid \mathcal{F}_S] = \mathbb{E}[\mathbb{E}[f \mid \mathcal{F}_{(S\cup T\setminus S)^c}] \mid \mathcal{F}_{S\cup T}]$. Since $T \setminus S \cap T = S \cup T \setminus S = T \setminus S$, (2.3.3) becomes

$$\mathrm{Var}(\mathbb{E}[f - \mathbb{E}[f \mid \mathcal{F}_{(T\setminus S)^c}] \mid \mathcal{F}_T]) \leq \mathrm{Var}(\mathbb{E}[f - \mathbb{E}[f \mid \mathcal{F}_{(T\setminus S)^c}] \mid \mathcal{F}_{S\cup T}]),$$

which always holds, because orthogonal projection cannot increase the variance (L^2-norm). □

The subgame v_U denotes the game v with its domain restricted to the subset $U \subseteq [n]$.

Lemma 2.3.5. *If v is convex and transitive game then $S \subseteq T$ implies*

$$\phi_i(v_S) \leq \phi_i(v_T)$$

Proof. We are going to show this when $T = S \cup \{j\}$. Let $|S| = k$ We have

$$\phi_i(v_S) = \frac{1}{k} \sum_{L \subseteq S \setminus \{i\}} \frac{v(L \cup \{i\}) - v(L)}{\binom{k-1}{|L|}} \leq \frac{1}{k+1} \sum_{L \subseteq T \setminus \{i\}} \frac{v(L \cup \{i\}) - v(L)}{\binom{k}{|L|}}$$

and

$$\phi_i(v_T) = \frac{1}{k+1} \left(\sum_{L \subseteq S \setminus \{i\}} \frac{v(L \cup \{i\}) - v(L)}{\binom{k}{|L|}} + \sum_{L \subseteq S \setminus \{i\}} \frac{v(L \cup \{i,j\}) - v(L \cup \{j\})}{\binom{k}{|L|+1}} \right)$$

It is a straightforward calculation to verify that for any $l \leq k$

$$\frac{1}{k} \frac{1}{\binom{k-1}{l}} = \frac{1}{k+1} \left(\frac{1}{\binom{k}{l}} + \frac{1}{\binom{k}{l+1}} \right)$$

and therefore, using that by supermodularity, $v(L \cup \{i\}) - v(L) \leq v(L \cup \{i,j\}) - v(L \cup \{j\})$, we get

$$\phi_i(v_S) = \frac{1}{k+1} \left(\sum_{L \subseteq S \setminus \{i\}} \frac{v(L \cup \{i\}) - v(L)}{\binom{k}{|L|}} + \sum_{L \subseteq S \setminus \{i\}} \frac{v(L \cup \{i\}) - v(L)}{\binom{k}{|L|+1}} \right) \leq \phi_i(v_{S \cup \{j\}})$$

□

Lemma 2.3.5 implies that for any $S \subset V$

$$v(S) = \sum_{i \in S} \phi_i(v_S) \leq \sum_{i \in U} \phi_i(v)$$

which shows that $\phi(v)$ is indeed in the core. It is clear that Theorem 2.1.6 and Theorem 2.2.1 follows immediately from Proposition 2.3.3 and Lemma 2.3.5.

Observe that for a transitive game with a non-empty core the Shapley value, i.e. the uniform vector will always be in the core. It is because the core is convex and itself is invariant under the group action. Therefore, one could weaken the condition of Proposition 2.3.3 by only requiring the non-emptiness of the core. A classical result in Cooperative Game Theory (see for example [BDT08] Theorem 2.4) gives necessary and sufficient conditions for this. It has to be said that on a practical level, the conditions of this Theorem are not very easy to verify.

Theorem 2.3.6 (Bondareva-Shapley)**.** *The core of the game v is non-empty if and only if for every $\alpha : 2^V \setminus \emptyset \to [0,1]$ such that for every $i \in V$*

$$\sum_{S \subseteq V \,:\, i \in S} \alpha(S) = 1$$

it holds that

$$\sum_{S \subseteq V} \alpha(S) v(S) \leq v(V)$$

2.4 Sparse Reconstruction for Planar Percolation

2.4.1 A Brief Introduction to Percolation Theory

Graph percolation theory arose historically in statistical mechanics in the 60s. The motivation was to understand the percolation of some liquid in a porous body. This is modeled by a random graph whose vertices, called sites, correspond to points in the body and the edges represent possible links between sites. Percolation theory aims at understanding some distributional properties of the connected components (referred to as clusters) in the above random graph.

There are a number of possible models depending on the randomization procedure. Here our focus is the most straightforward model, Bernoulli edge percolation. In this model each edge is open (meaning that the liquid can flow), independently from the status of any other edge. Edge percolation refers to the fact that the $0-1$ random variables that decide whether locally the fluid can pass, are assigned to edges contrary to site percolation, where they are assigned to the vertices of the graph.

As we want to have an automorphism invariant model we require that each edge of the graph has the same probability p to be open. In case the graph is infinite, it is a question of particular interest whether for a particular value of p the graph contains an infinite cluster or not. This is a tail event, consequently for any value of p either there is or there is not such a cluster, almost surely. A simple coupling argument shows that by increasing p we can only introduce an infinite cluster. It has been shown accordingly that any infinite connected graph admits a critical value p_c:

Definition 2.4.1 (Critical Probability). $p_c = \inf\{p : \mathbb{P}_p(\exists\, \infty \text{ cluster}) = 1\}$

It tuns out that the critical model in many graphs displays interesting, fractal-like features. There is a universality principle coming from statistical physics which connects the behaviour of various graph models around phase transition. The idea is that, although the low level description of models may differ, ultimately they all describe the same high-level phenomenon. Physicists believe that any graph percolation that comes from a nice d-dimensional lattice describes the same 'ideal' d-dimensional percolation at the critical probability p_c, only in possibly different frames.

This principle suggests the existence of so-called critical exponents, which describe the probability of important observables of the percolation at the phase transition (i.e. at $p = p_c$) via universal power laws. For example, although the value of p_c may vary from graph to graph, the Hausdorff dimension of the connectivity clusters at criticality is believed to be the same. Physicists can calculate the value of these exponents and they believe that these values are universal in the sense above. Nevertheless, from the point of view of the mathematician, little is actually known.

In this chapter we consider Bernoulli edge percolation on the $n \times n$ square lattice with $p = 1/2$. Our main focus will be the left right crossing event LR_n. This is the event that there exist to vertices x from the left boundary of the square and y from the right in such a way that there is a path consisting of open edges between x and y.

It is not too difficult to show that when $p = \frac{1}{2}$ then $\mathbb{P}[\mathsf{LR}_n]$ tends to $\frac{1}{2}$ as $n \to \infty$. Every percolation configuration on a square (or, in fact, any planar) lattice induces a percolation configuration on the dual lattice. In the dual lattice the sites are faces and two faces are connected in configuration if the two faces are bordered with an edge which is closed in the original percolation.

The proof uses two observations. First, that $n-1 \times n$ rectangles are self dual and second that there is a left-right crossing in the original configuration if and only if there is an up-down crossing in the dual configuration. In the sequel, we shall also make use of this rotational symmetry of the percolation model.

This suggests that the critical probability for $\mathbb{Z}^2$ is $p_c = \frac{1}{2}$. This is indeed the case, but it is far from trivial to prove this. It has to be noted that in case $p < \frac{1}{2}(p > \frac{1}{2})$ the probability of LR_n goes to $0(1)$ exponentially fast.

Critical planar percolation has been more extensively studied, and there has been some important developments in the last few decades. The main breakthrough was by Smirnov [Sm01], who showed that in the case of the triangular lattice the universality conjecture of the physicists holds, in particular the value of the critical exponent is as predicted. For the square lattice, however (and for any planar lattice) no similar result has been proved.

Arm exponents

There is a family of critical exponents that we would like to highlight since it plays an important part in the sequel.

The 1-arm event on $\mathbb{Z}^2$) (we only consider this lattice, but the arm events can be defined for any transitive planar lattice) $A_1(R)$ is the event that there is a path of open edges from 0 to a site (vertex) which is at graph distance R away from the origin. The event $A_1(R,r)$ is the event that there is path of open edges starting somewhere in distance at most r from 0 and ending at a site which is at distance R from the origin. It is conjectured based on the above universality principle that in any reasonable lattice (on $\mathbb{Z}^2$, in particular)

$$\alpha_1(R,r) := \mathbb{P}[A_1(R)] \asymp \left(\frac{r}{R}\right)^{\frac{5}{48}+o(1)}.$$

This was, in fact proven for site percolation on the triangular lattice in []. Up until today this is the only lattice where this exponent is verified.

There are other arm events that are of interest, for example the four arm event, which is closely related to the pivotals of the crossing event LR_n. In case of the four arm event we also require that every second arm needs to go trough dual edges. That is, two arms of open paths is separated by two dual arms on each side.

In our proof we are going to use another event, the 3-arm event in a half plane which we denote by $A_3^+(R,r)$. This is the event that there are two paths of open edges in the positive half plane $\mathbb{Z} \times \mathbb{N}$ starting at distance r from the origin and reaching until distance R, and the two open arms are separated with a similar arm consisting of dual edges (which is also entirely in the half plane $\mathbb{Z} \times \mathbb{N}$).

It turns out that the exponent of $A_3^+(R,r)$ is known for $\mathbb{Z}^2$. There is a combinatorial argument that does not rely on the universality conjecture. The heuristics is, strange as it may seem, that fractional arm exponents are hard, while integer arm exponents are approachable.

Proposition 2.4.1 ([LSW02]). *For the $\mathbb{Z}^2$ lattice*

$$\mathbb{P}[A_3^+(R,r)] = \alpha_3^+(R,r) \asymp \left(\frac{r}{R}\right)^{2+o(1)}.$$

2.4.2 Sparse Reconstruction for Planar Percolation

In this section we are going to show that the left-right crossing event in critical planar percolation cannot be reconstructed from a sparse subset of spins, by this answering a question posed by Itai Benjamini. Note first that we can use the frame work of Boolean functions, since a percolation configuration can be described as an element in $\{-1,1\}^E$, where E is the edge set of the graph. We shall denote by $\mathsf{LR}_n : \{-1,1\}^{E(\mathbb{Z}_n^2)} \longrightarrow \{-1,1\}$ the indicator function of the planar crossing event. More precisely, $\mathsf{LR}(\omega)$ is 1 if there is a left-right crossing of 1s in ω and -1 otherwise.

Theorem 2.1.1, however, does not apply for this question, since the left-right crossing is not a transitive event. Still we shall make use of the results of the previous section. We argue that the left-right crossing event is in fact not too far from being transitive.

Here is a brief summary of what we are going to do: Let us denote by LR the characteristic function of the left-right crossing. We will show that for every ϵ there is a sublattice (subgroup) $H_\epsilon \subseteq \mathbb{Z}_n^2$ whose size only depends on ϵ and $M^{H_\epsilon}[\mathsf{LR}_n]$, the projection of LR_n into the space of H_ϵ-invariant functions, is close to a transitive function (in the sense that $\mathrm{Corr}(\mathsf{LR}^{H_\epsilon}, g) \geq 1 - O(\epsilon)$ for some transitive function g). This g is in fact the projection of LR_n to the space of $\mathbb{Z}_n^2$-invariant functions. As we shall see, in case two functions are highly correlated there clue with respect to a particular subset is also close.

Now if the crossing event LR_n had uniformly positive clue with respect to some sequence of subsets U_n, the projection $M^{H_\epsilon}[\mathsf{LR}_n]$ would also have high clue with respect to the union of the original subset U_n and its H_ϵ-translates, which is still small since ϵ is fixed.

But this is impossible because then in turn the transitive function g being highly correlated with $M^{H_\epsilon}[\mathsf{LR}_n]$ would also have had uniformly positive clue with respect to a sparse sequence of subsets which is contradiction with Theorem 2.1.1.

While this question has not been investigated in this general form, in [GPS10] there have already been a number of partial results. concerning the information content of some particular sparse subsets. Based on a deep analysis of the Fourier spectrum of the percolation crossing event upper bounds for the clue of some particular sequences of small subsets of bits has been established.

Here are a few examples of these sort of results from [GPS10]. If U_n^c is a random set of bits of density $n^{-\frac{3}{4}+\epsilon}$, than $\mathsf{clue}(U_n) \to 0$. Also, it is known that if every disk of radius $n^{\frac{3}{8}-\epsilon}$ contains a bit from U_n^c, then $\mathsf{clue}(U_n) \to 0$. On the other hand, if U_n^c has a scaling limit of Hausdorff-dimension strictly less than $\frac{5}{4}$, then $\mathsf{clue}(U_n) \to 1$.

It is also known that there is a revealment algorithm for the crossing event of the percolation (on the triangular lattice) with revealment $\delta \sim n^{-\frac{1}{4}}$ ([SS10]). This, in particular, implies that any sequence of sets S_n of size $o(n^{\frac{1}{4}})$ is asymptotically clueless, since denoting the random set of queried bits by J_n, we have:

$$\mathbb{E}|S_n \cap J_n| = \sum_{i \in S_n} \mathbb{P}[i \in J_n] \leq \sum_{i \in S_n} \delta \sim |S_n| n^{-\frac{1}{4}} \to 0$$

whenever $|S_n| = o(n^{\frac{1}{4}})$. It is clear that if S_n is asymptotically disjoint from J_n, it cannot admit a large clue.

We now turn to the proof. Since our proof uses the fact that, roughly speaking the percolation crossing event is almost transitive, we need to define the projection operation that sends a function to the space of functions that are invariant under some specific group action.

In case we have a group G acting on E we shall also consider an invariant (Haar) probability measure on G. As we are in a finite setting this will be simply the uniform measure on a finite group G.

The space of Boolean functions over a given configuration space $\{-1,1\}^E$ and a corresponding probability measure (the uniform measure in this case) can be endowed with a Hilbert space structure via the scalar product $\langle f,g\rangle := \mathbb{E}[fg]$. There is a natural operation that turns an arbitrary function f into a G-invariant function on the same space:

$$M^G[f] := \frac{1}{|G|}\sum_{g\in G} f^g \tag{2.4.1}$$

In case the action of G is transitive than, of course, Theorem 2.1.1 applies for $M^G[f]$. We will now show that the operator M^G is the orthogonal projection onto the subspace of G-invariant functions. First, it is obvious that $M^G[f] = f$ whenever g is G-invariant.

For any $g \in G$

$$\sum_{h\in G} \mathbb{E}\left[f(x)f^h(x)\right] = \sum_{h\in G} \mathbb{E}\left[f^g(x)f^h(x)\right] \tag{2.4.2}$$

using that by the G-invariance of the measure $\mathbb{E}\left[f(x)f^h(x)\right] = \mathbb{E}\left[f^{g_1}(x)f^{hg_1}(x)\right]$, which allows us to write

$$\mathbb{E}[M^G[f]^2] = \frac{1}{|G|^2}\sum_{g\in G}\sum_{h\in G}\mathbb{E}\left[f^g(x)f^h(x)\right] = \frac{1}{|G|}\sum_{g\in G}\mathbb{E}\left[f^g(x)f(x)\right] = \mathbb{E}\left[M^G[f]f\right],$$

where we used 2.4.2. Therefore, we can conclude that

$$\mathbb{E}\left[(f - M^G[f])M^G[f]\right] = 0,$$

which shows that M^G is a projection operator.

The sum

$$S(f) = \sum_{g\in G}\mathbb{E}[ff^g]$$

is called the susceptibility of f and we will discuss it in more detail in Section 3.1.3. It is clear from the above that for any $f : \{-1,1\}^V \longrightarrow \mathbb{R}$

$$\mathbb{E}[M^G[f]^2] = \frac{1}{|G|}S(f) \tag{2.4.3}$$

In the same way as above by the G-invariance of the measure $\mathrm{Cov}(f^{g_1}, f^{g_2}) = \mathrm{Cov}(f, f^{g_2-g_1})$ and therefore

$$\mathrm{Var}(M^G[f]) = \frac{1}{|G|^2}\sum_{g\in G}\sum_{h\in G}\mathrm{Cov}(f^g, f^h) = \frac{1}{|G|}\sum_{g\in G}\mathrm{Cov}(f, f^g) \tag{2.4.4}$$

The following two simple technical lemmas will be useful to estimate how much a projection can distort correlations. The geometrical intuition is that in case the correlation of two functions is high and the projection is not too 'radical' (meaning here that it does not decrease the norm drastically), then the correlation will be preserved by the projection. Note that these results are completely general, we do not make use of the fact that there is an underlying product measure.

Lemma 2.4.2. *Let $f, g \in L^2(S,\mu)$ satisfying*

$$\mathrm{Corr}(f,g) \geq 1-\epsilon.$$

Let U be a subspace of $L^2(S,\mu)$ and let us denote by P the orthogonal projection onto this subspace. Assume that

$$\mathsf{clue}(f \mid U) \geq c, \quad \text{and } \mathsf{clue}(g \mid U) \geq c.$$

Then

$$\mathrm{Corr}(P[f], P[g]) \geq 1 - \frac{\epsilon}{c}.$$

Proof. Without loss of generality we may assume that $\mathbb{E}[f] = \mathbb{E}[g] = 0$ and $\mathrm{Var}(f) = \mathrm{Var}(g) = 1$, since both clue and correlation are invariant under linear transformations. As in this case they are equivalent, we may use $\| \ \|^2$ instead of the variance, depending on the context.

Using that the the variance of f and g are equal, we get

$$\begin{aligned}\|f-g\|^2 =&\mathrm{Var}(f) + \mathrm{Var}(g) - 2\sqrt{\mathrm{Var}(f)\mathrm{Var}(g)}\mathrm{Corr}(f,g) = \\ &2(1-\mathrm{Corr}(f,g)) \leq 2\epsilon,\end{aligned}$$

In a similar fashion, we get for the projected functions that

$$\|P[f]-P[g]\|^2 =$$
$$\sqrt{\mathrm{Var}(P[f])\mathrm{Var}(P[g])}\left(\frac{\mathrm{Var}(P[f])}{\mathrm{Var}(P[g])} + \frac{\mathrm{Var}(P[g])}{\mathrm{Var}(P[f])} - 2\mathrm{Corr}(P[f],P[g])\right)$$

Now noting that

$$\sqrt{\mathrm{Var}(P[f])\mathrm{Var}(P[g])} \geq c$$

and

$$\frac{\mathrm{Var}(P[f])}{\mathrm{Var}(P[g])} + \frac{\mathrm{Var}(P[g])}{\mathrm{Var}(P[f])} \geq 2,$$

we obtain that

$$\|P[f]-P[g]\|^2 \geq 2c(1-\mathrm{Corr}(P[f],P[g])).$$

Since P is a projection it can only decrease the L^2 norm and therefore:

$$\|f-g\|^2 \geq \|P[f]-P[g]\|^2.$$

Finally, putting together estimates for $\|f-g\|^2$ and $\|P[f]-P[g]\|^2$ we conclude that

$$2\epsilon \geq \|f-g\|^2 \geq \|P[f]-P[g]\|^2 \geq 2c(1-\mathrm{Corr}(P[f],P[g]).$$

After reordering the inequality the statement follows. □

Lemma 2.4.3. *Let $f, g \in L^2(S,\mu)$ with*

$$\mathrm{Corr}(f,g) \geq 1-\epsilon.$$

Let P denote the orthogonal projection onto the subspace U of $L^2(S,\mu)$ and suppose that

$$\mathsf{clue}(f \mid U) \geq c.$$

Under these conditions

$$\mathrm{Corr}(P[f], P[g]) \geq 1 - \frac{\epsilon}{c-2\epsilon}$$

and

$$\mathsf{clue}(g \mid U) \geq c - 2\epsilon.$$

Proof. Again, without loss of generality we may assume that $\mathbb{E}[f] = \mathbb{E}[g] = 0$ and $\mathrm{Var}(f) = \mathrm{Var}(g) = 1$ and therefore we may use $\| \quad \|^2$ instead of variance, like previously.

Using that $P[f]$ is the closest point to f in U for every $h \in U$

$$\|f - P[f]\|^2 \leq \|f - h\|^2.$$

Therefore with the triangle inequality we get

$$\|g - P[g]\|^2 \leq \|g - P[f]\|^2 \leq \|g - f\|^2 + \|f - P[f]\|^2. \tag{2.4.5}$$

Recall that, since P is an orthogonal projection for every $f \in L^2(S, \mu)$ we have

$$\|P[f]\|^2 + \|f - P[f]\|^2 = \|f\|^2 \tag{2.4.6}$$

As in Lemma 2.4.2 $\mathrm{Corr}(f, g) \geq 1 - \epsilon$ implies $\|g - f\|^2 \leq 2\epsilon$.

On the other hand, by our assumptions

$$\mathsf{clue}(f \mid U) = \frac{\|P[f]\|^2}{\|f\|^2} = \|P[f]\|^2 \geq c$$

and (2.4.6) shows that $\|f - P[f]\|^2 \leq 1 - c$. Therefore plugging in the estimates into (2.4.5) we can write (using that dividing by $\|g\|^2 = 1$ does not change the equation)

$$\frac{\|g - P[g]\|^2}{\|g\|^2} \leq 2\epsilon + (1 - c).$$

Using (2.4.6) again, we get

$$1 - \mathsf{clue}(g \mid U) \leq 2\epsilon + 1 - c$$

from which $\mathsf{clue}(g \mid U) \geq (c - 2\epsilon)$ is immediate.

We can apply Lemma 2.4.2 to get that $\mathrm{Corr}(P[f], |P[g]) \geq 1 - \frac{\epsilon}{c - 2\epsilon}$ □

In the sequel D_δ will denote the rectangle $[-\delta n, \delta n]^2 \subset \mathbb{Z}_n^2$. Obviously $|H_\delta| = \frac{1}{\delta^2}$ and $|D_\delta| = (\delta n)^2$.

Lemma 2.4.4. *Let $D_\delta := [-\delta n, \delta n]^2$ as above. Then there is a $K > 0$ such that for every $d_1, d_2 \in D_\delta$*

$$\mathrm{Corr}(\mathsf{LR}^{d_1}, \mathsf{LR}^{d_2}) \geq 1 - K\delta$$

Proof. Let $d =\in D_\delta$. We are going to show that

$$\mathbb{P}[\mathsf{LR}^0 \neq \mathsf{LR}^t] \leq O(\delta).$$

From this the statement of the lemma follows. Indeed, for any $d_1, d_2 \in D_\delta$

$$\mathrm{Corr}(\mathsf{LR}^{d_1}, \mathsf{LR}^{d_2}) = 1 - 2\mathbb{P}[\mathsf{LR}^{d_1} \neq \mathsf{LR}^{d_2}] = 1 - 2\mathbb{P}[\mathsf{LR} \neq \mathsf{LR}^d] \geq 1 - O(\delta).$$

Let us assume that $d = (0, t)$. Observe (see Figure **??**) that in both cases the event $\left\{\mathsf{LR}^0 \neq \mathsf{LR}^d\right\}$ entails a 3-arm event in a half plane from radius $O(\delta)$ to distance $O(1)$.

As the grid size approaches to zero, the probability of the 3-arm event is in quadratic order by Proposition 2.4.1:

$$\alpha_3^+(\delta, 1) = O(\delta^2)$$

The 3-arm event happens in one of $O(\frac{1}{\delta})$ different $\delta \times \delta$ boxes, so by the union bound

$$\mathbb{P}[\mathsf{LR}^0 \neq \mathsf{LR}^t] \leq \alpha_3^+(\delta, 1) O(\frac{1}{\delta}) = O(\delta)$$

In case $d = (t, 0)$ we have exactly the same argument exploiting the $\pi/2$ rotational symmetry of the model (switching to the dual lattice and using that LR_n does not happen if and only if there is a dual up-down crossing.

The case of a general $d \in D_\delta$ now easily follows. If $\left\{\mathsf{LR}^0 \neq \mathsf{LR}^t\right\}$ then either $\left\{\mathsf{LR}^0 \neq \mathsf{LR}^{d^1}\right\}$ or $\left\{\mathsf{LR}^0 \neq \mathsf{LR}^{d^2}\right\}$, where d^1 and d^2 are the projections of d onto the first and the second coordinates, respectively.

As a consequence, $\mathbb{P}[\mathsf{LR}^0 \neq \mathsf{LR}^d] \leq \mathbb{P}[\mathsf{LR}^0 \neq \mathsf{LR}^{d^1}] + \mathbb{P}[\mathsf{LR}^0 \neq \mathsf{LR}^{d^2}] \leq O(\delta)$. □

In the sequel we will use the function $M^{\mathbb{Z}_n^2}[\mathsf{LR}]$, which is obviously a transitive function, as $\mathbb{Z}_n^2$ acts transitively on the coordinates.

We shall also consider projections to a coarser grid of mesh size δ. The subgroup H_δ is the group generated by the two perpendicular elements $(0, \delta n)$ and $(\delta n, 0)$

Lemma 2.4.5. *Let $\epsilon > 0$. Then there is a $K > 0$ such that*

$$\mathrm{Corr}(M^{\mathbb{Z}_n^2}[\mathsf{LR}], M^{H_{\epsilon^2}}[\mathsf{LR}]) \geq 1 - K\epsilon.$$

Proof. We consider a new spin system σ on the $\mathbb{Z}_n^2$ torus which is a factor of the uniform Bernoulli percolation on the edges. At every vertex $v \in \mathbb{Z}_n^2$ $\sigma_v = \mathsf{LR}^v$.

The outline of the proof is as follows: First we observe that for a randomly chosen $\delta n \times \delta n$ square the value of σ is the same on the four vertices of the square, with probability $1 - O(\delta)$. For a fixed configuration we call a square on the δn-grid *good* if this is the case, *bad* otherwise.

The second step is to show that the event that there exist a point t inside a good square such that σ_t differs from the value of σ on the vertices of the square also happens with probability at most $1 - O(\delta)$. These two claims together suffice to show that the average on the $\frac{1}{\epsilon^2}$ grid already gives a good approximation about the average on the entire torus $\mathbb{Z}_n^2$.

Define the event $A =: \left\{\sigma_0 = \sigma_{(0,n\delta)} = \sigma_{(n\delta,0)} = \sigma_{(n\delta,n\delta)}\right\}$. By Lemma 2.4.4 and the union bound

$$\mathbb{P}[A^c] \leq 2(\mathbb{P}[\sigma_0 \neq \sigma_{(0,\delta)}] + \mathbb{P}[\sigma_0 \neq \sigma_{(\delta,0)}]) \leq O(\delta).$$

Because of the translation invariance of the measure this means that on average all except $O(\delta)$ portion of the $\frac{1}{\delta^2}$ small squares are 'good'. This translates to the the following bound on the number of bad $\delta n \times \delta n$ squares:

$$\mathbb{E}[|\text{bad squares}|] \leq O(\delta)\frac{1}{\delta^2} = O(\frac{1}{\delta}). \tag{2.4.7}$$

Let B denote the event that for every $t \in \mathbb{Z}_n^2 \cap [0, \delta n]^2$ the values σ_t are the same. We are going to show that $\mathbb{P}[B^c \cap [0, \delta n]^2$ is a good square$] \leq O(\delta)$. With other words, if a square on the δn grid has the same value on all of the four vertices of the square then with high probability this is the value everywhere inside the square.

First, observe that the event $B^c \cap \{[0, \delta n]^2$ is a good square$\}$ implies the existence of an 'alternating' triple t_1, t_2, t_3 on a vertical or horizontal line segment of length at most $n\delta$ on the torus such that $\sigma_{t_1} = \sigma_{t_3}$ but $\sigma_{t_1} \neq \sigma_{t_2}$. Indeed, if there is a t on the edges of the square such that $\sigma_t \neq \sigma_{v_i}$, we are ready, otherwise the point inside with its projections to two parallel edges of $[0, \delta n]^2$ will do.

This configuration, in a similar way as in Lemma 2.4.4, implies the existence of two 3-arm events in two disjoint half planes both from distance δ to $O(1)$ (see Figure **??**), and this enables us to give an upper bound on $\mathbb{P}[B^c \mid [0, \delta n]^2 \text{ is a good square}]$.

Let us denote by d the distance on the unit square of the two δ boxes where the two 3-arm events start. Clearly, there are two, independent 3-arm events for half plane from distance δn to $\frac{dn}{2}$ (as they are supported on disjoint bits) and also two 3-arm event for half plane from distance $\frac{dn}{2}$ to $O(1)n$. The former two are also independent as they are realized in two disjoint half planes so again they are supported on disjoint bits. Thus the probability of this is, by Proposition 2.4.1

$$\left(\alpha_3^+(\delta, \frac{d}{2})\right)^2 \left(\alpha_3^+(\frac{d}{2}, O(1))\right)^2 = O(\left(\frac{\delta}{d}\right)^2) O(\left(\frac{d}{1}\right)^2) = O(\delta^4),$$

independently from the distance d.

One of the 3-arm events can be started at any of $O(\frac{1}{\delta^2})$ different $\delta n \times \delta n$-boxes and once this is fixed, the second one can be chosen $O(\frac{1}{\delta})$ different ways (since it has to be δ close to the other one in at least one of the coordinates). Therefore the two 3-arm events can be realized in $\frac{1}{\delta^3}$ different ways, and the union bound gives that

$$\mathbb{P}[B^c \cap [0, \delta n]^2 \text{ is a good square}] \leq O(\frac{1}{\delta^3}) O(\delta^4) = O(\delta). \tag{2.4.8}$$

Let us call a $\delta n \times \delta n$ square *perfect*, if it is good and for any t in the box the σ_t values are the same as those in the vertices of the square. A square is *imperfect*, if it is not perfect. Using this expressions (2.4.8) says that the probability that a square is good, but not perfect is small. Therefore, putting together (2.4.7) and (2.4.8) we get that

$$\mathbb{P}[\ [0, \delta n]^2 \text{ is perfect}] \geq 1 - \mathbb{P}[\ [\delta n]^2 \text{ is bad}] - \mathbb{P}[B^c \cap [0, \delta n]^2 \text{ is good}] \geq 1 - O(\delta).$$

We are now ready estimate the correlation. For a given ϵ choose $\delta = \epsilon^2$. Applying Markov's inequality gives

$$\mathbb{P}[|\text{imperfect squares}| > \epsilon \frac{1}{\delta^2}] = \mathbb{P}[|\text{imperfect squares}| > \frac{1}{\epsilon^3}] \leq \frac{O(1/\epsilon^2)}{1/\epsilon^3} = O(\epsilon).$$

But on the event $\left\{|\text{imperfect squares}| \leq \epsilon \frac{1}{\delta^2}\right\}$ we have $\left| M^{\mathbb{Z}_n^2}[\mathsf{LR}] - M^{H_{\epsilon^2}}[\mathsf{LR}] \right| \leq \epsilon$. So

$$\mathbb{P}[|M^{\mathbb{Z}_n^2}[\mathsf{LR}] - M^{H_{\epsilon^2}}[\mathsf{LR}]| \leq \epsilon] \geq 1 - O(\epsilon),$$

which implies

$$\mathrm{Corr}(M^{\mathbb{Z}_n^2}[\mathsf{LR}], M^{H_{\epsilon^2}}[\mathsf{LR}]) \geq 1 - O(\epsilon).$$

because $|M^{\mathbb{Z}_n^2}[\mathsf{LR}] - M^{H_{\epsilon^2}}[\mathsf{LR}]| \leq 2$.

□

Now we are ready to prove the main result of this section.

Theorem 2.4.6. *There is no Sparse reconstruction for the left-right crossing in critical planar percolation*

Proof. Let $U_n \subseteq \mathbb{Z}_n^2$ be a sparse sequence of subsets, i.e., $\lim_n \frac{|U_n|}{n^2} = 0$. Indirectly, we assume that there is a $c > 0$ such that $\mathsf{clue}(\mathsf{LR}_n \mid U_n) > c$ for every large n.

We start by giving an outline of the proof. Fix two grid sizes a finer δ and a coarser one η so $0 < \delta < \eta$. We are going to show that the assumption that there is a sparse sequence of subsets with clue greater then $c > 0$ for the crossing event, then the average of the translated crossing events on the δ-grid $M^{H_\delta}[\mathsf{LR}_n]$ also has clue greater then $c' > 0$ for a larger, but still sparse sequence of subsets U_n^δ (where c' depends on η, but η does not depend on n).

But Lemma 2.4.5 shows that the average of the translates on the δ-grid and the average of all translates $M^{\mathbb{Z}_n^2}[\mathsf{LR}_n]$ are highly correlated. Therefore, the same sequence of sparse subsets also gives us positive amount of clue about $M^{\mathbb{Z}_n^2}[\mathsf{LR}_n]$. But this is in contradiction with Theorem 2.1.1, which claims that a sequence of sparse subsets cannot give positive clue about a transitive function.

We define the set $U_n^\delta = \cup_{t \in H_\delta} U^t$, where $U^t = \{u + t \,:\, u \in U\}$. So U_n^δ is just the union of all H_δ-translates of U. Clearly, $\mathsf{clue}(\mathsf{LR}_n \mid U_n^\delta) \geq c$.

We start by giving a lower bound on $\mathsf{clue}(M^{H_\delta}[\mathsf{LR}_n] \mid U_n^\delta)$. We will denote by P the projection (conditional expectation, from the probabilistic point of view) onto $\mathcal{F}_{U_n^\delta}$. With this notation

$$\mathsf{clue}(M^{H_\delta}[\mathsf{LR}_n] \mid U_n^\delta) = \frac{\mathrm{Var}(P[M^{H_\delta}[\mathsf{LR}_n]])}{\mathrm{Var}(M^{H_\delta}[\mathsf{LR}_n])}.$$

As the Bernoulli measure is $\mathbb{Z}_n^2$-invariant, we clearly have $\mathrm{Var}(\mathsf{LR}_n) = \mathrm{Var}(\mathsf{LR}_n^g)$ for every $g \in \mathbb{Z}_n^2$ and just like in (2.4.4)

$$\mathrm{Var}(M^{H_\delta}[\mathsf{LR}_n]) = \frac{1}{|H_\delta|}\sum_{h \in H_\delta} \mathrm{Cov}(\mathsf{LR}_n, \mathsf{LR}_n^h) =$$
$$\mathrm{Var}(\mathsf{LR}_n)\frac{1}{|H_\delta|}\sum_{h \in H_\delta} \mathrm{Corr}(\mathsf{LR}_n, \mathsf{LR}_n^h) \leq \mathrm{Var}(\mathsf{LR}_n).$$

We continue by giving a lower bound for the variance of $P[M^{H_\delta}[\mathsf{LR}_n]]$. Let $D = [-\eta n, \eta n] \times [-\eta n, \eta n]$.

$$\begin{aligned} &\mathrm{Var}(P[M^{H_\delta}[\mathsf{LR}_n]]) = \\ &\frac{1}{|H_\delta|^2}\sum_{h_1 \in H_\delta}\sum_{h_2 \in H_\delta} \mathrm{Cov}(P[\mathsf{LR}_n^{h_1}], P[\mathsf{LR}_n^{h_1+h_2}]) \geq \\ &\frac{1}{|H_\delta|^2}\sum_{h_1 \in H_\delta}\sum_{d \in D \cap H_\delta} \mathrm{Cov}(P[\mathsf{LR}_n^{h_1}], P[\mathsf{LR}_n^{h_1+d}]) = \\ &\frac{\mathrm{Var}(P[\mathsf{LR}_n])}{|H_\delta|^2}\sum_{h_1 \in H_\delta}\sum_{d \in D \cap H_\delta} \mathrm{Corr}(P[\mathsf{LR}_n^{h_1}], P[\mathsf{LR}_n^{h_1+d}]). \end{aligned} \tag{2.4.9}$$

For the inequality we used that LR_n is monotone and therefore by the FKG-inequality $\mathrm{Cov}(P[\mathsf{LR}_n^{h_1}], P[\mathsf{LR}_n^{h_2}] \geq 0$ and after that U_n^δ is H_δ-invariant and therefore $\mathrm{Var}(P[\mathsf{LR}_n]) = \mathrm{Var}(P[\mathsf{LR}_n^h])$ for any $h \in H_\delta$.

By Lemma 2.4.4 there exists a $K > 0$ such that

$$\mathrm{Corr}(\mathsf{LR}_n^t, \mathsf{LR}_n^{t+d}) \geq 1 - K\eta$$

for every $d \in D$ and $t \in \mathbb{Z}_n^2$. Applying Lemma 2.4.2 for LR_n^t, LR_n^{t+d} and P and choosing η small enough to assure $2K\eta < c/2$ we get that

$$\mathrm{Corr}(P[\mathsf{LR}_n^t], P[\mathsf{LR}_n^{t+d}]) \geq 1 - \frac{K\eta}{c - 2K\eta} \geq 1 - \frac{2K\eta}{c}.$$

By substituting this into (2.4.9), we obtain the following bound lower bound

$$\mathrm{Var}(P[M^{H_\delta}[\mathsf{LR}_n]]) \geq \ \mathrm{Var}(P[\mathsf{LR}_n])\frac{|D \cap H_\delta|}{|H_\delta|}\left(1 - \frac{2K\eta}{c}\right) = \eta^2 \mathrm{Var}(P[\mathsf{LR}_n])\left(1 - \frac{2K\eta}{c}\right),$$

where we used that $|H_\delta| = \frac{1}{\delta^2}$ and $|D \cap H_\delta| = \eta^2/\delta^2$, and thus $|D \cap H_\delta|/|H_\delta| = \eta^2$.

The lower bound for $\mathrm{Var}(P[M^{H_\delta}[\mathsf{LR}_n]])$ and the upper bound for $\mathrm{Var}(M^{H_\delta}[\mathsf{LR}_n])$ together yields (recall that $\mathrm{Var}(P[\mathsf{LR}_n]/\mathrm{Var}(\mathsf{LR}_n) \geq c)$

$$\begin{aligned}\mathsf{clue}(M^{H_\delta}[\mathsf{LR}_n] \mid U_n^\delta) = \frac{\mathrm{Var}(P[M^{H_\delta}[\mathsf{LR}_n]])}{\mathrm{Var}(M^{H_\delta}[\mathsf{LR}_n])} \geq \\ \frac{\mathrm{Var}(P[\mathsf{LR}_n]}{\mathrm{Var}(\mathsf{LR}_n)}\eta^2\left(1 - \frac{2K\eta}{c}\right) \geq \eta^2(c - 2K\eta).\end{aligned} \tag{2.4.10}$$

In order to simplify the notation we shall write $M[\ ]$ for the operator $M^{\mathbb{Z}_n^2}[\ ]$. By Lemma 2.4.5

$$\mathrm{Corr}(M^{H_\delta}[\mathsf{LR}_n], M[\mathsf{LR}_n]) \geq 1 - O(\sqrt{\delta}).$$

Applying Lemma 2.4.3 again with $M^{H_\delta}[\mathsf{LR}_n]$ and $M[\mathsf{LR}_n]$ we get from (2.4.10) that

$$\mathsf{clue}(M[\mathsf{LR}_n] \mid U_n^\delta) \geq \eta^2(c - 2K\eta) - O(\sqrt{\delta}).$$

But $M[\mathsf{LR}_n]$ is transitive and Theorem 2.1.1 tells us that

$$\frac{1}{\delta^2}\frac{|U_n|}{n} \geq \mathsf{clue}_{M[\mathsf{LR}_n]}(M[\mathsf{LR}_n] \mid U_n^\delta).$$

Here we used that obviously $|U_n^\delta| \leq \frac{1}{\delta^2}|U_n|$. Comparing the lower bound an the upper bound for $M[\mathsf{LR}_n]$

$$\frac{1}{\delta^2}\frac{|U_n|}{n} \geq \eta^2(c - O(\eta)) - O(\sqrt{\delta}).$$

After reordering and choosing $\delta = \eta^6$

$$\frac{1}{\delta^2\eta^2}\frac{|U_n|}{n} + \frac{O(\sqrt{\delta})}{\eta^2} + O(\eta) = \frac{1}{O(\eta^{14})}\frac{|U_n|}{n} + O(\eta) \geq c.$$

But this is a contradiction, since right hand side can be made arbitrary small by first choosing a sufficiently small η to make the term $O(\eta) < \epsilon/2$ and after selecting n large enough so that $\frac{1}{O(\eta^{14})}\frac{|U_n|}{n} < \epsilon/2$ as well. □

Chapter 3

Sparse Reconstruction in Spin Systems

3.1 Results for General Spin Systems

3.1.1 Introduction

We have seen in Chapter 2 (Theorem 2.1.1 and Theorem 2.1.6) that if we endow the configuration space with a product measure then sparse reconstruction is not possible for transitive functions. That is, for any sequence of transitive functions and any sequence of subsets of the coordinates the clue of the sequence vanishes as n goes to infinity. In the present Chapter we will investigate the same sort of questions for different sequences of probability measures on the hypercube.

In order to ensure that the question makes sense we will have to require that the probability measure in question is invariant under the action of some group Γ_n, where Γ_n acts transitively on the coordinate set V_n. We pose an additional requirement, namely that the sequence of probability measure has to be weekly convergent. We hope that certain properties of the limiting measure give information whether the sequence admits sparse reconstruction or not.

It turns out, however, that if we want to anchor our sequence to a limiting spin system, we need a somewhat stronger link then weak convergence of measures. We also have to ensure that the symmetries in the sequence and in the limit are consistent. We require that coordinate set V_n has a graph structure $G_n = (V_n, E_n)$ consistent with the symmetries in the sense that $\Gamma_n \leq \mathsf{Aut}(G_n)$, and the graph sequence converges locally to $G(V, E)$.

Local convergence of transitive graph means that for every $r \in \mathbb{N}$ there is an $N_r \in \mathbb{N}$ such that whenever $n \geq N_r$ then for any vertex v the r-neighbourhood of $v \in V_n$ (counted in graph distance) $B_r(v)$ is isomorphic to the r-neighbourhood of any vertex $v \in V$ from the vertex set of G.

There is one pitfall that we need to avoid. In case μ_n or μ is not invariant under the full automorphism group of G_n or G, respectively, we risk that calling two rooted neighbourhoods isomorphic, because they are isomorphic as rooted graphs, but they have different distributions (as the measure μ is invariant under the action of a smaller group).

In order to avoid such problems, we assume that the neighbourhoods $B_r(v)$ are edge-decorated and we only consider local isomorphisms that preserve decoration of the edges. In fact, instead of focusing on the whole $\Gamma \leq \mathsf{Aut}(G)$, we only care about the stabilizer subgroup $\Gamma_v \leq \mathsf{Aut}(G_n)_v$ for some $v \in V$ (as the action is transitive it does not matter which v one chooses). For a given stabilizer subgroup Γ_v we shall consider a decoration of the edges in such a way that the group of rooted automorphisms of the decorated rooted graph G, d, r is exactly Γ_v. The practical importance of this lies in the fact that in such a way cylinder events can be pulled back from the infinite graph onto the finite ones, by considering respective large balls isomorphic to each other. If μ_n converges to μ according to this conditions we will say that μ_n is locally convergent.

And why is this always possible? Here is a simple argument: :)...

Now that are framework is clear, we are going to list a few possible notions of Sparse Reconstruction. While these concepts are equivalent for sequences of product measures, (they all fail) when we allow for different sequences of measures, the picture becomes richer.

Definition 3.1.1 ((Weak) Sparse Reconstruction)**.** Let G_n be a sequence of finite transitive graphs with a transitive group action Γ_n on V_n. Let μ_n be a Γ_n-invariant sequence of probability measures on $\{-1,1\}^{V_n}$, and suppose that μ_n is weakly convergent.

Let $f_n : \{-1,1\}^{V_n} \longrightarrow \mathbb{R}$. There is Sparse Reconstruction for f_n in μ_n if there is a sequence of subsets of spins $U_n \subseteq V_n$ with $\lim_n \frac{|U_n|}{|V_n|} = 0$ such that

$$\mathsf{clue}(f_n \mid U_n) > c.$$

for some constant $c > 0$.

There is Weak Sparse Reconstruction (briefly: WSR) for μ_n if there exist a sequence of transitive functions $f_n : \{-1,1\}^{V_n} \longrightarrow \mathbb{R}$ such that there is Sparse Reconstruction for f_n.

There is Sparse Reconstruction (briefly: SR) for μ_n if there exist a sequence of transitive, non-degenerated Boolean functions f_n such that there is Sparse Reconstruction for f_n.

As we shall see in Corollary **??** if there is WSR, then there is also a sequence of Boolean functions for which there exists sparse reconstruction. The difference between WSR and SR lies in that in the latter case we also require non-degeneracy of the sequence.

We will give an example of a sequence of measures for which there is WSR, but no SR. (See the example under Corollary 3.1.9.) Of course, for sequences of product measures, there is neither SR nor WSR.

The first natural question to ask is whether the existence of Sparse Reconstruction is the attribute of the limiting measure. That is, if μ_n and ν_n sequences of measures share the same weak limit, then either both of the sequences have SR (WSR), or have not SR. The answer to this question is negative, in general.

It is possible to construct a sequence μ_n weakly convergent to a product measure which admits sparse reconstruction. Indeed, let μ_n be the following measure on $\{-1,1\}^{\mathbb{Z}_n}$. We choose a uniformly random $i \in \mathbb{Z}_n$ and around i in a neighborhood of size $\lfloor n^{\frac{2}{3}} \rfloor$ we flip a fair coin and make every spin in the interval $+1$ or -1 according to the coin flip. Outside this interval the spins are iid coin flips. Now it is easy to see that Majority can be reconstructed from this sequence.

One can choose U simply to be the multipliers of $n^{\frac{1}{2}}$. With high probability we can identify where is the long $+$ or $-$ interval and again with high probability whether it is $+$ or $-$ will tell us the Majority. At the same time this spin system weakly converges to the product measure on $\{-1,1\}^{\mathbb{Z}}$.

It seems that in general it is more difficult to tweak a sequence where SR naturally exists. For example, the following is true:

Proposition 3.1.1. *Let μ_n be a sequence of spin systems that locally converges to a Γ invariant measure μ which is not Γ-ergodic, then there is Sparse Reconstruction for μ_n from a set of constant size.*

Proof. As the limiting measure μ is non-ergodic there is a transitive event $\mathcal{A}$ with non trivial probability. As any measurable event can be approximated with an event depending on a finite subset of the coordinates (a cylinder event) with desired accuracy, for a given $\epsilon > 0$ one can choose an event A_ϵ satisfying

$$\mu[A \triangle A_\epsilon] < \epsilon,$$

where $A \triangle B$ is the symmetric difference of the events A and B. For a $\gamma \in \Gamma$ the translated event A_ϵ^γ is also a good approximation of A, using the invariance of A and therefore it is easy to see that $\mu[A_\epsilon^\gamma \triangle A_\epsilon] < 2\epsilon$.

Now choose a root $r \in V$ and choose $N \in \mathbb{N}$ large enough so that A_ϵ is $B_N(r)$-measurable (that is, the coordinates on which A_ϵ depends are inside the ball $B_N(r)$).

Now choose n large enough so that the $B_N(r)$ balls are isomorphic in G_n and G and that $|\mu_n[A_\epsilon] - \mu[A_\epsilon]| < \epsilon$ and consequently $\mu_n[A_\epsilon^\gamma \triangle A_\epsilon] < 4\epsilon$. Consider on the configuration space of G_n the event $\left\{\sum_{\gamma \in \Gamma_n} (2\mathbb{1}_{A_\epsilon^\gamma} - 1) > 0\right\}$, that is the majority of the translates of A_ϵ are satisfied. Let us denote the indicator of this event by $\mathsf{Maj}_n(A_\epsilon)$. Clearly, this is a transitive Boolean function for every n and if ϵ is small enough, it is also non-degenerated. At the same time, it is easy to verify that knowing A_ϵ already gives a positive clue about $\mathsf{Maj}_n(A_\epsilon)$ and A_ϵ depends on a constant number of coordinates. □

But can we have such a stability for an ergodic measure?

Question 3.1.2. *Is there an ergodic measure μ such that whenever μ_n converges weakly to μ_n there is SR for μ_n?*

Another line of questions is concerned about whether it is true if in some sense μ_n contains less randomness (or less information) then ν_n and ν_n admits SR, then is it true that μ_n admits SR as well. Of course, the important point here is how we make the expression 'contains less randomness' precise?

A natural attempt is to express the degree of randomness in a sequence with asymptotic entropy.

Definition 3.1.2 (asymptotic entropy)**.** Let μ_n be a sequence of measures. The asymptotic entropy of the sequence

$$\mathcal{H}(\mu_n) := \lim_{n\to\infty} \frac{H(\mu_n)}{n}$$

if it exists.

It turns out that $\mathcal{H}(\nu_n) > \mathcal{H}(\mu_n)$ and ν_n having SR does not imply that μ_n has SR. First, the above example of a spin system that weakly converges to a product space and still admits SR testifies that it can happen that the asymptotic entropy is 1 (as large as it can possibly be) but still there is Sparse Reconstruction. Still we believe that it does not hold in the opposite direction.

Conjecture 3.1.3. *If $\mathcal{H}(\mu_n) = 0$, then there is SR for μ_n.*

Question 3.1.4. *Is it true that for every $\epsilon > 0$ there is a weakly convergent sequence μ_n such that $\mathcal{H}(\mu_n) < \epsilon$ and there is no SR for μ_n ?*

Another way of expressing that μ has no more randomness then ν is to say that μ is a factor of ν. It turns out it is possible that there is SR for the sequence ν_n while the sequence μ_n, which is a factor of ν_n, does not admit SR. (Folyt...)

There is an alternative version of Sparse reconstruction that does not require any symmetry of the underlying μ_n measures.

Definition 3.1.3 (Sparse Reconstruction, with randomness)**.** Let E_n be a sequence of finite sets and let μ_n be a weakly convergent sequence of probability measures on $\{-1,1\}^{V_n}$. For every n let $\mathcal{U}_n$ be a random subset of V_n independent from the spin system with the property that

$$\delta_n = \max_{j \in V_n} \mathbb{P}[j \in \mathcal{U}] \to 0$$

There is Random Sparse Reconstruction (RSR) for μ_n if there is a sequence of Boolean functions $f_n : \{-1,1\}^{V_n} \longrightarrow \{-1,1\}$ and a $\mathcal{U}_n$ as above

$$\mathbb{E}[\mathsf{clue}_{f_n}(\mathcal{U}_n)] > c.$$

for some $c > 0$

It is easy to see that in case SR holds for a sequence μ_n then RSR as well. Instead of the the deterministic set U_n we define $\mathcal{U}_n$ as a uniformly random G_n translate of U_n. Because of transitivity of f_n the clue does not change by taking a translate, and $\delta_n = \frac{|U_n|}{|E_n|}$. Also, with some small modifications of the proof of Theorem 2.1.1 one can show that for product measures RSR does not hold either. We do not know, however, whether the two concepts are equivalent in general.

Question 3.1.5. *Suppose that for a transitive, weakly convergent sequence μ_n there is Random Sparse Reconstruction. Does this imply that there is also Sparse Reconstruction for μ_n?*

Remark 3.1.6. The spectral sample does not exist anymore in this general setting and therefore the short proof does not work for transitive functions. At first it looks like that we can use the exclusion-inclusion principle in just the same way as for product measures to get a spectral sample. The problem is that this measure in general assigns negative weights to certain subsets. Indeed, it is easy to find spin systems where transitive functions can be reconstructed from the values of a small subset.

3.1.2 Different Measures of Clue

We use (at least) two different concept of clue and therefore it is important to show that - at least in the most important cases - sparse reconstruction according to one of them is equivalent with sparse reconstruction with respect the other.

Proposition 3.1.7 (L^2 and Information Theoretic clue)**.** *Let μ be a measure on $\{-1,1\}^n$ and $\sigma = (\sigma_1, \dots \sigma_n)$ a spin system distributed according to μ.*

Let $f : \{-1,1\}^n \longrightarrow \{-1,1\}$ satisfying $k < |f(x)| \le K$ and let $Z = f(\sigma)$. For any $U \subseteq [n]$

$$\frac{\mathrm{Var}(\mathbb{E}[Z \mid \mathcal{F}_U])}{\mathrm{Var}(Z)} \le K \frac{I(Z, \sigma_U)}{H(Z)}. \tag{3.1.1}$$

Proof. First we show that

$$\mathrm{Var}(\mathbb{E}[Z \mid \mathcal{F}_U]) \le 2K^2 I(Z, \sigma_U).$$

The argument follows Lemma 4.4 in [Tao05]. First we fix some notations

$$p_z := \mathbb{P}[Z = z]$$

$$p_u := \mathbb{P}[\sigma_U = u]$$

$$p_{z|u} := \mathbb{P}[Z = z \mid \sigma_U = u]$$

Now, with this notation we have

$$\mathrm{Var}(\mathbb{E}[Z \mid \mathcal{F}_U]) = \sum_{u \in \sigma_U} p_u (\mathbb{E}[Z] - \mathbb{E}[Z \mid \sigma_U = u])^2$$

and for a fixed $u \in \sigma_U$

$$(\mathbb{E}[Z] - \mathbb{E}[Z \mid \sigma_U = u])^2 = \sum_{z \in R(f)} (p_z z - p_{z|u} z)^2 = \sum_z z^2 (p_z - p_{z|u})^2 \le K^2 \sum_{z \in R(f)} (p_z - p_{z|u})^2.$$

So, finally we get that

$$\mathrm{Var}(\mathbb{E}[Z \mid \mathcal{F}_U]) \le K^2 \sum_{u \in \sigma_U} \sum_z (p_z - p_{z|u})^2.$$

With the notation $h(x) := -x \log x$ for $x \in [0, 1]$ we can write the mutual information as

$$I(Z, \sigma_U) = H(Z) - H(Z|\sigma_U) = \sum_z \Big(h(p_z) - \sum_{u \in \sigma_U} p_u h(p_{z|u})\Big) \tag{3.1.2}$$

Using linear Taylor expansion with error term around p_z for $h(p_{z|u})$ we get the following estimate

$$h(p_{z|u}) \le h(p_z) + h'(p_z)(p_{z|u} - p_z) - \frac{1}{2p^*_{z|u}} (p_{z|u} - p_z)^2$$

with some $p^*_{z|u} \in (p_z, p_{z|u})$ (or $\in (p_{z|u}, p_z)$), using for the error term that $h''(x) = -\frac{1}{x}$. Substituting this estimate into (3.1.2), and observing that the term $h'(p_z)$ cancels, since for any $z \in R(f)$ we have $\sum_{u \in \sigma_U} p_u (p_{z|u} - p_z) = p_z - p_z = 0$, we obtain

$$\sum_{u \in \sigma_U} \sum_z \frac{(p_z - p_{z|u})^2}{p^*_{z|u}} \le 2I(Z, \sigma_U).$$

As $0 < p^*_{z|u} < 1$ we can conclude that

$$\mathrm{Var}(\mathbb{E}[Z \mid \mathcal{F}_U]) \le K^2 \sum_{u \in \sigma_U} \sum_z \frac{(p_z - p_{z|u})^2}{p^*_{z|u}} \le 2K^2 I(Z, \sigma_U)$$

In the sequel we show that under the conditions of the lemma

$$H(Z) \le C \mathrm{Var}(Z).$$

In case f is Boolean and thus Z takes on ± 1 almost surely, the entropy can be expressed as a function of $x = \mathbb{E}[Z]$. A quadratic Taylor expansion around 0 gives the following asymptotics:

$$-\left(\frac{1-x}{2}\log\frac{1-x}{2}+\frac{1+x}{2}\log\frac{1+x}{2}\right) = 1-\frac{1}{\ln 4}x^2+O(x^4),$$

So in case $|\mathbb{E}[Z]| \leq 1-c$ a simple calculation shows that

$$H(Z) \approx 1-\frac{1}{\ln 4}\mathbb{E}[Z]^2 \leq \frac{1}{c}(1-\mathbb{E}[Z]^2) = \frac{1}{c}\mathrm{Var}(Z).$$

In case f is still binary valued and $\min(f) - \max(f) \geq L$ for some $k > 0$:

$$H(Z) \leq \frac{1}{c\min(l^2,1)}\mathrm{Var}(Z)$$

We continue by induction on $|R(f)|$, the cardinality of the range of f. In the case when $|R(f)| \leq 2$, we already proved that our claim is true.

Now let $|R(f)| > 2$ For a $\emptyset \neq I \subset R(f)$ we define the event $A := \{Z \in I\}$. Using the law of total variance $\mathrm{Var}(Z) = \mathrm{Var}(\mathbb{E}[Z \mid \mathcal{F}_A]) + \mathbb{E}[\mathrm{Var}(Z \mid \mathcal{F}_A)]$, where $\mathcal{F}_A$ is the σ-algebra generated by A. In a similar way, $H(Z) = H(Z, \mathbb{1}_A) = H(\mathbb{1}_A) + H(Z \mid \mathbb{1}_A)$, using that A is Z-measurable.

In case $\frac{c}{2} < \mathbb{P}[A] < 1-\frac{c}{2}$, using that both $|\mathbb{E}[Z \mid A]|$ and $|\mathbb{E}[Z \mid A^c]| > k$

$$H(A) \leq \frac{1}{2\min(l^2,1)}\mathrm{Var}(\mathbb{E}[Z \mid \mathcal{F}_A]).$$

Conditioned on A and A^c the range of f is smaller then $R(f)$, respectively so by the induction hypothesis we have

$$H(Z \mid A) \leq \frac{1}{2c\min(k^2,1)}\mathrm{Var}(Z \mid A)$$

together with the respective upper bound for $H(Z \mid A^c)$. So we get

$$H(Z \mid \mathbb{1}_A) = \mathbb{P}[A]H(Z \mid A) + \mathbb{P}[A^c]H(Z \mid A^c) \leq \frac{1}{2c\min(k^2,1)}\mathbb{E}[\mathrm{Var}(Z \mid \mathcal{F}_A)]$$

$\square$

3.1.3 Reconstruction from random sets

In this section we state some general results. The setup is as before. We consider a sequence $\{\sigma^n : n \in \mathbb{N}\}$ where σ^n is a $\{-1,1\}^{V_n}$-valued random variables with law μ_n. We also assume that for each n there is a group G_n acting transitively on V_n and the law μ_n of σ^n is invariant under this group action. In particular, for every $j \in V_n$ the distribution of σ^n_j is the same, where we denote by σ^n_j the projection of σ^n to the jth coordinate, a ± 1-valued random variable.

We will sometimes consider this setup with the modification that the random variables are not binary, but $\mathbb{R}$-valued. In order to point out the difference in this case we will denote our sequence with ϕ^n instead of σ^n. Clearly, if a statement or definition works with ϕ^n it also does so with σ^n.

We introduce the notation $m_n := |V_n|$. We define the magnetization operator as

$$M_n[\phi] := \frac{1}{m_n} \sum_{j \in V_n} \phi_j^n. \tag{3.1.3}$$

For the next definition it is useful to specify one vertex of V_n as root denoted by 0 (as all vertices look the same this choice is arbitrary). We define the susceptibility of ϕ^n function as

$$S_n(\phi) := \sum_{j \in V_n} \mathrm{Cov}(\phi_0^n, \phi_j^n). \tag{3.1.4}$$

The term 'magnetization' comes from statistical physics, more specifically the Ising model (see section 3.2), a spin model which is central in this work. The value of the spins in this model are thought of as the charge of a particle, and the magnetization as the charge of the whole field. the concept of susceptibility originates from the Ising model as well. It can be shown that for the Ising model this quantity measures the change in the magnetic field of the system upon a small change in the external magnetic field, hence the name.

Recall that because of the translation invariance of the measure we have the following relationship between $M[\phi]$ and $S(\phi)$:

$$\mathrm{Var}(M_n[\phi]) = \frac{S_n(\phi)}{m_n}.$$

This is discussed in more detail in Section 2.4.2, see (2.4.3). Also, using again translation invariance,

$$\frac{1}{m_n} \frac{S_n(\phi)}{\mathrm{Var}(\phi_k^n)} = \frac{1}{m_n} \sum_{j \in V_n} \mathrm{Corr}(\phi_0^n, \phi_j^n) = \overline{\mathrm{Corr}}(\phi_0^n, \phi_j^n). \tag{3.1.5}$$

where $\overline{\mathrm{Corr}}(\phi_0^n, \phi_j^n)$ is the average correlation between the function and its translates.

The upcoming proposition states that in case for a spin system the average correlation is sufficiently high then the magnetization can be reconstructed from a random set with high probability. We will state in a slightly larger generality, with $\mathbb{R}$-valued random variables, as we do not use the fact that spins are taking only two values.

Proposition 3.1.8 (Sparse Reconstruction from random sets)**.** *Let $\{\phi^n \ : \ n \in \mathbb{N}$ be a sequence of $\mathbb{R}^{V_n}$-valued random variables with distribution invariant under the group action of G_n on V_n*

Suppose that

$$\overline{\mathrm{Corr}}(\phi_0^n, \phi_j^n) \gg \frac{1}{|V_n|}$$

then there is a sequence of numbers $k_n = o(m_n)$ such that for a uniform random subset $\mathcal{H}^{k_n}$ of size k_n and for any $\epsilon > 0$

$$\lim_{n \to \infty} \mathbb{P}[\mathsf{clue}(M_n[\phi] \mid \mathcal{H}^{k_n}) > 1 - \epsilon] = 1.$$

Proof. Let us introduce the shorthand notations $M_n := M_n[\phi]$ and $S_n := S_n(\phi)$. We give a lower bound for $\mathbb{E}\left[\mathrm{Corr}(M_n, \mathbb{E}[M^{\mathcal{H}}[\sigma^n] \mid \mathcal{H}])\right]$, the average correlation between the total magnetization and the magnetization of a uniformly random subset of k_n spins.

For any subset $U_n \subseteq V_n$, define the random variable

$$M_n^{U_n} := \frac{1}{|U_n|} \sum_{j \in U_n} \phi_j^n$$

We have

$$\mathrm{Cov}(M_n, M_n^{U_n}) = \frac{1}{m_n|U_n|}\sum_{j\in U_n}\sum_{i\in V_n}\mathrm{Cov}(\phi_j^n, \phi_i^n) = \frac{1}{m_n}S_n.$$

Recall that $\mathrm{Var}(M_n) = \frac{1}{m_n}S_n$ as well. So $\mathrm{Corr}(M_n, M_n^U)$ depends only on $\mathrm{Var}(M^{U_n})$. We now consider a uniformly random set of spins of size $k_n = |U_n|$ and write the average correlation between the magnetization of the system and the magnetization of the random set. Let $\mathcal{H}$ denote the random set of k_n spins we know.

Observe that by Jensen's inequality

$$\mathbb{E}\left[\mathrm{Corr}(M_n, \mathbb{E}[M_n^{\mathcal{H}} \mid \mathcal{H}])\right] = \sqrt{\frac{S_n}{m_n}}\mathbb{E}\left[\frac{1}{\sqrt{\mathrm{Var}(M_n^{\mathcal{H}} \mid \mathcal{H})}}\right] \geq \sqrt{\frac{S_n}{m_n}}\frac{1}{\sqrt{\mathbb{E}[\mathrm{Var}(M_n^{\mathcal{H}} \mid \mathcal{H})]}}, \tag{3.1.6}$$

and therefore it is enough to estimate the expected variance of the magnetization of a uniform random subset of k_n elements. So we can write

$$\mathbb{E}\left[\mathrm{Var}(M^{\mathcal{H}} \mid \mathcal{H})\right] = \frac{1}{k_n^2}\mathbb{E}\left[\mathbb{E}\left[\sum_{i,j\in V_n}\mathrm{Cov}(\phi_i^n, \phi_j^n)\mathbb{1}_{i\in\mathcal{H}}\mathbb{1}_{j\in\mathcal{H}} \mid \mathcal{H}\right]\right] = \\ = \frac{1}{k_n^2}\sum_{i,j\in V_n}\mathbb{E}\left[\mathbb{1}_{i\in\mathcal{H}}\mathbb{1}_{j\in\mathcal{H}}\mathrm{Cov}(\phi_i^n, \phi_j^n)\right].$$

Since

$$\mathbb{E}\left[\mathbb{1}_{i\in\mathcal{H}}\mathbb{1}_{j\in\mathcal{H}}\mathrm{Cov}(\phi_i^n, \phi_j^n)\right] = \begin{cases} \frac{k_n}{m_n}\mathrm{Var}(\phi_j^n) & \text{if } i = j \\ \frac{k_n(k_n-1)}{m_n(m_n-1)}\mathrm{Cov}(\phi_i^n, \phi_j^n) & \text{if } i \neq j \end{cases},$$

we get, using the notation $\mathrm{Var}(\phi_j^n) = s_n$ (because of invariance it does not depend on j):

$$\mathbb{E}[\mathrm{Var}(M^{\mathcal{H}} \mid \mathcal{H})] = \frac{1}{m_n k_n}\sum_{i\in V_n}s_n + \frac{k_n-1}{k_n}\frac{1}{m_n(m_n-1)}\sum_{i\neq j}\mathrm{Cov}(\phi_i^n, \phi_j^n) = \\ \frac{k_n-1}{k_n}\frac{1}{m_n(m_n-1)}\sum_{i,j\in V_n}\mathrm{Cov}(\phi_i^n, \phi_j^n) + s_n\left(\sum_{i\in V_n}\frac{1}{m_n k_n} - \frac{k_n-1}{k_n}\frac{1}{m_n(m_n-1)}\right) = \\ \frac{k_n-1}{k_n}\frac{1}{m_n(m_n-1)}m_n S_n + s_n\left(\frac{1}{k_n}\left(1 - \frac{k_n-1}{m_n-1}\right)\right) = \\ \frac{k_n-1}{k_n}\frac{1}{m_n-1}S_n + s_n\left(\frac{1}{k_n}\left(1 - \frac{k_n-1}{m_n-1}\right)\right)$$

Now we can give a lower bound for the average correlation over all subsets of size k_n.

Substituting back into (3.1.6), we get

$$\mathbb{E}\left[\mathrm{Corr}(M_n, \mathbb{E}[M_n^{\mathcal{H}} \mid \mathcal{H}])\right] \geq \frac{\sqrt{S_n}}{\sqrt{m_n}\sqrt{\frac{k_n-1}{k_n}\frac{1}{m_n-1}S_n + s_n\left(\frac{1}{k_n}\left(1-\frac{k_n-1}{m_n-1}\right)\right)}} =$$

$$\left(\frac{k_n-1}{k_n}\frac{m_n}{m_n-1} + \frac{s_n}{S_n}\left(\frac{m_n}{k_n}\left(1-\frac{k_n-1}{m_n-1}\right)\right)\right)^{-\frac{1}{2}} \asymp \left(1+\frac{s_n}{S_n}\frac{m_n}{k_n}\right)^{-\frac{1}{2}} =$$

$$\left(1+\frac{1}{k_n\overline{\mathrm{Corr}}(\phi_0^n, \phi_j^n)}\right)^{-\frac{1}{2}},$$

exploiting that $\frac{k_n-1}{m_n-1} \to 0$, by assumption and that by (3.1.5), $\frac{S_n}{s_n m_n} = \overline{\mathrm{Corr}}(\phi_0^n, \phi_j^n)$.

So if $k_n\overline{\mathrm{Corr}}(\phi_0^n, \phi_j^n) \to \infty$ then the right hand side tends to 1 as n goes to ∞. Since by assumption $\overline{\mathrm{Corr}}(\phi_0^n, \phi_j^n) \gg \frac{1}{m_n}$, we can choose a sequence k_n such that

$$\overline{\mathrm{Corr}}(\phi_0^n, \phi_j^n) \gg \frac{1}{k_n} \gg \frac{1}{m_n}.$$

In this case $\mathbb{E}\left[\mathrm{Corr}(M_n, \mathbb{E}[M_n^{\mathcal{H}} \mid \mathcal{H}])\right] \to 1$ and therefore, as correlations can be at most 1, the correlation $\mathrm{Corr}(M_n, \mathbb{E}[M_n^{\mathcal{H}} \mid \mathcal{H}])$, and thus its square, the clue (see (2.1.1)) tends to 1 with high probability. □

We would like to highlight the special case when $\phi_0^n = \sigma_0^n$ is uniform $\{-1, 1\}$-valued for all n.

Corollary 3.1.9. *Suppose that* $\{\sigma^n \ : \ n \in \mathbb{N}$ *is a sequence of* $\{-1,1\}^{V_n}$*-valued random variables with distribution invariant under the group action of* G_n *on* V_n *and* $\mathrm{Var}(\sigma_0^n) = 1$

If $S_n(\sigma) \to \infty$*, or equivalently* $\mathrm{Var}(M_n[\sigma^n]) \gg \frac{1}{m_n}$ *then there is a sequence of numbers* $k_n = o(m_n)$ *such that for a uniform random subset* $\mathcal{H}^{k_n}$ *of size* k_n *and for any* $\epsilon > 0$

$$\lim_{n\to\infty} \mathbb{P}[\mathsf{clue}(M_n[\sigma] \mid \mathcal{H}^{k_n}) > 1-\epsilon] = 1.$$

Proof. It is straightforward to check that $S_n(\sigma) \to \infty$ is equivalent to $\overline{\mathrm{Corr}}(\phi_0^n, \phi_j^n) \gg \frac{1}{m_n}$, when s_n is constant. □

In order to conclude SR we need to reconstruct non-degenerate Boolean functions, and therefore it is an important question whether Sparse Reconstruction of the total magnetization implies Sparse Reconstruction for the Majority function. In fact, in case the Magnetization is not concentrated there is no reason for this implication to hold. We might think about the following example:

Let us take the convex combination of an iid spin system and a system in which all the spins are +1 or all the spin are −1 with probability $\frac{1}{2}$, respectively. With probability $\frac{1}{\sqrt{n}}$ we choose the ±-system and with probability $1-\frac{1}{\sqrt{n}}$ we choose the iid system. Now it is clear that in this mixed system $\mathrm{Var}(M_n) \gg n$ so by Theorem 3.1.8 the magnetization can be reconstructed, but Majority (or any other non-degenerated Boolean function) cannot. So in this sequence of measure there is Weak Sparse Reconstruction, but no Sparse Reconstruction.

The following proposition gives sufficient conditions under which Maj can also be reconstructed.

Proposition 3.1.10. *Let σ^n be a sequence of spin systems as above. Suppose there is a sequence of naturals a_n such that $\frac{a_n}{\sqrt{m_n}} \to \infty$ and for every large n it holds that*

$$\mathbb{P}[|\sum_{j \in V_n} \sigma_j^n| \geq K a_n] > c \tag{3.1.7}$$

for some $c > 0$. Then there is a sequence $p_n \to 0$ such that for the random set $\mathcal{B}^{p_n}$ (in which every element is chosen independently with probability p_n) and arbitrary $\epsilon > 0$

$$\mathbb{P}\left[\mathsf{clue}(\mathsf{Maj} \mid \mathcal{B}^{p_n}) > 1 - \epsilon\right] > c.$$

Proof. Conditionally on the event $A = \{|M_n| \geq K a_n\}$ the expectation of the total magnetization in a Bernoulli sample can be bounded as follows. (We call total magnetization the unnormalized sum of the spins.)

$$\mathbb{E}\left[\mathbb{E}\left[|\sum_{j \in \mathcal{B}^{p_n}} \sigma_j^n| \;\Big|\; \mathcal{B}^{p_n}\right] \;\Big|\; A\right] =$$

$$\mathbb{E}\left[|\sum_{j \in V_n : \sigma_j^n = 1} \mathbb{1}_{j \in \mathcal{B}^{p_n}} - \sum_{k \in V_n : \sigma_k^n = -1} \mathbb{1}_{k \in \mathcal{B}^{p_n}}| \;\Big|\; A\right] = p_n \mathbb{E}[|\sum_{j \in V_n} \sigma_j^n|] \geq K a_n p_n$$

Now we compute its variance, using that the events $\{j \in \mathcal{B}^{p_n}\}$ and $\{k \in \mathcal{B}^{p_n}\}$ are independent, whenever $k \neq j$:

$$\mathrm{Var}\left(\mathbb{E}\left[\sum_{j \in \mathcal{B}^{p_n}} \sigma_j^n \;\Big|\; \mathcal{B}^{p_n}\right] \;\Big|\; A\right) =$$

$$\mathrm{Var}\left(\sum_{j \in V_n : \sigma_j^n = 1} \mathbb{1}_{j \in \mathcal{B}^{p_n}} - \sum_{k \in V_n : \sigma_k^n = -1} \mathbb{1}_{k \in \mathcal{B}^{p_n}} \;\Big|\; A\right) = m_n p_n (1 - p_n)$$

This means that for every ϵ there exists a $C > 0$ such that

$$\mathbb{P}\left[\left|\mathbb{E}\left[\sum_{j \in \mathcal{B}^{p_n}} \sigma_j^n \;\Big|\; \mathcal{B}^{p_n}\right]\right| > K a_n p_n - C_\epsilon \sqrt{m_n p_n} \;\Big|\; A\right] > 1 - \epsilon$$

since the total magnetization of the sample follows binomial distribution.

In case one chooses p_n to satisfy $a_n p_n \gg \sqrt{m_n p_n}$ then, conditioned on A the fluctuations of the random sample are small compared to the sample magnetization. Therefore, conditioned on A, with high probability the majority of the sample coincides with the majority of the original system.

Formally, choose n large enough so that $C_\epsilon \sqrt{m_n p_n} \leq \frac{K}{2} a_n p_n$. Then we have

$$\mathbb{P}\left[\left|\mathbb{E}\left[\sum_{j \in \mathcal{B}^{p_n}} \sigma_j^n \;\Big|\; \mathcal{B}^{p_n}\right]\right| > \frac{K}{2} a_n p_n \;\Big|\; A\right] > 1 - \epsilon,$$

and this of course entails $\{\mathsf{Maj} = \mathsf{Maj}(\mathbb{E}[\sigma^n \mid \mathcal{B}^{p_n}])\}$ (the latter random function is the majority on the random bits of $\mathcal{B}^{p_n}$). Therefore, conditioned on A with high probability the magnetization can be reconstructed from $\mathcal{B}^{p_n}$.

It remains t verify that the condition $a_n p_n \gg \sqrt{m_n p_n}$ is consistent with our assumptions. Indeed, equivalently we can write

$$p_n \gg \frac{m_n}{a_n^2}.$$

which means that p_n is of order $o(1)$ by the assumption that $\frac{a_n}{\sqrt{m_n}} \to \infty$. Therefore, $\mathcal{B}^{p_n}$ is sparse with high probability. In particular, there exits also a sequence of subsets U_n with density tending to 0 and the majority has uniformly positive clue with respect to this sequence. $\square$

Moreover, with small additional cost - a couple of independent samples - we can learn with high probability whether A holds or not, thus we know if the magnetization of the random set gives a good guess for the total magnetization or not.

3.1.4 The 3-Correlation Lemma

First we need a slight generalization of the concepts of magnetization and susceptibility. Let us consider a spin system σ distributed according to μ, with coordinate set V and group action G, as before.

Recall that for a function $f : \{-1,1\}^V \longrightarrow \mathbb{R}$ f^g denotes the g-translated version of f. Since μ is G-invariant, $Z := f(\sigma)$ and $Z^g := f^g(\sigma)$ has the same distribution. One can define magnetization (as we have already done in Chapter 2 see (2.4.1)) and susceptibility for arbitrary function on the configuration space by:

$$M[Z,\mu] := \frac{1}{|G|}\sum_{g\in G} Z^g$$

and

$$S(Z,\mu) := \frac{1}{|G_v|}\sum_{g\in G} \mathrm{Cov}(Z,Z^g),$$

. where G_v is the stabilizer subgroup of a vertex. In case the action of G on V is not free, that is the stabilizer subgroup of a vertex is not trivial then for every Z we count every covariance $|G_v|$ many times. Indeed, as $|G| = |V||G_v|$ we have $|G_v|$ times too many terms in the susceptibility. Warning: in case Z has additional symmetries it is possible that there are still repetitions in the sum of $S(Z,\mu)$ and it is perfectly fine. For example when Z itself is transitive, that is G-invariant, $\mathrm{Cov}(Z,Z^g) = \mathrm{Var}(Z)$ for every g and thus $S(Z) = \frac{|G|}{|G_v|} = |V|\mathrm{Var}(Z)$.

In the sequel, to avoid this technical difficulty we will assume that the action of the group on the coordinate set is free (thus we omit the coefficient $\frac{1}{|G_v|}$). We emphasize, however, that all the results are true without this additional condition.

Also, along the lines of (3.1.5) we have

$$\frac{1}{|G_n|}\frac{S_n(Z)}{\mathrm{Var}(Z)} := \frac{1}{|G_n|}\sum_{j\in V_n} \mathrm{Corr}(Z,Z^g) := \overline{\mathrm{Corr}}(Z,Z^g), \tag{3.1.8}$$

In case there is no room for ambiguity we are going to omit the dependence on the measure to simplify notation. Observe that for any f the system of random variables $\{Z^g \ : \ g \in G\}$ is a G-invariant family (although possibly the same random variables appear multiple times), so in fact the definitions in (3.1.3) and (3.1.4) already cover this case.

The following statement, although it follows from some elementary facts by straightforward calculations, has some interesting consequences.

Lemma 3.1.11 (3-Correlation Lemma). *Let $\sigma = \{\sigma_j : j \in V\}$ be a spin system with G-invariant distribution, where G acts transitively on V. Let $f : \{-1,1\}^V \longrightarrow \mathbb{R}$ be a transitive function and let $Z := f(\sigma)$. Then*

$$\mathrm{Corr}(Z, M[\mathbb{E}[Z \,|\, \mathcal{F}_U]])\mathrm{Corr}(\mathbb{E}[Z \,|\, \mathcal{F}_U], M[\mathbb{E}[Z \,|\, \mathcal{F}_U]]) = \mathrm{Corr}(Z, \mathbb{E}[Z \,|\, \mathcal{F}_U])$$

Proof. As in (2.1.1), we have:

$$\mathrm{Corr}(Z, \mathbb{E}[Z \,|\, \mathcal{F}_U]) = \frac{\mathrm{Var}(\mathbb{E}[Z \,|\, \mathcal{F}_U])}{\sqrt{\mathrm{Var}(Z)}\sqrt{\mathrm{Var}(\mathbb{E}[Z \,|\, \mathcal{F}_U])}} = \sqrt{\frac{\mathrm{Var}(\mathbb{E}[Z \,|\, \mathcal{F}_U])}{\mathrm{Var}(Z)}}$$

Now we turn to the left hand side. First, observe that

$$\mathrm{Cov}(Z, M[\mathbb{E}[Z \,|\, \mathcal{F}_U]]) = \frac{1}{|V|}\sum_{g \in G} \mathrm{Cov}(Z, \mathbb{E}[Z \,|\, \mathcal{F}_{U^g}]) = \mathrm{Var}(\mathbb{E}[Z \,|\, \mathcal{F}_U]),$$

using that Z is transitive and therefore $\mathrm{Cov}(Z, \mathbb{E}[Z \,|\, \mathcal{F}_{U^g}])$ is G-invariant. Using that $\mathrm{Var}(M[\mathbb{E}[Z \,|\, \mathcal{F}_U]]) = S(\mathbb{E}[Z \,|\, \mathcal{F}_U])/|V|$, we get

$$\mathrm{Corr}(f, M[\mathbb{E}[Z \,|\, \mathcal{F}_U]]) = \frac{\mathrm{Var}(\mathbb{E}[Z \,|\, \mathcal{F}_U])}{\sqrt{\mathrm{Var}(Z)}}\sqrt{\frac{|V|}{S(\mathbb{E}[Z \,|\, \mathcal{F}_U])}}. \tag{3.1.9}$$

As for the other term, we can estimate the covariance as follows

$$\mathrm{Cov}(\mathbb{E}[Z \,|\, \mathcal{F}_U], M[\mathbb{E}[Z \,|\, \mathcal{F}_U]]) = \\ \frac{1}{|V|}\sum_{j \in V} \mathrm{Cov}(\mathbb{E}[Z \,|\, \mathcal{F}_U], \mathbb{E}[Z \,|\, \mathcal{F}_{U^j}]) = \frac{S(\mathbb{E}[Z \,|\, \mathcal{F}_U])}{|V|}.$$

So we get for the respective correlation:

$$\mathrm{Corr}(\mathbb{E}[Z \,|\, \mathcal{F}_U], M[\mathbb{E}[Z \,|\, \mathcal{F}_U]]) = \\ \frac{S(\mathbb{E}[Z \,|\, \mathcal{F}_U])/|V|}{\sqrt{\mathrm{Var}(\mathbb{E}[Z \,|\, \mathcal{F}_U])S(\mathbb{E}[Z \,|\, \mathcal{F}_U])/|V|}} = \sqrt{\frac{S(\mathbb{E}[Z \,|\, \mathcal{F}_U])}{|V|\mathrm{Var}(\mathbb{E}[Z \,|\, \mathcal{F}_U])}} \tag{3.1.10}$$

It is now easy to see that when multiplying (3.1.10) with (3.1.9), one gets $\mathrm{Corr}(Z, \mathbb{E}[Z \,|\, \mathcal{F}_U])$ as stated. □

Substitute in any σ-measurable random variable Z in the place of $\mathbb{E}[Z \,|\, \mathcal{F}_U]$ and compare it with (3.1.8). This gives rise to the following, strange looking identity:

$$\overline{\mathrm{Corr}}(Z, Z^g) = \mathrm{Corr}^2(Z, M[Z]) \tag{3.1.11}$$

Corollary 3.1.12. *If in a spin system σ^n there is weak sparse reconstruction , then there is also weak sparse reconstruction with clue tending to* 1.

Proof. By assumption, there exist a sequence of subsets $U_n \subseteq V_n$ with $|U_n| = o(V_n)$ and a sequence of functions of $f_n : \{-1,1\}^{V_n} \longrightarrow \mathbb{R}$ with

$$\mathsf{clue}_{\sigma^n}(f_n \mid U_n) > c$$

for some $c > 0$. Let $Z_n = f_n(\sigma^n)$. Recalling that $\mathsf{clue}(f_n \mid U_n) = \mathrm{Corr}^2(Z_n, \mathbb{E}[Z_n \,|\, \mathcal{F}_{U_n}])$ it follows, using Lemma 3.1.11 that

$$c < \mathrm{Corr}^2(Z_n, \mathbb{E}[Z_n \,|\, \mathcal{F}_{U_n}]) \leq \mathrm{Corr}^2(\mathbb{E}[Z_n \,|\, \mathcal{F}_{U_n}], M[\mathbb{E}[f_n \,|\, \mathcal{F}_{U_n}]]).$$

This means, according to (3.1.11) that

$$c < \overline{\mathrm{Corr}}(\mathbb{E}[Z_n \,|\, \mathcal{F}_{U_n}], \mathbb{E}[Z_n \,|\, \mathcal{F}_{U_n^g}]). \tag{3.1.12}$$

Now we consider the spin system $\phi^n_g := \mathbb{E}[Z_n \,|\, \mathcal{F}_{U_n^g}]$ indexed by G and apply the argument in Proposition 3.1.8. We recall from the proof of Proposition 3.1.8 that the expected correlation with respect to $\mathcal{H}^{k_n}$, a uniformly random subset of coordinates with k_n elements is given by

$$\mathbb{E}\left[\mathrm{Corr}(M[\phi^n], \mathbb{E}[M[\phi^n]^{\mathcal{H}} \mid \mathcal{H}])\right] \geq \left(1 + \frac{1}{k_n \overline{\mathrm{Corr}}(\phi^n, \phi^n_g)}\right)^{-\frac{1}{2}}.$$

So taking into account (3.1.12) it follows that

$$\mathbb{E}\left[\mathrm{Corr}(M[\phi^n], \mathbb{E}[M[\phi^n]^{\mathcal{H}} \mid \mathcal{H}])\right] \geq \left(1 + \frac{1}{k_n c}\right)^{-\frac{1}{2}}.$$

Let k_n be a sequence of integers such that $k_n \to \infty$, but $|U_n| k_n \ll |V_n|$. From this choice it is immediate that $\mathbb{E}\left[\mathrm{Corr}(M[\phi^n], \mathbb{E}[M[\phi^n]^{\mathcal{H}} \mid \mathcal{H}])\right] \to 1$. On the other hand, for a fixed set sampled from $\mathcal{H}^{k_n}$, the function $\mathbb{E}[M[\phi^n]^{\mathcal{H}} \mid \mathcal{H}]$ depends on k_n coordinates of ϕ^n, and ultimately on at most $|U_n| k_n$ coordinates of σ^n (since each ϕ^n_g depends on U_n coordinates of σ^n), which is sparse, by our choice of k_n.

Since the expected correlation tends to 1, there is a sequence of k_n-element subsets which reconstructs $M[\mathbb{E}[Z_n \,|\, \mathcal{F}_{U_n}]]$ with high probability. □

We continue with another consequence of Lemma 3.1.11, which gives a potential tool to show that there is no SR for a particular spin system.

Corollary 3.1.13. *For a sequence of spin system σ^n there is no sparse reconstruction if and only if there is an $\epsilon > 0$ such that for every sequence of subsets $U_n \subseteq V_n$ with $U_n \ll V_n$ and every Z_n sequence of $\mathcal{F}_{U_n}$-measurable random variables*

$$\overline{\mathrm{Corr}}(Z_n, Z^g_n) < 1 - \epsilon \tag{3.1.13}$$

for every $n \geq N$.

Proof. Indirectly, assume that 3.1.13 holds but there exists a sequence of subsets $U_n \subseteq [n]$ and a sequence of transitive functions f_n with $\liminf \mathsf{clue}_{\sigma^n}(f_n \,| U_n) = c > 0$. By Corollary 3.1.12 we may assume that $\lim_n \mathsf{clue}_{f_n}(U_n) = 1$.

Set $Z_n = f_n(\sigma^n)$. If n is large enough

$$1 - \epsilon \leq \mathrm{Corr}^2(Z_n, \mathbb{E}[Z_n \,|\, \mathcal{F}_{U_n}]) \leq$$
$$\mathrm{Corr}^2(\mathbb{E}[Z \,|\, \mathcal{F}_U], M[\mathbb{E}[Z \,|\, \mathcal{F}_U]]) = \overline{\mathrm{Corr}}(\mathbb{E}[Z_n \,|\, \mathcal{F}_{U_n}], \mathbb{E}[Z_n \,|\, \mathcal{F}_{U_n^g}]).$$

where we first used Lemma 3.1.11 and after (3.1.11). As $\mathbb{E}[Z_n \,|\, \mathcal{F}_{U_n}]$ is trivially $\mathcal{F}_{U_n}$-measurable, this is in contradiction with our assumptions, so there is no SR on σ^n. □

This result allows us to give yet another proof for Theorem 2.1.6. We need the following:

Lemma 3.1.14. *Let $f : \{-1,1\}^V \longrightarrow \mathbb{R}$ be a function on the n-dimensional hypercube with the uniform measure, G a group acting on V transitively. If f is $\mathcal{F}_U$-measurable for some $U \subseteq [n]$ then $S(f) \leq |U|$.*

Proof. Observe that for $g \in G$

$$f^g = \sum_{S \subseteq V} \widehat{f}(S)\chi_{S^g} = \sum_{S \subseteq V} \widehat{f}(S^{-g})\chi_S.$$

and therefore

$$\widehat{f^g}(S) = \widehat{f}(S^{-g}).$$

We can now express the susceptibility of f in terms of the Fourier-Walsh transform of f.

$$S(f) = \sum_{g \in G} \mathrm{Cov}(f(\omega), f^g(\omega)) = \sum_{g \in G} \sum_{S \subseteq V} \widehat{f}(S)\widehat{f}(S^{-g}) = \sum_{S \subseteq V} \sum_{g \in G} \widehat{f}(S)\widehat{f}(S^{-g})$$

The sum can be partitioned according to G-orbits of subsets. Let $\mathcal{O}$ denote the set of G-orbits of the subsets of V. Then

$$S(f) = \sum_{G \cdot S \in \mathcal{O}} \sum_{g,h \in G} \widehat{f}(S^h)\widehat{f}(S^{h-g}) = \sum_{G \cdot S \in \mathcal{O}} \left(\sum_{g \in G} \widehat{f}(S^g) \right)^2$$

For a particular $u \in U$ there are exactly $|U|$ translations such that $g \cdot u \in U$ as well. Because f is $\mathcal{F}_U$-measurable $\widehat{f}(S^g)$ can have nonzero coefficients only if $S^g \subseteq U$. So each orbit $G \cdot S$ contains at most $|U|$ subsets with non-zero Fourier coefficient and therefore, by the Cauchy-Schwartz inequality:

$$\left(\sum_{g \in G} \widehat{f}(S^g) \right)^2 \leq |U| \sum_{g \in G} \widehat{f}^2(S^g), \tag{3.1.14}$$

and thus we get

$$S(f) = \sum_{G \cdot S \in \mathcal{O}} \left(\sum_{g \in G} \widehat{f}(S^g) \right)^2 \leq |U| \mathrm{Var}(f(\omega)).$$

□

Combining the above result with Corollary 3.1.13 we immidiately get the promised alternative proof for Theorem 2.1.1. Indeed

$$\overline{\mathrm{Corr}} f_n(\omega), f_n^g(\omega)) = \frac{S(f_n)}{|V_n| \mathrm{Var}(f_n(\omega))} = \frac{|U_n|}{|V_n|} \to 0.$$

Remark 3.1.15. It is straightforward to generalise the above result to general product measures (Theorem 2.1.6) if one replaces the Fourier-Walsh transform with the Efron-Stein decomposition (See Theorem 2.1.5).

Remark 3.1.16. In equation 3.1.14 there is equality when $f = \sum_{j \in U} \omega_j$ and therefore the inequality of Lemma 3.1.14 is sharp.

An Algorithmic Method

Lemma 3.1.11 suggests an algorithmic method to find functions with high clue. We introduce the notation

$$\pi_U[f] := \mathbb{E}[f \mid \mathcal{F}_U]$$

and let $T[f] := M[\pi_U[f]]$ Now we can rewrite the statement of Lemma 3.1.11. For every $\mathbb{Z}_n^d$-transitive function f :

$$\mathrm{Corr}(f, T[f])\mathrm{Corr}(\pi_U[f], T[f]) = \mathrm{Corr}(f, \pi_U[f])$$

In case $\mathrm{Corr}(f, T[f]) < 1$ we have $\mathrm{Corr}(\pi_U[f], T[f]) > \mathrm{Corr}(f, \pi_U[f])$. Since $\mathrm{Corr}(\pi_U[f], T[f]) \leq \mathrm{Corr}(\pi_U[T[f]], T[f])$ (as $\pi_U[T[f]]$ is the function that maximizes the correlation with $T[f]$ among $\mathcal{F}_U$-measurable ones), we get that

$$\mathrm{Corr}(\pi_U[T[f]], T[f]) > \mathrm{Corr}(f, \pi_U[f])$$

This means that iteratively applying the operator T to a given function we can increase the clue whenever $\mathrm{Corr}(f, T[f]) < 1$. In case $\mathrm{Corr}(f, T[f]) = 1$ that is, if f is an eigenfunction of T, the iteration comes to an end. So if T admits an eigenbasis then it is sufficient to calculate the clue of the eigenfunctions since in that case $T^n[f]$ converges to the linear combination of some eigenfunctions belonging to the same eigenspace.

Indeed, it is exactly what happens in the i.i.d. case, where there is an eigenfunction corresponding to every $\mathbb{Z}_n^d$-orbit. For simplicity we discuss the case of the uniform hypercube.

Let L be a $\mathbb{Z}_n^d$-orbit of S and $\chi_L = \sum_{S \in L} \chi_S$. Then

$$\pi_U[\chi_{O(S)}] = \sum_{j \in \mathbb{Z}_n^d,\, S^j \subseteq U} \chi_{S^j}$$

and

$$T[\chi_{O(S)}] = C(O(S), U)\chi_L,$$

where

$$C(O(S), U) := |\{j \in \mathbb{Z}_n^d \;:\; S^j \subseteq U\}|$$

denotes the number of translations of the subset S which are contained in U. Obviously, the functions f_L form an orthogonal basis for the space of $\mathbb{Z}_n^d$-invariant functions. Indeed, every transitive function f can be represented in this basis as $f = \sum_{O \in \mathcal{O}} \widehat{f}(O)\chi_O$.

Moreover, $\lim_{n\to\infty} T^n[f]$ is contained in the eigenspace corresponding to the the largest eigenvalue with nonzero coefficient in f. In particular, if f has non-zero energy on level 1 (that is, a linear part) then $T^n[f]$ tends to the magnetization. The reason is that for any given $U \subseteq [n]$ the eigenvalue $C(O(S), U)$ maximized by the singletons.

Proposition 3.1.17. *Let f be a transitive function and $U \subseteq [n]$. Then*

$$\mathrm{Var}(\mathbb{E}[f \mid \mathcal{F}_U]) = \frac{1}{n} \sum_{L \in \mathcal{O}} C(L, U)\widehat{f}^2(L) \tag{3.1.15}$$

Proof. For convenience suppose that $\mathrm{Var}(f) = 1$.

$$\mathrm{Var}(\mathbb{E}[f \mid \mathcal{F}_U]) = \mathbb{P}[\mathscr{S} \subseteq U \mid \mathscr{S} \neq \emptyset] = \sum_{L \in \mathcal{O}} \mathbb{P}[\mathscr{S} \in L]\mathbb{P}[\mathscr{S} \subseteq U \mid \mathscr{S} \in L] \tag{3.1.16}$$

By the orbit counting lemma the orbit O has $\frac{n}{|\mathbf{Stab}(O)|}$ many elements and therefore

$$\mathbb{P}[\mathscr{S} \in O] = \frac{n}{|\mathbf{Stab}(O)|} \sum_{S \in O} \widehat{f}^2(S) = \left(\sum_{S \in O} (\widehat{f}(S) \right)^2 = \widehat{f}^2(L),$$

while

$$\mathbb{P}[\mathscr{S} \subseteq U \,|\, \mathscr{S} \in O] = \frac{C(O,U)}{n}$$

□

3.2 Sparse Reconstruction for the Ising Model

3.2.1 The Curie-Weiss model

A slight strengthening of the argument above is enough to show that there is no sparse reconstruction for the subcritical Curie-Weiss model (we know that magnetization can be reconstructed on any critical and supercritcal Ising model by Theorem **??**).

Theorem 3.2.1. *There is no sparse reconstruction for the subcritical Curie-Weiss model.*

We divide the proof of the Theorem into a few steps.

Lemma 3.2.2. *Let $\sigma[n]$ be a sequence of spin systems and suppose that there is a $C > 0$ such that for every n*

$$H(\sigma[n]) = n - C \tag{3.2.1}$$

then there is no sparse reconstruction for $\sigma[n]$.

Proof. The proof repeats that of Lemma 2.2.3 and Theorem 2.2.1. First observe that

$$\sum_j^L H(\sigma(S_j)) \leq \sum_j \sum_{i \in S_j} H(\sigma(i)) = k \sum_{i \in [n]} H(\sigma(i)) \leq k(H(\sigma[n]) + C) \tag{3.2.2}$$

were, for the last inequality we used the condition of the Lemma.

In turn, together with the Shearer inequality as in 2.2.5, we obtain that

$$\sum_j^L I(Z, \sigma(S_j)) \leq k(I(Z, \sigma[n]) + C) \tag{3.2.3}$$

Now we can use this inequality just as in the proof of Theorem 2.2.1 to get that

$$nI(Z, \sigma_U) \leq |U|(I(Z, \sigma[n]) + C) \tag{3.2.4}$$

obviously, $\frac{|U|(I(Z,\sigma[n])+C)}{n} = o(1)$ which is exactly what we wanted to show. □

Theorem 3.2.3 (Tail of subcritical Curie-Weiss)**.** *If $\beta < \beta_c = 1$ then*

$$\lim_n \Pr[M_n > C\sqrt{n}] = \sqrt{\frac{1-\beta}{2\pi}} \int_x^\infty \exp -\frac{1-\beta}{2} t^2 dt \tag{3.2.5}$$

where $M_n := \sum_{i=1}^n \sigma(i)$ is the total magnetisation

For a proof of this result see for example [].

Lemma 3.2.4. *Let $\sigma[n]$ denote a subcritical Curie-Weiss model on n spins and let $M_k = \sum_{i=1}^{k} \sigma(i)$. Then for every $t > 0$ and $0 < i \leq n$ positive integer*

$$\Pr[M_i > t\sqrt{n}] \leq \frac{e^{-Ct^2}}{1 - 4e^{-\frac{t^2}{4}}} \tag{3.2.6}$$

for some positive constant C

Proof. First we are going to show that for every $t > 0$ and $0 < i \leq n$ we have

$$\Pr[M_n \leq \frac{t}{2}\sqrt{n} \ \mid \ M_i > t\sqrt{n}] \leq 4e^{-\frac{t^2}{4}}. \tag{3.2.7}$$

Let us now fix an $1 \leq i \leq n$. Conditioned on the event $\{M_i > C\sqrt{n}\}$ we may consider a coupling between the process M_{i+k} and the simple random walk $S_k \ : \ (k = 1, 2, \ldots n-i)$, where each time M_{i+k} decreases (that is, $\sigma_k = -1$) S_k decreases as well.

As long as $M_{i+k} \geq 0$ such a coupling exists because σ_{k+1} conditioned on the magnetization of the first $i + k$ spins already revealed is a Bernoulli random variable with expectation $m_{n-i-k}(\beta, \frac{M_{i+k}}{2n}) > 0$ independent from the value of any of the individual spins revealed before.

Therefore, using the above coupling:

$$\Pr[M_n \leq \frac{t}{2}\sqrt{n} \mid \ M_i > t\sqrt{n}] \leq \tag{3.2.8}$$

$$\leq \Pr[\min_k \{M_{i+k}\} \leq 0 \mid \ M_i > t\sqrt{n}] + \Pr[M_n \leq \frac{t}{2}\sqrt{n} \ \text{ and } \ \min_k \{M_{i+k}\} > 0 \mid \ M_i > t\sqrt{n}] \leq \tag{3.2.9}$$

$$\leq \Pr[\min S_1, S_2, \ldots S_{n-i} \leq -t\sqrt{n}] + \Pr[S_{n-i} \leq -\frac{t}{2}\sqrt{n}] \leq \tag{3.2.10}$$

$$\leq 2\Pr[S_{n-i} > t\sqrt{n}] + \Pr[S_{n-i} = t\sqrt{n}] + \Pr[S_{n-i} \geq \frac{t}{2}\sqrt{n}] \leq 4e^{-\frac{t^2}{4}} \tag{3.2.11}$$

For the second inequality we used the monotone coupling between M_{i+k} and the simple random walk S_k, while in the second one we used the symmetry of the SRW with respect to the origin and the standard result that $\Pr[\{\max S_1, S_2, \ldots S_{n-i}\} \geq l] = 2\Pr[S_{n-i} > l] + \Pr[S_{n-i} = l]$. Finally in the last row we used the Gaussian estimation for the tail of a binomially distributed random variable.

After using the definition of conditional probability and rearranging (3.2.7)

$$\Pr[M_i > t\sqrt{n}] \leq \frac{\Pr[M_n > \frac{t}{2}\sqrt{n} \ \text{ and } \ M_i > t\sqrt{n}]}{1 - 4e^{-\frac{t^2}{4}}} \tag{3.2.12}$$

Using that by Theorem 3.2.3

$$\Pr[M_n > \frac{t}{2}\sqrt{n} \ \text{ and } \ M_i > t\sqrt{n}] \leq \Pr[M_n > \frac{t}{2}\sqrt{n}] \leq e^{-Ct^2} \tag{3.2.13}$$

we obtain

$$\Pr[M_i > t\sqrt{n}] \leq \frac{e^{-Ct^2}}{1 - 4e^{-\frac{t^2}{4}}} \tag{3.2.14}$$

□

Now we are ready to show that the condition of Lemma 3.2.2 is satisfied for the subcritical Curie-Weiss model.

Lemma 3.2.5. *The subcritical Curie-Weiss model with a fixed temperature β and with $h = 0$ satisfies the conditions of Lemma 3.2.2, that is denoting the Curie-Weiss model on n spins by $\sigma_\beta[n]$, there exist a positive constant C such that for all large enough n*

$$H(\sigma_\beta([n])) \geq n - C \tag{3.2.15}$$

Proof. According to the chain rule of entropy

$$H(\sigma_\beta[n]) = \sum_{k=0}^{n-1} H(\sigma(k+1) \mid \sigma[k]) \tag{3.2.16}$$

Because of the lack of geometry all the information is encoded in the sum of the spins, i.e. the magnetization. Therefore we can write:

$$H\left(\sigma(k+1) \mid \sigma([k])\right) = \sum_t \Pr\left[M_k = t\right] H\left(\sigma(k+1) \mid M_k = t\right). \tag{3.2.17}$$

where again $M_k = \sum_{i=1}^{k} \sigma(i)$.

Since $\sigma(k)$ is a Bernoulli random variable its conditional distribution, and thus its conditional entropy is determined by the conditional expected value $\mathbb{E}[\sigma(k+1) = 1 \mid M_k = t]$. That is

$$H\left(\sigma(k+1) \mid M_k = t\right) = h\left(\mathbb{E}[\sigma(k+1) = 1 \mid M_k = t]\right) \tag{3.2.18}$$

where

$$h(x) := \frac{1-x}{2} \log \frac{1-x}{2} + \frac{1+x}{2} \log \frac{1+x}{2} = 1 - \frac{1}{\ln 4} x^2 + O(x^4) \tag{3.2.19}$$

using the Taylor expansion of h around 0. Let us compute the Hamiltonian conditioned on the event that sum of the first k spins is t:

$$\begin{aligned}
\mathbf{H_{n,0}}(\sigma \mid M_k = t) &= -\frac{1}{2n} \sum_{i,j>k} \sigma(i)\sigma(j) - \sum_{i \leq k} \sigma(i)) \sum_{l>k} \sigma(l) - \frac{1}{2n} \sum_{i,j \leq k} \sigma(i)\sigma(j) \\
&= -\frac{t}{2n} \sum_{i>k} \sigma(i) - \frac{1}{2n} \sum_{i,j>k} \sigma(i)\sigma(j) - \frac{t^2}{2n}
\end{aligned}$$

This shows that conditioned on the event $\{M_k = t\}$ the spin system $\sigma[n] \setminus [k]$ has the law of a Curie-Weiss model on $n - k$ spins with parameters $(\beta, \frac{t}{2n})$. As a consequence

$$\mathbb{E}\left[\sigma(k+1) \mid M_k = t\right] = m_{n-k}\left(\beta, \frac{t}{2n}\right) \tag{3.2.20}$$

where $m_n(\beta, h) := \frac{1}{n}\mathbb{E}[M_n(\sigma_{\beta,h}[n])]$ is the expected magnetization per site.

Using a first order approximation for $m_{n-k}\left(\beta, \frac{t}{2n}\right)$ around $h = 0$ we get that

$$m_{n-k}(\beta, \frac{t}{2n}) = m_{n-k}(\beta, 0) + \frac{t}{2n}\frac{\partial m}{\partial h} + O\left(\frac{t^2}{n^2}\right). \tag{3.2.21}$$

Note that $\frac{\partial m_n}{\partial h} \leq \frac{\partial m_n}{\partial h} = \beta\chi$ where χ denotes the susceptibility. It is known (see []) that the susceptibility is finite in the subcritical (high temperature) regime. Obviously $m_n(\beta, 0) = 0$, so the first order approximation says that for every $t \geq 0$:

$$\mathbb{E}\left[\sigma(k+1) \mid M_k = t\right] = \beta\chi\frac{t}{2n} + O\left(\frac{t^2}{n^2}\right). \tag{3.2.22}$$

From Equation 3.2.18, taking into account the expansion of h as in 3.2.19 we obtain that

$$H\left(\sigma(k+1) \mid M_k = t\right) = 1 - C\frac{t^2}{n^2} + O\left(\frac{t^3}{n^3}\right) \tag{3.2.23}$$

We introduce the following notation:

$$f(t) = \Pr\left[M_k = t\right] \tag{3.2.24}$$

$$F(t) = \sum_{s=0}^{t} \Pr\left[M_k = s\right] = \Pr\left[0 \leq M_k \leq t\right] \tag{3.2.25}$$

$$h(t) = 1 - H\left(\sigma(k) \mid M_k = t\right) \tag{3.2.26}$$

with this we can rewrite (3.2.17)

$$H\left(\sigma(k+1) \mid \sigma[k]\right) = \sum_t \Pr\left[M_k = t\right] H\left(\sigma(k+1) \mid M_k = t\right) = \tag{3.2.27}$$

$$= 1 - \sum_t \Pr\left[M_k = t\right]\left(1 - H\left(\sigma(k+1) \mid M_k = t\right)\right) = 1 - \sum_{t=-k}^{k} f(t)h(t) \tag{3.2.28}$$

In what follows we are going to give an upper bound on $\sum_{t=0}^{k} f(t)h(t)$ which will result in a lower bound for $H\left(\sigma(k+1) \mid \sigma[k]\right)$ and, in turn, for $H(\sigma[n])$.

According to summation by parts, we have:

$$\sum_{t=0}^{k} f(t)h(t) = F(k)h(k) - F(0)h(0) + \sum_{t=0}^{k-1} F(t)(h(t+1) - h(t)) = \tag{3.2.29}$$

$$= F(k)h(k) + \sum_{t=0}^{k-1} (F(k) - \Pr\left[M_k > t\right])(h(t+1) - h(t)) = \tag{3.2.30}$$

$$= \sum_{t=0}^{k-1} \Pr\left[M_k > t\right]\left(h(t+1) - h(t)\right) \tag{3.2.31}$$

where we first used that $F(0) = 0$ and after that $\sum_{t=0}^{k-1} F(k)(h(t+1) - h(t)) = F(k)h(k)$. Now we split the above sum into three parts and bound them separately.

$$\sum_{t=0}^{k-1} \Pr[M_k > t](h(t+1) - h(t)) = \sum_{t=0}^{L\sqrt{n}-1} (\dots) + \sum_{t=L\sqrt{n}}^{n^{\frac{3}{4}}-1} (\dots) + \sum_{t=n^{\frac{3}{4}}}^{k-1} (\dots) \tag{3.2.32}$$

Let us start with the first sum:

$$\sum_{t=0}^{L\sqrt{n}-1} \Pr[M_k > t](h(t+1) - h(t)) \leq \sum_{t=0}^{L\sqrt{n}-1} (h(t+1) - h(t)) = h(L\sqrt{n}) - h(0) \tag{3.2.33}$$

Using that $h(0) = 0$ and $h(L\sqrt{n}) = C\frac{L^2 n}{n^2} + O\left(\frac{n^{3/2}}{n^3}\right)$ by the approximation of (3.2.23), so

$$\sum_{t=0}^{L\sqrt{n}-1} \Pr[M_k > t](h(t+1) - h(t)) \leq CL^2\frac{1}{n} + O\left(\frac{1}{n^{3/2}}\right) \tag{3.2.34}$$

Now we turn to the second sum. By Lemma 3.2.4

$$\sum_{t=L\sqrt{n}}^{n^{\frac{3}{4}}-1} (\Pr[M_k > t](h(t+1) - h(t)) \leq \sum_{t=L\sqrt{n}}^{n^{\frac{3}{4}}-1} (h(t+1) - h(t))\frac{e^{-\frac{Ct^2}{n}}}{1 - 4e^{-\frac{t^2}{4n}}} \tag{3.2.35}$$

First note that $t = o(n)$, so we can still use the first order approximation to get $h(t+1) - h(t) = \frac{2t+1}{n^2} + O\left(\frac{t^2}{n^3}\right) = \frac{Ct+o(t)}{n^2}$.

One can choose L large enough so that for every large n both

$$e^{-\frac{Ct^2}{n}} \leq \left(\frac{t^2}{2n}\right)^{-2} \tag{3.2.36}$$

and

$$1 - 4e^{-\frac{t^2}{4n}} \geq \frac{1}{2} \tag{3.2.37}$$

are satisfied whenever $t \geq L\sqrt{n}$. With such an L we have:

$$\sum_{t=L\sqrt{n}}^{n^{\frac{3}{4}}-1} (h(t+1) - h(t))\frac{e^{-\frac{Ct^2}{n}}}{1 - 4e^{-\frac{t^2}{4n}}} \leq \frac{C}{n^2}\sum_{t=L\sqrt{n}}^{n^{\frac{3}{4}}-1} t\left(\frac{t}{2\sqrt{n}}\right)^{-4} = C'\sum_{t=L\sqrt{n}}^{n^{\frac{3}{4}}-1} t^{-3} \tag{3.2.38}$$

Now approximating the sum with the respective integral, we get that

$$\sum_{t=L\sqrt{n}}^{n^{\frac{3}{4}}-1} \Pr[M_k > t](h(t+1) - h(t)) \leq C"(L\sqrt{n})^{-2} - n^{-\frac{3}{2}}) \leq C"\frac{1}{L^2 n} \tag{3.2.39}$$

Finally, using the tail estimation of Lemma 3.2.4 and (3.2.37):

$$\sum_{t=n^{\frac{3}{4}}}^{k-1} \Pr[M_k > t](h(t+1) - h(t)) \leq \frac{1}{2}\sum_{t=n^{\frac{3}{4}}}^{k-1} e^{-\frac{Ct^2}{2n}} = o\left(\frac{1}{n}\right) \tag{3.2.40}$$

where we used the trivial bound $h(t+1)-h(t)\leq 1$.

Now we can put everything together, using that $\sum_{t=1}^{k} f(t)h(t)=\sum_{t=-k}^{-1} f(t)h(t)$.

$$1-H\left(\sigma(k)|\sigma([k-1])\right)\leq 2\sum_{t=0}^{k-1}\Pr\left[M_{k-1}>t\right]\left(h(t+1)-h(t)\right)\leq C\left(L^2+\frac{1}{L^2}\right)\frac{1}{n}+o\left(\frac{1}{n}\right) \tag{3.2.41}$$

and therefore, substituting this estimate into the chain rule we have that for some constant $K>0$

$$H(\sigma_\beta([n]))=\sum_{k=1}^{n} H(\sigma(k)|\sigma([k-1]))\geq n\left(1-K\frac{1}{n}+o\left(\frac{1}{n}\right)\right)=n-K+o\left(1\right) \tag{3.2.42}$$

□

3.2.2 General results for the Ising model

Theorem 3.2.6 (Van den Berg - Steif, 1999). *For $\beta<\beta_c$, the unique Ising measure μ on $\mathbb{Z}^d$ is a finitary factor of* $\mathrm{Unif}[0,1]^{\mathbb{Z}^d}$*, with coding radius $\mathbb{P}[R>t]<\exp(-ct)$.*

Now we are ready to prove that small enough sets are clueless with respect to the subcritical Ising measure. In view of Theorem 3.2.6 a transitive function of the Ising spins on a finite torus can be also regarded as a transitive function of iid bits, in which case we have a good control on the clue of subsets.

For the critical Ising, however, sparse reconstruction is possible:

Theorem 3.2.7 (Sparse Reconstruction at Critical Ising). *At $\beta=\beta_c$ on $\mathbb{Z}_n^2$, the total magnetization $M_n(\sigma):=\sum_x\sigma(x)$ can be guessed with high precision from the sparse magnetization $M_n^\epsilon(\sigma):=\sum_{n^\epsilon|x}\sigma(x)$,as long as $\epsilon<7/8$. This implies* $\mathsf{clue}_{M_n}(n^\epsilon\text{-grid})=1-o(1)$.

Proof. We know from **??** that $\mathbb{E}\sigma(x)\sigma(y)\asymp c\left\|x-y\right\|^{-\frac{1}{4}}$ and thus we can compute the order of magnitude of the variance of magnetization. We use a standard trick: We divide the square $\mathbb{Z}_n^2$ into logarithmically increasing anullii.

$$\mathrm{Var}M_n\asymp n^2+n^2\sum_{x\in\mathbb{Z}_n^2}\mathbb{E}\sigma(0)\sigma(x)=n^2+n^2\sum_{k=1}^{\log n}\left(2^k-2^{k-1}\right)\left(2^k\right)^{-\frac{1}{4}}\asymp n^2+n^2O(n^{2-\frac{1}{4}})$$

and in a similar way, for the magnetization of the sparse grid we get

$$\begin{aligned}\mathrm{Var}M_n^\epsilon &\asymp (\frac{n}{\epsilon})^2+(\frac{n}{\epsilon})^2\sum_{k=1}^{\log n}\left(2^{k(2-2\epsilon)}-2(k-1)(2-2\epsilon)\right)\left(2^k\right)^{-\frac{1}{4}}\\ &\asymp n^{2-2\epsilon}+n^{2-2\epsilon}O(n^{-\frac{1}{4}}n^{2-2\epsilon})\asymp O(n^{2-2\epsilon}+n^{4-4\epsilon-\frac{1}{4}}).\end{aligned}$$

Finally

$$\begin{aligned}\mathrm{Cov}(M_n,M_n^\epsilon)=\sum_{x\in\mathbb{Z}_n^2,y\in}\mathbb{E}\sigma(x)\sigma(y)&\asymp n^{2-2\epsilon}\sum_{x\in\mathbb{Z}_n^2}\mathbb{E}\sigma(0)\sigma(x)\\ &\asymp n^{2-2\epsilon}O(n^{2-\frac{1}{4}})\asymp O(n^{4-2\epsilon-\frac{1}{4}}).\end{aligned}$$

If we choose $\epsilon < 1 - \frac{1}{8} = \frac{7}{8}$, then in $\mathrm{Var} M_n^\epsilon$ the second term wins, so $\mathrm{Var} M_n^\epsilon = O(n^{4-4\epsilon-\frac{1}{4}})$. Putting all these together we get an estimation about the correlation:

$$\mathrm{Corr}(M_n, M_n^\epsilon) = \frac{O(n^{4-2\epsilon-\frac{1}{4}})}{O(n^{2-\frac{1}{8}})O(n^{2-2\epsilon-\frac{1}{8}})} = O(1)$$

□

3.3 Factor of IID measures

In this section we investigate sequences of spin systems that converges to finitary factor of IID systems. As this is a class of measures that are relatively approachable it is an obvious choice trying to understand them. Moreover some of the Ising models can also be described in this framework.

Finitary factor of IID systems are also interesting because basically they describe the type of measures that we have can investigate with experimental tools, by simulations. As the computational power of computers keeps increasing simulations has become an important (although not strictly mathematical) tool to understand the behavior of some systems. This is, however, possible only when there is an efficient way to sample from the distribution of the spin system we want to understand. Finitary factor of IID measures (spin systems) are those that can be sampled by a local algorithm from an IID spin system.

Definition 3.3.1 ((Finitary) Factor of IID systems)**.** Let $G = (V, E)$ be a transitive graph. A spin system on $\{-1, +1\}^V$ with distribution μ is a factor of IID, if there is a measurable map $\psi : [0, 1]^V \to \{-1, +1\}^V$ such that if $X \sim \mathrm{Unif}[0, 1]^V$ then the spin system defined by

$$\sigma_v := \psi(X_v) \quad v \in V(G)$$

is distributed w.r.t. μ.

A factor map is called *finitary*, if additionally, there is a random coding radius $R < \infty$ almost surely ,for which it holds that $\psi(X_v)$ is determined by $\{X_u : u \in B_R(v)\}$, including the value of R.

From a practical point of view the additional condition of being finitary guarantees that one can sample from the spin system, since σ_v is actually determined by a finite neighbourhood of X_v. Nevertheless, again from a practical point of view, if we have no control of the number of vertices $u \in V(G)$ for which X_u needs to be revealed, this condition is still not enough.

Therefore those finitary factor if IID systems where the coding volume (i.e the number of uniform random variables one needs to know to learn the value of a particular spin) has finite expected value bear special importance.

We would like to investigate under what conditions can we conclude that there is or there is not sparse reconstruction (or some of its variant) for a sequence μ_n converging to a Finitary factor of IID. We have to point out, in light of some negative results presented in Section **??**, that it is not at all clear that we can expect such results. Therefore we would like to narrow down the setup. We shall only consider such sequences of spin systems which themselves are factors of IIDs and in particular are generated by (a possibly truncated version of) the same local algorithm that we see in the limit.

Let $G(V,E)$ be an infinite, edge-decorated transitive graph and μ an $\mathsf{Aut}(G)$-invariant finitary factor of IID measure on $\{-1,+1\}^V$. We consider a sequence of transitive edge-decorated graphs $G_n(V_n,E_n)$ that converges to G locally.

From μ_n, however, we expect more than weak convergence: whenever n is large enough so that all N-neighbourhoods in G_n are the same as in G (such n exists thanks to the local convergence), and the coding radius $R \le N$ then we generate σ_v according to ϕ as a factor of IID. In case either of these conditions does not hold σ_v can be generated with some alternative local algorithm $\phi_n : [0,1]^{V_n} \to \{-1,+1\}^{V_n}$. It is clear that as $n \to \infty$ the probability that a given vertex σ_v is obtained with ϕ, tends to 1.

Let $S^+(f) := \sum_{g\in\mathsf{Aut}(G)} |\mathrm{Cov}(f,f^g)|$. We will need the following fact to give a simple sufficient condition on when we have WSR for a ffIID system.

Lemma 3.3.1. *Let μ be a ffIID spin system on G. Let $F \subseteq V(G)$ finite and let $f : \{-1,+1\}^F \to \mathbb{R}$. If μ_n converges to μ_n in the above sense, then*

$$S^+(f) = \infty \to \lim_{n\to\infty} S_n^+(f_n) = \infty$$

for some $f_n : \{-1,+1\}^{F_n} \to \mathbb{R}$, where $F_n \subseteq V_n$ satisfying $|F_n| = |F|$.

moreover, if μ_n satisfies the condition

$$\lim_{n\to\infty} \mathbb{P}[\exists\ u \in V_n : \quad R_u > \mathrm{Diam}(G_n)] = 0$$

(where $\mathrm{Diam}(G_n)$ is the diameter of the graph) then it is also holds that

$$\lim_{n\to\infty} S(f_n) = S(f)$$

whenever $S(f)$ is absolutely convergent.

Proof. First we define f_n for all n large enough so that the $d = \mathrm{Diam}(F)$-ball on G_n and on G are isomorphic. Observe that it is sufficient to give an injection from F onto G_n. Pick an arbitrary $r \in F$, and again an arbitrary $r' \in V_n$. Using the (rooted) isomorphism between the d-balls of r in G and r' in G_n, we can find a bijection between the vertices of a subset F_n in G_n and F.

We start by proving the first statement. Define the sequence of ffIID measures ν_n on G as follows. For every $v \in V$ we run the algorithm ϕ restricted to the ball $B_n(v)$. If the spin value σ_v can be calculated from $B_n(v)$ (where σ is distributed according to μ), then we write the respective spin value, otherwise we flip a coin independent from everything else, according to the distribution of σ_v conditioned on $B_n(v)$. It is clear that $\nu_n \xrightarrow{a.s.} \mu$ and that $\lim_{n\to\infty} S_{\nu_n}(f) = S_\mu(f)$. Therefore we can choose n such that $S_{\nu_n}(f) > L$

Fix a large number L and choose a finite subset $H \subset \mathsf{Aut}(G)$ in such a way, that $\sum_{g\in H} |\mathrm{Cov}(f,f^g)| > L$. For an arbitrary $v \in F$ we may choose a large, but fix r in such a way that $\bigcup_{g\in H} F^g \subseteq B_r(v)$.

Now choose $K > r$ large enough so that $\mathbb{P}[R_v > K] < \epsilon/|B_r(v)|$. Using the union bound, we have

$$\mathbb{P}[\forall\ u \in B_r(v) : \quad R_u \le K] = 1 - |B_r(v)|\mathbb{P}[\quad R_v > K] \le 1 - \epsilon.$$

Let $A := \{\forall\ u \in B_r(v) : \quad R_u \le K\}$. Note that for large enough n the ball $B_K(v)$ looks identical in G and in G_n and we now consider such an n. So we have the following estimate for the susceptibility of f_n on G_n:

$$S^+(f_n) \geq \sum_{g\in H} |\mathrm{Cov}(f_n, f_n^g)| \geq$$
$$\mathbb{P}[A] \sum_{g\in H} |\mathrm{Cov}(f_n, f_n^g)| \;\Big|\; A) = (1-\epsilon) \sum_{g\in H} |\mathrm{Cov}(f, f^g)| > (1-\epsilon)L.$$

Since L and ϵ was arbitrary, we are done.

For the second statement the same argument (here we choose $H \subset \mathsf{Aut}(G)$ such that $S^+(f) - \sum_{g\in H} |\mathrm{Cov}(f, f^g)| < \epsilon$) yields that $\liminf_n S(f_n) \geq S(f)$.

In order to see the other inequality let us denote by $B_n = \{\forall\ u \in V_n : \quad R_u \leq \mathrm{Diam}(G_n)\}$ and observe that conditioned on B_n $\mathrm{Cov}(f_n, f_n^g) = \mathrm{Cov}(f, f^g)$ and therefore

$$S^+(f_n) = \sum_{g\in \mathsf{Aut}(G_n)} |\mathrm{Cov}(f_n, f_n^g)| \leq \mathbb{P}[B_n] \sum_{g\in \mathsf{Aut}(G_n)} |\mathrm{Cov}(f, f^g)| + \mathbb{P}[B_n^c]s^2 \leq (1-\epsilon)S(f) + \epsilon s^2,$$

where $s^2 = \mathrm{Var}(f)$. □

In the sequel we will mostly focus on finitary factor of IID measures on the infinite d-dimensional lattice $G = \mathbb{Z}^d$. We note that most of the results can be extended to amenable graphs or polynomial growth graphs.

Lemma 3.3.2. *[BS99] If X is a finitary factor of an i.i.d. process on $\mathbb{Z}^d$ with $|X_i| \leq K$ almost surely. Let N_0 denote the coding radius of X_0. Suppose that the expected coding volume $\mathbb{E}[(N_0)^d]$ is finite. Then there is a constant C that only depends on K and d such that*

$$S^+(X) \leq C\mathbb{E}[(N_0)^d] \tag{3.3.1}$$

Proof.

$$\mathrm{Cov}(X_0, X_j) = \sum_{\max(k,l)\geq \frac{|j|}{2}-1} \mathrm{Cov}(X_0\mathbf{1}_{N_0=k}, X_j\mathbf{1}_{N_j=l}) + \sum_{\max(k,l)< \frac{|j|}{2}-1} \mathrm{Cov}(X_0\mathbf{1}_{N_0=k}, X_j\mathbf{1}_{N_j=l}) \leq$$

Let $\overline{X}_j := X_j - \mathbb{E}[X_j]$

$$\leq 4K^2 \sum_{\max(k,l)\geq \frac{|j|}{2}-1} \Pr[N_0 = k, N_j = l] + \sum_{\max(k,l)< \frac{|j|}{2}-1} \mathbb{E}[\overline{X}_0\mathbf{1}_{N_0=k}]\mathbb{E}[\overline{X}_j\mathbf{1}_{N_j=k}]$$

$$\leq 8K^2 \sum_{k\geq \frac{j}{2}-1} \Pr[N_0 = k] + \left(\sum_{k=0}^{\frac{|j|}{2}-2} \mathbb{E}[\overline{X}_0\mathbf{1}_{N_0=k}]\right)^2$$

Since $\overline{X}_0$ has 0 expected value

$$\left(\sum_{k=0}^{\frac{|j|}{2}-2} \mathbb{E}[\overline{X}_0\mathbf{1}_{N_0=k}]\right)^2 = \left(\sum_{k\geq \frac{|j|}{2}-1} \mathbb{E}[\overline{X}_0\mathbf{1}_{N_0=k}]\right)^2 \leq K^2 \sum_{k\geq \frac{|j|}{2}-1} \Pr[N_0 = k]$$

and therefore

$$\mathrm{Cov}(X_0, X_j) \le 9K^2 \sum_{k \ge \frac{|j|}{2}-1} \Pr[N_0 = k] =$$

$$K^2 \sum_{k\,:\,2k+1 \ge |j|} \Pr[2N_0 + 1 = 2k+1] \le K^2 \sum_{i \ge |j|} \Pr[2N_0 + 1 = i]$$

Observing that the number of $u \in \mathbb{Z}^d$ such that $|u| = m$ equals $C_0 m^{d-1}$ (since it is the boundary of a d-dimensional hypercube with side length $2m$) we get

$$\sum_{j \in \mathbb{Z}^d} \mathrm{Cov}(X_0, X_j) \le 9K^2 \sum_{j \in \mathbb{Z}^d} \sum_{i \ge |j|} \Pr[2N_0 + 1 = i] \le$$

$$C' \sum_{m=0}^{\infty} m^{d-1} \sum_{i \ge m} \Pr[2N_0 + 1 = i] = C' \sum_{i=1}^{\infty} \Pr[2N_0 + 1 = i] \sum_{m \le i} m^{d-1}$$

For every d there is a constant C_1 such that $\sum_{m \le i} m^{d-1} \le C_1 i^d$ thus

$$\sum_{j \in \mathbb{Z}^d} \mathrm{Cov}(X_0, X_j) \le C'' \sum_{j \in \mathbb{Z}^d} \Pr[2N_0 + 1 = i] i^d \le C\mathbb{E}[(N_0)^d]$$

□

Corollary 3.3.3. *If μ is a ffIID spin system on $\mathbb{Z}^d$ with $\mathbb{E}[(N_0)^d] = \infty$, where $(N_0)^d$ is the coding volume. Let μ_n be a sequence of ffIID spin systems converging to μ_n in the prescribed sense. Then there is weak sparse reconstruction for μ_n.*

Proof. According to Lemma 3.3.2, infinite coding volume implies that the absolute susceptibility of the system (of one spin) is infinite. In turn by Lemma 3.3.1, this means that the absolute susceptibility in μ_n tends to ∞ as n goes to ∞. Now we only need to show that in that case there is also a sequence of functions with $S(f_n)$ tending to ∞. Indeed, then by Corollary 3.1.9, it follows that there is sparse reconstruction for μ_n. □

Corollary 3.3.3 naturally raises the question, whether the converse is true. Can there be Sparse reconstruction for a sequence that converges to a ffIID spin system with finite expected coding volume? The following Theorem that relies on the IID case, gives a partial answer.

Theorem 3.3.4. *For any sequence of transitive, non-degenerated Boolean functions $f_n : \{-1,1\}^{V_n} \longrightarrow \{-1,1\}$. Let μ be a finitary factor of IID on $\mathbb{Z}^[d]$ with the property that $\mathbb{P}[R > t] < \exp(-ct)]$, where R is the coding radius, and μ_n is a factor of IID sequence converging to μ (in the sense specified above). Then for any subset $U_n \subseteq V_n$ satisfying $|U_n| = o(n^d/\log^d n)$,*

$$\mathsf{clue}(f_n \mid U_n) \to 0.$$

Proof. Let $U_n \subseteq V_n$ and for $u \in U_n$ let R_u denote the (random) coding radius of the spin σ_u. For any r be a positive integer, by the union bound

$$\mathbb{P}[\forall\, u \in U_n \;:\; R_u < r] = 1 - \mathbb{P}[\exists\, u \in U_n \;:\; R_u \ge r] > 1 - |U_n| \exp(-cr)$$

By definition, whenever $\{\forall\, u \in U_n : \quad R_u < r\}$ happens the spins in U_n can be calculated from at most $|U_n|r^d$ independent uniformly distributed variables. Denote the set of this bits by $J_n = \bigcup_{u\in U_n} B_r(u)$
Let us choose the sequence of integers r_n in such a way that

$$|J_n| \le |U_n| r_n^d \ll |V_n| = n^d, \tag{3.3.2}$$

and at the same time

$$\lim_n 1 - |U_n| \exp(-c r_n) = 1. \tag{3.3.3}$$

Now we can consider $g_n := f_n(\psi(X))$, the same function as f_n but interpreted as a function of the uniform IID variables X. Obviously g_n is transitive as well. On the one hand, it follows from Theorem 2.1.6 using the condition (3.3.2) that $\mathsf{clue}_{\mathrm{Unif}}\left(g_n \mid \mathcal{F}_{J_n(X)}\right) \to 0$.

On the other hand, conditioned on the event $\{\forall\, u \in U_n : \quad R_u < r\}$,

$$\mathbb{E}[f_n | \mathcal{F}_{U_n(\sigma)}] = \mathbb{E}_{\mu_n}[g_n | \mathcal{F}_{J_n(X)}] \tag{3.3.4}$$

and by (3.3.3) this happens with high probability.

Observe that if one chooses $r_n = K \log n^d$ with sufficiently large K (3.3.2) and (3.3.3) are both satisfied (for the latter using our assumption on the size of U_n).

Let us denote by $\mathcal{G}$ the minimal σ-algebra for which both the random set of uniform variables necessary to compute the spins of U_n and J_n are measurable. So $\mathbb{E}[g_n \mid \mathcal{G}] = \mathbb{E}[f_n \mid \mathcal{F}_{U_n(\sigma)}]$. Since $\mathcal{F}_{J_n(X)} \subseteq \mathcal{G}$ by the definition of $\mathcal{G}$, we have by Pythagoras's Theorem

$$\left\|\mathbb{E}[g_n \mid \mathcal{G}]\right\|^2 = \left\|\mathbb{E}[g \mid \mathcal{F}_{J_n(X)}]\right\|^2 + \left\|\mathbb{E}[g \mid \mathcal{G}] - \mathbb{E}[g_n \mid \mathcal{F}_{J_n(X)}]\right\|^2.$$

Subtracting the common squared expectation we get

$$\mathrm{Var}\left(\mathbb{E}[g \mid \mathcal{G}]\right) = \mathrm{Var}\left(\mathbb{E}[g \mid \mathcal{F}_{J_n(X)}]\right) + \left\|\mathbb{E}[g \mid \mathcal{G}] - \mathbb{E}[g \mid \omega_{V_n}]\right\|^2.$$

Since f_n is non-degenerate $\mathrm{Var}\left(\mathbb{E}[g \mid \mathcal{F}_{J_n(X)}]\right) \to 0$ (as we already pointed out) by Theorem 2.1.6 and the second term is smaller then $2\mathbb{P}[\mathbb{E}[f_n|\mathcal{F}_{U_n(\sigma)}] \ne \mathbb{E}[g_n|\mathcal{F}_{J_n(X)}]\ ]$ which again tends to 0 by (3.3.4). □

The reader might wonder whether this Theorem can be improved. Three natural direction comes into mind to improve the result. First, can we right $o(n^d$ instead of $= o(n^d/\log^d n)$ in the condition of the Theorem ? Second, can we substitute the exponential decay of the coding volume with some weaker condition (like finite expected coding volume)? Third, do we really need to assume non-degeneracy of the sequence?

We start by answering the third question, positively. We show an ffIID sequence converging to a ffIID spin system on $\mathbb{Z}$ in which one can reconstruct a sequence of functions surely, from a set of coordinates of constant size. The local algorithm is as follows: We read the bits starting from 0 going to the right, and we stop when we find two consecutive bits with equal value and this value will be the spin that we write at 0. In case we

3.4 Generalised DaC measures

Let N be a linear subspace of $\mathbb{F}_2^n$. Define the event $A_N\,\{\chi_S = 1\ : S \in N\}$ The measure $\Pr_N$ is defined as the uniform measure conditioned on A_N. Pr[n] Let $\mathcal{N}$ be a random linear subspace of $\mathbb{F}_2^n$. Then $\mathcal{N}$ induces a probability measure on $\{-1,1\}^n$ by

Definition 3.4.1. $\Pr_{\mathcal{N}}[\sigma = x] := \mathbb{E}[\Pr_{\mathcal{N}}[\sigma = x | A_{\mathcal{N}}]]$

We can generalise the Fourier-Walsh transform for these kind of measures. Let $f(x) = \sum_{T \subset [n]} \widehat{f}(T)\chi_T(x)$. It is straightforward to see that for every $N \in \mathcal{N}$

$$\Pr[\chi_S(\omega) = \chi_{S \oplus N}(\omega) | A_N] = 1.$$

With other words $\{-1,1\}^n$ is divided into $\mathcal{N}$-congruence classes, and conditioned on $A_{\mathcal{N}}$ and for any $\omega \in \{-1,1\}^n$ the value of $\chi_S(\omega)$ only depends on the congruence class to which S belongs to.

Let $\mathcal{F}(\mathcal{N}) = \mathcal{P}(A_{\mathcal{N}})$. For a fixed subspace $\mathcal{N}$ we can now write for any $x \in A_{\mathcal{N}}$

$$f(x) = \sum_{C \in \mathbb{F}_2^n / \mathcal{N}} \widehat{f_{\mathcal{N}}}(C)\chi_C(x),$$

where

$$\widehat{f_{\mathcal{N}}}(C) = \sum_{T \in C} \widehat{f}(T) = \sum_{N \in \mathcal{N}} \widehat{f}(T \oplus N). \tag{3.4.1}$$

Here T is any subset that belongs to the congruence class C. We note that instead of a linear subspace $\mathcal{N}$ of $\mathbb{F}_2^n$ we can also determine the measure by a linear transformation $\Pi : \mathbb{F}_2^n \to \mathbb{F}_2^n$ satisfying $\mathscr{K}(\Pi) = \mathcal{N}$ so that $\Pi(S) = \Pi(T)$ if and only if $S \oplus T \in \mathcal{N}$.

We can write 3.4.1 in a slightly different way. With an abuse of notation we denote by $\widehat{f_{\mathcal{N}}}(T)$ the Fourier-coefficient of the congruence class of T. For any $x \in A_{\mathcal{N}}$ we have

$$f(x) = \frac{1}{2^{\dim \mathcal{N}}} \sum_{T \subset [n]} \widehat{f_{\mathcal{N}}}(T)\chi_T(x).$$

There are $2^{\dim \mathcal{N}}$ subset in every $\mathcal{N}$-congruence class we counted every subset this many times, hence the normalization factor.

We also introduce the notation

$$\widehat{\widehat{f}}_{\mathcal{N}}(T) := \frac{\widehat{f_{\mathcal{N}}}(T)}{2^{\dim \mathcal{N}}}$$

to be able to write

$$f(x) = \sum_{T \subset [n]} \widehat{\widehat{f}}_{\mathcal{N}}(T)\chi_T(x).$$

Using that trivially $\mathbb{E}[\chi_T | A_{\mathcal{N}}] = 1$ if and only if $T \in \mathcal{N}$ and otherwise $\mathbb{E}[\chi_T | A_{\mathcal{N}}] = 0$

$$\mathbb{E}[f | A_{\mathcal{N}}] = \widehat{f_{\mathcal{N}}}(\emptyset)$$

Conditionally on $A_{\mathcal{N}}$ either $S \oplus T \in \mathcal{N}$, in which case $\chi_S = \chi_T$ or if $S \oplus T \notin \mathcal{N}$ then χ_S and χ_T are independent. In order to see this, note that for fixed $\mathcal{N}$ $\mathbb{E}[\chi_S | \mathcal{N}] = 0$, whenever $S \notin \mathcal{N}$. So $\mathrm{Cov}(\chi_S, \chi_T | \mathcal{N}) = 0$. Hence, after expanding any two function f and g we get

$$\mathbb{E}[fg | A_{\mathcal{N}}] = \sum_{C \in \mathbb{F}_2^n / \mathcal{N}} \widehat{f_{\mathcal{N}}}(C)\widehat{g_{\mathcal{N}}}(C) = \frac{1}{2^{\dim \mathcal{N}}} \sum_{T \subset [n]} \widehat{f_{\mathcal{N}}}(T)\widehat{g_{\mathcal{N}}}(T) \tag{3.4.2}$$

Where for the equality we again used the fact that when the sum is taken by subsets, each term of the form $\widehat{f}_{\mathcal{N}}(C)\widehat{g}_{\mathcal{N}}(C)$ is counted $2^{\dim \mathcal{N}}$ many times.

In particular:

$$\mathbb{E}[f^2|A_{\mathcal{N}}] = \sum_{C\in\mathbb{F}_2^n/\mathcal{N}} \widehat{f}^2_{\mathcal{N}}(C) = \frac{1}{2^{\dim \mathcal{N}}} \sum_{T\subset[n]} \widehat{f}^2_{\mathcal{N}}(T)$$

These observations allow to express the first of second moment of functions according to a DaC measure induced by a random linear subspace with the random Fourier coefficients $\widehat{f}_{\mathcal{N}}(T)$.

$$\mathbb{E}[f] = \mathbb{E}[\widehat{f}_{\mathcal{N}}(\emptyset)] = \mathbb{E}[\sum_{N\in\mathcal{N}} \widehat{f}(N)] = \sum_{T\subset[n]} \widehat{f}(T)\Pr[T\in\mathcal{N}]$$

Proposition 3.4.1. *Let μ be a DaC measure and let $\mathcal{N}$ be a corresponding random linear subspace. Then for any two functions $f,g:\{-1,1\}^n\mapsto\mathbb{R}$*

$$\mathbb{E}_\mu[f(\sigma)g(\sigma)] = \sum_{T\subseteq[n]} \mathbb{E}[\frac{1}{2^{\dim\mathcal{N}}}\widehat{f}_{\mathcal{N}}(T)\widehat{g}_{\mathcal{N}}(T)] \tag{3.4.3}$$

Proof. By the tower property of conditional expectation

$$\mathbb{E}[f(\sigma)g((\sigma)] = \mathbb{E}[\mathbb{E}[fg(\sigma)|\mathcal{N}]]$$

For a fixed $\mathcal{N}$, the Fourier expansion of f and g gives Equation 3.4.2. Now taking expectation on both sides yields the statement. ☐

Proposition 3.4.2. *Let μ be a DaC measure and let $\mathcal{N}$ be a corresponding random linear subspace. Then for any function $f:\{-1,1\}^n\mapsto\mathbb{R}$ with $\mathbb{E}_\mu[f]=0$*

$$S_\mu(f) = \frac{1}{n^d}\sum_{T\in[n^d]} \mathbb{E}\left[\frac{1}{2^{\dim\mathcal{N}}}\left(\sum_{j\in\mathbb{Z}_n^d}\widehat{f}_{\mathcal{N}^j}(T^j)\right)^2\right] \tag{3.4.4}$$

Proof. Since $\mathbb{E}_\mu[f]=0$ also $\mathbb{E}_\mu[f^j]=0$, because of the translation invariance of μ. Therefore $\mathrm{Cov}_\mu(f,f^j)=\mathbb{E}_\mu[ff^j]$. Therefore we can use Proposition 3.4.1.

First note that $(\chi_T)^j(\sigma)=\chi_T(\sigma^{-j})=\prod_{k\in T}\sigma_{k-j}=\chi_{T^{-j}}(\sigma)$. Therefore,

$$f^j = \sum_{T\subset[n]}\widehat{f}(T)\chi_{T^{-j}} = \sum_{T\subset[n]}\widehat{f}(T^j)\chi_T$$

that is, $\widehat{f^j}=\widehat{f}(T^j)$

For a fixed subspace $\mathcal{N}$ and a fixed subset T and $j\in\mathbb{Z}_n^d$ we have

$$\widehat{f^j}_{\mathcal{N}}(T) = \sum_{N\in\mathcal{N}}\widehat{f^j}(T\oplus N) = \sum_{N\in\mathcal{N}}\widehat{f}(T^j\oplus N^j) = \widehat{f}_{\mathcal{N}^j}(T^j)$$

where we used the simple fact that $(T\oplus N)^j=(T^j\oplus N^j)$.

We are now ready to express the susceptibility of f with the spectrum.

$$S_\mu(f)=\sum_{T\subseteq[n]}\sum_{k\in\mathbb{Z}_n^d}\mathbb{E}\left[\frac{1}{2^{\dim\mathcal{N}}}\widehat{f}_{\mathcal{N}}(T)\widehat{f^k}_{\mathcal{N}}(T)\right]=\sum_{T\subseteq[n]}\sum_{k\in\mathbb{Z}_n^d}\mathbb{E}\left[\frac{1}{2^{\dim\mathcal{N}}}\widehat{f}_{\mathcal{N}}(T)\widehat{f}_{\mathcal{N}^k}(T^k)\right]=$$
$$=\frac{1}{n^d}\sum_{k,l\in\mathbb{Z}_n^d}\sum_{T\subseteq n]}\mathbb{E}\left[\frac{1}{2^{\dim\mathcal{N}}}\widehat{f}_{\mathcal{N}^l}(T)\widehat{f}_{\mathcal{N}^{l+k}}(T^k)\right]$$

For the last equality we exploited the fact that the measure μ together with the random subspace $\mathcal{N}$ is transitive and therefore averaging over all translates of $\mathcal{N}$ returns the original measure.

Now let us divide the sum according to orbits of subsets. Since we enumerate all translates T^i in the orbit of $[T]$, every term is counted $|\mathbf{Stab}([T])|$ many times. Therefore we get for an orbit that $[T]$

$$\frac{1}{n^d|\mathbf{Stab}([T])|}\sum_{k,l,j\in\mathbb{Z}_n^d}\mathbb{E}\left[\frac{1}{2^{\dim\mathcal{N}}}\widehat{f}_{\mathcal{N}^l}(T^j)\widehat{f}_{\mathcal{N}^{l+k}}(T^{j+k})\right]=$$
$$\frac{1}{n^d|\mathbf{Stab}([T])|}\sum_{i,l,m\in\mathbb{Z}_n^d}\mathbb{E}\left[\frac{1}{2^{\dim\mathcal{N}}}\widehat{f}_{\mathcal{N}^{i+j}}(T^j)\widehat{f}_{\mathcal{N}^{i+m}}(T^m)\right]$$

where we used the substitutions $i=l-j$ and $m=j+k$. So we can write the above sum in a more concise form:

$$\frac{1}{n^d|\mathbf{Stab}([T])|}\sum_{i\in\mathbb{Z}_n^d}\mathbb{E}\left[\frac{1}{2^{\dim\mathcal{N}}}\left(\sum_{j\in\mathbb{Z}_n^d}\widehat{f}_{\mathcal{N}^{i+j}}(T^j)\right)^2\right]=$$
$$\frac{1}{|\mathbf{Stab}([T])|}\mathbb{E}\left[\frac{1}{2^{\dim\mathcal{N}}}\left(\sum_{j\in\mathbb{Z}_n^d}\widehat{f}_{\mathcal{N}^j}(T^j)\right)^2\right]$$

For the equality we again used that the measure is transitive.

Now if we sum according to all subsets T, every term as above is counted $|[T]|$ times. So we get

$$S_\mu(f)=\sum_{T\in[n^d]}\frac{1}{|\mathbf{Stab}([T])||[T]|}\mathbb{E}\left[\frac{1}{2^{\dim\mathcal{N}}}\left(\sum_{j\in\mathbb{Z}_n^d}\widehat{f}_{\mathcal{N}^j}(T^j)\right)^2\right]$$

Noting that by the Orbit Counting Lemma $|\mathbf{Stab}([T])||[T]|=n^d$ for every $T\subseteq[n^d]$ finishes the proof. □

Now we will try to establish some upper bounds for $S_\mu(f)$. In the sequel we are going to fix a $U\subseteq[n^d]$ and we will assume that f is $\mathcal{F}(U)$-measurable. This is equivalent to the Fourier coefficients of f being supported on subsets of U.

For every subset $T\subseteq[n^d]$ define a random subset $F(T)$ as follows:

$$F_{\mathcal{N}}(T):=\left\{l\in\mathbb{Z}_n^d\ :\ (T\oplus\mathcal{N})^l\cap\mathcal{P}(U)\neq\emptyset\right\}$$

Where $T\oplus\mathcal{N}=\{T\oplus N\ :\ N\in\mathcal{N}\}$.

Obviously, $F_{\mathcal{N}}(T)$ also depends on U, but to ease the notation we are going to suppress this dependence. Note that whenever $(T\oplus\mathcal{N})^j\cap\mathcal{P}(U)=\emptyset$ then $\widehat{f}_{\mathcal{N}^j}(T)=0$ since f is

$\mathcal{F}(U)$-measurable and thus all its Fourier coefficients $\widehat{f}_S$ are 0 in case $S \not\subseteq U$. Therefore, for fixed $\mathcal{N}$ the Cauchy-Schwarz inequality gives that:

$$\left(\sum_{j\in\mathbb{Z}_n^d} \widehat{f}_{\mathcal{N}^j}(T^j)\right)^2 \leq \sum_{j\in\mathbb{Z}_n^d} \widehat{f}^2_{\mathcal{N}^j}(T^j) \sum_{j\in\mathbb{Z}_n^d} \mathbf{1}^2_{\{j\in F_{\mathcal{N}}(T))\}} = |F_{\mathcal{N}}(T)| \sum_{j\in\mathbb{Z}_n^d} \widehat{f}^2_{\mathcal{N}^j}(T^j)$$

So we have for any $\mathcal{F}(U)$-measurable function f

$$\begin{aligned} S_\mu(f) \leq & \frac{1}{n^d} \sum_{T\in[n^d]} \mathbb{E}\left[\frac{1}{2^{\dim \mathcal{N}}} |F_{\mathcal{N}}(T)| \sum_{j\in\mathbb{Z}_n^d} \widehat{f}^2_{\mathcal{N}^j}(T^j)\right] \\ = & \frac{1}{n^d} \sum_{j\in\mathbb{Z}_n^d} \sum_{T\in[n^d]} \mathbb{E}\left[\frac{1}{2^{\dim \mathcal{N}}} |F_{\mathcal{N}}(T)| \widehat{f}^2_{\mathcal{N}^j}(T^j)\right] = \sum_{T\in[n^d]} \mathbb{E}\left[\frac{1}{2^{\dim \mathcal{N}}} |F_{\mathcal{N}}(T)| \widehat{f}^2_{\mathcal{N}(T)}\right] \end{aligned}$$

Where in the last equality we used again that our measure is transitive.

Now the question is how large $F_{\mathcal{N}}(T)$ can be? Let us start with an example: Suppose that $\mathcal{N}$ and T are such that there exist an l satisfying $l \in T \oplus N$ for all $N \in \mathcal{N}$. Now if $k \in F_{\mathcal{N}}(T)$ that is, there is $N \in \mathcal{N}$ such that $(T \oplus N)^k \subseteq U$, then obviously $l^k = l + k \in U$. But this may happen for at most $|U|$ many $k \in \mathbb{Z}_n^d$ and therefore we have the upper bound $|F_{\mathcal{N}}(T)| \leq |U|$.

We slightly generalise this idea: a set $L \subseteq [n]$ is a covering set for $(T, \mathcal{N})$ if $(T \oplus N) \cap L \neq \emptyset$ for every $N \in \mathcal{N}$. Following the argument above, in case $k \in F_{\mathcal{N}}(T)$ then for some $l \in L$ it holds that $l + k \in U$. This implies that $|F_{\mathcal{N}}(T)| \leq |L||U|$.

Definition 3.4.2 (Size of minimal covering set)**.**

$$\beta_N(T) = \begin{cases} \min |L| \ : \ (T \oplus N) \cap L \neq \emptyset \ \ \forall N \in \mathcal{N} & \text{if } T \notin \mathcal{N} \\ 0 & \text{if } T \in \mathcal{N}. \end{cases}$$

Observe that if $\emptyset \in T \oplus \mathcal{N}$ then there is no covering set. This happens if and only if $T \in \mathcal{N}$. So in this case we have to look for other methods to bound the susceptibility. Now we split the sum accordingly:

$$\begin{aligned} S_\mu(f) \leq & \sum_{T\in[n^d]} \mathbb{E}\left[\frac{1}{2^{\dim \mathcal{N}}} |F_{\mathcal{N}}(T)| \widehat{f}^2_{\mathcal{N}}(T) \mathbf{1}_{T\in\mathcal{N}}\right] + \sum_{T\in[n^d]} \mathbb{E}\left[\frac{1}{2^{\dim \mathcal{N}}} |F_{\mathcal{N}}(T)| \widehat{f}^2_{\mathcal{N}}(T) \mathbf{1}_{T\notin\mathcal{N}}\right] \\ \leq & n^d \mathbb{E}[\widehat{f}^2_{\mathcal{N}}(\emptyset)] + |U| \sum_{T\subseteq[n^d]} \mathbb{E}\left[\frac{1}{2^{\dim \mathcal{N}}} \beta_N(T) \widehat{f}^2_{\mathcal{N}(T)}\right] \end{aligned}$$

Note that $\mathbb{E}[\widehat{f}^2_{\mathcal{N}}(\emptyset)] = \mathbb{E}[(\mathbb{E}[f|\mathcal{N}])^2] - \mathbb{E}^2[\mathbb{E}[f|\mathcal{N}]] = \mathrm{Var}(\mathbb{E}[f|\mathcal{N}])$ using the fact that $\mathbb{E}^2[\mathbb{E}[f|\mathcal{N}]] = \mathbb{E}[f] = 0$ by assumption.

3.4.1 FK-Ising

It is well know that the Ising model can be described as a DaC model. Namely on any graph G one performs an edge percolation according to the law

$$\phi_{p,2}(\omega) = \frac{1}{Z} \prod_{e\in E} p^{\omega(e)} (1-p)^{1-\omega(e)} 2^{k(\omega)}$$

where $p \in (0,1)$ and $k(\omega)$ is the number of connectivity clusters.

For a given configuration ω let $C_1, C_2, \ldots C_{k(\omega)}$ denote the sizes of the connectivity clusters. It is easy to see that $\dim \mathcal{N} = \sum_i (C_i - 1) = n - k(\omega)$. Therefore

$$\mathbb{E}_\phi[f(\sigma)g(\sigma)] = \frac{1}{2^n} \sum_{T \subseteq [n]} \mathbb{E}_\phi[2^{k(\omega)} \widehat{f}_{\mathcal{N}}(T) \widehat{g}_{\mathcal{N}}(T)]$$

Now let $\mathscr{B}$ denote the Dac measure induced by the Bernouilli(p) percolation on G. Then by the definition of $\phi_{p,2}$:

$$\mathbb{E}_{\phi_{p,2}}[f] = \mathbb{E}_{\mathscr{B}}[2^k f]$$

so we have:

$$\mathbb{E}_\phi[f(\sigma)g(\sigma)] = \frac{1}{2^n} \sum_{T \subseteq [n]} \mathbb{E}_{\mathscr{B}}[2^{2k(\omega)} \widehat{f}_{\mathcal{N}}(T) \widehat{g}_{\mathcal{N}}(T)] = 2^n \sum_{T \subseteq [n]} \mathbb{E}_{\mathscr{B}}[\widehat{\widehat{f}}_{\mathcal{N}}(T) \widehat{\widehat{g}}_{\mathcal{N}}(T)]$$

Bibliography

[AGMT] D. Ahlberg, S. Griffiths, R. Morris, V. Tassion. Quenched Voronoi percolation. *Preprint*, `arXiv:1501.04075 [math.PR]`

[ALy07] D. Aldous and R. Lyons. Processes on unimodular random networks. *Electron. J. Probab.* **12** (2007), Paper 54, 1454–1508. `http://128.208.128.142/~ejpecp/viewarticle.php?id=1754`

[BKS99] I. Benjamini, G. Kalai and O. Schramm. Noise sensitivity of Boolean functions and applications to percolation. *Inst. Hautes tudes Sci. Publ. Math.* **90** (1999), 5–43. `[arXiv:math/9811157]`

[BS01] I. Benjamini and O. Schramm. Recurrence of distributional limits of finite planar graphs. *Electron. J. Probab.* **6** (2001), no. 23, 13 pp. (electronic). `[arXiv:math.PR/0011019]`

[BL89] M. Ben-Or and N. Linial. Collective Coin Flipping. *Advances in Computing Research* **5** (1989): 91–115.

[BS99] J. Van den Berg, and J. Steif . On the Existence and Nonexistence of Finitary Codings for a Class of Random Fields. *The Annals of Probability*, no. 3 (1999) 1501–522.

[Boy14] R. Bouyrie. On quantitative noise stability and influences for discrete and continuous models. *Preprint*, `arXiv:1401.7337 [math.PR]`.

[BDT08] R. Branzei, D. Dimitrov, S. Tijs. *Models in cooperative game theory.* Springer-Verlag Berlin Heidelberg, 2008.

[F18] M. P. Forsström. Denseness of volatile and nonvolatile sequences of functions. *Stochastic Processes and their Applications.* (2018), 128(11): 3880–3896.

[FK96] E. Friedgut and G. Kalai. Every monotone graph property has a sharp threshold. *Proc. Amer. Math. Soc.* **124** (1996), 2993–3002.

[Ga12] Gavinsky, Dmitry et al. A Tail Bound for Read-k Families of Functions. *Random Struct. Algorithms* **47** (2012): 99-108.

[GPS10] C. Garban, G. Pete and O. Schramm. The Fourier spectrum of critical percolation. *Acta Mathematica* **205** (1) (2010), 19–104. `arXiv:0803.3750 [math.PR]`

[GS15] C. Garban and J. E. Steif. *Noise sensitivity of Boolean functions and percolation.* Cambridge University Press, 2015. Preliminary version: `arXiv:1102.5761 [math.PR]`.

[JS16] J. Jonasson, J. E. Steif. Volatility of Boolean functions. *Stochastic Processes and their Applications.* (2016), 126(10): 2956–2975.

[KKL88] Jeff Kahn, Gil Kalai, and Nathan Linial. The influence of variables on Boolean functions. *29th Annual Symposium on Foundations of Computer Science*, **68-80**, 1988.

[K05] G. Kalai. Noise Sensitivity and Chaos in Social Choice Theory. *Bolyai Society Mathematical Studies.* **20** (2005), 10.1007/978-3-642-13580-_8.

[KK13] N. Keller, G. Kindler. Quantitative relationship beetween noise sensitivity and influences. *Combinatorica* **33** (2013), no. 1, 45–71.

[LSW02] G. F. Lawler, O. Schramm and W. Werner. One-arm exponent for critical 2D percolation. *Electronic Journal of Probability* **7** (2002), no. 2, 1–13. `arXiv:1003.1839 [math.CO]`

[Lub94] A. Lubotzky. *Discrete groups, expanding graphs and invariant measures.* Progress in Mathematics, 125. Birkhäuser Verlag, Basel, 1994. With an appendix by J. D. Rogawski.

[Mo10] Elchanan Mossel. *Gaussian bounds for noise correlation of functions . Geom. Funct. Anal.* Vol. 19 (2010) 17131756

[OD14] Ryan O'Donnell. *Analysis of Boolean functions.* Cambridge University Press, 2014. `http://get.analysisofbooleanfunctions.org`

[PSS09] Y. Peres, O. Schramm, and J. E. Steif. Dynamical sensitivity of the infinite cluster in critical percolation. *Annales de l'Institut Henri Poincaré, Probab. et Stat.* **45** (2009), 491–514. `arXiv:0708.4287 [math.PR]`

[Pet] G. Pete. *Probability and geometry on groups.* Book in preparation, `http://www.math.bme.hu/~gabor/PGG.pdf`.

[SW01] S. Smirnov and W. Werner. Critical exponents for two-dimensional percolation In *Math. Res. Lett.*, **8** (2001) pp. 729–744, `arXiv::[math.PR/0109120]`

[Ste09] J. E. Steif. A survey of dynamical percolation. In *Fractal geometry and stochastics IV*, Birkhäuser, pp. 145–174, 2009. `arXiv:0901.4760 [math.PR]`.

[SS10] O. Schramm and J. E. Steif. Quantitative noise sensitivity and exceptional times for percolation. *Ann. Math.* **171(2)** (2010), 619–672. `[arXiv:math.PR/0504586]`

[Ste09] J. E. Steif. A survey of dynamical percolation. In *Fractal geometry and stochastics IV*, Birkhäuser, pp. 145–174, 2009. `arXiv:0901.4760 [math.PR]`.

[Sm01] S. Smirnov. Critical percolation in the plane: conformal invariance, Cardy's formula, scaling limits. *C. R. Acad. Sci. Paris Sér. I Math.* **333** (2001), 239–244

[Tao05] Terence Tao Szemerédi's regularity lemma revisited *Contributions to Discrete Mathematics* **125** (2003), 195–224

[W09] Wendelin Werner. Lectures on two-dimensional critical percolation. In Statistical mechanics, IAS/Park City Math. Ser. **16** (2009), 297–360.