Noise sensitivity, pivotality and clue for transitive functions on spin 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

Acknowledgements

First of all, I would like to express my gratitude to my Ph. D. supervisor, Gábor Pete, who decided on one of his weak moments to accept me as his student. I learned a lot from him. While his frighteningly vast knowledge in several areas of mathematics was always at hand I could learn from him the process of thinking: how to attack and a research question. His enthusiasm and his relentless passion for mathematics has been contagious and it helped me through the difficult periods. Thanks to his sense of humor and his direct and informal style, it was also entertaining to work together. He gave me a lot of freedom, but whenever needed he was available for help literally 24 hours a day.

I would like to express my thanks to a few excellent researchers who supported my research with their ideas. To Hugo Duminil-Copin for his assistance on the proof of Proposition 3.4.4, to Christophe Garban, for his helpful comments on Theorem 1.2.3 and his help and insights on Theorem 3.4.5, to Tom Hutchcroft, for his inspiration that helped to complete Theorem 1.2.3, to Balázs Szegedy for his suggestion that lead to Proposition 3.2.8, to Ádám Timár, for suggesting Question 3.3.9 (and for some educative discussions in Jakarta and in Kebun Raya Bogor). Last but not least, I would like to express how grateful I am to Itai Benjamini and Gil Kalai for sharing their exciting questions.

I would also like to thank to Károly Böröczky, Head of the Department of Mathematics. Always relaxed, helpful and reassuring, ready to cope with anything that comes in the way. With him in the background one felt safe at any time. I am also grateful to Elvira Kadvány and Melinda Balázs, for their pro-activity, help and patience about practical or administration matters and their kindness, in general.

Finally, I would like to thank my family, who were helpful and patient and they were next to me and it was good that they were next to me. My wife, Hülya, with whom we went through the whole thing together and from whom I got plenty of encouragement, our cats, Sarıkız and Sezar who stayed always relaxed, my parents and my sister Zsuzska, who were always supportive and understanding, and my grandmother whom I lost on the way.

Abstract

In Chapter 1 we give a brief introduction into the analysis of Boolean functions. Several equivalent definitions of noise sensitivity are discussed. We highlight the complex relationship between noise sensitivity/stability and the pivotal set. In particular, answering a question of G. Kalai, we construct a noise stable sequence of monotone, transitive Boolean functions which have many pivotals with high probability. This part of the chapter is based on the solo paper [Ga19+].

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 in U_n gives us a non-vanishing amount of 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 in different ways, measuring information content in L^2 and with entropy. We also highlight some interesting connections with cooperative game theory. Furthermore, we show that the left-right crossing event for critical planar percolation on the square lattice does not admit sparse reconstruction either. These results answer questions posed by I. Benjamini.

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 than $1/|V_n|$, 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.

Chapters 2 and 3 are based on joint research with Gábor Pete.

Chapter 1

Noise Sensitivity and the Pivotal Set

1.1 Introduction to Noise Sensitivity and Noise Stability

1.1.1 Basic Definitions

Noise Sensitivity for Boolean functions was introduced in the seminal work of Benjamini, Kalai and Schramm [BKS99]. One of the main motivations behind this concept was 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. Here comes the formal definition:

Definition 1.1.1 (Noise Sensitivity)**.** Let ϵ be a positive real number. For a uniform random vector $\omega \in \{-1,1\}^{k_n}$ denote ω^ϵ the random vector which we obtain from ω by resampling each of its bits independently with probability ϵ. A sequence of non-degenerate functions $f_n : \{-1,1\}^{k_n} \longrightarrow \{-1,1\}$ is noise sensitive if and only if for every $\epsilon > 0$

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

The parity of the number of -1s is a noise sensitive function. The most renowned example of a noise sensitive Boolean function, which initiated the whole investigation is the left right crossing event in critical planar percolation (see Section 2.5.1).

We note that the above definition naturally extends to $\mathbb{R}$-valued functions in case in the definition we substitute asymptotic decorrelation with asymptotic independence. (For binary-valued function decorrelation is equivalent to independence.)

Our main interest will be non-degenerate sequences. A sequence of Boolean functions $f_n : \{-1,1\}^{V_n} \longrightarrow \{-1,1\}$ is called non-degenerate if there exists an $\epsilon > 0$ such that for all n

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

where the expectation is taken according to the uniform measure on $\{-1,1\}^{V_n}$.

Remark 1.1.1. The usual definition of noise sensitivity is slightly different. In [GS15] a sequence of Boolean functions is said to be noise sensitive if for every $\epsilon > 0$

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

This basically means that the expected covariance 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 usual definition and ours are equivalent for non-degenerate functions, see Theorem 1.1.4. We have a preference for this form since it features the notion of clue (Definition 2.1.1), one of the central concepts in this work. Using the concept of clue, Definition (1.1.1) states that $\lim_n \mathsf{clue}(f_n \mid \omega^\epsilon) = 0$

The notion of noise sensitivity has applications in complexity theory and social choice theory. In an influential article ([MOO05]) it was proved that among low influence sequences of Boolean functions Majority is the most noise stable — see the definition below — ('Majority is the stablest' theorem). This fact has some far reaching consequences in complexity theory. It turns out that assuming the Unique Games Conjecture it follows that it is NP-hard to better approximate the Max-Cut problem then the best known algorithm (Goemans-Williamson algorithm, [GW95]).

In the social choice theory framework 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 setup noise sensitivity of a voting system means that even if a small ϵ ratio of the votes are corrupted there is a reasonable chance that this will change the outcome of the election. See [K05] for applications of noise sensitivity in the social choice theory setting.

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

Definition 1.1.2 (Noise Stability)**.** A sequence of functions $f_n : \{-1,1\}^{k_n} \longrightarrow \{-1,1\}$ is noise stable if and only if

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

The most important examples of noise stable functions are the majority function defined as

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

and the dictator which equals ω_i for a specific coordinate $i \in V$ (i.e., the dictator).

1.1.2 The Fourier-Walsh expansion

We introduce a function transform on the hypercube which turns out to be an essential tool in the analysis of Boolean functions. We still consider the uniform measure $\mathbb{P}_{1/2} := (\frac{1}{2}\delta_{-1} + \frac{1}{2}\delta_1)^{\otimes V_n}$. We can introduce the natural inner product $(f,g) = \mathbb{E}[fg]$ on the space of real functions on the hypercube.

Definition 1.1.3 (Fourier-Walsh expansion)**.** For any $f \in L^2(\{-1,1\}^V, \mathbb{P}_{1/2})$ and $\omega \in \{-1,1\}^V$

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

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 in fact the characters of $\mathbb{Z}_2^V$.

It is straightforward to check that the functions χ_S form an orthonormal basis with respect to the inner product so 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 in T. So $\mathcal{F}_T$ expresses knowing the coordinates in T. It turns out that the conditional expectation of any function $f : \{-1,1\}^n \longrightarrow \mathbb{R}$ 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 $S \subseteq T$ then $\mathbb{E}[\chi_S \mid \mathcal{F}_T] = \chi_S$ in any other case $\mathbb{E}[\chi_S \mid \mathcal{F}_T] = 0$.

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 is re-randomised according to the uniform measure. It is easy to see that after time t the joint distributions of (ω^0, ω^t) and $(\omega, \omega^\epsilon)$ are the same with the conversion $\epsilon = 1 - e^{-t}$.

So, on the one hand, we have the original interpretation of noise sensitivity and stability reminiscent to information theory. That is, we try to compute a piece of information (represented by the value of f_n) but the input is corrupted with noise. The question addressed by noise sensitivity is the following: can we recover the original information?

On the other hand, we have a more geometric/probabilistic kind of interpretation. We perform a simple random walk on the discrete hypercube and we have a subset of vertices $\mathcal{A}_n$ (represented via $f_n = \mathbb{1}_{\mathcal{A}_n}$). The question is whether after an arbitrary small but fixed amount of time we can remember if we started the walk from a vertex in $\mathcal{A}_n$ or not.

One can also think about the Fourier-Walsh expansion in a more probabilistic way. Observe the functions χ_S are the eigenfunctions of the simple random walk on the hypercube or.

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

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

Let us introduce the operator

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

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

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

This shows that the function χ_S is eigenfunction of the operator $T_\epsilon[\,]$.

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. *[GS15] 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 We can normalize this measure to get a probability measure. The random subset $\mathscr{S}_f$ distributed accordingly is called the spectral sample.

Definition 1.1.4 (Spectral Sample)**.** Let $f \in L^2(\{-1,1\}^V, \mathbb{P}_{1/2})$. The spectral sample $\mathscr{S}_f$ of f is a random subset of V chosen according to the distribution

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

The advantage of this concept is that it introduces a new and rather compact language, where the concepts that we introduced so far admit straightforward translations. 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. So we get the following (which is, in fact no more than a rephrasing of Theorem 1.1.2:

Proposition 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 $|\mathscr{S}_{f_n}| \neq 0$, $|\mathscr{S}_{f_n}| \to \infty$ 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 the 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 concepts 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, as we shall see is one of the key steps in the (first) proof Theorem 2.1.1.

1.1.3 Equivalent Characterisations of Noise Sensitivity

Here we collect a few statements that are equivalent to noise sensitivity. 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 a set V we introduce the level p Bernoulli random subset $\mathcal{B}(V)^p$ of V. That is, each $i \in V$ is in $\mathcal{B}(V)^p$ with probability p, independently from what happens to the other elements.

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. *Let $\mathbb{E}[\mathsf{clue}(f_n \mid \mathcal{B}(V_n)^p)] := \mathbb{E}\left[\frac{\mathrm{Var}(\mathbb{E}[f_n \mid \mathcal{B}(V_n)^p])}{\mathrm{Var}(f_n)}\right]$ (See Definition 2.1.1), where $\mathcal{B}(V_n)^p$ is the Bernoulli p random subset of V_n. For every $p \in (0,1)$*
$$\lim_{n\to\infty} \mathbb{E}[\mathsf{clue}(f_n \mid \mathcal{B}(V_n)^p)] = 0.$$
4. *Let $\mathbb{P}_{f_n}$ be the uniform measure on $\{f_n = 1\}$ and consider a simple random walk $\{X^t \ : t \in [0,\infty)\}$ on the hypercube with initial distribution $\mathbb{P}_f$. Denote by $\mathbb{P}^t_f[\omega]$ the measure according to the distribution of X^t. Then for every $t > 0$*
$$\lim_{n\to\infty} \|\mathbb{P} - \mathbb{P}^t_f\|_1 = 0.$$
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 $\epsilon \in (0,1)$*
$$\mathbb{E}[f_n(\omega^\epsilon) \mid \omega] \xrightarrow{\mathbb{P}} \mathbb{E}[f].$$

Proof. $(1 \Leftrightarrow 2)$

Note that $\mathbb{E}[\chi_S(\omega)\chi_S(\omega^\epsilon)] = \prod_{i\in S} \mathbb{E}[\omega(i)\omega_i^\epsilon] = (1-\epsilon)^{|S|}$. 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 $(\omega^\epsilon, \omega)$ has the same distribution as $(\omega, \omega^\epsilon)$ 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-degenerate, therefore $1/\mathrm{Var}(f_n)$ is just a constant factor and the equivalence follows.

$(1 \Rightarrow 3)$

$$\mathbb{E}[\mathrm{Var}(\mathbb{E}[f \mid \mathcal{B}(V_n)^p])] = \mathbb{E}[\sum_{\emptyset \neq S \subset V} \widehat{f}(S)^2 \mathbb{1}_{S \subseteq \mathcal{B}(V_n)^p}]$$
$$= \sum_{\emptyset \neq S \subset V} \widehat{f}(S)^2 \mathbb{P}[S \subseteq \mathcal{B}(V_n)^p] = \sum_{\emptyset \neq S \subset V} \widehat{f}(S)^2 p^{|S|}.$$

Using $p = 1 - \epsilon$ and noting again that the variance of f_n is of constant order, we get the desired equivalence.

$(1 \Rightarrow 4)$

We note that from a probabilistic perspective $\frac{1}{2}\| \quad \|_1$ is the total variation distance of measures.

Observe that the Radon-Nikodym derivative $\frac{d\mathbb{P}_{f_n}}{d\mathbb{P}}$ is $\frac{2}{\mathbb{E}[f_n]+1}$ if $f_n(\omega) = 1$ and 0 otherwise, thus for any $\omega \in \{-1,1\}^{V_n}$

$$\frac{d\mathbb{P}_{f_n}}{d\mathbb{P}}(\omega) = \frac{f_n(\omega)+1}{\mathbb{E}[f_n]+1}.$$

Similarly,

$$\frac{d\mathbb{P}^t_{f_n}}{d\mathbb{P}}(\omega) = \frac{\mathbb{E}[f_n(\omega^t) + 1 \mid \omega]}{\mathbb{E}[f_n]+1}.$$

So we can write

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

Using the Cauchy-Schwarz inequality we get

$$\|\mathbb{P} - \mathbb{P}^t_{f_n}\|_1 \leq \sum_{\omega \in \{-1,1\}^{V_n}} \sqrt{\frac{1}{2^{|V_n|}}} \left(\sqrt{\frac{1}{2^{|V_n|}}} \left| 1 - \frac{\mathbb{E}[f_n(\omega^t) \mid \omega] + 1}{\mathbb{E}[f_n]+1} \right| \right) \leq \left\| 1 - \frac{\mathbb{E}[f_n(\omega^t) \mid \omega] + 1}{\mathbb{E}[f_n]+1} \right\|_2.$$

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

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

$(4 \Rightarrow 5)$

Let X_t as before the simple random walk with initial distribution $\mathbb{P}_{f_n}$ and t such that $1 - e^{-t} = \epsilon$.

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

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

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

$(5 \Rightarrow 6)$

Indirectly assume, that for a $\delta > 0$ and for all n

$$\mathbb{P}[|\mathbb{E}[f_n(\omega^\epsilon) \mid \omega] - \mathbb{E}[f_n]| > \delta] > c,$$

for some $c > 0$. But then $\mathbb{E}[|\mathbb{E}[f_n(\omega^\epsilon) \mid \omega] - \mathbb{E}[f_n]|] > \delta c$, so either $|\mathbb{E}[f_n(\omega^\epsilon) \mid f_n(\omega) = 1] - \mathbb{E}[f_n]|$ or $|\mathbb{E}[f_n(\omega^\epsilon) \mid f_n(\omega) = -1] - \mathbb{E}[f_n]|$ is greater than δc for all n, which is in contradiction with our assumption.

$(6 \Rightarrow 1)$

Since with high probability $\mathbb{E}[f_n(\omega^\epsilon) \mid \omega]$ is δ-close to $\mathbb{E}[f_n]$ and otherwise $|\mathbb{E}[f_n(\omega^\epsilon) \mid \omega] - \mathbb{E}[f_n]| \leq 2$, it is clear that $\mathrm{Var}(\mathbb{E}[f_n(\omega^\epsilon) \mid \omega])$ tends to 0 as n goes to ∞. □

Again we remind the reader that the usual definition of noise sensitivity is not normalized with the variance. In the original article [BKS99] noise sensitivity is defined according to Property 6 above, while in [GS15] it is defined with Property 2. This means that in the usual setting degenerate sequences are automatically noise sensitive.

The intuition behind this is that with high probability we know everything about a degenerate sequences of functions and what else a probabilist can ask for? Our point is that it might be meaningful to differentiate between degenerate sequences as well, depending on the speed of decorrelation. Also, defining noise sensitivity via covariance or conditional variance is semantically vague, as these notions – in contrast with clue or correlation – are not dimensionless concepts expressing information content.

It might be of interest to point out that some of the above statements do not admit a straightforward generalization for the degenerate case which is equivalent to the definition we suggest. Statements 1 and 3 are obviously equivalent even for degenerate sequences, while statement 2 only needs to be rescaled by the variance.

The other statements, however, are not that easy to extend to the degenerate case. As for statement 4, if the set $\{f_n = 1\}$ is small then there is no chance that after a constant amount of time $\mathbb{P}_f^t$ will be close to the stationary measure $\mathbb{P}$. This suggests that one needs to rescale time according to the density of $\{f_n = 1\}$. Similar problems arise for statements 5 and 6, which are vacuous for degenerate functions. Perhaps one needs to impose some additional requirements on the speed of convergence to differentiate noise sensitive functions from the rest.

Here we defined noise sensitivity with respect to the uniform measure on the hypercube. But one may wonder if this concept can be extended to different probability spaces. One of the difficulties of such an extension lies in finding a suitable notion of noise that can be quantified and preserves the probability measure.

In case is there is a natural dynamics with stationary measure $\mathbb{Q}$, noise can be thought of as running the dynamics for some fixed time according to some suitably chosen time scale. This idea is used in [F15] to define noise sensitivity for Markov Chains. However, there are many cases, for example the low temperature Ising models, for which it is not easy to find a suitable process.

We would like to point out that Statement 3 from Theorem 1.1.4 could be the basis of noise sensitivity concept that does not rely on some dynamics. The average clue of a level p Bernoulli subset can be extended without any difficulty for any probability measure.

One difficulty might be to find the right scaling. If fro example the probability measure in question has a lower entropy and there is a lots of structure, it is possible that after learning the bits of a $1-\epsilon$-density subset we know the whole configuration with a non-vanishing probability. In this way all function can have non-vanishing clue, thus there are no noise sensitive sequences. One might fix this by relaxing the conditions of Statement 3. Instead of requiring $\lim_{n\to\infty}\mathbb{E}[\mathsf{clue}(f_n \mid \mathcal{B}(V_n)^p)] = 0$ for all $0<p<1$, one may specify some it to hold for some particular sequences of p_n.

1.2 Noise Sensitivity versus Pivotals

1.2.1 Pivotal set, Influence and the Spectrum

In [KKL88], way before noise sensitivity, a very natural discrete partial derivative concept had been introduced and studied for Boolean functions. We introduce the influence of variables and the pivotal set 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.

The 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 $I(f) := \sum_{j\in V} I_j(f)$.

Perhaps 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 by calculating the Fourier-Walsh expansion of $\partial_i f = f(\omega^j) - f(\omega)$. 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* $i,\ j\ \in V$

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

For a proof, see [GS15] Corollary IX.7. The fact that the one dimensional marginals are equal follows 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 which originally appeared in [LMN93], but can also be found in [GS15] (Proposition IX.5).

The sudden idea that the two random sets might follow the same distribution can be easily discarded as already on three bits one can find counterexamples. Still, these observations raise the possibility of characterising noise sensitivity and stability with the help of influences or the pivotal set. According to Proposition 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 both the Fourier spectrum and the pivotal set is typically on sets larger than n^ϵ for some $\epsilon > 0$. See [GPS10] or Chapter X in [GS15] for further details on the similarities and differences between the distribution of the spectrum and the pivotal set.

Another important example is the Majority Maj_{2n+1}. 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 on the order of $\sqrt{n}$ (this can be easily verified). 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. their expected size in general does not indicate 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 functions (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 are 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 efficient way to sample the spectral sample (although it is worth mentioning that according to [BV93] show that there is at least an efficient quantum algorithm for computing the Fourier-Walsh transform of certain functions).

Still there is 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 the parity function, but it is true 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. This interpretation will show up again. As we shall see, the covariance of two functions can be expressed as an integral of the size of the intersection of two p-dependent pivotal sets, see (2.1.5).

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 $f_n : \{-1,1\}^{V_n} \longrightarrow \{-1,1\}$ such that f_n is transitive, monotone and noise stable, but at the same time $\mathbb{P}[\mathscr{P}_n(\omega) \neq \emptyset] > c$ for some constant $c > 0$ for all $n \in \mathbb{N}$?

Let Γ be a group acting 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 $\gamma \in \Gamma$ we denote by $f^\gamma(x) := f(x^{-\gamma})$ the action of Γ on the function f.

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

In the language of social choice theory the transitivity of a voting scheme can be interpreted as each voters are treated the same.

We are going to show that the answer to the question above 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 of a sequence 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. Here is a sketch of a sequence of monotone functions which is transitive, not noise sensitive and the pivotal set is nonempty with a uniformly positive probability.

Take a noise sensitive sequence g_n of monotone, transitive, non-degenerate Boolean functions on k_n bits with the property that its pivotal set is non-empty with a probability larger then a $c > 0$ for all large n. For example the standard Tribes_n function satisfies these conditions. Let Maj_{k_n} be the Majority function on the same k_n bits. Now let

$$f_n = \begin{cases} \mathsf{Tribes}_n & \text{if } \mathsf{Maj}_{k_n} = -1 \\ 1 & \text{if } \mathsf{Maj}_{k_n} = 1. \end{cases}$$

It is easy to verify that f_n is monotone, transitive and admits pivotals with a positive probability. At the same time it is asymptotically positively correlated with Maj_{k_n} and therefore cannot be noise sensitive. (Observe that a noise sensitive and a noise stable sequence must be asymptotically uncorrelated.)

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

(Recall that 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 the votes of some people who can turn the result in a particular direction.

It might be also interesting to think of the bribable sequence geometrically. We may consider the subset of the hypercube on which the value of the bribable function is nonzero. This results in a sequence of invariant and monotone subsets of the hypercube with density going to 0, but with the property that almost any vertex of the hypercube is a neighbour, meaning that it can be reached from the set by an edge.

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

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

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

From now on we will write m_n/l_n instead of k_n. The question we would like to answer is the following: What conditions we need to impose on the sequences m_n and p_n so that we can find an appropriate sequence $l_n \ll m_n$ that both (1.2.6) and (1.2.7) are satisfied?

First we note that in case $0 < \inf_n p_n \leq \sup_n p_n < 1$ the argument used in the uniform case keeps working. So we are going to 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

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

Taking logarithm from (1.2.7), we get

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

omitting the term $\log q_n \to 0$. Expressing $\log m_n$ from (1.2.9) and plugging it into (1.2.8) we obtain

$$d_n - \log\frac{1}{p_n} - \log l_n = d_n - \log\frac{l_n}{p_n} \to -\infty. \tag{1.2.10}$$

First, in case $\log\frac{1}{p_n} \geq o(\log m_n)$, we can set l_n so that $l_n \log\frac{1}{p_n} \asymp \log m_n$. Like this we have $d_n \asymp \log\frac{1}{p_n}$ and thus (1.2.10) becomes $-\log l_n$ and both conditions are satisfied.

On the other hand, in n case $\log\frac{1}{p_n} = c\log m_n + o(\log m_n)$ for some $c \in (0,1)$, and suppose that $c = \frac{p}{q} + \alpha$ where p does not divide q and $\alpha < \frac{1}{q}$. Under these conditions let us set $l_n = \lfloor q/p \rfloor + 1$. Then $d_n = (\frac{q \mod p}{q} - \alpha)\log m_n - o(\log m_n)$. From our assumptions it is clear that $(\frac{q \mod p}{q} - \alpha) > 0$ and thus (1.2.8) is satisfied. For (1.2.10) we have $(\frac{q \mod p - p}{q} - \alpha)\log m_n - \log\lfloor q/p \rfloor$ which clearly tends to infinity, so again our conditions are satisfied.

This fails to work in general, when $c = 1/k$. If $\log\frac{1}{p_n} = 1/k\log m_n + o(\log m_n)$ with the $o(\log m_n)$ tending to infinity we set $l_n - 1 = k$, and we are OK. But if the error term is bounded, (surprisingly) we don't know how to handle the case.

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

$$\log m_n - \log l_n - q_n l_n \to -\infty. \tag{1.2.11}$$

While taking logarithm from (1.2.7) we get

$$d_n := \log m_n - \log\frac{1}{q_n} - q_n l_n \to \infty \tag{1.2.12}$$

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 (1.2.11):

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

If we choose d_n as for example $\frac{1}{2}\log\log(q_n m_n)$, then (1.2.11) and (1.2.12) 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, $l_n \to \infty$.

This analysis shows that if there is an $o < \alpha 1$ such that $p_n, q_n \geq 1/n^\alpha$ (except for $p_n, q_n = n^{-1/k}$ for $k \in \mathbb{N}$) then we still have a sequence of Boolean functions with the properties of Theorem 1.2.3.

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 $A_n := \{|\mathscr{P}_n| \le a_n\}$. It is clear that $\mathbb{E}[\int_0^1 \mathbb{1}_{X_n(t)\in A_n} dt] = \mathbb{P}[|\mathscr{P}_n| \le a_n] \to 0$ 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 $e^{-(1-\epsilon)a_n} < \epsilon$

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

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 slightly 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.

1.2.4 An alternative construction of a bribable sequence

Here we sketch a completely different way of constructing a bribable sequence. Its disadvantage is that it is non-monotone therefore it only implies a somewhat weaker result. 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 $j = 1, 2, \dots L$ 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 $j = 1, 2, \dots L$ 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, $\sum_{j \in [L]} \sigma_j \leq -2k$).

1.2.5 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. Once noise sensitivity is established, one can take different sequences $\epsilon_n \to 0$ and investigate whether the asymptotic decorrelation survives or not if the original input is perturbed with a small ϵ_n noise.

A randomized algorithm $\mathcal{A}$ for a Boolean function $\{-1,1\}^V \longrightarrow \{-1,1\}$ queries the bits ω_j for $j \in V$ one by one, in such a way that the decision of the bit to ask next might be made based on the outcome of the values already learned and on external randomness as well.

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 $J_{\mathcal{A}} \subseteq V$ denote the random set of coordinates queried by the algorithm $\mathcal{A}$ until it learns the value of f. Let R denote all possible random algorithms on $\{-1,1\}^V$ The revealment of f is

$$\delta_f = \inf_{\mathcal{A} \in R} \max_{j \in V} \mathbb{P}[j \in J_{\mathcal{A}}] \tag{1.2.13}$$

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

Theorem 1.2.13 (Revealment and Noise Sensitivity)**.** *Let $f : \{-1,1\}^V \longrightarrow \{-1,1\}$ then*

$$\frac{\sum_{|S|=k} \widehat{f}(S)^2}{\|f\|_2^2} \leq \delta_f k \tag{1.2.14}$$

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 L^2-Clue and 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 itself invariant under a transitive group of 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 bits of vertices in U.

Definition 2.1.1 (L^2-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. This concept first appeared under this name in [GPS10].

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 a 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 an 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.

We continue formalising the informal question posed at the beginning. We introduce the concept of sparse reconstruction which essentially formalizes the question we asked for general functions. We should emphasize that this is one of the central concept of this thesis.

Definition 2.1.2 (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?

The answer turns out to be negative. The following theorem provides a sharp upper bound on the clue of not only Boolean, but general real-valued transitive functions. The proof 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 1.1.2).

Theorem 2.1.1 (Clue of Transitive Functions)**.** *If $f : \{-1,1\}^V \longrightarrow \mathbb{R}$ 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:

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

where $\tilde{\mathbb{P}}$ denotes the probability measure conditioned on $\{\mathscr{S} \neq \emptyset\}$. □

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

It is worth to point out 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.

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 the 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 $\epsilon > 0$

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

holds. The Tribes function $\mathsf{Tribes}(l_n, k_n)$ — which has already been defined in Chapter 1 just before Proposition 1.2.5 — 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 $V_n = M_n \cup T_n$ with $|M_n| = m_n$ and $|T_n| = t_n$. Now we define our function as follows:

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

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 $\frac{\log t_n}{t_n} = \frac{1}{m_n^{2/3}}$, 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, $\lim_n \mathsf{clue}_{f_n}(T_n) = 1$.

Remark 2.1.4. Here we point out an interpretation of the random element X of the spectral sample appearing in the proof of Theorem 2.1.1. This setup also has some interesting connections with one of the key lemmas in Chatterjee's book on superconcentration and chaos [C14].

For a function $f : \{-1,1\}^V \longrightarrow \mathbb{R}$ we define the stability of f at level p as

$$\mathbf{Stab}_f(p) := \sum_{S \subseteq V_n} \widehat{f}(S)^2 p^{|S|}.$$

Stability has two interpretations. On the one hand, it measures the noise stability of f: that is, if f is defined on the p-correlated bit sets x and y, then $\mathbf{Stab}_f(p) = \mathbb{E}[f(x)f(y)]$. On the other hand, it is also the closely related to the expected clue of a Bernoulli random set of coordinates $\mathcal{B}^p$ of density p: $\frac{\mathbf{Stab}_f(p)}{\mathrm{Var}(f)} = \mathbb{E}[\mathsf{clue}(f \mid \mathcal{B}^p)]$.

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

$$\frac{\mathbf{Stab}_f(\mathbf{x})}{\mathrm{Var}(f)} = \frac{1}{\mathrm{Var}(f)} \sum_{S \subseteq V} \widehat{f}(S)^2 \prod_{i \in S} x_i$$

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

Denote by $\overline{p}$ the vector with all of its coordinates is equal to p and for a $j \in V$ take the partial derivative of $\mathbf{Stab}_f(\overline{p})$ with respect to the jth coordinate. We obtain that

$$\frac{\partial \mathbf{Stab}_f(\overline{p})}{\partial p_j} = \sum_{S \ni j} \widehat{f}(S)^2 p^{|S|-1}.$$

Now here is the relationship with X, the uniformly random element of the spectral sample:

$$\int_0^1 \frac{\partial \mathbf{Stab}_f(\overline{p})}{\partial p_j} dp = \sum_{S \ni j} \widehat{f}(S)^2 \frac{1}{|S|} = \mathrm{Var}(f)\mathbb{P}[X = j]. \tag{2.1.3}$$

The above quantity can be understood as the average increase in clue over all p values, induced by a small increase in the probability of selecting j into the random set. This interpretation becomes even more explicit in the cooperative game theory framework (see Proposition 2.4.1 below).

Now we get to the connection with Chatterjee's work. Let $f, g : \{-1,1\}^V \longrightarrow \{-1,1\}$ be monotone Boolean functions. We start by expressing $\mathbb{P}[j \in \mathscr{P}_f(\omega) \cap \mathscr{P}_g(\omega^{1-p})]$ in terms of the Fourier-Walsh transform.

Observe that for any monotone $f : \{-1,1\}^V \longrightarrow \{-1,1\}$, we have

$$\nabla_j f(\omega) = f(\omega|\omega_j = 1) - f(\omega|\omega_j = -1) = \sum_{S \ni j} \widehat{f}(S)\chi_{S\setminus j}(\omega).$$

As j is in $\mathscr{P}_f(\omega)$ if and only if $\nabla_j f(\omega) = 2$ and otherwise $\nabla_j f(\omega) = 0$, we get that

$$\mathbb{1}_{i \in \mathscr{P}_f(\omega)} = \frac{1}{2} \sum_{S \ni j} \widehat{f}(S)\chi_{S\setminus j}(\omega).$$

Now recall that

$$\mathbb{E}[\chi_T(\omega)\chi_S(\omega^{1-p})] = \begin{cases} 0 & \text{if } T \neq S, \\ p^{|S|} & \text{if } T = S, \end{cases}$$

and thus, whenever f and g are monotone, we have

$$\mathbb{P}[j \in \mathscr{P}_f(\omega) \cap \mathscr{P}_g(\omega^{1-p})] = \mathbb{E}[\mathbb{1}_{j \in \mathscr{P}_f(\omega)} \mathbb{1}_{j \in \mathscr{P}_g(\omega^{1-p})}] = \frac{1}{4} \sum_{S \ni j} \widehat{f}(S)\widehat{g}(S)p^{|S|-1}. \tag{2.1.4}$$

(We note that this formula is almost a generalization of Lemma 2.7 in [RS18].) Using that $\sum_{j \in V} \mathbb{P}[j \in \mathscr{P}_f(\omega) \cap \mathscr{P}_g(\omega^{1-p})] = \mathbb{E}[|\mathscr{P}_f(\omega) \cap \mathscr{P}_g(\omega^{1-p})|]$, we get from (2.1.4) that

$$\int_0^1 \mathbb{E}[|\mathscr{P}_f(\omega) \cap \mathscr{P}_g(\omega^{1-p})|] dp = \frac{1}{4} \sum_{j \in V} \left(\sum_{S \ni j} \widehat{f}(S)\widehat{g}(S) \frac{1}{|S|} \right) = \frac{1}{4}\mathrm{Cov}(f, g). \tag{2.1.5}$$

This is essentially a special case of Lemma 2.1 from [C14] (referred to as "covariance lemma"), where the Markov process is the random walk on the hypercube. At the same, time setting $g = f$, by (2.1.3) we have

$$\mathbb{P}[X = j] = \frac{1}{\mathrm{Var}(f)} \int_0^1 \frac{\partial \mathbf{Stab}_f(\overline{p})}{\partial p_j} dp = \frac{4}{\mathrm{Var}(f)} \int_0^1 \mathbb{P}[j \in \mathscr{P}_f(\omega) \cap \mathscr{P}_f(\omega^{1-p})] dp,$$

which is a coordinate-wise localized version of the covariance lemma.

One may ask whether a result similar to Theorem 2.1.1 can be derived in case we replace the $\{-1,1\}$ space in the domain with something more complicated or if we replace the product measure with some other measure. A natural idea in this direction is to try to generalize the concept of spectral sample. We might take again Equation (1.1.7) as a starting point.

Observe that the quantity $\mathsf{clue}(f|U)$ is well defined for any $U \subseteq V$ on any product space X^V, no matter what the probability measure is. So one would like to use equation (1.1.7) as the definition for a generalised spectral sample. As the probabilities $\mathbb{P}[\mathscr{S} \subseteq U]$ are known for all U, one can also calculate the probabilities $\mathbb{P}[\mathscr{S} = T]$ for all T. Once we have this generalised spectral sample in hand (depending on the function, the space and the underlying measure) we might be able to repeat the argument in the proof of Theorem 2.1.1.

Unfortunately this strategy fails in general. The problem is that nothing guarantees that the quantities $\mathbb{P}[\mathscr{S} = T]$ that we get from the Möbius inversion are non-negative. Nevertheless, in case the underlying measure is a product measure, the above strategy works as the quantities $\mathbb{P}[\mathscr{S} = T]$ turn out to be non-negative. As we will show, this follows directly from the so-called Efron-Stein decomposition ([OD14], Section 8.2), a generalization of the Fourier-Walsh transform for product measures.

We will need the following simple observation, which turns out to be crucial. In fact, as we shall see Efron-Stein decomposition as well as the possibility of a spectral sample, ultimately depends on Fubini's Theorem.

Lemma 2.1.5. *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 2.1.6 (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 Mü bius 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

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

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 2.1.5 twice gives that

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

We used that $T \cap (L \cup \{i\}) = T \cap L$, since $i \notin L$ and $i \notin T$. This shows that $\mathbb{E}[gf^{=S}] = 0$. 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 2.1.7 (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 Other approaches to measuring "clue"

2.2.1 Significance and Influence of subsets

We would like to make a small detour to discuss some possible alternatives to "clue" as defined in Definition 2.1.1. Given a Boolean function $f : \{-1,1\}^V \longrightarrow \{-1,1\}$ and an underlying probability measure $\mathbb{P}$, we basically want to quantify the amount of information a subset of the coordinates gives us about the function f. We will denote the size of the coordinate set V by n.

We start with a sort of dual to clue.

Definition 2.2.1. The significance of a subset $U \subseteq V$ is

$$\mathsf{sig}(f \mid U) = \frac{\mathbb{E}[\mathrm{Var}(f \mid \mathcal{F}_{U^c})]}{\mathrm{Var}(f)}$$

We call it a dual, because we have $\mathsf{sig}(f \mid U) = 1 - \mathsf{clue}(f \mid U^c)$. It expresses how much additional information we are still missing on average if we know the values of the bits outside of U. We have the following description of $\mathsf{sig}(f \mid U)$ in terms of the spectral sample:

$$\mathsf{sig}(f \mid U) = \mathbb{P}[\mathscr{S}_f \cap U \neq \emptyset].$$

This shows that for product measures $\mathsf{sig}(f \mid U) \geq \mathsf{clue}(f \mid U)$. In general, this inequality does not hold. ($\mathsf{sig}(f \mid U) > \mathsf{clue}(f \mid U)$ whenever $\mathsf{clue}(f \mid U) + \mathsf{clue}(f \mid U^c) > 1$, which can easily happen if the underlying measure has lots of dependencies.) Also Theorem 2.1.1 is not true if we replace clue by sig. For example, any subset $U \subseteq V$ has significance 1 with respect to the parity function χ_V, which is obviously transitive. The famous "It Aint Over Till Its Over" Theorem proved in [MOO05] which can be stated that for sequences of functions with low maximal influence for arbitrary small (but fixed) ϵ the average significance of a Bernoulli random subset of level $1 - \epsilon$ is not vanishing.

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

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

Influence is, however, much weaker then sig (in the sense that it is easier to have high influence than to have high significance). 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, the significance rather quantifies the average gain of information (measured in variance) for a uniformly random configuration of U.

We can again take the dual concept of influence, the combinatorial equivalent of clue, which is the probability that the subset U is a witness. For a Boolean function $f : \{-1, 1\}^V \longrightarrow \{-1, 1\}$ and a configuration $\omega \in \{-1, 1\}^V$ a subset $W \subseteq V$ is a witness for f if ω_W already decides the value of f.

$$W(f \mid U) = \mathbb{P}[f \text{ is determined by the bits on } U]$$

Obviously, $I(f \mid U) \geq W(f \mid U)$ holds irrespective of what the measure is, and also $I(f \mid U) \geq \mathsf{sig}(f \mid U)$, which entails after taking the dual in both sides, that $\mathsf{clue}(f \mid U) \geq W(f \mid U)$.

There are still many questions to be investigated. For the left-right crossing event LR_n for critical planar percolation, when U_n is a sub-square, it is proved in [GPS10] that $I(\mathsf{LR}_n \mid U_n) \asymp \mathsf{sig}(\mathsf{LR}_n \mid U_n)$. For Maj_n on the other hand, this is not the case. As it is easy to check, $I(\mathsf{Maj}_n \mid U) \gg \mathsf{sig}(\mathsf{Maj}_n \mid U)$ for any sequence of subsets with constant density.

Question 2.2.1. *Characterise sequences of Boolean functions such that for any sequence of subsets with constant density $I(f_n \mid U) \gg \mathsf{sig}(f_n \mid U)$ holds, or where $I(f_n \mid U_n) \asymp \mathsf{sig}(f_n \mid U_n)$, respectively.*

2.2.2 Clue via entropy

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 play 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 $f : \{-1,1\}^V \longrightarrow \{-1,1\}$), but all the argument we present here work in this slightly more general framework.

For sake of completeness we start with some classical definitions.

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

Definition 2.2.2 (Entropy)**.** Let X be a discrete 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 2.2.3 (Conditional Entropy)**.** Let X and Y be two discrete random variables defined on the same probability space. 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 2.2.4 (Mutual Information)**.** Let X and Y be two discrete random variables defined on the same probability space. 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. Although are main focus is still Boolean functions on the hypercube, we define the notion more generally

Definition 2.2.5 (I-Clue)**.** Let $X_v : \quad v \in V$ be a finite family of discrete real valued random variables defined on the same probability space and for some $f : \mathbb{R}^V \longrightarrow \mathbb{R}$ lets consider the random variable $Z = f(X_V)$. The information theoretic clue (I-clue) of f with respect to $U \subseteq V$ 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.

As for the cases discussed before, here too we can introduce the dual (which expresses again how much information we are missing if we don't know the coordinates in U.)

$$\mathsf{sig}^I(f \mid U) = 1 - \frac{I(Z : \omega_{U^c})}{H(Z)} = \frac{H(Z \mid \omega_{U^c})}{H(Z)}.$$

I-clue versus L^2-clue

In the sequel, besides the L^2 version of clue, we shall also use the entropy based I-clue. Hence the following result is of some importance.

Proposition 2.2.2. *Let μ be a measure on $\{-1,1\}^n$ and $\sigma = (\sigma_1, \dots \sigma_n)$ a spin system distributed according to μ.*

Let $f : \{-1,1\}^n \longrightarrow \mathbb{R}$ with $|f(x)| \leq K$, and let $Z = f(\sigma)$. For any $U \subseteq [n]$

$$\frac{\mathrm{Var}(\mathbb{E}[Z \,|\, \mathcal{F}_U])}{\mathrm{Var}(Z)} \leq \frac{K^2 \min((\delta_f/2)^2, 1)}{p_{\min}} \frac{I(Z, \sigma_U)}{H(Z)}, \tag{2.2.2}$$

where $\delta_f = \min\{x - y : \ x, y \in \mathsf{Ran}(f), \ x \neq y\}$, and $p_{\min} := \min\{\mathbb{P}(Z = x) : \ x \in \mathsf{Ran}(f), \ \mathbb{P}(Z = x) > 0\}$.

We note that this estimate is very poor in many cases, in particular, for a sequence of functions with the cardinality of the range going to infinity. Nevertheless, we do not see a clear way to improve it in general. For our main focus, Boolean functions, however, the inequality is sharp.

Proof. First we show that

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

Our argument follows Lemma 4.4 in [Tao05]. First we fix some notations. Let $z \in \mathsf{Ran}(f)$ and $u \in \{-1,1\}^U$. Then

$$p_z := \mathbb{P}[Z = z], \quad p_u := \mathbb{P}[\sigma_U = u], \quad p_{z|u} := \mathbb{P}[Z = z \,|\, \sigma_U = u].$$

Now, with this notation we have

$$\mathrm{Var}(\mathbb{E}[Z \,|\, \mathcal{F}_U]) = \sum_{u \in \{-1,1\}^U} p_u (\mathbb{E}[Z] - \mathbb{E}[Z \,|\, \sigma_U = u])^2,$$

and for a fixed $u \in \{-1,1\}^U$

$$(\mathbb{E}[Z] - \mathbb{E}[Z \,|\, \sigma_U = u])^2 = \sum_{z \in \mathsf{Ran}(f)} (p_z z - p_{z|u} z)^2 =$$
$$\sum_{z \in \mathsf{Ran}(f)} z^2 (p_z - p_{z|u})^2 \leq K^2 \sum_{z \in \mathsf{Ran}(f)} (p_z - p_{z|u})^2.$$

So we get that

$$\mathrm{Var}(\mathbb{E}[Z \,|\, \mathcal{F}_U]) \leq K^2 \sum_{u \in \{-1,1\}^U} \sum_z (p_z - p_{z|u})^2. \tag{2.2.3}$$

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

$$I(Z,\sigma_U) = H(Z) - H(Z|\sigma_U) = \sum_z \left(h(p_z) - \sum_{u\in\{-1,1\}^U} p_u h(p_{z|u}) \right) \tag{2.2.4}$$

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

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

with some $p^*_{z|u}$ between $p_{z|u}$ and p_z, using for the error term that $h''(x) = -\frac{1}{x}$. Substituting this estimate into (2.2.4), we observe that the terms with $h'(p_z)$ cancel, since for any $z \in \mathsf{Ran}(f)$ we have $\sum_{u\in\{-1,1\}^U} p_u(p_{z|u} - p_z) = p_z - p_z = 0$. Therefore we obtain that

$$\sum_{u\in\{-1,1\}^U} \sum_z \frac{(p_z - p_{z|u})^2}{p^*_{z|u}} = 2I(Z,\sigma_U).$$

As $0 < p^*_{z|u} < 1$ we can conclude, using (2.2.3) that

$$\mathrm{Var}(\mathbb{E}[Z \mid \mathcal{F}_U]) \le K^2 \sum_{u\in\{-1,1\}^U} \sum_z \frac{(p_z - p_{z|u})^2}{p^*_{z|u}} \le 2K^2 I(Z,\sigma_U).$$

Now we turn to the denominator. We have to show that the entropy can be bounded by the variance, that is $H(Z) \le C\mathrm{Var}(Z)$, with $C = (2p_{\min}\min((\delta_f/2)^2, 1))^{-1}$.

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:

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

where log denotes base 2 logarithm.

Recall that $p_{\min} = \min(\mathbb{P}[f = 1], \mathbb{P}[f = -1])$. Then $|\mathbb{E}[Z]| = 1 - 2p_{\min}$. At the same time, a simple calculation shows that if $|\mathbb{E}[Z] \le 1 - c$, that is, $c \le 2p_{\min}$ then

$$H(Z) \le 1 - \frac{1}{\ln 4}\mathbb{E}[Z]^2 \le \frac{1}{c}(1 - \mathbb{E}[Z]^2) \le \frac{1}{2p_{\min}}\mathrm{Var}(Z).$$

In case f is still binary valued and $\max(f) - \min(f) = \delta_f$ we can get a a ± 1-valued function by an affine transformation. So we have :

$$H(Z) \le \frac{1}{2p_{\min}\min((\delta_f/2)^2, 1)}\mathrm{Var}(Z). \tag{2.2.5}$$

We continue by induction on $|\mathsf{Ran}(f)|$, the cardinality of the range of f. We have just proved the claim when $|\mathsf{Ran}(f)| \le 2$.

Now let $|\mathsf{Ran}(f)| > 2$ and for some $x \in \mathbb{R}$ we define the event $A := \{Z > x\}$. We choose an x in such a way that both $\mathbb{P}[A]$ and $\mathbb{P}[A^c]$ are positive.

According to 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. At the same time we decompose the entropy according to A as well:

$$H(Z) = H(Z, \mathbb{1}_A) = H(\mathbb{1}_A) + H(Z \mid \mathbb{1}_A),$$

using that A is Z-measurable.

Obviously, $p_{\min} \leq \min(\mathbb{P}[A], \mathbb{P}[A^c])$ and it is also clear that $|\mathbb{E}[Z \mid A] - \mathbb{E}[Z \mid A^c]| \geq \delta_f$, hence using (2.2.5) for $\mathbb{E}[Z \mid \mathcal{F}_A]$ (which is indeed binary valued), we get that

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

Conditioned on either A or A^c the range of f is smaller then $\mathsf{Ran}(f)$, so by the induction hypothesis we have

$$H(Z \mid A) \leq \frac{1}{2p_{\min} \min((\delta_f/2)^2, 1)} \mathrm{Var}(Z \mid A),$$

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

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

□

In particular, this result shows that for a non-degenerated sequence of Boolean functions, no sparse reconstruction with respect to clue^I implies no sparse reconstruction with respect to clue.

2.2.3 Clue via distances between probability measures

In this section we introduce a somewhat different approach to measure the information content of a subset of coordinates about a Boolean function. We can interpret a Boolean function as the density of a probability measure, the uniform measure conditioned on the set of configurations $\{\omega \in \{-1,1\}^V : f(\omega) = 1\}$. If we think about a Boolean function as a function mapping to $\{0,1\}$ instead of $\{-1,1\}$, it is indeed treating f like a (probability) density function.

More generally, every $f : \{-1,1\}^n \longrightarrow \mathbb{R}_{\geq 0}$ with $\mathbb{E}[f] > 0$ can be interpreted as a density, and can be used to define another probability measure on the same space by

$$\mathbb{Q}[\omega] := \frac{1}{\mathbb{E}[f]} f(\omega) \mathbb{P}[\omega], \quad \omega \in \{-1,1\}^n. \tag{2.2.6}$$

Now that we turned our function to a measure we can think about the information content in terms of distance of probability measures. For example, total variation distance between two measures on a discrete space is given by

$$\delta(\mathbb{P}, \mathbb{Q}) := \frac{1}{2} \sum_{\omega \in \Omega} |\mathbb{P}[\omega] - \mathbb{Q}[\omega]|. \tag{2.2.7}$$

In this framework, the total information content of the function is measured by its distance from the original measure $\mathbb{P}$. How can we now quantify the information content of a subset of coordinates? We can simply take the marginals (projections) of the respective measures $\mathbb{P}$ and $\mathbb{Q}$ with respect to a subset of coordinates U and take their distance.

$$\mathsf{clue}^{TV}(f \mid U) = \frac{\delta(\mathbb{P}|_U, \mathbb{Q}|_U)}{\delta(\mathbb{P}, \mathbb{Q})}.$$

We can express this in a somewhat more familiar form, using L^1 distance. Let f be the density of $\mathbb{Q}$, then with $Z := f(\omega)$ we have $\delta(\mathbb{P}, \mathbb{Q}) = \frac{1}{2\mathbb{E}[Z]}\mathbb{E}[|Z - \mathbb{E}[Z]|]$, as it can be easily calculated if we substitute (2.2.6) into (2.2.7). In a similar way we get that

$$\delta(\mathbb{P}|_U, \mathbb{Q}|_U) = \frac{1}{2\mathbb{E}[Z]} \sum_{\omega \in \{-1,1\}^U} \mathbb{P}[\omega] \left|\mathbb{E}[Z|\mathcal{F}_U](\omega) - \mathbb{E}[Z]\right| = \frac{1}{2\mathbb{E}[Z]}\mathbb{E}[\mathbb{E}[Z|\mathcal{F}_U](\omega) - \mathbb{E}[Z]],$$

and, as a consequence,

$$\mathsf{clue}^{TV}(f \mid U) = \frac{\mathbb{E}[|\mathbb{E}[Z|\mathcal{F}_U] - \mathbb{E}[Z]|]}{\mathbb{E}[|Z - \mathbb{E}[Z]|]}, \tag{2.2.8}$$

which is clearly just an L^1 version of the original definition (Definition 2.1.1). Without going into details we mention that again one can define a dual notion by $\mathsf{sig}^{TV}(f \mid U) = 1 - \mathsf{clue}^{TV}(f \mid U^c)$.

Let us compare clue and clue^{TV} for Boolean functions. More precisely let $f : \{-1,1\}^n \longrightarrow \{-1,1\}$ and let $Z' = (Z+1)/2$ the corresponding $\{0,1\}$-valued random variable. A simple calculation shows that $\mathrm{Var}(Z) = 2\mathbb{E}[|Z' - \mathbb{E}[Z']|]$. At the same time, using Cauchy-Schwarz inequality:

$$\mathbb{E}[|\mathbb{E}[Z'|\mathcal{F}_U] - \mathbb{E}[Z']|] = \sum_{\omega \in \{-1,1\}^U} \sqrt{\mathbb{P}[\omega]}\sqrt{\mathbb{P}[\omega]} \left|\mathbb{E}[Z'|\mathcal{F}_U](\omega) - \mathbb{E}[Z]\right|$$

$$\leq \sum_{\omega \in \{-1,1\}^U} \mathbb{P}[\omega] \left(\mathbb{E}[Z'|\mathcal{F}_U](\omega) - \mathbb{E}[Z]\right)^2 = \mathrm{Var}(\mathbb{E}[Z'|\mathcal{F}_U]) = \frac{1}{4}\mathrm{Var}(\mathbb{E}[Z|\mathcal{F}_U]),$$

and consequently

$$\mathsf{clue}^{TV}(f' \mid U) \leq \frac{1}{8}\mathsf{clue}(f \mid U).$$

We emphasize that this is true irrespective of the underlying measure. In particular, this means that Theorem 2.1.1 remains true if we replace clue by clue^{TV}.

There is another way to measure distance between probability measures which relies on concepts from information theory. The Kullback-Liebler (KL) divergence or relative entropy is defined as follows.

Definition 2.2.6 (Relative entropy)**.** Let $\mathbb{Q}$ and $\mathbb{P}$ measures on the same discrete probability space Ω, where $\mathbb{Q} \ll \mathbb{P}$. the relative entropy between $\mathbb{Q}$ and $\mathbb{P}$ is

$$D(\mathbb{Q}||\mathbb{P}) = -\sum_{x \in \Omega} \mathbb{Q}(x) \log \frac{\mathbb{Q}(x)}{\mathbb{P}(x)}.$$

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

Now we can define yet another concept of clue according to the same logic as for the total variation distance. Let $f : \{-1,1\}^V \longrightarrow \mathbb{R}_{\geq 0}$ and let $\mathbb{Q}$ denote the probability measure on $\{-1,1\}^V$ with density f. If $D(\mathbb{Q}||\mathbb{P})$ expresses the total 'information distance' between $\mathbb{Q}$ and $\mathbb{P}$, we can interpret the quantity $D(\mathbb{Q}_U||\mathbb{P}_U)$ as the 'information distance' restricted to the respective subset of coordinates. Hence we define the KL-divergence clue.

$$\mathsf{clue}^{KL}(f \mid U) := \frac{D(\mathbb{Q}_U||\mathbb{P}_U)}{D(\mathbb{Q}||\mathbb{P})}.$$

One can easily derive that

$$D(\mathbb{Q}||\mathbb{P}) = \mathsf{Ent}(Z),$$

where $Z = f(\omega)$ as usual and $\mathsf{Ent}(Z) := \mathbb{E}[Z \log Z] - \mathbb{E}[Z] \log \mathbb{E}[Z]$ (the expectation is taken with respect to $\mathbb{P}$ and in case $f(\omega) = 0$ we have, by continuity, $f(\omega) \log f(\omega) = 0$). Similarly, we have $D(\mathbb{Q}||\mathbb{P}) = \mathsf{Ent}(\mathbb{E}[Z|U])$, and thus we obtain the following formula reminiscent to (2.2.8)

$$\mathsf{clue}^{KL}(f \mid U) = \frac{\mathsf{Ent}(\mathbb{E}[Z|U])}{\mathsf{Ent}(Z)}.$$

We mention that $\mathsf{Ent}(Z)$ and $\mathrm{Var}(Z)$ together with the respective concepts of clue can be examined in the general framework of Φ-entropies (see for example [BLM13], Chapter 14 and 15). The main idea is that for a convex function $\mathbb{R}_{\geq 0} \to \mathbb{R}$ one can assign a respective Φ-entropy for every integrable random variable X by

$$H_\Phi(X) = \mathbb{E}[\Phi(X)] - \Phi(\mathbb{E}[X]).$$

It turns out that under some general analytic conditions on Φ many important properties we require from an information measure remains valid for $H_\Phi(X)$ (for example, it is always non-negative because of Jensen's inequality). In particular, we get $H_\Phi(X) = \mathrm{Var}(X)$ when $\Phi(x) = x^2$ and $H_\Phi(X) = \mathsf{Ent}(X)$ with $\Phi(x) = x \log x$.

2.3 Sparse Reconstruction with respect to I-clue and KL-clue

In this section we show some equivalents of Theorem 2.1.7 for the I-clue and KL-clue. We note that the following theorem, 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). The notation and setup follows Section 2.2.2.

Theorem 2.3.1. *Let $\{X_v : v \in V\}$ be discrete valued, i.i.d, random variables with finite entropy. Let $f : \Omega^n \longrightarrow \mathbb{R}$ be a transitive function and $Z = f(\{X_v : v \in V\})$. Then*

$$\mathsf{clue}^I(f \mid U) \leq \frac{|U|}{n}. \tag{2.3.1}$$

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

Theorem 2.3.2 (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*

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

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

Lemma 2.3.3. *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*

$$\sum_{j}^{L} I(Z : X_{S_j}) \leq kI(Z : X_{[n]}) \tag{2.3.2}$$

Proof. Without loss of generality we can assume that each i appears in exactly k sets. Indeed, if this is not the case, we can always add some additional subsets so that this condition is satisfied. While adding new sets the right hand side of the inequality does not change and the left hand side can only increase.

Since the variables X_i are independent:

$$\sum_{j}^{L} H(X_{S_j}) = \sum_{j} \sum_{i \in S_j} H(X_i) = k \sum_{i \in [n]} H(X_i) = kH(X_{[n]}) \tag{2.3.3}$$

On the other hand, using Shearer's inequality

$$-\sum_{j}^{L} H(X_{S_j}|Z) \leq -kH(X_{[n]}|Z) \tag{2.3.4}$$

Using that $I(Z : X_{S_j}) = H(X_{S_j}) - H(X_{S_j}|Z)$ and adding up (2.3.3) and (2.3.4) completes the proof. □

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 2.3.3 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. □

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 their variance is $\Omega(1)$. In particular, for a non-degenerated sequence of Boolean functions no sparse reconstruction with respect to I-clue implies no sparse reconstruction with respect to the original, L^2 version. This follows from Proposition 2.2.2. 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.

Interestingly enough, along the same logic one can prove the respective version of Theorem 2.1.7 and Theorem 2.3.1 for clue^{KL}. We should emphasize that, in contrast with mutual information, relative entropy is a concept that remains meaningful for continuous random variables as well. So Theorem 2.3.5 holds for all product measures like 2.1.7. The following Shearer-type inequality holds:

Lemma 2.3.4. *Let $\mathbb{P}$ be a product measure on the hypercube (in fact, any product measure will do) and and μ another probability measure on the same space satisfying $\mu \ll \mathbb{P}$.*

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

$$\sum_{j}^{L} D(\mu_{S_i} || \mathbb{P}_{S_i} \leq k D(\mu || \mathbb{P}).$$

In our application, of course μ is the measure with density f. It is easy to recognise that Lemma 2.3.4 is a close relative of Lemma 2.3.3. The proof of this Lemma is also a straightforward consequence of Shearer's inequality (Theorem 2.3.2), for a proof see [GLSS12]. The corresponding clue theorem follows in the same way as Lemma 2.3.3 implies Theorem 2.3.1.

Theorem 2.3.5. *Let $\{X_v \,:\, v \in V\}$ be Ω- valued, i.i.d, random variables . Let $f : \Omega^n \longrightarrow \mathbb{R}$ be a transitive function and $Z = f(\{X_v \,:\, v \in V\})$. Then*

$$\mathsf{clue}_f^{KL}(U) \leq \frac{|U|}{n} \tag{2.3.5}$$

It is worth noting that for sequences of transitive Boolean functions on the hypercube there is no sparse reconstruction no matter which version of clue we wish to choose. Indeed clue^{TV} and I (influence) are dominated by clue so Theorem 2.1.1 applies, while for clue^{I} and clue^{KL} it has been shown in the present section (Theorem 2.3.1 and Theorem 2.3.5).

2.4 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 2.4.1 (Shapley value)**.**

$$\phi_i(v) = \frac{1}{|V|} \sum_{S \subseteq V \setminus \{i\}} \frac{v(S \cup \{i\}) - v(S)}{\binom{|V|-1}{|S|}} \tag{2.4.1}$$

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 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 2.4.1. *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 are $\binom{n-1-|T|}{k-|T|}$ k-element subsets S which contain 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

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

Now we apply the so called Hockey-stick identity — $\sum_{k=|T|}^{n-1} \binom{k}{|T|} = \binom{n}{|T|+1}$ — and we get the desired formula. □

Given how naturally the Shapley value arises 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 game theory and Combinatorics. The advantage of this approach is that it makes it more clear the conditions under which a small clue theorem can be true. It should also be 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 2.4.2 (Core)**.** The core of a cooperative game v with set of playeres V is the set $C(v) \subseteq \mathbb{R}^{|V|}$ in such a way that $x \in C(v)$ if and only if

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

and for every $S \subset V$

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

We have the following simple observation.

Lemma 2.4.2. *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 Lemma 2.4.2.

Definition 2.4.3 (Convex games)**.** A cooperative game v is convex if the characteristic function is supermodular. That is, for every subset of players $S, T \subseteq [n]$

$$v(S) + v(T) \leq v(S \cup T) + v(S \cap T). \tag{2.4.2}$$

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 2.2.5):

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

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

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

For the L^2 version, the supermodularity of $\mathrm{Var}(\mathbb{E}[f\,|\,\mathcal{F}_U])$ follows easily from the spectral description. Here we present an argument that does not require Fourier-Walsh expansion or Efron-Stein decomposition.

Proposition 2.4.3. *Let $f : X^V \longrightarrow \mathbb{R}$, where X^V is endowed with a product measure. The set function (cooperative game) $v_f(S) = \mathrm{Var}(\mathbb{E}[f\,|\,\mathcal{F}_S])$ for ($S \subseteq V$) is supermodular (convex).*

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

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

and therefore (2.4.2) can be rewritten as

$$\mathrm{Var}(\mathbb{E}[f\,|\,\mathcal{F}_T] - \mathbb{E}[f\,|\,\mathcal{F}_{S\cap T}]) \leq \mathrm{Var}(\mathbb{E}[f\,|\,\mathcal{F}_{S\cup T}] - \mathbb{E}[f\,|\,\mathcal{F}_S]). \tag{2.4.3}$$

Fix $S, T \subseteq V$ such that $S \subseteq T$. Using Lemma 2.1.5 for $(T \setminus S)^c$ and T, we get

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

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 $\mathbb{E}[f\,|\,\mathcal{F}_{S\cap T}] = \mathbb{E}[\mathbb{E}[f\,|\,\mathcal{F}_{(T\setminus(S\cap T))^c}]\,|\,\mathcal{F}_T]$ and $\mathbb{E}[f\,|\,\mathcal{F}_S] = \mathbb{E}[\mathbb{E}[f\,|\,\mathcal{F}_{((S\cup T)\setminus S)^c}]\,|\,\mathcal{F}_{S\cup T}]$. Since $T \setminus (S \cap T) = (S \cup T) \setminus S = T \setminus S$, (2.4.3) becomes

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

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 2.4.4. *If v is a 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

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

□

Now Lemma 2.4.4 implies that if v is a convex game then for any $S \subset V$

$$v(S) = \sum_{i\in S}\phi_i(v_S) \leq \sum_{i\in U}\phi_i(v), \tag{2.4.4}$$

and therefore $\phi(v)$ is indeed in the core. By Proposition 2.4.3 for any real function f the game v_f is convex (if the underlying measure is a product measure), so using (2.4.4) we see that Lemma 2.4.2 applies and we obtain Theorem 2.1.7 (and Theorem 2.3.1, if we replace v_f with v_f^I in the argument above).

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 2.4.2 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, however, that on a practical level, the conditions of this theorem are not very easy to verify.

Theorem 2.4.5 (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

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

2.5 Sparse Reconstruction for Planar Percolation

2.5.1 A Brief Introduction to Percolation Theory

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, which 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 2.5.1 (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 can be shown 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 [Sm01]. Up until today this is the only lattice where the value of this (and many other) exponents are 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 2.5.1 ([LSW02])**.** *For the $\mathbb{Z}^2$ lattice*

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

Our result

In the sequel 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.

First observe that we can use the framework of Boolean functions, since an edge percolation configuration can be described as an element in $\{-1, 1\}^E$, where E is the edge set of the graph on which we percolate.So $\mathsf{LR}_n : \{-1, 1\}^{E(R)} \longrightarrow \{-1, 1\}$ denotes the indicator function of the planar crossing event, where R is the $n \times n$ square. More precisely, $\mathsf{LR}(\omega)$ is 1 if there is a connected path of open edges (1s) connecting some vertex on the left side with some vertex on the right side of the $n \times n$ square. in ω and -1 otherwise.

In the sequel we will use a slight modification of this setup. We are going to the embed the $n \times n$ square into the torus $\mathbb{Z}_n^2$) and think about the crossing event as a Boolean function $\mathsf{LR}_n : \{-1,1\}^{E(\mathbb{Z}_n^2)} \longrightarrow \{-1,1\}$. For LR_n we simply ignore the extra edges of the torus, that is if $e \notin E(R)$ then the value of ω_e does not influence LR_n.

The reason for this embedding is that we shall use the symmetries of the torus, in order to make us of the results of the previous section. In particular, now we can translate LR_n with some element of $\mathbb{Z}_n^2$ and still get a function defined on $\{-1,1\}^{E(\mathbb{Z}_n^2)}$. One may object that Theorem 2.1.1 does not apply directly since the left-right crossing is not a transitive event. Nevertheless, we are going to argue that the left-right crossing event is not too far from being transitive.

Here is a brief summary of what we are going to do: let us denote by LR_n the characteristic function of the left-right crossing . We will show that for every ϵ there is a corresponding sublattice $H_\epsilon \subseteq \mathbb{Z}_n^2$ the size of which only depends on ϵ with the following property: $M^{H_\epsilon}[\mathsf{LR}_n]$, the average of the LR_n translates on the H_ϵ lattice, is close to a transitive (that is, $\mathbb{Z}_n^2$-invariant) function $M[\mathsf{LR}_n]$ in the sense that $\mathrm{Corr}(M^{H_\epsilon}[\mathsf{LR}_n], M[\mathsf{LR}_n]) \geq 1 - O(\sqrt{\epsilon})$. This will be shown in Lemma 2.5.5. The function $M[\mathsf{LR}_n]$ is, in fact, the projection of LR_n onto the space of $\mathbb{Z}_n^2$-invariant functions. As we shall see, Lemma 2.5.3 tells us that in case two functions are highly correlated their clues with respect to any 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 function $M^{H_\epsilon}[\mathsf{LR}_n]$ 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 the size of H_ϵ does not grow with n. But this is impossible because then in turn the transitive function $M[\mathsf{LR}_n]$, 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 in 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 this 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 $\delta \sim n^{-(\frac{1}{4}+o(1))}$ ([SS10]). This, in particular, implies that any sequence of sets S_n of size $n^{\frac{1}{4}-\epsilon}$ is asymptotically clueless, since denoting the random set of queried bits by J_n, we have:

$$\mathbb{P}[S_n \cap J_n \neq \emptyset] \leq \mathbb{E}|S_n \cap J_n| = \sum_{i \in S_n} \mathbb{P}[i \in J_n] \leq \sum_{i \in S_n} \delta = |S_n| n^{-\frac{1}{4}+o(1)} \to 0$$

whenever $|S_n| \asymp n^{\frac{1}{4}-\epsilon}$. That is, with high probability, S_n is disjoint from J_n and therefore conditioned on this event the bits in S_n do not influence the value of the function. From this it is not difficult to conclude that S_n cannot have a large clue.

2.5.2 Projections and Clue

In this section we present some general results that are necessary to prove that there is no Sparse Reconstruction for the crossing event. We are going to use the upcoming two simple lemmas to estimate how much a projection can distort correlations. The geometric 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 projection will roughly preserve the correlation. Note that these results are completely general, i.e., we do not make use of the fact that the underlying measure is a product measure.

The space of functions over a given configuration space $\{-1,1\}^V$ and a corresponding probability measure (the uniform measure in this case) can be endowed with a Hilbert space structure via the inner product $\langle f, g\rangle := \mathbb{E}[fg]$. In order to state the following Lemmas in full generality, we introduce a generalization of clue for closed linear subspaces, making use of the Hilbert space structure. It is the logical extension of clue as it was defined previously for σ-algebras (See Definition 2.1.1).

Let μ be a probability measure and let $\mathcal{H}$ be a closed subspace of $L^2(S,\mu)$. (In our applications we always have $S = \{-1,1\}^V$ for some finite set V.) Denote by $P_{\mathcal{H}}$ the orthogonal projection onto this subspace. For any $f \in L^2(S,\mu)$ we define the clue of f with respect to the subspace $\mathcal{H}$ as

$$\mathsf{clue}(f \mid \mathcal{H}) = \frac{\mathrm{Var}(P_{\mathcal{H}}[f])}{\mathrm{Var}(f)}.$$

Lemma 2.5.2. *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

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

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 $\mathrm{Var}(f) = \mathrm{Var}(g) = 1$, we get

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

In a similar fashion, we get for the respective projections that

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

Using first that $\mathrm{Var}(f) = \mathrm{Var}(g) = 1$ and after our assumption, we get

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

On the other hand

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

so we conclude that

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

Finally, putting together estimates for $\|f - g\|^2$ and $\|P[f] - P[g]\|^2$ and using that P being a a projection cannot only increase the L^2 norm 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 this inequality the statement follows. □

Lemma 2.5.3. *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,

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

and

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

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

$$\|g - P[g]\|^2 \leq \|g - P[f]\|^2 \leq \|g - f\|^2 + \|f - P[f]\|^2. \tag{2.5.1}$$

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

$$\|P[f]\|^2 + \|f - P[f]\|^2 = \|f\|^2 \tag{2.5.2}$$

As in Lemma 2.5.2, $\mathrm{Corr}(f, g) \geq 1 - \epsilon$ implies $\|g - f\|^2 \leq 2\epsilon$.

On the other hand, by our assumptions,

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

Thus (2.5.2) shows that $\|f - P[f]\|^2 \leq 1 - c$. Plugging the estimates into (2.5.1) 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 (2.5.2) 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 2.5.2 to get that $\mathrm{Corr}(P[f], |P[g]) \geq 1 - \frac{\epsilon}{c-2\epsilon}$ □

2.5.3 No Sparse Reconstruction for critical planar percolation

Let $0 < \delta < 1$. For a $\mathbf{t} \in \mathbb{Z}_n^2$ we will denote the rectangle $\mathbf{t} + [-\lfloor \delta n \rfloor, \lfloor \delta n \rfloor]^2 \subset \mathbb{Z}_n^2$ by $R_\delta(\mathbf{t})$. It is straightforward to see that $4(\delta n - 1)^2 \leq |R_\delta(t)| \leq 4(\delta n)^2$.

Lemma 2.5.4. *Let $R_\delta := R_\delta(\mathbf{0}) = [-\lfloor \delta n \rfloor, \lfloor \delta n \rfloor]^2$ as above. Then there is a $K > 0$ such that for every $\mathbf{d}_1, \mathbf{d}_2 \in R_\delta$*

$$\mathrm{Corr}(\mathsf{LR}_n^{\mathbf{d}_1}, \mathsf{LR}_n^{\mathbf{d}_2}) \geq 1 - K\delta$$

Proof. Let $\mathbf{d} \in R_\delta$. We are going to show that

$$\mathbb{P}[\mathsf{LR}_n \neq \mathsf{LR}_n^{\mathbf{d}}] \leq O(\delta).$$

From this the statement of the lemma follows. Indeed, for any $\mathbf{d}_1, \mathbf{d}_2 \in R_\delta$

$$\mathrm{Corr}(\mathsf{LR}_n^{\mathbf{d}_1}, \mathsf{LR}_n^{\mathbf{d}_2}) = 1 - 2\mathbb{P}[\mathsf{LR}_n^{\mathbf{d}_1} \neq \mathsf{LR}_n^{\mathbf{d}_2}] = 1 - 2\mathbb{P}[\mathsf{LR}_n \neq \mathsf{LR}_n^{\mathbf{d}_2 - \mathbf{d}_1}] \geq 1 - O(\delta).$$

Let us assume that $\mathbf{d} = (0, t)$. Observe (see Figure 2.5.3) that the event $\{\mathsf{LR}_n \neq \mathsf{LR}_n^{\mathbf{d}}\}$ entails a 3-arm event in a half plane from radius $O(\delta)$ to distance $O(1)$.

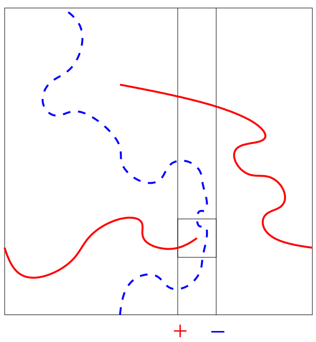

Figure 2.5.1: If $\mathsf{LR} \neq \mathsf{LR}_n^{\mathbf{d}}$, then we have a 3-arm event in a half plane.

As the grid size approaches to zero, the probability of the 3-arm event is in quadratic order by Proposition 2.5.1:

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

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

$$\mathbb{P}[\mathsf{LR}_n \neq \mathsf{LR}_n^{\mathbf{d}}] \leq \alpha_3^+(\delta, 1) O(\frac{1}{\delta}) = O(\delta).$$

In case $\mathbf{d} = (t, 0)$ 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 $\mathbf{d} \in R_\delta$ now easily follows. If $\left\{\mathsf{LR}_n \neq \mathsf{LR}_n^{\mathbf{d}}\right\}$, then either $\left\{\mathsf{LR}_n \neq \mathsf{LR}_n^{\mathbf{d}_x}\right\}$ or $\left\{\mathsf{LR}_n \neq \mathsf{LR}_n^{\mathbf{d}_y}\right\}$, where $\mathbf{d}_x$ and $\mathbf{d}_y$ are the projections of $\mathbf{d}$ onto the first and the second coordinates, respectively.

As a consequence, $\mathbb{P}[\mathsf{LR}_n \neq \mathsf{LR}_n^{\mathbf{d}}] \leq \mathbb{P}[\mathsf{LR}_n \neq \mathsf{LR}_n^{\mathbf{d}_x}] + \mathbb{P}[\mathsf{LR}_n \neq \mathsf{LR}_n^{\mathbf{d}_y}] \leq O(\delta)$. □

Our proof uses the idea that the percolation crossing event is almost transitive, so we define a linear operator that maps any function to a transitive one. Let V a finite set and let Γ be a group acting on V. Recall that Γ also acts on the configuration space $\{-1,1\}^V$ by $\omega_v^\gamma := \omega_{v^{-\gamma}}$ and, in turn, this extends to an action of Γ on the functions $f : \{-1,1\}^V \longrightarrow \mathbb{R}$. For any $\gamma \in \Gamma$ we have $f^\gamma(\omega) := f(\omega^{-\gamma})$.

Let us assume that Γ acts transitively on V. We define a natural averaging operator that turns an arbitrary function $f : \{-1,1\}^V \longrightarrow \mathbb{R}$ into a Γ-invariant and thus transitive function (Definition 1.2.3) on the same space:

Definition 2.5.2 (Magnetization)**.** Let $f : \{-1,1\}^V \longrightarrow \mathbb{R}$ and Γ a group acting transitively on V. Then the magnetization of f is

$$M[f] := \frac{1}{|\Gamma|} \sum_{\gamma \in \Gamma} f^\gamma. \tag{2.5.3}$$

In fact, as we shall see in Section 3.2.3, magnetization is the orthogonal projection onto the space of transitive functions. We will also use the following notation: for a $H \subset \Gamma$ and $f : \{-1,1\}^V \longrightarrow \mathbb{R}$ let

$$M^H[f] := \frac{1}{|H|} \sum_{\gamma \in H} f^\gamma,$$

that is, the average of H-translates of f.

For a number $\delta > 0$ we shall also consider a coarser lattice of mesh size δ. More precisely, we define the sublattice $H_\delta := \{\lfloor n\delta \rfloor, 2\lfloor n\delta \rfloor, \ldots, L\lfloor n\delta \rfloor\}^2$, where L is the largest integer such that $L\lfloor n\delta \rfloor < n$ (that is, $L = \lfloor \frac{n}{\lfloor n\delta \rfloor} \rfloor$). Obviously, $\frac{n}{n\delta} - 1 \leq \frac{n}{\lfloor n\delta \rfloor} - 1 \leq L \leq \frac{n}{\lfloor n\delta \rfloor} \leq \frac{n}{n\delta - 1}$, so $1/\delta - 1 \leq L \leq 1/\delta + O(1/n)$ and therefore

$$\left(\frac{1}{\delta} - 1\right)^2 \leq |H_\delta| \leq \frac{1}{\delta^2} + O\left(\frac{1}{n}\right).$$

In the following lemma we compare the function $M^{\mathbb{Z}_n^2}[\mathsf{LR}]$, which is obviously transitive, as $\mathbb{Z}_n^2$ acts transitively on the vertices of the torus, and $M^{H_\delta}[\mathsf{LR}]$, the average of the translates of the crossing event over the δn-lattice.

Lemma 2.5.5. *Let $\delta > 0$. Then*

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

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. Namely, at every vertex $\mathbf{v} \in \mathbb{Z}_n^2$, we set $\sigma_{\mathbf{v}} = \mathsf{LR}_n^{\mathbf{v}}$.

The outline of the proof is as follows: First we observe that for any $\delta n \times \delta n$ square on the δ-lattice 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-lattice *good* if this is the case, *bad* otherwise.

The second step is to show that the event that there exists 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 δn-lattice already gives a good approximation about the average on the entire torus $\mathbb{Z}_n^2$.

Define the event $A := \left\{\sigma_{\mathbf{0}} = \sigma_{(0,n\delta)} = \sigma_{(n\delta,0)} = \sigma_{(n\delta,n\delta)}\right\}$. By Lemma 2.5.4 and the union bound,

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

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, that is, we have:

$$\mathbb{E}[|\text{bad squares}|] \leq O(\delta)|H_\delta| \leq O(\delta)\, O\left(\frac{1}{\delta^2}\right) = O\left(\frac{1}{\delta}\right). \tag{2.5.4}$$

Let B denote the event that for every $\mathbf{t} \in \mathbb{Z}_n^2 \cap [0, \delta n]^2$ the values σ_t are the same. We are going to show that $\mathbb{P}[B^c \cap [0, \delta n]^2 \text{ is a good square}] \leq O(\delta)$. In other words, if a square on the δn lattice 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 $B^c \cap \{[0, \delta n]^2 \text{ is a good square}\}$ 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 boundary of the square such that $\sigma_t \neq \sigma_{v_i}$, the statement is true. If this is not the case, there is a vertex t inside the square $\sigma_t \neq \sigma_{v_i}$, but for any vertex b on the boundary of the square $\sigma_t \neq \sigma_b$ so again the statement holds.

This configuration, in a similar way to Lemma 2.5.4, implies the existence of two 3-arm events in two disjoint half planes both from distance δ to $O(1)$ (see Figure 2.5.3), 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

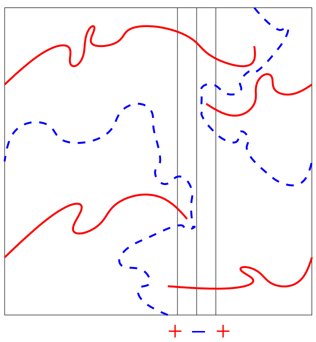

Figure 2.5.2: $\sigma_{t_1} \neq \sigma_{t_2}$ and $\sigma_{t_2} \neq \sigma_{t_3}$ results in two 3-arm events in two disjoint half planes

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 2.5.1,

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

independently from the distance d.

One of the 3-arm events can be started at any of $O(1/\delta^2)$ different $\delta n \times \delta n$-boxes and once this is fixed, the second one can be chosen $O(1/\delta)$ 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 $O(1/\delta^3)$ different ways, and the union bound gives that

$$\mathbb{P}[B^c \cap [0, \delta n]^2 \text{ is a good square}] \leq O(1/\delta^3)O(\delta^4) = O(\delta). \tag{2.5.5}$$

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. So expression (2.5.5) says that the probability that a square is good, but not perfect is small. Therefore, putting together (2.5.4) and (2.5.5) 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 to the correlation. We first use Markov's inequality to bound the probability that at least $\sqrt{\delta}$ ratio of all squares are imperfect:

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

Note that the respective magnetizations restricted onto an imperfect square are equal, so the difference between $M^{\mathbb{Z}_n^2}[\mathsf{LR}]$ and $M^{H_\delta}[\mathsf{LR}]$ comes only from imperfect squares. Therefore, on the event $\left\{|\text{imperfect squares}| \leq \sqrt{\delta}(1/\delta)^2\right\}$ we have

$$\left|M^{\mathbb{Z}_n^2}[\mathsf{LR}] - M^{H_\delta}[\mathsf{LR}]\right| \leq O(\sqrt{\delta}),$$

using that magnetization on any set is between -1 and 1, and thus the respective differences are at most 2 on imperfect squares. So

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

which implies

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

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

□

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

Theorem 2.5.6. *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 an arbitrary small $\delta > 0$. Using the indirect assumption that there is a sparse sequence of subsets with clue greater than $c > 0$, we are going to show that the average of the translated crossing events on the δ-lattice $M^{H_\delta}[\mathsf{LR}_n]$ also has clue greater than $c' > 0$ for a larger, but still sparse sequence of subsets U_n^δ (where c' depends only on δ, but does not depend on n).

At the same time Lemma 2.5.5 shows that the average of the translates on the δ-lattice 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 $M^{\mathbb{Z}_n^2}[\mathsf{LR}_n]$. Nevertheless, this is in contradiction with Theorem 2.1.1, which claims that a sequence of sparse subsets cannot give asymptotically positive clue about a transitive function.

For a given δ, we define the set $U_n^\delta = \cup_{\mathbf{t}\in H_\delta} U^{\mathbf{t}}$, where $U^{\mathbf{t}} = \{\mathbf{u}+\mathbf{t} \,:\, \mathbf{u} \in U\}$. So U_n^δ is the union of all H_δ-translates of U. Clearly, $\mathsf{clue}(\mathsf{LR}_n \mid U_n^\delta) \geq c$. We shall choose the appropriate value of δ at the end of the proof.

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

$$\begin{aligned}\mathrm{Var}(M^{H_\delta}[\mathsf{LR}_n]) =& \frac{1}{|H_\delta|^2}\sum_{\mathbf{h},\mathbf{g}\in H_\delta} \mathrm{Cov}(\mathsf{LR}_n^{\mathbf{h}}, \mathsf{LR}_n^{\mathbf{g}})\\ =&\mathrm{Var}(\mathsf{LR}_n)\frac{1}{|H_\delta|^2}\sum_{\mathbf{h},\mathbf{g}\in H_\delta} \mathrm{Corr}(\mathsf{LR}_n^{\mathbf{h}}, \mathsf{LR}_n^{\mathbf{g}})\\ \leq&\mathrm{Var}(\mathsf{LR}_n).\end{aligned}$$

We are now ready to bound $\mathsf{clue}(M^{H_\delta}[\mathsf{LR}_n] \mid U_n^\delta)$ from below. We will denote by P the projection (conditional expectation, from the probabilistic point of view) onto $\mathcal{F}_{U_n^\delta}$. Let $\mathbf{h}_1$ and $\mathbf{h}_2 \in \mathbb{Z}_n^2$. Then

$$\begin{aligned}&\frac{\mathrm{Cov}(P[\mathsf{LR}_n^{\mathbf{h}_1}], P[\mathsf{LR}_n^{\mathbf{h}_2}])}{\mathrm{Var}(\mathsf{LR}_n)}\\ =&\frac{\sqrt{\mathrm{Var}(P[\mathsf{LR}_n^{\mathbf{h}_1}])\mathrm{Var}(P[\mathsf{LR}_n^{\mathbf{h}_2}])}}{\mathrm{Var}(\mathsf{LR}_n)}\mathrm{Corr}(P[\mathsf{LR}_n^{\mathbf{h}_1}], P[\mathsf{LR}_n^{\mathbf{h}_2}])\\ =&\sqrt{\mathsf{clue}(\mathsf{LR}_n^{\mathbf{h}_1} \mid U_n^\delta)\mathsf{clue}(\mathsf{LR}_n^{\mathbf{h}_2} \mid U_n^\delta)}\mathrm{Corr}(P[\mathsf{LR}_n^{\mathbf{h}_1}], P[\mathsf{LR}_n^{\mathbf{h}_2}])\\ \geq& c\,\mathrm{Corr}(P[\mathsf{LR}_n^{\mathbf{h}_1}], P[\mathsf{LR}_n^{\mathbf{h}_2}]),\end{aligned} \tag{2.5.6}$$

where used that $\mathsf{clue}(\mathsf{LR}_n^{\mathbf{h}} \mid U_n^\delta) > c$ for any $\mathbf{h} \in H_\delta$.

We fix another grid size θ, which is coarser than δ so $0 < \delta < \theta$. Now we have

$$\begin{aligned}&\mathsf{clue}(M^{H_\delta}[\mathsf{LR}_n] \mid U_n^\delta)\\ &= \frac{\mathrm{Var}(P[M^{H_\delta}[\mathsf{LR}_n]])}{\mathrm{Var}(M^{H_\delta}[\mathsf{LR}_n])}\\ &\geq \frac{1}{|H_\delta|^2}\frac{\sum_{\mathbf{h}_1,\mathbf{h}_2\in H_\delta}\mathrm{Cov}(P[\mathsf{LR}_n^{\mathbf{h}_1}], P[\mathsf{LR}_n^{\mathbf{h}_2}])}{\mathrm{Var}(\mathsf{LR}_n)}\\ &\geq \frac{c}{|H_\delta|^2}\sum_{\mathbf{h}_1,\mathbf{h}_2\in H_\delta}\mathrm{Corr}(P[\mathsf{LR}_n^{\mathbf{h}_1}], P[\mathsf{LR}_n^{\mathbf{h}_2}])\\ &\geq \frac{c}{|H_\delta|^2}\sum_{\mathbf{h}\in H_\delta}\sum_{\mathbf{d}\in R_\theta(\mathbf{h})\cap H_\delta}\mathrm{Corr}(P[\mathsf{LR}_n^{\mathbf{h}}], P[\mathsf{LR}_n^{\mathbf{d}}])\end{aligned} \tag{2.5.7}$$

We remind the reader that $R_\theta(\mathbf{h})$ is the square with side length $2\theta n$ around $\mathbf{h}$. In the estimation above we first used the upper bound for $\mathrm{Var}(M^{H_\delta}[\mathsf{LR}_n])$ after (2.5.6) and finally that LR_n is monotone, and therefore, by the FKG-inequality $\mathrm{Cov}(P[\mathsf{LR}_n^{\mathbf{h}_1}], P[\mathsf{LR}_n^{\mathbf{h}_2}] \geq 0$.

By Lemma 2.5.4 there exists a $K > 0$ such that

$$\mathrm{Corr}(\mathsf{LR}_n^{\mathbf{h}}, \mathsf{LR}_n^{\mathbf{d}}) \geq 1 - K\theta$$

for every $\mathbf{h} \in \mathbb{Z}_n^2$ and $\mathbf{d} \in R_\theta(\mathbf{h})$. Applying Lemma 2.5.2 for $\mathsf{LR}_n^{\mathbf{h}}$, $\mathsf{LR}_n^{\mathbf{d}}$ and P, and choosing θ small enough so that $2K\theta < c/2$, we get that

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

Plugging this back into (2.5.7), and using that $|H_\delta| = \frac{1}{\delta^2}$ and $|R_\theta(\mathbf{h}) \cap H_\delta| = |R_\theta \cap H_\delta| = 4\theta^2/\delta^2$, and thus $|R_\theta(\mathbf{h}) \cap H_\delta|/|H_\delta| = \theta^2$ for any $\mathbf{h}$, we obtain the following bound:

$$\begin{aligned} \mathsf{clue}(M^{H_\delta}[\mathsf{LR}_n] \mid U_n^\delta) \geq & \frac{c}{|H_\delta|^2}|H_\delta||R_\theta \cap H_\delta| \left(1 - \frac{2K\theta}{c}\right) \\ = & \frac{|R_\theta \cap H_\delta|}{|H_\delta|} c \left(1 - \frac{2K\theta}{c}\right) \\ = & \theta^2(c - 2K\theta) \\ \geq & \theta^2 \frac{c}{2}. \end{aligned} \tag{2.5.8}$$

At the same time, by Lemma 2.5.5, there is some $L > 0$ such that

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

Now choose $\delta \leq \theta^4 \frac{c^2}{16L}$ so that $L\sqrt{\delta} \leq \theta^2 \frac{c}{4}$. Applying Lemma 2.5.3 again with $M^{H_\delta}[\mathsf{LR}_n]$ and $M[\mathsf{LR}_n]$ we get from (2.5.8) that for all $n \in \mathbb{N}$

$$\mathsf{clue}(M[\mathsf{LR}_n] \mid U_n^\delta) \geq \theta^2 \frac{c}{2} - L\sqrt{\delta} \geq \theta^2 \frac{c}{4}. \tag{2.5.9}$$

But $M[\mathsf{LR}_n]$ is transitive and $|U_n^\delta| = \frac{1}{\delta^2}|U_n| = o(n^2)$ and therefore Theorem 2.1.1, tells us that $\mathsf{clue}(M[\mathsf{LR}_n] \mid U_n^\delta) \longrightarrow 0$, which is in contradiction with (2.5.9). □

Chapter 3

Sparse Reconstruction in Spin Systems

3.1 Introduction

In Chapter 2 we introduced the concept of sparse reconstruction for a sequence of (Boolean) functions. We have seen that in case the underlying measure is a product measure the clue of a transitive function with respect to a subset of coordinates is bounded by the density of the subset. In particular, there is no sparse reconstruction for sequences of transitive functions.

In this Chapter we broaden our view. We shall investigate what happens if we replace the product measure with some different measure. We typically consider a sequence of locally convergent finite transitive graphs $G_n(V_n, E_n)$ and a corresponding sequence of measures $\mathbb{P}_n$ on $\{-1,1\}^{V_n}$ which are invariant under the automorphisms of the corresponding graph G_n. (Our definition allows for somewhat greater generality, but most of the interesting cases fits in this framework.) Under what conditions can we ensure that some version of Theorem 2.1.7 remains true?

Perhaps unsurprisingly, we have only few results that apply for general sequences of spin systems (an alternative way to call the transitive measures on the hypercube). The general intuition is that if in a spin system there is not too much dependency, i.e., in some sense it is close to a product measure, we expect that there is no sparse reconstruction, while if a spin system admits lots of dependency, sparse reconstruction for some transitive function is possible.

Our guinea pig spin system in this work will be the Ising model on a sequence of transitive graphs, in particular, on the d-dimensional torus $\mathbb{Z}_n^d$. For Ising measures, the intuition above comes down as follows: For high temperature models we expect that no sparse reconstruction is possible for a transitive function, while for low temperature models, there is enough dependency to make sparse reconstruction possible. As we shall see, it is not difficult to show that below the critical temperature the magnetization, and the majority as well can be reconstructed from a sparse sequence of subset. This follows from the fact that the low temperature Ising model is not ergodic (See Proposition 3.2.3).

More interestingly, it turns out that — at least on the tori $\mathbb{Z}_n^d$ — the magnetization and the majority as well can be reconstructed on the sequence of critical Ising measures too. This follows from Corollary 3.2.7, which claims that if the susceptibility of a sequence of spin systems (see (3.2.2)) approaches infinity then sparse reconstruction is possible from a sequence of sparse random set of spins. It is a well known fact in Statistical Physics that susceptibility explodes at the critical temperature, hence the result.

As for the high temperature models, we conjecture that no sparse reconstruction is possible, but at present we cannot prove it in full generality. Nevertheless, we have some partial results. On the one hand, in Section 3.4.3 we use the concept of I-clue (Definition 2.2.5) and entropy estimates to show that for the high temperature Curie-Weiss model (i.e., the Ising model on the complete graph) there is no sparse reconstruction.

In Section 3.3 we investigate sequences tending to finitary factor of IID measure on $\mathbb{Z}^d$. Applying Theorem 2.1.7 we can show that no sparse reconstruction is possible from a sequence of subsets with density much lower then $(1/\log n)^d$, for a finitary factor of IID measure with exponentially decaying coding radius. As high temperature Ising is known to be a finitary factor of IID with the required conditions (as proven in [BS99]), this theorem applies to the high temperature Ising model.

Susceptibility seems to be a key concept for factor of IID spin models as well. If the expected coding volume of a factor of IID system is finite then there is no sparse reconstruction for the magnetization from a spin system converging to this measure (in some appropriate sense, which we define in the sequel). This follows from the fact that finite expected coding volume implies finite susceptibility (Lemma 3.3.1). Therefore the key question here is whether finite expected coding volume implies no sparse reconstruction. A positive answer would be another indication that susceptibility plays a crucial role in these kind of problems.

It is, however, possible that the susceptibility is a distraction. This is why it would be crucial to understand the low temperature + Ising measure, i.e, the Ising measure conditioned to have positive total magnetization. In the Statistical Physics community it is usually regarded to be similar to the high temperature model. In particular, susceptibility is finite for the + Ising measure. Still, there are some important differences. First, in contrast with high temperature models, the magnetization has subexponential tail (because the Ising model is a Markov field and we only need to pay the price of the negative spins on the boundary of a cluster of $-$s). Second, it was proven in [BM02], (for not unrelated reasons) that the usual Glauber dynamics for the low temperature + Ising measure has no spectral gap (again contrary to high temperature Ising).

We hoped to be able to reconstruct a low probability event, for example, the event that the size of the largest volume $-$ cluster is larger than its median from a sparse grid. Should our efforts be successful, it would be the first natural example where susceptibility is finite, still sparse reconstruction is possible.

3.2 Results for General Spin Systems

3.2.1 Basic concepts, facts and questions

We have seen in Chapter 2 (Theorem 2.1.1 and Theorem 2.1.7) 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 this present Chapter we will investigate the same sort of questions for different sequences of probability measures.

In order to ensure that the question makes sense we will have to require certain conditions from the sequence of probability measures $\mathbb{P}_n$ in question. In general, the measure $\mathbb{P}_n$ on $\{-1,1\}^{V_n}$ is invariant under the action of some group Γ_n, where Γ_n acts transitively on the coordinate set V_n, although we also introduce a notion of sparse reconstruction that does not rely on symmetries (see Definition 3.2.2).

We pose an additional requirement, namely that the sequence of probability measure has to be weakly convergent. This is partly to discard very irregular sequences of measures and partly in hope that certain properties of the limiting measure will indicate 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 stronger link than weak convergence of measures. We also need to ensure that the symmetries in the sequence and in the limit are consistent. In case $\mathbb{P}_n$ or $\mathbb{P}$ is not invariant under the full automorphism group of G_n or G, respectively, we may falsely call two rooted neighbourhoods isomorphic, because they might be isomorphic as rooted graphs, nevertheless they have different distributions (as the measure $\mathbb{P}$ is invariant under a different group action).

The setup we choose here is not the most general possible, but it includes all the situations we are interested in and most of those we can think of. So let $G = (V, E)$ be a transitive, edge-labelled graph (with possible orientation of the edges). With other words, we assume $\mathsf{Aut}(G)$ act on G transitively (preserving the labels as well). The edge labels and directed edges are necessary to distinguish neighbourhoods that are isomorphic as rooted graphs labels, but fail to have the same distribution.

We now consider a sequence $\{G_n(V_n, E_n)\ : n \in \mathbb{N}\}$ of edge-labelled, directed graphs, that converges to G in the Benjamini–Schramm sense. This means that for every finite, rooted and edge-labelled directed graph B, the probability that the R-ball around a uniformly randomly chosen vertex from G_n is isomorphic to B converges as $n \to \infty$ (see [BS01]). This is, in general defined for a random limit, a distribution on the possible realisations of the R-ball, for all $R \in \mathbb{N}$. In our case, where G is a deterministic and transitive graph, it means that more and more of the R-neighbourhoods in the graph sequence become isomorphic to the R-neighbourhood on G, as $n \to \infty$. If, as it is the case in most of our applications, the graph sequence consists of deterministic, transitive graphs, this means that for large enough n the R-balls in G_n stabilize according to G (also known as local convergence of graphs). Nevertheless, this setup allows convergence to G via sequences of non-transitive or random graphs.

For each n we define a (usually $\mathsf{Aut}(G_n)$-invariant) measure $\mathbb{P}_n$ on $\{-1,1\}^{V_n}$. Note that any particular configuration $\sigma \in \{-1,1\}^V$ on the vertex set can be seen as an edge and vertex-labelled graph. Thus G_n, its vertex configuration endowed with the measure $\mathbb{P}_n$, can be identified with a vertex- and edge-labelled random graph, which we shall represent with the pair $(G_n, \mathbb{P}_n)$.

The sequence of finite vertex- and edge-labelled, directed random graphs $\{(G_n, \mathbb{P}_n) : n \in \mathbb{N}\}$ is required to converge to a transitive, $\mathsf{Aut}(G)$-invariant, vertex- and edge-labelled directed graph $(G, \mathbb{P})$ in the Benjamini–Schramm sense. It is, in fact just the same as weak convergence of measures, but it is a convenient to interpret this convergence in the Benjamini–Schramm setup. We will call such a sequence $\{(G_n, \mathbb{P}_n) : n \in \mathbb{N}\}$ a convergent spin system, where $\{G_n\}$ alone, and also decorated with vertex labels as $\{(G_n, \mathbb{P}_n)\}$ are both Benjamini–Schramm convergent. This will be the object of our investigations throughout this chapter.

Often, however, when we do not explicitly make use of the limit of the spin system, it is more natural to think about $\mathbb{P}_n$ as a measure on $\{-1, 1\}^{V_n}$, where V_n is the vertex set of a finite (in most cases) transitive graphs. Therefore, in the sequel we shall often talk about simply a sequence of measures $\mathbb{P}_n$, and omit the random graph interpretation lurking in the background.

Now that our framework is set, 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 3.2.1 ((Weak) Sparse Reconstruction)**.** Let $\{(G_n, \mathbb{P}_n) : n \in \mathbb{N}\}$ be a sequence of finite transitive spin systems that Benjamini–Schramm converges to the spin system $(G, \mathbb{P})$. Let $f_n : \{-1, 1\}^{V_n} \longrightarrow \mathbb{R}$. There is Sparse Reconstruction for f_n in $\mathbb{P}_n$ 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 $\{(G_n, \mathbb{P}_n) : n \in \mathbb{N}\}$ if there exists 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 $\mathbb{P}_n$ if there exist a sequence of transitive, non-degenerate Boolean functions f_n such that there is Sparse Reconstruction for f_n.

The difference between WSR and SR lies in that in the latter case we also require non-degeneracy of the sequence. It is not difficult to show (based on Corollary 3.2.10) that whenever there is WSR, there is also a possibly degenerate sequence of Boolean functions for which there exists sparse reconstruction.

We will give an example of a sequence of measures for which there is WSR, but no SR. (See the example under Corollary 3.2.7.) Of course, for sequences of product measures, there is neither SR, nor WSR.

There is an alternative version of sparse reconstruction that does not require any symmetry of the sequence of graphs or the corresponding measures $\mathbb{P}_n$.

Definition 3.2.2 (Random Sparse Reconstruction)**.** Let $\{(G_n, \mathbb{P}_n) : n \in \mathbb{N}\}$ be a sequence of finite spin systems that Benjamini–Schramm converges to the spin system $(G, \mathbb{P})$. For every n let $\mathcal{U}_n$ be a random subset of V_n independent from the spin system with the property that

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

The quantity δ_n is called the revealment of $\mathcal{U}_n$.

There is random sparse reconstruction (briefly, RSR) for $\{(G_n, \mathbb{P}_n) : n \in \mathbb{N}\}$ if there is a sequence of Boolean functions $f_n : \{-1,1\}^{V_n} \longrightarrow \{-1,1\}$ and a $\mathcal{U}_n$ as above such that

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

for some $c > 0$.

Remark 3.2.1. It is easy to see that in case SR holds for a sequence $(\mathbb{P}_n, G_n)$ then RSR as well (in case SR is meaningful). Instead of the the deterministic set U_n we define $\mathcal{U}_n$ as a uniformly random $\mathsf{Aut}(G_n)$ translate of U_n. Because of the transitivity of f_n the clue is invariant under automorphisms, and $\delta_n = \frac{|U_n|}{|V_n|}$. Also, with some minor modifications of the proof of Theorem 2.1.1 one can show that for product measures RSR does not hold either.

The two concepts are, however, not equivalent in general. We give an example to demonstrate this. Let $G_n = \mathbb{Z}_{2n}$ and consider the following $\mathsf{Aut}(\mathbb{Z}_{2n})$-invariant measure. We first choose a uniformly random edge and after we sample the n spins to the left from this edge as IID Bernoulli($\frac{3}{4}$) variables and the remaining spins as IID Bernoulli($\frac{1}{4}$), respectively. First observe that there is no WSR in this system. Indeed, take a sparse sequence of subsets $U_n \subseteq \mathbb{Z}_{2n}$ and consider the σ-algebra $\mathcal{F}_n$ generated by $\mathcal{F}_{U_n}$ and the uniformly random edge that determines the "border". For any sequence of transitive functions we have $\mathsf{clue}(f_n \mid U_n) \to 0$ by Theorem 2.1.7, since conditioned on the "border" edge, the spin system is distributed as a product measure (with the factors having equal variances), and therefore a sparse sequence of subsets has a vanishing clue. Since $\mathsf{clue}(f_n \mid U_n) \leq \mathsf{clue}(f_n \mid \mathcal{F}_n)$, WSR is not possible in this system.

At the same time a single bit can be reconstructed from a Bernoulli random set of density $\log n/n$. Indeed, observe that with probability $(\log n - 2)/\log n \to 1$ we learn whether we in the $\frac{3}{4}$-side or the $\frac{1}{4}$-side of the cycle. This in turn gives positive amount of information on the value of the bit. Applying this construction with $\mathrm{Ber}(1-\epsilon)$ and $\mathrm{Ber}(\epsilon)$, we can get clue arbitrarily close to 1.

In the above example the limiting spin system is $\mathrm{Ber}(\epsilon)$ or $\mathrm{Ber}(1-\epsilon)$ on $\mathbb{Z}$ with probability 1/2–1/2, which is not ergodic. We do not know whether this can happen in case the limiting measure $\mathbb{P}$ is ergodic. More generally we can ask the following

Question 3.2.2. *Suppose that for a transitive, convergent sequence of spin systems $\mathbb{P}_n$ there is Random Sparse Reconstruction. Under what condition does this imply that there is also Sparse Reconstruction for $\mathbb{P}_n$?*

There are two reasons why RSR is of interest. First, as already mentioned, it allows us to investigate sparse reconstruction for sequences of spin systems which do not exhibit symmetries. One example is to take the sequence of random d-regular graphs on n vertices, and put finitary factor of IID measures as $\mathbb{P}_n$s (see Definition 3.3.1). This sequence is known to BS-converge to the d-regular (infinite) tree (see [LyP], Chapter 6). Another point is that RSR can be directly applied to G, without defining a sequence of converging measures (or with other words $G_n = G$, for every $n \in \mathbb{N}$). Indeed one can take a sequence of random finite set $\mathcal{U}_n \subseteq V$ on the infinite vertex set of G with vanishing revealment. Therefore one can hope that by understanding the relation between SR and RSR, we have better chances to link SR to the properties of the limiting spin system.

Indeed, one of the first natural question that comes into mind is whether the existence of Sparse Reconstruction is the attribute of the limiting measure. That is, if $\mathbb{P}_n$ and $\mathbb{Q}_n$

are sequences of measures sharing the same weak limit, then either both of the sequences admit SR (WSR), or both do not. The answer to this question is, in general, negative.

It is possible to construct a sequence $\mathbb{P}_n$ weakly converging to a product measure, which admits sparse reconstruction. Indeed, let $\mathbb{P}_n$ 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 the long $+$ or $-$ interval is and again with high probability, whether it is $+$ or $-$ will tell us the Majority. At the same time, as it is easy to check, this spin system weakly converges to the product measure on $\{-1,1\}^{\mathbb{Z}}$.

It seems intuitive that if the limiting spin system is not ergodic, then there is some sort of sparse reconstruction in the sequence. This is, however, not true in general. First, the example in Remark 3.2.1 shows a sequence weakly converging to a not ergodic measure without (weak) sparse reconstruction. There is an even simpler example for a sequence without even RSR: Let G_n be a path of $2n$ vertices, and put IID Bernoulli($\frac{3}{4}$) measure on the first n vertices and IID Bernoulli($\frac{1}{4}$) to the remaining vertices. It is obvious that the limiting measure is the same, but now there is no RSR. Nevertheless, in case G_n is an expander sequence, we have RSR. (For a review on expanders, see [Lu94].)

First we give the relevant definitions. For a finite graph G the Cheeger constant is

$$h(G) = \min\{\frac{|\partial(W)|}{|W|} : W \subseteq V(G),\ |W| \leq \frac{1}{2}|V(G)|\},$$

where $\partial(W) = \{(x,y) \in E(G) :\ x \in W,\ y \notin W\}$.

Definition 3.2.3 (Expander sequence)**.** A sequence $\{G_n(V_n, E_n) :\ n \in \mathbb{N}\}$ of bounded degree graphs with $|V_n| \to \infty$ is an h-expander for some $h > 0$, if

$$h(G_n) > h$$

for every $n \in \mathbb{N}$.

Proposition 3.2.3. *Let $\{(G_n, \mathbb{P}_n)\ : n \in \mathbb{N}\}$ be a sequence of spin systems, where $\{G_n\}$ is an h-expander sequence with maximum degree D and let $(G, \mathbb{P})$ the limiting spin system.*

If the $\mathsf{Aut}(G)$-invariant measure $\mathbb{P}$ is not $\mathsf{Aut}(G)$-ergodic, then there is (Random) Sparse Reconstruction for $\{(G_n, \mathbb{P}_n)\ : n \in \mathbb{N}\}$.

Proof. As the limiting measure $\mathbb{P}$ is not 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 any $\delta > 0$ one can choose such an event A_δ satisfying

$$\mathbb{P}[A \triangle A_\delta] < \delta,$$

where $A \triangle B$ is the symmetric difference of the events A and B. Now pick a root $o \in V$ and choose $R \in \mathbb{N}$ large enough so that A_δ is $B_R(o)$-measurable (that is, the coordinates on which A_δ depends are inside the ball $B_R(o)$. Since G is transitive, all R-balls are isomorphic on G, so for any $x \in V$ we can define the $\mathcal{F}_{B_R(o)}$-measurable event A_δ^x. Using that A is invariant, it is clear that A_δ^x is an equally good approximation of A and thus

$$\mathbb{P}[A_\delta^x \triangle A_\delta^y] < 2\delta$$

for any $x, y \in V$. Suppose that x and y are neighbors (so $(x, y) \in E(G)$). We can choose R_2 large enough so that both A^x_δ and A^y_δ are $\mathcal{F}_{B_{R_2}(x)}$-measurable.

Now choose n large enough so that the R_2 balls are isomorphic both in G_n and G and in $(G_n, \mathbb{P}_n)$ and $(G, \mathbb{P})$ (as vertex decorated graphs) with probability at least $1 - \delta$ — this is possible because of the Benjamini–Schramm convergence.

For every $v \in V_n$ one can identify the ball $B_{R_2}(x)$ on G with the ball $B_{R_2}(v)$ on G_n via a rooted isomorphism with large probability. Using the same isomorphism one can define the $B_N(v)$-measurable events A^v_δ and A^u_δ on $\{-1, 1\}^{V_n}$ for any $(u, v) \in E(G_n)$.

Observe that for any $u, v \in V_n$, with $(u, v) \in E$ we have by the choice of n

$$\begin{aligned}
&\mathbb{P}_n[\mathbb{1}_{A^v_\delta} \neq \mathbb{1}_{A^u_\delta}] \\
&= \mathbb{P}_n[A^v_\delta \triangle A^u_\delta] < \mathbb{P}_n[A^x_\delta \triangle A^y_\delta] + \delta + \mathbb{P}[B_{R_2}(v) \text{ is not isomorphic to } B_{R_2}(x)] \\
&< 4\delta.
\end{aligned}$$

Define the random variable J_n as the number of edges (u, v) in G_n such that $\{\mathbb{1}_{A^v_\delta} \neq \mathbb{1}_{A^u_\delta}\}$. Fix an $\epsilon > 0$. Clearly, $\mathbb{E}[J_n] < 4\delta|E_n|$ and therefore by Markov's inequality

$$\mathbb{P}_n\left[J_n \geq \frac{2h\epsilon}{D}|E_n|\right] \leq 4\delta/(\frac{2h\epsilon}{D}),$$

which, by a suitable choice of δ can be made arbitrary small as h and D and ϵ are constants.

Now observe that the event $\{J_n < \frac{2h\epsilon}{D}|E_n|\}$ implies $|\sum_{v \in V_n}(2\mathbb{1}_{A^v_\delta} - 1)| > (1 - \epsilon)|V_n|$. Indeed, if this is not the case then both $\{v \in V_n : A^v_\delta\}$ and its complement has at least $\epsilon|V_n|$ element. So by the expansion property, there are at least $|V_n|h\epsilon \geq 2|E_n|h\epsilon/D$ edges between $\{v \in V_n : A^v_\delta\}$ and its complement (we used that $D|V_n| \geq 2|E_n|$). But for all these edges (u, v) we have $\mathbb{1}_{A^v_\delta} \neq \mathbb{1}_{A^u_\delta}$ and therefore $J_n \geq \frac{2h\epsilon}{D}|E_n|$ as well.

Observe that there is RSR for $\mathbb{1}_{A^u_\delta}$ from the random set $B_R(v)$ for a uniformly chosen $v \in V_n$ (recall that A^v_δ is $B_R(v)$ -measurable). As the probability that $|\sum_{v \in V_n}(2\mathbb{1}_{A^v_\delta} - 1)| > (1 - \epsilon)|V_n|$ is close to 1 the randomly chosen $\mathbb{1}_{A^u_\delta}$ coincides with A^v_δ with probability $1 - \epsilon$. At the same time, since G_n is of bounded degree, it is clear that the revealment of this random set goes to 0.

Moreover, in case $(G_n, \mathbb{P}_n)$ is a transitive sequence, the event $\left\{\sum_{v \in V_n}(2\mathbb{1}_{A^v_\epsilon} - 1) > 0\right\}$, that is, the majority of the events A^v_ϵ can be reconstructed from $B_R(v)$ for a fixed $v \in V_n$. The argument is essentially the same as for RSR. □

Using the first part of the argument it is easy to show that in case $(G, \mathbb{P})$ is not $\mathsf{Aut}(G)$-ergodic, there is Random Sparse Reconstruction for $(G, \mathbb{P})$ (that is, for the sequence $(G_n, \mathbb{P}_n) = (G, \mathbb{P})$).

We have seen that for a non ergodic measure on $(G, \mathbb{P})$ it is possible that a sequence of spin systems converging to $(G, \mathbb{P})$ admits SR or does not. The same holds for product measures. Can we always disturb a sequence to change its "natural" behavior? We hope that such a pathological behavior can only happen when the limit is a not ergodic measure.

Question 3.2.4. *Is there an ergodic* $\mathsf{Aut}(G)$*-invariant spin system* $(G, \mathbb{P})$ *such that whenever* $(G_n, \mathbb{P}_n)$ *converges to* $(G, \mathbb{P})$ *(in the above sense) there is SR for* $(G_n, \mathbb{P}_n)$*?*

An interesting example is the critical Ising model on $\mathbb{Z}^2$, which is ergodic, but admits sparse reconstruction. The magnetization, i.e., the sum of spins can be reconstructed (see Proposition 3.4.1). At the same time, one can take the sequence of critical Ising measures on the torus $\mathbb{Z}_n^2$ conditioned on $M_n = 0$. It can be shown that this spin system also converges to the critical Ising on $\mathbb{Z}^2$, but the magnetization clearly cannot be reconstructed. In this model, however, we suspect that other transitive functions can still be reconstructed.

Another line of questions concerns whether it is true that if in some sense $\mathbb{P}_n$ contains less randomness (or less information) then $\mathbb{Q}_n$ and $\mathbb{Q}_n$ admits SR, then is it true that $\mathbb{P}_n$ 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 3.2.4 (asymptotic entropy)**.** Let $\{\mathbb{P}_n\}$ be a sequence of measures. The asymptotic entropy of the sequence is

$$\mathcal{H}(\{\mathbb{P}_n\}) := \lim_{n\to\infty} \frac{H(\mathbb{P}_n)}{nH(\mathbb{P}_n[\ |\sigma_0])}$$

if it exists, where $\mathbb{P}_n[\ |\sigma_0]$ is the distribution of an individual spin.

It turns out that $\mathcal{H}(\{\mathbb{Q}_n\}) > \mathcal{H}(\{\mathbb{P}_n\})$ and $\{\mathbb{Q}_n\}$ having SR does not imply that $\{\mathbb{P}_n\}$ 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.

At the same time, there exits also a sequence $(G_n, \mathbb{P}_n)$ with $\mathcal{H}(\{\mathbb{P}_n\}) = 0$ in such a way that $(G_n, \mathbb{P}_n)$ admits no WSR. The example is again a version of Remark 3.2.1. The only modification is that after cutting the cycle, we sample the two halves according to $\text{Ber}(\epsilon_n)$ and $\text{Ber}(1-\epsilon_n)$, respectively, for some sequence $\epsilon_n \to 0$. This takes care of the asymptotic entropy. Using the chain rule for entropy one gets:

$$H(\mathbb{P}_n) = H(\mathbb{P}_n \mid \mathsf{cut}) + H(\mathsf{cut}) = \epsilon_n n + \log n = o(n),$$

where cut is the uniformly random place where the cycle is cut.

We do not know if such an example exists for ergodic measures.

Question 3.2.5. *Let $(G_n, \mathbb{P}_n)$ converge to the $\mathsf{Aut}(G)$-ergodic spin system $(G, \mathbb{P})$. If $\mathcal{H}(\{\mathbb{P}_n\}) = 0$, then is it true that there is SR for $(G_n, \mathbb{P}_n)$?*

Another way of expressing that $\mathbb{P}$ has no more randomness then $\mathbb{Q}$ is to say that $\mathbb{P}$ is a factor of $(\mathbb{Q}, H_n)$. In fact, according to Ornstein-Weiss theory , on amenable graph this is equivalent to the question above (see [OW87]). Indeed, it is possible that there is SR for the sequence $\mathbb{Q}_n$ while the sequence $\mathbb{P}_n$, which is a factor of $\mathbb{Q}_n$, does not admit SR. For example one can take take k_n independent copies of critical Ising model on the torus, for some $k_n \to \infty$. As we shall see (Theorem 3.4.5) there is sparse reconstruction for the the two-dimensional critical Ising, and it is easy to see that if there is sparse reconstruction from a model, then there is also sparse reconstruction for the spin system with arbitrary number of independent copies of it. However, if the factor spin system is the product of the spins of all the k_n copies at a given vertex, it is easy to believe that the law of this spin system converges to the uniform measure.

We see that sparse reconstruction is not monotone with respect to the most common ways of measuring information. It would be interesting to find some invariant of sequences of spin systems which behaves well with sparse reconstruction.

3.2.2 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 variable with law $\mathbb{P}_n$. We also assume that for each n there is a group Γ_n acting transitively on V_n and the law $\mathbb{P}_n$ of σ^n is invariant under this group action. In particular, for every $j \in V_n$ the distribution of σ_j^n is the same, where we denote by σ_j^n the projection of σ^n to the jth coordinate, a ± 1-valued random variable.

We will sometimes consider this setup with a slight generalization: we allow that the random variables σ_j^n 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 with σ^n.

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

$$M_n[\phi] := \frac{1}{m_n} \sum_{j \in V_n} \phi_j^n. \tag{3.2.1}$$

Compare this with Definition 2.5.2. The only difference is that here we are concerned with the magnetization of a spin system, so here $f = \mathrm{Id}$ and the magnetisation is a property of the underlying measure $\mathbb{P}_n$.

The term 'magnetization' comes from statistical physics, more specifically the Ising model (see section 3.4), a spin model which is central in this work. In the Ising model the value of the spins can be thought of as the charge of a particle, and in this framework the magnetization is interpreted as the charge of the whole field.

For the next definition it is useful to assign one vertex of V_n as root, denoted by 0 (as all vertices look the same the choice of the root does not change anything). We define the susceptibility of the random variable ϕ^n as

$$S_n(\phi) := S(\phi^n) = \sum_{j \in V_n} \mathrm{Cov}(\phi_0^n, \phi_j^n). \tag{3.2.2}$$

The concept of susceptibility is borrowed 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. Note that if absolute convergent, the susceptibility can also be defined for (countably) infinite spin systems as well.

By the Γ_n-invariance of the measure, $\sum_{j \in V_n} \mathrm{Cov}(\phi_0^n, \phi_j^n) = \sum_{j \in V_n} \mathrm{Cov}(\phi_k^n, \phi_j^n)$ for any $k \in V_n$ and therefore

$$\mathrm{Var}(M_n[\phi]) = \frac{1}{m_n^2} \sum_{k \in V_n} \sum_{j \in V_n} \mathrm{Cov}(\phi_k^n, \phi_j^n) = \frac{1}{m_n} \sum_{j \in V_n} \mathrm{Cov}(\phi_0^n, \phi_j^n).$$

Thus we have the following relationship between $M[\phi]$ and $S(\phi)$:

$$\mathrm{Var}(M_n[\phi]) = \frac{S_n(\phi)}{m_n}. \tag{3.2.3}$$

Furthermore, using translation invariance again,

$$\frac{1}{m_n} \frac{S_n(\phi)}{\mathrm{Var}(\phi_0^n)} = \frac{1}{m_n} \sum_{j \in V_n} \mathrm{Corr}(\phi_0^n, \phi_j^n) = \overline{\mathrm{Corr}}(\phi_0^n, \phi^n). \tag{3.2.4}$$

where $\overline{\mathrm{Corr}}(\phi_0^n, \phi^n)$ is the average correlation between the random variable ϕ_0^n and its translates.

The next proposition states that in case the average correlation defined above is sufficiently high for a spin system then the magnetization can be reconstructed from a sparse random set with high probability.

Proposition 3.2.6 (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 Γ_n on V_n. Suppose that*

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

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, and therefore once we fix the size of U_n then $\mathrm{Corr}(M_n, M_n^{U_n})$ depends only on the quantity $\mathrm{Var}(M^{U_n})$. 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

$$\begin{aligned}\mathbb{E}\left[\mathrm{Corr}(M_n, \mathbb{E}[M_n^{\mathcal{H}} \mid \mathcal{H}])\right] &= \sqrt{\frac{S_n}{m_n}}\,\mathbb{E}\left[\frac{1}{\sqrt{\mathrm{Var}(M_n^{\mathcal{H}} \mid \mathcal{H})}}\right] \\ &\geq \sqrt{\frac{S_n}{m_n}} \frac{1}{\sqrt{\mathbb{E}[\mathrm{Var}(M_n^{\mathcal{H}} \mid \mathcal{H})]}},\end{aligned} \tag{3.2.5}$$

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

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

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):

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

Now we can give a lower bound for the average correlation over all subsets of size k_n. Substituting back into (3.2.5), we get

$$\begin{aligned} \mathbb{E}\left[\mathrm{Corr}(M_n,\mathbb{E}[M_n^{\mathcal{H}} \mid \mathcal{H}])\right] &\geq \frac{\sqrt{S_n}}{\sqrt{m_n}\sqrt{\frac{k_n-1}{k_n}\frac{1}{m_n-1}S_n + s_n\left(\frac{1}{k_n}\left(1-\frac{k_n-1}{m_n-1}\right)\right)}} \\ &= \left(\frac{k_n-1}{k_n}\frac{m_n}{m_n-1} + \frac{s_n}{S_n}\left(\frac{m_n}{k_n}\left(1-\frac{k_n-1}{m_n-1}\right)\right)\right)^{-\frac{1}{2}} \\ &\asymp \left(1+\frac{s_n}{S_n}\frac{m_n}{k_n}\right)^{-\frac{1}{2}} \asymp \left(1+\frac{1}{k_n\overline{\mathrm{Corr}}(\phi_0^n,\phi^n)}\right)^{-\frac{1}{2}}, \end{aligned}$$

using that on the one hand, $(k_n-1)/(m_n-1) \longrightarrow 0$, by assumption and that on the other hand, $S_n/s_n m_n = \overline{\mathrm{Corr}}(\phi_0^n,\phi^n)$, by (3.2.4).

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

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

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 3.2.7. *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 Γ_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 $\epsilon>0$

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

Proof. It is straightforward to check that $S_n(\sigma) \to \infty$ is equivalent to $\overline{\mathrm{Corr}}(\phi_0^n, \phi^n) \gg \frac{1}{m_n}$, when $\mathrm{Var}(\sigma_0^n)$ 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 a uniform IID measure and a measure 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 $\mathrm{Var}(M_n) \gg n$, and consequently, by Theorem 3.2.6 the magnetization can be reconstructed, while Majority (or any other non-degenerate Boolean function) cannot. So in this sequence of measures there is Weak Sparse Reconstruction, but no Sparse Reconstruction.

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

Proposition 3.2.8. *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*

$$\mathbb{P}\left[|\sum_{j \in V_n} \sigma_j^n| \geq K a_n \right] > c, \tag{3.2.6}$$

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 $A = \left\{ |\sum_{j \in V_n} \sigma_j^n| \geq K a_n \right\}$ the expectation of the sum $|\sum_{j \in \mathcal{B}^{p_n}} \sigma_j^n|$ in a Bernoulli sample can be bounded as follows:

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

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 $k \neq j$:

$$\begin{aligned}
&\mathrm{Var}\left(\mathbb{E}\left[\sum_{j \in \mathcal{B}^{p_n}} \sigma_j^n \;\middle|\; \mathcal{B}^{p_n} \right] \;\middle|\; A \right) \\
=&\mathrm{Var}\left(\sum_{j \in V_n : \sigma_j^n = 1} \mathbb{1}_{j \in \mathcal{B}^{p_n}} - \sum_{k \in V_n : \sigma_k^n = -1} \mathbb{1}_{k \in \mathcal{B}^{p_n}} \;\middle|\; A \right) \\
=&m_n p_n (1 - p_n).
\end{aligned}$$

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

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

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, still conditioned on A, the majority of the sample coincides with the majority of the original system with high probability.

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

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

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, the magnetization can be reconstructed from $\mathcal{B}^{p_n}$ with high probability.

It remains to 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. □

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.

3.2.3 The 3-Correlation Lemma

First we discuss a slight generalization of the concepts of magnetization and susceptibility. Let us consider a spin system σ distributed according to $\mathbb{P}$, with coordinate set V and group action Γ, as before.

Recall that for a function $f : \{-1, 1\}^V \longrightarrow \mathbb{R}$ f^γ denotes the γ-translated version of f. Since $\mathbb{P}$ is Γ-invariant, $Z := f(\sigma)$ and $Z^\gamma := f^\gamma(\sigma)$ has the same distribution. One can define magnetization (as we have already done in Chapter 2 see (2.5.2)) and susceptibility for arbitrary function on the configuration space by:

$$M[f, \mathbb{P}] = M[Z] := \frac{1}{|\Gamma|}\sum_{\gamma\in\Gamma} Z^\gamma,$$

and

$$S_{\mathbb{P}}(f) = S(Z) := \frac{1}{|\Gamma_v|}\sum_{\gamma\in\Gamma} \mathrm{Cov}(Z, Z^\gamma),$$

where Γ_v is the stabilizer subgroup of an arbitrary vertex. In case the action of Γ on V is not free, that is the stabilizer subgroup of a vertex is not trivial then for every Z we count every term of the form $\mathrm{Cov}(Z, Z^\gamma)$ exactly $|\Gamma_v|$ many times. Indeed, as $|\Gamma| = |V||\Gamma_v|$ we have $|\Gamma_v|$ times too many terms in the susceptibility, hence the factor $1/|\Gamma_v|$. Warning: in case Z has additional symmetries it is possible that there are still repetitions in the sum of $S(Z)$ and it is perfectly fine. For example when Z itself is transitive, that is Γ-invariant, thus $\mathrm{Cov}(Z, Z^\gamma) = \mathrm{Var}(Z)$ for every γ and in this case $S(Z) = \frac{|\Gamma|}{|\Gamma_v|} = |V|\mathrm{Var}(Z)$.

It is easy to verify that the identity (3.2.3) continues to hold in this more general setting:

$$\mathrm{Var}(M[Z]) = \frac{S(Z)}{|V|}.$$

Along the lines of (3.2.4) (recalling that $|\Gamma| = |V||\Gamma_v|$), we again have

$$\frac{1}{|V|}\frac{S(Z)}{\mathrm{Var}(Z)} = \frac{1}{|\Gamma|}\sum_{\gamma\in\Gamma}\mathrm{Corr}(Z, Z^\gamma) = \overline{\mathrm{Corr}}_\Gamma(Z, Z). \tag{3.2.7}$$

In the sequel, to simplify notation we are going to avoid the factor $1/|\Gamma_v|$, that is, we implicitly assume that the action of the group on the coordinate set is free. We emphasize, however, that all the results are true without assuming a free action.

Observe that for any measurable Z the system of random variables $\{Z^\gamma\ \gamma \in \Gamma\}$ is a Γ-invariant family (although possibly the same random variables appear multiple times), so Equation (3.2.1) in fact already covers this case.

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

Lemma 3.2.9 (3-Correlation Lemma)**.** *Let $\sigma = \{\sigma_j\ :\ j \in V\}$ be a spin system with Γ-invariant distribution, where Γ acts transitively on V. Let $f : \{-1, 1\}^V \longrightarrow \mathbb{R}$ be a Γ-invariant (thus 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

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

observing that Z is transitive and therefore $\mathrm{Cov}(Z, \mathbb{E}[Z\,|\,\mathcal{F}_{U^\gamma}])$ is the same for any $\gamma \in \Gamma$. Since by (3.2.3) $\mathrm{Var}(M[\mathbb{E}[Z\,|\,\mathcal{F}_U]]) = S(\mathbb{E}[Z\,|\,\mathcal{F}_U])/|V|$, we get

$$\mathrm{Corr}(Z, M[\mathbb{E}[Z\,|\,\mathcal{F}_U]]) = \frac{\mathrm{Var}(\mathbb{E}[Z\,|\,\mathcal{F}_U])}{\sqrt{\mathrm{Var}(Z)}}\sqrt{\frac{|V|}{S(\mathbb{E}[Z\,|\,\mathcal{F}_U])}}. \tag{3.2.8}$$

As for the second factor, we can write the covariance as follows

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

So we get for the respective correlation:

$$\begin{aligned}&\operatorname{Corr}(\mathbb{E}[Z\,|\,\mathcal{F}_U], M[\mathbb{E}[Z\,|\,\mathcal{F}_U]])\\ =&\frac{S(\mathbb{E}[Z\,|\,\mathcal{F}_U])/|V|}{\sqrt{\operatorname{Var}(\mathbb{E}[Z\,|\,\mathcal{F}_U])S(\mathbb{E}[Z\,|\,\mathcal{F}_U])/|V|}}\\ =&\sqrt{\frac{S(\mathbb{E}[Z\,|\,\mathcal{F}_U])}{|V|\operatorname{Var}(\mathbb{E}[Z\,|\,\mathcal{F}_U])}}\end{aligned} \tag{3.2.9}$$

It is now easy to check that upon multiplying (3.2.9) with (3.2.8), one gets $\operatorname{Corr}(Z, \mathbb{E}[Z\,|\,\mathcal{F}_U])$ as stated. □

We may write any σ-measurable random variable Z in the place of $\mathbb{E}[Z\,|\,\mathcal{F}_U]$ in (3.2.9) and compare it with (3.2.7). This gives rise to the following, somewhat bizarre identity:

$$\overline{\operatorname{Corr}}_\Gamma(Z,Z) = \operatorname{Corr}^2(Z, M[Z]) \tag{3.2.10}$$

Now we state a few consequences of Lemma 3.2.9.

Corollary 3.2.10. *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 $\mathsf{clue}(f_n \mid U_n) = \operatorname{Corr}^2(Z_n, \mathbb{E}[Z_n\,|\,\mathcal{F}_{U_n}])$, it follows, using Lemma 3.2.9 that

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

According to (3.2.10) this means that

$$c < \overline{\operatorname{Corr}}(\mathbb{E}[Z_n\,|\,\mathcal{F}_{U_n}], \mathbb{E}[Z_n\,|\,\mathcal{F}_{U_n^\gamma}]). \tag{3.2.11}$$

Now we consider the spin system $\{\mathbb{E}[Z_n\,|\,\mathcal{F}_{U_n^\gamma}] : \gamma \in \Gamma\}$ and apply the argument in Proposition 3.2.6. We introduce the notation $\phi^n_\gamma := \mathbb{E}[Z_n\,|\,\mathcal{F}_{U_n^\gamma}]$.

Now recall from the proof of Proposition 3.2.6 that the expected correlation with respect to $\mathcal{H}^{k_n}$, a uniformly random subset of coordinates with k_n elements is given by

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

So taking into account (3.2.11) it follows that

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

Let k_n be a sequence of integers such that $k_n \longrightarrow \infty$, but $|U_n| k_n \ll |V_n|$. From this choice it is immediate that $\mathbb{E}\left[\operatorname{Corr}(M[\phi^n], \mathbb{E}[M[\phi^n]^{\mathcal{H}} \mid \mathcal{H}])\right] \longrightarrow 1$. 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 3.2.9, which provides a potential tool to show that there is no WSR for a particular spin system.

Corollary 3.2.11. *For a sequence of Γ_n-invariant spin system σ^n there is no weak 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*

$$\overline{\mathrm{Corr}}_\Gamma(Z_n, Z_n) < 1 - \epsilon \tag{3.2.12}$$

for every $n \geq N$, where the average is taken over all $\gamma \in \Gamma_n$.

Proof. Indirectly, assume that 3.2.12 holds but there exists a sequence of subsets $U_n \subseteq [n]$ and a sequence of transitive functions f_n with $\liminf \mathsf{clue}(f_n \mid U_n) = c > 0$. By Corollary 3.2.10, we may assume that $\lim_n \mathsf{clue}(f_n \mid U_n) = 1$.

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

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

where we first used Lemma 3.2.9, and after (3.2.10). As $\mathbb{E}[Z_n \,|\, \mathcal{F}_{U_n}]$ is obviously $\mathcal{F}_{U_n}$-measurable, this is in contradiction with our assumptions, so there is no WSR for the sequence σ^n. □

In order to show how Corollary 3.2.11 can be applied we give yet another proof for Theorem 2.1.7. For this we need the following:

Lemma 3.2.12. *Let $\mathbb{P}$ be the uniform measure on $\{-1,1\}^V$ and let $f : \{-1,1\}^V \longrightarrow \mathbb{R}$. Let Γ 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 $\gamma \in \Gamma$

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

and therefore

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

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

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

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

$$S(f) = \sum_{\Gamma \cdot S \in \mathcal{O}} \sum_{\gamma_1, \gamma_2 \in G} \widehat{f}(S^{\gamma_1})\widehat{f}(S^{\gamma_1 - \gamma_2}) = \sum_{\Gamma \cdot S \in \mathcal{O}} \left(\sum_{\gamma \in \Gamma} \widehat{f}(S^\gamma) \right)^2.$$

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

$$\left(\sum_{\gamma\in\Gamma} \widehat{f}(S^\gamma)\right)^2 \le |U| \sum_{\gamma\in\Gamma} \widehat{f}^2(S^\gamma), \tag{3.2.13}$$

and thus we get

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

□

Combining the above result with Corollary 3.2.11 we immediately get the promised alternative proof for Theorem 2.1.1. Indeed, for any sequence of $\mathcal{F}_{U_n}$-measurable functions f_n, one has

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

Remark 3.2.13. It is straightforward to generalise the above result to general product measures (Theorem 2.1.7), if one replaces the Fourier-Walsh transform with the Efron-Stein decomposition (see Theorem 2.1.6).

Remark 3.2.14. In equation 3.2.13 there is equality when $f = \sum_{j\in U} \omega_j$ and therefore the inequality of Lemma 3.2.12 is sharp.

Further observations

First, we want to point out that whenever Γ is a group acting on V (the action does not need to be transitive here), then the operator M^Γ is the orthogonal projection onto the space of Γ-invariant functions with respect to the inner product $\langle f, g\rangle := \mathbb{E}[fg]$. This can be proved as follows. Let Z be a random variable and recall that

$$M^\Gamma[Z] := \frac{1}{|\Gamma|} \sum_{\gamma\in\Gamma} Z^\gamma.$$

First, it is obvious from the definition that $M^\Gamma[Z] = Z$ whenever Z is Γ-invariant.

For any $\sigma \in \Gamma$

$$\sum_{\gamma\in\Gamma} \mathbb{E}\,[Z\ Z^\gamma] = \sum_{\gamma\in\Gamma} \mathbb{E}\,[Z^\sigma Z^\gamma], \tag{3.2.14}$$

since, by the Γ-invariance of the measure, $\mathbb{E}\,[Z^\sigma Z^\gamma] = \mathbb{E}\,[Z\ Z^{\gamma-\sigma}]$. This allows us to write

$$\mathbb{E}[M^\Gamma[Z]^2] = \frac{1}{|\Gamma|^2} \sum_{\gamma\in\Gamma}\sum_{\sigma\in\Gamma} \mathbb{E}\,[Z^\gamma Z^\sigma] = \frac{1}{|\Gamma|} \sum_{\gamma\in\Gamma} \mathbb{E}\,[Z^\gamma Z] = \mathbb{E}\left[Z M^\Gamma[Z]\right].$$

So we conclude that

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

which shows that M^Γ is indeed a projection operator.

This indicates that the question sparse reconstruction can be formulated from a more general perspective. Let $\mathcal{F}$ and $\mathcal{G}$ be two σ-subalgebras of a probability measure space.

Then one can ask how much do the information content of these σ-algebras overlap. One way to measure this is as follows:

$$C(\mathcal{F},\mathcal{G}) := \sup\{\mathrm{Corr}(X,Y) \ : \ X \ \mathcal{F}\text{-measurable}, \ Y \ \mathcal{G}\text{-measurable}\}.$$

We have to immediately point out out that this concept does not behave at all like (an inverted) distance. For example, both $\mathcal{F}$ and $\mathcal{G}$ has zero distance from the σ-algebra generated by $\mathcal{F}\cup\mathcal{G}$, while $\mathcal{F}$ and $\mathcal{G}$ can be arbitrary far from each other (meaning that $C(\mathcal{F},\mathcal{G})$ is close to 0.)

The relation to (weak) sparse reconstruction is the following: For a sequence of Γ_n-invariant spin systems let $\mathcal{T}_n$ denote the σ-algebra generated by Γ_n invariant functions. Now it is straightforward to see that no WSR on this spin system from the sequence of subsets U_n is equivalent with $\lim_n C(\mathcal{T}_n,\mathcal{F}_{U_n}) = 0$. Apart from a slightly more abstract view on the question, this language also reveals that there is an implicit symmetry in our sparse reconstruction question. We could as well ask if there is sparse reconstruction from $\mathcal{T}_n$ with respect to a function which depends only on coordinates from U_n.

We mention that $C(\mathcal{F},\mathcal{G})$ could be defined via a different information distance measure. For example, instead of $\mathrm{Corr}(X,Y)$ one could use $I(X,Y)$ (see Definition 2.2.4).

Lemma 3.2.9 suggests an iterative method to find functions with high clue. Let us introduce the notations

$$\mathrm{P}_U[Z] := \mathbb{E}[Z \,|\, \mathcal{F}_U], \quad \text{and} \quad T_U[Z] := M[\mathrm{P}_U[Z]].$$

Now we can rewrite the statement of Lemma 3.2.9 in this language. For every Γ_n-invariant random variable Z it holds that

$$\mathrm{Corr}(Z,T_U[Z])\mathrm{Corr}(\mathrm{P}_U[Z],T_U[Z]) = \mathrm{Corr}(Z,\mathrm{P}_U[Z]).$$

In case $\mathrm{Corr}(Z,T_U[Z]) < 1$, this obviously implies $\mathrm{Corr}(\mathrm{P}_U[Z],T_U[Z]) > \mathrm{Corr}(Z,\mathrm{P}_U[Z])$. Since $\mathrm{P}_U[T_U[Z]]$ is the function that (among $\mathcal{F}_U$-measurable variables) maximizes the correlation with $T_U[Z]$, one has $\mathrm{Corr}(\mathrm{P}_U[Z],T_U[Z]) \le \mathrm{Corr}(\mathrm{P}_U[T_U[Z]],T_U[Z])$. This two observations results in the following inequality:

$$\mathrm{Corr}(\mathrm{P}_U[T_U[Z]],T_U[Z]) > \mathrm{Corr}(Z,\mathrm{P}_U[Z]).$$

This means that applying the operator T_U to a given function we can increase the clue whenever $\mathrm{Corr}(Z,T_U[Z]) < 1$. So starting from an arbitrary Γ_n-invariant random variable Z we can increase the clue by iteratively applying the operator T_U. In case $\mathrm{Corr}(Z,T_U[Z]) = 1$, that is, if Z is an eigenfunction of T_U, the iteration comes to an end. So when T_U admits an eigenbasis it is sufficient to calculate the clue of the eigenfunctions since in that case $T_U^n[Z]$ converges to the linear combination of some eigenfunctions belonging to the same eigenspace.

Unfortunately, for most spin systems it is not easy to find such eigenfunctions. In order to at least illustrate the idea, we again turn to the product measure. For simplicity, we discuss the case of the uniform measure on $\{-1,1\}^V$. In this case, there is an eigenfunction corresponding to every Γ-orbit.

Indeed, let $\Gamma\cdot S$ be a Γ-orbit of S, as before and define $\xi_{\Gamma\cdot S} := \frac{1}{\sqrt{|\Gamma\cdot S|}}\sum_{T\in\Gamma\cdot S}\chi_T$ (we add the normalization to ensure that $\xi_{\Gamma\cdot S}$ has variance 1). Then

$$\mathrm{P}_U[\xi_{\Gamma\cdot S}] = \sum_{\gamma\in\Gamma:\ S^\gamma\subseteq U}\chi_{S^\gamma}.$$

and therefore

$$T_U[\xi_{\Gamma\cdot S}] = M^\Gamma[\mathrm{P}_U[\xi_{\Gamma\cdot S}]] = C(\Gamma\cdot S, U)\xi_{\Gamma\cdot S},$$

where

$$C(\Gamma\cdot S,\ U) := \frac{|\{\gamma\in\Gamma\ :\ S^\gamma\subseteq U\}|}{|\Gamma|}$$

denotes the fraction of translations of the subset S which are contained in U.

One can easily check the functions $\xi_{\Gamma\cdot S}$ form an orthonormal basis for the space of Γ-invariant functions. Indeed, every transitive function f can be represented in this basis as

$$f = \sum_{\Gamma\cdot S\in\mathcal{O}} \widehat{f}(\Gamma\cdot S)\xi_{\Gamma\cdot S},$$

where $\mathcal{O}$ stands for the set of Γ-orbits of 2^V and $\widehat{f}(\Gamma\cdot S)$ is the respective coefficient of f in this basis.

Moreover, $\lim_{n\to\infty} T_U^n[f]$ is contained in the eigenspace corresponding to the largest eigenvalue with nonzero coefficient in f. In particular, if f has non-zero energy on level 1 (the linear part) then $T_U^n[f]$ tends to the magnetization. The reason is, as it is easy to verify, that for any given $U\subseteq[n]$ the eigenvalue $C(\Gamma\cdot S,\ U)$ is maximized by the orbit of the singletons. So yet again we witness the extremal role played by the magnetisation

We note that one can express the conditional variance of transitive functions in a neat way using the basis $\{\xi_{\Gamma\cdot S}\ :\ \Gamma\cdot S\in\mathcal{O}\}$:

Proposition 3.2.15. *Let f be a transitive function and let $U\subseteq[n]$. Then*

$$\mathrm{Var}(\mathbb{E}[f\,|\,\mathcal{F}_U]) = \sum_{\Gamma\cdot S\in\mathcal{O}} C(\Gamma\cdot S,\ U)\widehat{f}^2(\Gamma\cdot S), \tag{3.2.15}$$

where the expectation and variance are taken according to the uniform measure.

Proof. For convenience, assume that $\mathrm{Var}(f)=1$. Now, using (1.1.7) we have

$$\begin{aligned}\mathrm{Var}(\mathbb{E}[f\,|\,\mathcal{F}_U]) &= \mathbb{P}[\mathscr{S}\subseteq U\,|\,\mathscr{S}\neq\emptyset]\\ &= \sum_{\Gamma\cdot S\in\mathcal{O}} \mathbb{P}[\mathscr{S}\in\Gamma\cdot S,\ \mathscr{S}\neq\emptyset]\mathbb{P}[\mathscr{S}\subseteq U\,|\,\mathscr{S}\in\Gamma\cdot S,\ \mathscr{S}\neq\emptyset].\end{aligned}$$

The first factor can be expressed in terms of the transitive basis:

$$\mathbb{P}[\mathscr{S}\in\Gamma\cdot S,\ \mathscr{S}\neq\emptyset] = \sum_{\emptyset\neq T\in\Gamma\cdot S}\widehat{f}(T)^2 = |\Gamma\cdot S|\widehat{f}(S)^2 = \widehat{f}(\Gamma\cdot S)^2.$$

Observing, that by the orbit counting lemma, $|\Gamma\cdot S||\Gamma_S| = |\Gamma|$, where Γ_S is the stabilizer of S, we obtain that

$$\mathbb{P}[\mathscr{S}\subseteq U\,|\,\mathscr{S}\in\Gamma\cdot S,\ \mathscr{S}\neq\emptyset] = \frac{|\{T\in\Gamma\cdot S\ :\ T\subseteq U\}|}{|\Gamma\cdot S|} = C(\Gamma\cdot S,\ U).$$

□

3.3 Factor of IID measures

3.3.1 Introduction

In this section we investigate sequences of spin systems that converge 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.

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 $X \sim \text{Unif}[0, 1]^V$ 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 almost surely exists a random coding radius $R < \infty$, for which it holds that $\psi(X_v)$ is determined by $\{X_u : u \in B_R(v)\}$, including the value of R.

Finitary factor of IID systems are measures that we can generate with a local algorithm that stops after a finite running time. The motivation to investigate these measures originally comes from ergodic theory, as finitary factor maps are in fact the continuous measure preserving maps between spin systems (see, for example [BS99]). At the same time as computational power 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 distributed way to sample from the distribution we want to understand. For finitary factor of IID measures (spin systems) thanks to the local algorithm, can be sampled by a distributed system from an IID measure. 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 over the vertices $u \in V(G)$ for which X_u needs to be revealed to learn σ_v, this condition is still not enough.

Therefore, those finitary factor if IID systems where the coding volume $\text{Vol}_v = |B_R(v)|$ (i.e, the number of uniform random variables one needs to reveal in order 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 no sparse reconstruction (or some of its variant) for a sequence μ_n converging to a Finitary factor of IID. In light of some negative results presented in Section 3.2.1 it is not clear that such results exist at all. Therefore, we narrow down the setup and introduce a stronger concept of convergence. In this section we only consider sequences of spin systems which themselves are factors of IID and, in particular, are generated by (a possibly truncated version of) the same local algorithm that we see in the limit.

For $n \in \mathbb{N}$ we generate the spin system σ^n as follows: For a vertex $v \in V$ we try to generate σ_v according to ϕ as a factor of IID, using independent uniformly distributed vertex labels on V. Let ρ_v be the largest r such that $B_v(r)$ is isomorphic to the r-ball on G. If for the coding radius R_v, we have $R_v \leq \rho_v$, we generate σ_v according to ϕ based on the labels. Otherwise, we use some alternative local algorithm that uses only the labels on vertices in $B_v(\rho_v)$. It is clear that as $n \to \infty$, the probability that any given vertex

σ_v is obtained with ϕ, tends to 1, thanks to the Benjamini–Schramm convergence. Also we note that $\{G_n, \mathbb{P}_n\}$, where $\mathbb{P}_n$ is the distribution of σ^n, converges to $(G, \mathbb{P})$ in the Benjamini–Schramm sense.

We first present a naive approach to argue that in case $\{(\mu_n, G_n) :\ n \in \mathbb{N}\}$ converges to a ffIID measure with finite expected coding volume, there should not be sparse reconstruction. The idea is that for each $u \in U_n$ one can learn the value of σ_v by asking at most Vol_v IID variables on average, which is a finite number. So we have to reveal at most $\mathrm{Vol}_v|U_n|$ IID variables in total which is much less then the total number of IID variables in the system, so by Theorem 2.1.7, there is no sparse reconstruction.

The reason why this idea cannot be directly implemented can be exemplified by Theorem 1.2.13, which tells us that a transitive function can be reconstructed from a low revealment set of coordinates in case the coordinates are queried iteratively in an adaptive way. This is indeed the case for the ffIID spins, where the subset of IID random variables are revealed according to a prescribed algorithm.

At the same time the setup here is quite different from Theorem 1.2.13, which suggests that from this adaptive set there will be no reconstruction. First, while the subset we query is adaptive, it is not our choice to decide which vertex to reveal next. Second, since the algorithm is localised, with high probability we will ask a deterministic sparse subset, the union of large enough balls around the vertices in U_n. In fact, as we shall see this naive idea is behind the proof of Theorem 3.3.8, which states, though under stronger conditions, that there is no SR for an ffIID measure with finite expected coding volume.

3.3.2 Convergence of susceptibility

In order to show that for ffIID measures there really is a connection beween the sequence and the limit, we have to show that the susceptibility converges. Let $S^+(f) := \sum_{g \in \mathsf{Aut}(G)} |\mathrm{Cov}(f, f^g)|$. The following result, which states that for a ffIID measure finite expected coding volume implies finite susceptibility, is an easy generalisation of Theorem 4.3 in [BS99].

Lemma 3.3.1. *Let X be a ffIID process with $|X_i| \le K$ almost surely on a transitive graph G of polynomial growth (that is, $|B_o(r)| \le Cr^d$ for some $d > 0$), and let R_o denote the coding radius of X_0. Suppose that the expected coding volume $\mathbb{E}[\mathrm{Vol}_0]$ is finite. Then*

$$S(X) \le S^+(X) \le 9K^2 3^d \mathbb{E}[\mathrm{Vol}]. \tag{3.3.1}$$

Proof. Let $\overline{X}_j := X_v - \mathbb{E}[X_v]$, the centered version of the original spin. For $v \in G$ let $d(o,v) = \max(|j_1|, |j_2| \ldots, |j_d|)$.

$$\begin{aligned}
&|\mathrm{Cov}(X_o, X_v)| \\
&\le \sum_{\max(k,l) \ge \frac{d(o,v)}{2} - 1} |\mathrm{Cov}(X_o \mathbb{1}_{R_o = k}, X_v \mathbb{1}_{R_v = l})| + \sum_{\max(k,l) < \frac{d(o,v)}{2} - 1} |\mathrm{Cov}(X_o \mathbb{1}_{R_o = k}, X_v \mathbb{1}_{R_v = l})| \\
&\le 4K^2 \sum_{\max(k,l) \ge \frac{d(o,v)}{2} - 1} \mathbb{P}[R_o = k, R_v = l] + \sum_{\max(k,l) < \frac{d(o,v)}{2} - 1} \mathbb{E}[\overline{X}_0 \mathbb{1}_{R_o = k}] \mathbb{E}[\overline{X}_j \mathbb{1}_{R_v = k}] \\
&\le 8K^2 \sum_{k \ge \frac{d(o,v)}{2} - 1} \mathbb{P}[R_o = k] + \left(\sum_{k=0}^{\frac{d(o,v)}{2} - 2} \mathbb{E}[\overline{X}_0 \mathbb{1}_{R_o = k}] \right)^2 .
\end{aligned}$$

Since $\overline{X}_0$ has 0 expected value

$$\left(\sum_{k=0}^{\frac{d(o,v)}{2}-2} \mathbb{E}[\overline{X}_0 \mathbb{1}_{R_o=k}]\right)^2 = \left(\sum_{k\geq \frac{d(o,v)}{2}-1} \mathbb{E}[\overline{X}_0 \mathbb{1}_{R_o=k}]\right)^2 \leq K^2 \sum_{k\geq \frac{d(o,v)}{2}-1} \mathbb{P}[R_o = k],$$

and therefore

$$\begin{aligned}|\mathrm{Cov}(X_o, X_v)| &\leq 9K^2 \sum_{k\geq \frac{d(o,v)}{2}-1} \mathbb{P}[R_o = k] \\ &= K^2 \sum_{k\,:\,2k+1\geq d(o,v)} \mathbb{P}[2R_o + 1 = 2k+1] \\ &\leq K^2 \sum_{i\geq d(o,v)} \mathbb{P}[2R_o + 1 = i].\end{aligned}$$

Summing this for all $v \in V$ we get that

$$\begin{aligned}&\sum_{v\in G} |\mathrm{Cov}(X_o, X_v)| \leq 9K^2 \sum_{v\in G} \sum_{i\geq d(o,v)} \mathbb{P}[2R_o + 1 = i] \\ &= 9K^2 \sum_i \sum_{v\in G:d(o,v)\leq i} \mathbb{P}[2R_o + 1 = i] \\ &= 9K^2 \sum_i |B_o(i)| \mathbb{P}[|B_o(2R_o+1)| = |B_o(i)|] \\ &= 9K^2 \mathbb{E}[|B_o(2R_o+1)|] \leq 9K^2 \mathbb{E}[|B_o(3R_o)|] \\ &\leq 9K^2 3^d \mathbb{E}[\mathrm{Vol}].\end{aligned}$$

□

Now we are ready to prove that under the same conditions the susceptibility of finitely supported functions converges.

Lemma 3.3.2. *Let σ be a ffIID spin system on a graph G distributed according to μ. Suppose that the sequence of spin systems $\{(\mu_n, G_n) : n \in \mathbb{N}\}$ converges to (μ, G) in the above, ffIID sense.*

Let $F \subseteq V(G)$ finite and let $f : \{-1,+1\}^F \to \mathbb{R}$. Then, there is a sequence of subsets $F_n \subseteq V_n$ with $|F_n| = |F|$ and a sequence $f_n : \{-1,+1\}^{F_n} \to \mathbb{R}$, such that $f_n \xrightarrow{w} f$. If $S(f) = \infty$, we have

$$\lim_{n\to\infty} S(f_n) = \infty.$$

Moreover, if $\mathbb{E}_\mu[\mathrm{Vol}_v] < \infty$ and G is of polynomial growth, then

$$\lim_{n\to\infty} S(f_n) = S(f).$$

Proof. First we define the sequence $F_n \subseteq V_n$ and the corresponding sequence of functions f_n. Choose a vertex $o \in F$ and large enough radius R so that $F \subseteq B_o(R)$. If n is large enough then the ball $B_o(R)$ in G and $B_{o'}(R)$ in G_n for some $o' \in V_n$ are isomorphic. We choose a subset $F_n \subseteq B_{o'}(R)$ as the image of such an isomorphism. Now f_n is just the composition of f and the isomorphism between the corresponding R-balls.

We continue by showing that for every $\gamma \in \mathsf{Aut}(G)$

$$\lim_{n\to\infty} \mathrm{Cov}_{\mu_n}(f, f^\gamma) = \mathrm{Cov}_\mu(f, f^\gamma).$$

First we note that if F is a finite subset in $V(G)$, then we can define the (random) coding radius R_F of F, as the smallest radius such that the union of R_F-balls around the elements of F determine the value of all the spins in F. Note that we have $R_F = \max_{v\in F} R_v \le |F|R_0$ (here R_0 is the coding radius of a spin).

Choose first N large enough so that $\mathbb{P}[\max(R_F, R_{F^\gamma}) > N] \le 2|F|\mathbb{P}[R_0 > N] < \epsilon/\max(f^2)$ and after choose $K \in \mathbb{N}$ such that $K \ge 2(N+R) + d(o, o^\gamma)$. (Recall that $o \in F$ and $F \subseteq B_o(R)$.) Now with probability $1 - \epsilon/\max(f^2)$ both f and f^γ can be computed from the uniform labels $\{X_u \ : \ u \in B_o(K)\}$. If now n is large enough so that the K-ball is isomorphic on G_n and on G, decomposing the covariance on μ_n with respect to $A := \{\max(R_F, R_{F^\gamma}) \le n\}$ we get

$$\mathrm{Cov}_{\mu_n}(f, f^\gamma) = \mathbb{P}[A]\mathrm{Cov}_\mu(f, f^\gamma|A) + \mathbb{P}[A^c]\mathrm{Cov}_{\mu_n}(f, f^\gamma|A^c),$$

and consequently, using that $\mathbb{P}[A^c] < \epsilon/\max(f^2)$ and that $\mathrm{Cov}(f, f^\gamma|A^c) \le \max(f^2)$ both according to μ and μ_n, we obtain that

$$|\mathrm{Cov}_{\mu^n}(f, f^\gamma) - \mathrm{Cov}_\mu(f, f^\gamma)| \le \mathbb{P}[A^c]2\max(f^2) \le 2\epsilon.$$

In case $S_\mu(f) = \infty$, then by Fatou's lemma, we have

$$S_\mu(f) = \sum_\gamma \lim_n \mathrm{Cov}_{\mu_n}(f, f^\gamma) \le \lim_n \sum_\gamma \mathrm{Cov}_{\mu_n}(f, f^\gamma) = \lim_n S_{\mu_n}(f) \tag{3.3.2}$$

so, $\lim_n S_{\mu_n}(f) = \infty$ as well.

For a subset $F \subseteq V$ we define $B_F(k) := \cup_{v\in F} B_v(k)$. Obviously, $|B_F(k)| \le |F||B_v(k)| = |F|\mathrm{Vol}_\sigma$. Now we can consider the spin system $\{f^\gamma \ : \ \gamma \in \Gamma\}$ and applying Lemma 3.3.1 we get that

$$S_\mu(f) \le C\mathbb{E}_\mu[\mathrm{Vol}_f] \le C|F|\mathbb{E}_\mu[\mathrm{Vol}_\sigma],$$

for some $C \in \mathbb{R}$. Observe that $\mathbb{E}_{\mu_n}[\mathrm{Vol}_{\sigma_n}] \le \mathbb{E}_\mu[\mathrm{Vol}_\sigma]$ for any $n \in \mathbb{N}$, since the algorithm generating μ_n queries at most as many vertex values as does μ, and therefore, applying Lemma 3.3.1 this time for the finite graph G_n, we have

$$S_{\mu_n}(f) \le C\mathbb{E}_{\mu_n}[\mathrm{Vol}_{\sigma_n}] \le C\mathbb{E}_\mu[\mathrm{Vol}_\sigma],$$

and thus in (3.3.2) we have equality by the dominated convergence theorem, hence the second statement. □

With this stronger notion of convergence, we finally have a few results where the limiting sequence decides whether there is sparse reconstruction or not:

Corollary 3.3.3. *If (μ, G) be an ffIID measure with $S_\mu = \infty$. Let $\{(\mu_n, G_n) : \ n \in \mathbb{N}\}$ be a sequence of ffIID spin systems converging to (μ, G) in the prescribed sense. Then there is weak sparse reconstruction for μ_n.*

Proof. By Lemma 3.3.2, this means that the absolute susceptibility in μ_n tends to ∞ as n goes to ∞. By Corollary 3.2.7, it follows that in this case there is sparse reconstruction for μ_n. □

Proposition 3.3.4. *Let $F \subseteq V(G)$ and f_n be a sequence of functions depending on the spins in F_n, where $|F_n| \leq M$ for every $n \in \mathbb{N}$, and assume that $\{(\mu_n, G_n) : n \in \mathbb{N}\}$ be a sequence of ffIID spin systems converging in the prescribed sense to (μ, G), an ffIID system with bounded expected coding radius. Then there is no random sparse reconstruction (see Definition 3.2.2) for f_n on μ_n.*

In particular, the proposition holds if $\mu_n = \mu$, where μ is an ffIID with finite expected coding radius.

Proof. First we note that if F is a finite subset in $V(G)$, then the (random) coding radius R_F of F, that is, the smallest radius such that the union of R_F-balls around the elements of F determine the value of the spins in F, has also finite expected value. Indeed, this follows from the trivial bound $R_F = \max_{v \in F} R_v \leq |F| R_0$ (here R_0 is the coding radius of a spin). Recall that $B_F(k) = \cup_{v \in F} B_v(k)$ and that clearly $|B_F(k)| \leq |F||B_v(k)|$. Therefore, taking into account that the coding radius on the finite graph is at most the coding radius in the infinite one, we get that $R_{F_n} \leq M R_0$ for every $n \in \mathbb{N}$ (recall that M is the universal bound on the size of F_n).

Let $\mathcal{H}_n$ be a random subset of V_n with revealment δ_n tending to 0. Let $Z_m = f_n(\sigma_n)$. For a change we will use information theoretic clue here but one can replace $I(Z_n, \mathcal{H}_n)$ with $\mathrm{Var}(\mathbb{E}[Z_n \,|\, \mathcal{F}_{U_n}])$ (and $H(Z_n)$ with $\mathrm{Var}(Z_m)$) and the argument works in the same way.

$$\begin{aligned}\mathbb{E}[I(Z_n, \mathcal{H}_n)] &= \sum_{H \subseteq V_n} \mathbb{P}[\mathcal{H}_n = H] I(Z_n, H) \\ &= \sum_{R_{F_n} = k} \sum_{H \subseteq V_n} \mathbb{P}[R_{F_n} = k, \mathcal{H}_n = H] \mathbb{1}_{\{B_F(k) \cap H \neq \emptyset\}} I(Z_n, H),\end{aligned}$$

where we used the fact that conditioned on the event $\{R_{F_n} = k,\ B_F(k) \cap H = \emptyset\}$, we have $I(Z_n, H) = 0$. Since $I(Z_n, H) \leq H(Z_n)$ for any $H \subseteq V_n$ and $\mathbb{P}[R_{F_n} = k, \mathcal{H}_n = H] = \mathbb{P}[R_{F_n} = k]\mathbb{P}[\mathcal{H}_n = H]$ we get

$$\begin{aligned}\mathbb{E}[I(Z_n, \mathcal{H}_n)] &\leq H(Z_n) \sum_{H \subseteq V_n} \mathbb{P}[\mathcal{H}_n = H] \left(\sum_{k \geq d(H,F)} \mathbb{P}[R_{F_n} = k] \right) \\ &= H(Z_n) \sum_r \mathbb{P}[d(\mathcal{H}_n, F) = r] \left(\sum_{k \geq r} \mathbb{P}[R_{F_n} = k] \right).\end{aligned}$$

At the same time, from the union bound and that the revealment of $\mathcal{H}_n$ is δ_n, we get the following estimate

$$\mathbb{P}[d(\mathcal{H}_n, F) = r] = \mathbb{P}[\exists h \in \mathcal{H}_n \ :\ h \in B_F(r)] \leq \sum_{v \in B_F(r)} \mathbb{P}[v \in \mathcal{H}_n] \leq \delta_n |B_F(r)|. \quad (3.3.3)$$

Since $\mathbb{E}[R_{F_n}] < \infty$, one can choose a large enough $L \in \mathbb{N}$ such that

$$\sum_{k \geq L} \mathbb{P}[R_{F_n} = k] \leq M\ \mathbb{E}[R_v \mathbb{1}_{\{R_v \geq L\}}] < \epsilon.$$

Now we split the sum above according to L:

$$\begin{aligned}
&\sum_r \mathbb{P}[d(\mathcal{H}_n, F) = r]\left(\sum_{k\geq r}\mathbb{P}[R_{F_n} = k]\right) \\
&\leq \sum_{r=1}^{L}\mathbb{P}[d(\mathcal{H}_n, F) = r]\mathbb{E}[R_{F_n}] + \sum_{r=L+1}^{\infty}\mathbb{P}[d(\mathcal{H}_n, F) = r]\epsilon \\
&\leq \delta_n \sum_{r=1}^{L}|B_F(r)| + \epsilon,
\end{aligned}$$

where we used that (3.3.3) to bound the first term and that trivially $\sum_{r=L+1}^{\infty}\mathbb{P}[d(\mathcal{H}_n, F) = r] \leq 1$ for the second term. Since $\delta_n \to 0$, for large enough n the first term will be at most ϵ as well. So,

$$\frac{\mathbb{E}[I(Z_n, \mathcal{H}_n)]}{H(Z_n)} \leq \frac{2\epsilon H(Z_n)}{H(Z_n)} = 2\epsilon$$

for every large n. □

3.3.3 Finite expected coding volume and magnetisation

In this section we are going to show that the magnetization cannot be reconstructed from an ffIID sequence converging to an ffIID measure with finite expected coding volume, whenever $S(f_n) \neq 0$. We will use Lemma 3.3.1 to show that in that case the magnetization is close to the magnetization of a block factor of IID system, but the latter admits no SR.

If we have a ffIID system σ, for a positive integer L one can consider an L block factor of IID σ^L that approximates (the distribution of) σ as follows: First, $\sigma_v^L = \sigma_v$ whenever the local algorithm ψ generating σ stops before going outside from the ball $B_v(L)$. Otherwise, we sample σ_v^L according to the distribution of σ_v conditioned on $B_v(L)$ independent from everything outside of $B_v(L)$. This block factor can be put on a finite graph G_n approximating G in the same way as the original ffIID.

Lemma 3.3.5. *Let X be a finitary factor of an IID spin system with finite expected coding volume and $|X_i| \leq K$ a.s. on a transitive graph G of polynomial growth. Let R_0 denote the coding radius of X_0. For $L \in \mathbb{N}$ let X^L be the L block factor approximation of X. Let $E^L := X - X^L$ (which is obviously also a ffIID spin system). Then*

$$S(E^L) \leq C\mathbb{E}[\mathrm{Vol}\mathbb{1}_{R_0>L}] \tag{3.3.4}$$

where the constant C depends only on K and the dimension d.

Proof. Observe that $\mathrm{Cov}(E_0^L, E_j^L) = 0$ whenever $\min(R_0, R_v) \leq R$, since $E_j^L \equiv 0$ whenever $R_v \leq R$. After this straightforward observation we just repeat the calculation as in Lemma 3.3.1.

We can differentiate two cases. If $L < \frac{d(o,v)}{2} - 1$ then we simply copy the calculation of Lemma 3.3.1. In case $L \geq \frac{d(o,v)}{2} - 1$ we have

$$\mathrm{Cov}(E_0^L, E_j^L) = \sum_{\max(k,l)>L}\mathrm{Cov}(E_0^L\mathbb{1}_{R_0=k}, E_j^L\mathbb{1}_{R_v=k}) \leq 9K^2\sum_{k>L}\mathbb{P}[R_0 = k].$$

Therefore, upon summing the covariances we get

$$\begin{aligned}
&\sum_{v\in G}\mathrm{Cov}(E_0^L, E_j^L)\\
&\le 9\sum_{v\in G}\sum_{k\ge\max(L+1,\frac{d(o,v)}{2}-1)}^{\infty}\mathbb{P}[R_0=k]\le 9K^2\mathbb{E}[|B_o(2R_o+1)|\mathbb{1}_{R_0>L}]\\
&\le 9K^2\mathbb{E}[|B_o(3R_o)|]\le 9K^2 3^d\mathbb{E}[\mathrm{Vol}\mathbb{1}_{R_0>L}].
\end{aligned}$$

□

We shall use the following estimate from [GPS10], which we present without proof.

Lemma 3.3.6. *[GPS] If* $\epsilon\max(\|f\|,\|g\|)\ge\|f-g\|$ *then*

$$\sum_{S\subseteq[N]}\left|\frac{\widehat{f}(S)^2}{\|f\|^2}-\frac{\widehat{g}(S)^2}{\|g\|^2}\right|\le\frac{4\epsilon}{(1-\epsilon)^2} \tag{3.3.5}$$

Proposition 3.3.7. *Let G be a graph of polynomial growth and (μ, G) a finitary factor of i.i.d. measure with finite expected coding volume. Let $\{(\mu_n, G_n)\}$ be a sequence of ffIID spin systems converging to (μ, G). Let X distributed according to μ. If X_v is almost surely bounded and $S_\mu\ne 0$ then there is no sparse reconstruction for the magnetization on μ_n.*

Proof. Let $\{X_n\}$ be a sequence of spin systems distributed according to μ_n and let M_n denote the corresponding magnetization. Recall that $\mathrm{Var}(M_n)=S_{\mu_n}/n$. Observe that for every large $n\in\mathbb{N}$ we have $c\le S_{\mu_n}\le C$ for some positive constants c and C. Indeed, on the one hand, $S_\mu\ne 0$ and from Lemma 3.3.1 we know that finite expected coding volume implies finite susceptibility. So $0<S_mu<\infty$. On the other hand, Lemma 3.3.2 states that $\lim_n S_{\mu_n}=S_\mu$. Consequently,

$$\frac{b}{n}\le\mathrm{Var}(M_n)\le\frac{B}{n}.$$

Let us denote by X_n^L the L-block factor approximation of X_n and with M_n^L the corresponding magnetization. Obviously, a similar lower and upper bound holds for $\mathrm{Var}(M_n^L)$ as well.

Since X has finite expected coding volume, by Lemma 3.3.5, for every ϵ there is an L such that for every large n we have $S(E^L)\le b\epsilon$, where $E^L=X-X^L$. Also let us define $E_n^L=X_n-X_n^L$ the version of E^L on G_n. According to Lemma 3.3.2 $S(E_n^L)\longrightarrow S(E^L)$, so $S(E_n^L)\le 2b\epsilon$ for every large n. Therefore, using the estimate for $\mathrm{Var}(M_n[E^L])$ in the proof of Lemma 2.5.2 we get

$$\sqrt{\mathrm{Var}(M_n)\mathrm{Var}(M_n^L)}2(1-\mathrm{Corr}(M_n,M_n^L))\le\mathrm{Var}(M_n-M_n^L)=\frac{S(E_n^L)}{n}\le\frac{2b\epsilon}{n}.$$

Recall that $\mathrm{Var}(M_n)$ and $\mathrm{Var}(M_n^L)$ are both at least $\frac{b}{n}$, so we get that

$$1-\mathrm{Corr}(M_n,M_n^L)\le\epsilon \tag{3.3.6}$$

Indirectly, suppose that there is a sequence of subsets $U_n\subseteq V_n$ with $|U_n|=o(|V_n|)$ such that $\mathsf{clue}(M_n\mid U_n)\ge c$ for some $c>0$ for all large n. Then Lemma 2.5.3 would imply that $\mathsf{clue}(M_n^L \mid U_n) \geq c - 2\epsilon$. But if we choose L large enough so that $\epsilon < c/4$ small, we get that

$$\mathsf{clue}(M_n^L \mid U_n) \geq c/2 > 0.$$

But this would mean that on the sequence of spin systems X_n^L there is sparse reconstruction for the magnetization. Nevertheless, this is impossible: X^L is a block-factor, which means that independently from n every spin X_v can be computed almost surely from the IID variables in $B_v(L)$ the ball of radius L around v. For every large n the cardinality of $B_v(L)$ is a constant: $K := |B_v(L)|$. A transitive function of X_n^L is also a transitive function of the IID variables and the spin values of U_n can be always computed if we know IID variables in $B_u(R)$ for every $u \in U_n$, which is at most $K|U_n|$ variables. So this would imply that for a sequence of product measures there is sparse reconstruction for a transitive function from a sequence of subsets of $K|U_n| = o(|V_n|)$ variables, which contradicts Theorem 2.1.7.

We sketch a different argument which might be also of interest. We can think of the above functions as functions of the i.i.d bits instead of that of the factor. It is clear that the size of the Fourier spectrum of M_n^L is at most R. According to Lemma 3.3.6, together with (3.3.6) shows that at least $1 - O(\epsilon)$ proportion of the Fourier energy of M_n is concentrated on sets of size at most R as well.

Now take a random translate U^k of U, which because of symmetry contains the same information as U and query all the bits which are necessary to learn the spins in U^k. This algorithm is symmetric on the bits and it queries on average at most $|U|\mathbb{E}[\mathrm{Vol}]$ bits, where Vol is the coding volume. If we denote by J the random set of queried bits we have:

$$n\mathbb{P}[i \in J] = \mathbb{E}[d(o,v)] \leq |U|\mathbb{E}[\mathrm{Vol}_0] = o(n) \tag{3.3.7}$$

By the Revealment Theorem (Theorem 1.2.13), for every fixed k

$$\sum_{|S|=k} \frac{\widehat{M_n}^2(S)}{\|M_n\|^2} \leq \frac{k|U|\mathbb{E}[\mathrm{Vol}_0]}{n} \to 0. \tag{3.3.8}$$

But this is in contradiction with the fact that a large part of the Fourier energy is concentrated on bounded degree (below L) . □

It is not difficult to extend this result. Take a sequence of uniformly bounded functions f_n with uniformly bounded support F_n. Without much effort one can prove that in case $\liminf S(f_n) > 0$ there is no SR for the sequence $M[f_n]$. The additional requirement $\liminf S(f_n) \neq 0$ is slightly disturbing, but up to now we could not find a clear way to eliminate.

3.3.4 A partial no SR result

We know from Corollary 3.3.3 that if the susceptibility for a factor of IID is infinite, then there is WSR. At the same time Lemma 3.3.1 says (using Lemma 3.3.2) that if the expected coding volume is finite then the susceptibility cannot grow to infinity and thus as we have just learned, there is no SR for the magnetisation (unless the susceptibility is 0) . But does this mean that there is no SR at all in this case? The following theorem, which relies on the IID case, gives a partial answer to this question.

Theorem 3.3.8. *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). Let $f_n : \{-1,1\}^{V_n} \longrightarrow \{-1,1\}$ be sequence of transitive, non-degenerate Boolean functions.*

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

$$\mathbb{P}[\forall\, u \in U_n \ :\ R_u < r] = 1 - \mathbb{P}[\exists\, u \in U_n \ :\ R_u \geq r] > 1 - |U_n| \exp(-cr)$$

By definition, whenever $\{\forall\, u \in U_n :\ R_u < r\}$ 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 $J_n(r) = \bigcup_{u \in U_n} B_r(u)$
Let us choose a sequence of integers r_n in such a way that

$$|J_n(r)| \leq |U_n| r_n^d \ll |V_n| = n^d, \tag{3.3.9}$$

and at the same time

$$\lim_n 1 - |U_n| \exp(-cr_n) = 1. \tag{3.3.10}$$

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 2.1.7 using the condition (3.3.9) that $\mathsf{clue}_{\mathrm{Unif}}\left(g_n \mid \mathcal{F}_{J_n(r)(X)}\right) \to 0$.

On the other hand, conditioned on the event $\{\forall\, u \in U_n :\ R_u < r\}$,

$$\mathbb{E}[f_n | \mathcal{F}_{U_n(\sigma)}] = \mathbb{E}_{\mu_n}[g_n | \mathcal{F}_{J_n(r)(X)}] \tag{3.3.11}$$

and by (3.3.10) this happens with high probability.

Observe that if one chooses $r_n = K \log n^d$ with sufficiently large K, (3.3.9) and (3.3.10) 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 $J_n(r)$ and the random set of uniform variables necessary to compute the spins of U_n are measurable. So $\mathbb{E}[g_n \mid \mathcal{G}] = \mathbb{E}[f_n \mid \mathcal{F}_{U_n(\sigma)}]$. Since $\mathcal{F}_{J_n(r)} \subseteq \mathcal{G}$ by the definition of $\mathcal{G}$, we have, by Pythagoras's Theorem

$$\left\|\mathbb{E}[g_n \mid \mathcal{G}]\right\|^2 = \left\|\mathbb{E}[g_n \mid \mathcal{F}_{J_n(r)}]\right\|^2 + \left\|\mathbb{E}[g_n \mid \mathcal{G}] - \mathbb{E}[g_n \mid \mathcal{F}_{J_n(r)(X)}]\right\|^2.$$

Subtracting the common squared expectation we get

$$\mathrm{Var}\left(\mathbb{E}[g_n \mid \mathcal{G}]\right) = \mathrm{Var}\left(\mathbb{E}[g_n \mid \mathcal{F}_{J_n(r)}]\right) + \left\|\mathbb{E}[g_n \mid \mathcal{G}] - \mathbb{E}[g_n \mid \omega_{V_n}]\right\|^2. \tag{3.3.12}$$

By Theorem 2.1.7, $\mathsf{clue}(g_n \mid \mathcal{F}_{J_n(r)}) \to 0$ and thus $\mathrm{Var}\left(\mathbb{E}[g \mid \mathcal{F}_{J_n(r)}]\right) \to 0$ as well, while the second term is smaller then $2\mathbb{P}[\mathbb{E}[f_n|\mathcal{F}_{U_n(\sigma)}] \neq \mathbb{E}[g_n|\mathcal{F}_{J_n(X)}]\ ]$ which tends to 0 by (3.3.11).

We should point out that it is at this point we use the fact that f_n, and thus g_n is a non-degenerate sequence. Indeed, if $\mathrm{Var}(g_n) \to 0$ then $\mathrm{Var}\left(\mathbb{E}[g_n \mid \mathcal{G}]\right) = o(\mathrm{Var}(g_n))$ does not follow from the fact that the second term in (3.3.12) goes to 0. $\square$

The reader might wonder whether this result can be improved. Three natural direction comes into mind. First, can we write $o(n^d)$ 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 (for instance, with 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: Read the bits starting from 0 going to the right along the x-axis, until one finds two consecutive bits with equal value. This value will be the spin σ_0 that we write at 0.

On the one hand this simple ffIID system admits a finite expected coding volume. It also have a special structure, one can quickly verify, that, almost surely, the string $-+-$ and $+-+$ is not contained in a configuration. Let us define the same rule on finite cycles, and in case the algorithm does not stop inside the cycle, then set $\sigma_0 = \omega_0$.

Observe that on $\mathbb{Z}_n$ one can reconstruct the event $\{R_0 > n\}$ from 3 consecutive bits. Indeed, we can see $-+-$ or $+-+$ if and only if $\{R_0 > n\}$ holds. Of course, this is a highly degenerate sequence of events, as the probability is exponentially small.

However, one can tweak this example to reconstruct a non-degenerate sequence that can be reconstructed. Define the same algorithm on $\mathbb{Z}^2$. So the algorithm only asking bits to the right direction on the x-axis. Now let's take a sequence of graphs $G_n = \mathbb{Z}_n \times \mathbb{Z}_{2^n}$ and put the same ffIID measure on it. On each copy $\mathbb{Z}_n$ the $\mathbb{P}[R_y > n] = 2^{-n}$, so the expected number of y coordinates on which this happens in 1, and therefore the sequence of events $\{\exists y : \ R_y > n\}$ is non-degenerate and it can be reconstructed from $3n$ spins. Ironically, the density of this subset is not much lower then the upper bound $n^2/\log^2 n$ in Theorem 3.3.8, so roughly speaking Theorem 3.3.8 turns out to be sharp.

Still, this example feels unsatisfactory, because it undermines an idea implicit in our project. We would like to have or not have sparse reconstruction based on the properties of the limiting ffIID measure, which is meant to concentrate all relevant information from a consistent sequence of finite spin systems. In this case, however, it is the very error of the approximation that we can reconstruct. Therefore we suggest a strengthening of the concept of convergence for ffIID spin systems.

For a sequence of ffIID spin systems $\{(\mathbb{P}_n, G_n) \ : \ n \in \mathbb{N}\}$ let ρ_n be the largest radius such that for every $v \in V_n$ the ball $B_v(\rho_n)$ is isomorphic to the corresponding ρ_n-ball on G. We say that a sequence of ffIID spin systems $(\mathbb{P}_n, G_n)$ converges to the ffIID $(\mathbb{P}, G)$ regularly if the sequence of approximations $\mathbb{P}_n$ satisfies

$$\lim_{n\to\infty} \mathbb{P}[\exists v \in V_n \ : \ R_v > \rho_n] = 0.$$

That is, we require that the approximation is regular enough so that the probability of any error to occur should go to 0. Using this condition, we can restate the main question concerning ffIID measures:

Question 3.3.9. *Suppose that a sequence of ffIID spin systems $(\mathbb{P}_n, G_n)$ converges to the ffIID $(\mathbb{P}, G)$ regularly, where $(\mathbb{P}, G)$ has finite expected coding volume. Is it true that there is no sparse reconstruction on $(\mathbb{P}_n, G_n)$?*

3.4 Sparse Reconstruction for the Ising Model

3.4.1 A short introduction to the Ising model

The Ising-model is one of the simplest and certainly the best understood and investigated model in Statistical Physics (for an introduction see [Pet] Section 13.1 or [T11]). It models a magnetic field, in which particles can have positive (+1) or negative (−1) charges randomly, but particles that are close to each other have a tendency to have the same (or in the antiferromagnetic case, the opposite) charge . Formally, there is an underlying finite graph G, which describes the geometry of being close. There is an energy function of the form $-\beta H(\sigma)$, on the configuration space $\{-1,1\}^V$ called the Hamiltonian, which is inversely proportional with the probability :

$$\mu_\beta[\sigma] := \frac{1}{Z_\beta} e^{-\beta H(\sigma)}.$$

Here $\beta > 0$ is interpreted as the inverse temperature and the partition function Z_β is a normalization factor that makes μ a probability measure. The Hamiltonian H : $\{-1,1\}^{V_n} \longrightarrow \mathbb{R}$ is defined as:

$$H(\sigma) = -J \sum_{(x,y)\in E} \sigma_x \sigma_y - h \sum_{x \in V} \sigma_x,$$

where the parameter J describes the strength of the interaction, while h is the external magnetic field.

One of the most interesting features of the Ising model is that on many graph sequences it exhibits phase transition (similarly to the Bernoulli percolation model), that is, there are some critical inverse temperature values (denoted by β_v), where the behavior of the model changes radically.

The phase transition, however, can be clearly observed only in on infinite model. It turns out that the Ising model can be defined as a weak limit of finite models on certain infinite graphs. The laws of finite models, however, may differ according to the boundary conditions. One of the important features of the phase transition is, that typically (in particular on $\mathbb{Z}^d$) for $\beta > \beta_c$, there are more than one possible Ising limiting measures. On $\mathbb{Z}^d$ for example, if one takes the weak limit of finite Ising models on the discrete hypercubes $[-n,n]^d$ with + boundary conditions we get a different limiting measure than with − boundary conditions. This is not the case for $\beta \leq \beta_c$, where there is a unique limiting Ising measure.

Another remarkable fact is that at the critical temperature the susceptibility of the model blows up. Therefore high temperature models have finite, while low temperature ones have infinite susceptibility. As we already mentioned, this concept has a special significance in the Ising model thanks to the formula

$$\beta S = \frac{\partial M}{\partial h}, \tag{3.4.1}$$

where M is the magnetization. The interpretation is, and hence the name, that it expresses the reactivity of the magnetization to an external magnetic field. This is in accordance with the two different measures: as the susceptibility becomes infinite, an arbitrary small external field can determine whether the model will turn out to be in the + measure or in the − one (this phenomenon is called spontaneous magnetization).

It is an important fact that the Ising model can be also described via a coupling with a percolation model. This is referred to as the FK–Ising model (FK stands for Fortuin and Kasteleyn), or more generally, the random cluster model. On a finite 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.

It turns out that if one assigns an independent fair coin flip for each connectivity cluster of the resulting random graph, the law of the spin system will be that of an Ising model with inverse temperature β, where $p = 1 - e^{-\beta J}$. This connection often proves to be useful and some concepts of the Ising model gain a new interpretation in the random cluster model. For instance, as one can easily verify, the susceptibility of the Ising model is just the expected cluster size of a uniformly chosen vertex.

The random cluster model can also be defined for infinite graphs via a weak limit and the above correspondence with the Ising model remains true as well. The phase transition can also be interpreted in terms of the random cluster model: the critical inverse temperature β_c is the critical point of the corresponding percolation, i.e., the point where an infinite open cluster is created.

3.4.2 Results for the Ising model on $\mathbb{Z}_n^d$

In this section we collect the results we have about sparse reconstruction on the Ising models. We start by the case of low temperature models, which is the most straightforward to handle.

Proposition 3.4.1. *Let $\{\sigma^n\}$ be the sequence of Ising models on $\mathbb{Z}_n^d$. Then the* Maj_n *sequence of Boolean function can be reconstructed from a sequence of sparse subsets in case $\beta > \beta_c$ (low temperature).*

Proof. It can be shown, that when $\beta > \beta_c$, then the sequence $\{\sigma^n\}$ on the torus converges to the Ising measure $\frac{1}{2}\mu_+ + \frac{1}{2}\mu$ on $\mathbb{Z}^d$, where μ_+ and μ_- are the weak limits of Ising measures on d-dimensional hypercubes with $+$ and $-$ boundary conditions, respectively. Thus the limiting measure is a convex combination of two measures, thus non-ergodic. This in itself is not enough though to conclude the existence of sparse reconstruction (see the discussion before Proposition 3.2.3).

Let us consider the coupling of σ_n with the corresponding random cluster model. Since the random cluster model is also supercritical, there is an infinite cluster almost surely in the limiting infinite random cluster measure. Moreover, it was shown in [Pi96] (Theorem 1.1) that with high probability there is one giant cluster of size at least Cn^d for some C while the remaining clusters are much smaller, of size $o(n^d)$. (In fact, this was only proven for β below the so-called slab threshold for the FK model when $d > 2$, but it was shown in [Bo05] that this coincides with the critical β.)

Switching back to the Ising model this shows that Maj_n is determined by the bit assigned to the giant cluster, with high probability. So on $\mathbb{Z}_n^d$, $|M_n| > Cn^d$, with high probability and therefore, by Proposition 3.2.8, the majority can be reconstructed from a sparse random set. It is not difficult to show that Maj_n can be reconstructed from a deterministic sparse sequence of spins as well. □

As for the high temperature models, our result is based on the fact that the high temperature Ising turn out to be ffIID measure (see [BS99]).

Theorem 3.4.2 (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 $\mathbb{P}[R > t] < \exp(-ct)$.*

This allows us to use the results of Section 3.3, in particular Theorem 3.3.8 which holds exactly for ffIID spin systems with exponential decay of the coding radius. So, we get as a corollary of Theorem 3.3.8:

Theorem 3.4.3. *Let $\{\mu_n\}$ be a sequence of subcritical Ising measures (i.e., $\beta < \beta_c$) on the tori $\mathbb{Z}_n^d$. Then there is no reconstruction on $\{\mu_n\}$ from sequences of subsets $U_n \subseteq \mathbb{Z}_n^d$ with $|U_n| = o(n^d/\log^d n)$. Furthermore, there is no sparse reconstruction for the Maj_n.*

Proof. The first claim is a direct consequence of Theorem 3.3.8, the second one follows from Proposition 3.3.7. □

For critical models, however, sparse reconstruction is possible.

Proposition 3.4.4 (Sparse Reconstruction for Critical Ising from random subsets). *Let $\{\mu_n\}$ be a sequence of critical Ising measures (i.e., $\beta = \beta_c$) on the tori $\mathbb{Z}_n^d$. Then the magnetization and the function Maj_n can be reconstructed with high* clue *from a sparse random subset of spins with a uniformly positive probability.*

Proof. According to [Si80], we have $\lim_{n\to\infty} S(\sigma_n^f) = \infty$ at $\beta = \beta_c$ where $sigma_n^f$ is the Ising model on $[-n,n]^d$ with free boundary conditions (that is, no magnetization on the boundaries of the square). In order to transfer this result to the torus, we need the fact that there is only one infinite Ising measure at critical temperature. This was shown in [Y52] for $d = 2$ first, in [AF86] for $d > 3$, and recently in [ADS13] for $d = 3$.

This means that the sequence of Ising models on the torus converges weakly to the same measure as the free one, which satisfies $S(\sigma_n^f) \to \infty$. This implies that the respective covariances converge to the same value as well. Since all covariances are non-negative we can apply the dominated convergence theorem, and thus we obtain that $S(\sigma_n) \to \infty$ as well. Therefore, by Corollary 3.2.7, the magnetization can be reconstructed from a sparse random subset of spins.

We are going to show that

$$E[M_n^4] \leq 3E[M_n^2]^2.$$

Indeed, then the Paley-Zygmund inequality implies that for arbitrary $\epsilon > 0$

$$\mathbb{E}[|M_n| > \epsilon\mathbb{E}[M_n^2]] = \mathbb{E}[M_n^2 > \epsilon^2\mathbb{E}[M_n^2]] \geq \frac{(1-\epsilon^2)^2}{3},$$

and by Proposition 3.2.8 with probability $1/3 - \epsilon^2$ we have

$$\mathsf{clue}(\mathsf{Maj}_n \mid \mathcal{B}^{p_n}) > 0.99$$

for every large n. First, the so-called Lebowitz inequality (or Gaussian bound, see [Le74]) says that for any $x, y, u, v \in \mathbb{Z}_n^d$ (in fact, on finite graphs in general)

$$\mathbb{E}[\sigma_x\sigma_y\sigma_u\sigma_v] \leq \mathbb{E}[\sigma_x\sigma_y]\mathbb{E}[\sigma_u\sigma_v] + \mathbb{E}[\sigma_x\sigma_u]\mathbb{E}[\sigma_y\sigma_v] + \mathbb{E}[\sigma_x\sigma_v]\mathbb{E}[\sigma_u\sigma_y].$$

We can sum the above inequalities for all possible quadruples $x, y, u, v \in \mathbb{Z}_n^d$, and thus we get that

$$E[M_n^4] \leq 3E[M_{2n}^2]^2,$$

as promised. □

Using some state-of-the-art results on two dimensional critical Ising model, we can say something stronger on $\mathbb{Z}_n^2$.

Theorem 3.4.5 (Sparse Reconstruction for Critical Ising on $\mathbb{Z}_n^2$)**.** *At $\beta = \beta_c$ on $\mathbb{Z}_n^2$, the total magnetization M_n can be reconstructed with high clue from a sublattice H^{s_n} of grid size s_n as long as $s_n = o(n^{\frac{7}{4}})$. That is,*

$$\lim_{n\to\infty} \mathsf{clue}(M_n \mid H^{s_n}) = 1.$$

Moreover,

$$\lim_{n\to\infty} \mathsf{clue}(\mathsf{Maj}_n \mid H^{s_n}) = 1.$$

Proof. It was shown in [CHI15] that

$$\mathbb{E}\sigma(x)\sigma(y) \to c\,\|x-y\|^{-\frac{1}{4}}$$

for some $c > 0$ (here $\|\ \|$ stands for the Euclidean norm), and this allows us to compute the correlation between the magnetisation M_n and the sparse magnetisation $M_n^{s_n}$ on the s_n-lattice, up to factors of size $o(1)$.

We consider a sequence of sublattices of the original square $[-n,n]^2$. Such a lattice H^{s_n} can be either described by the size s_n of the lattice or by number of lattice points along a line segment of length n, which we denote by k_n. We have the relation $n \le k_n s_n < n + s_n$, we allow some overlap to make sure that there is a vertex in every interval of length s_n. We will use the shorthand notation $M_n^{s_n}$ for the magnetization of H^{s_n}.

Just like in the proof of Theorem 3.2.6, we have $\mathrm{Var}(M_n) = \mathrm{Cov}(M_n, M_n^{s_n}) = S_n/n^2$, and therefore

$$\mathrm{Corr}(M_n, M_n^{s_n}) = \sqrt{\frac{S_n}{n^2 \mathrm{Var}(M_n^{s_n})}}$$

We use a standard trick: We divide the square $\mathbb{Z}_n^2$ into logarithmically increasing annuli with respect to Euclidean distance. Let us define two sequences: $\eta_n > 0$ with $\eta_n \to 0$ and $0 < \delta_n \to \infty$, but $\delta = o(n)$. We will estimate the susceptibility S_n on every concentric annulus between the radii $\delta_n(1+\eta_n)^k$ and $\delta_n(1+\eta_n)^{k+1}$ and the central circle of radius $\delta_n n$. Since we need a lower bound for S_n, we are going to estimate $\mathbb{E}[\sigma(0)\sigma(x)]$ for all x in the kth annulus by $c\left(\delta_n(1+\eta_n)^{k+1}\right)^{-\frac{1}{4}}$. Thus we have

$$\begin{aligned}
&\sum_{x\in\mathbb{Z}_n^2:\ \|x\|\le n} \mathbb{E}[\sigma(0)\sigma(x)] \\
&\ge 1 + c\pi\delta_n^{\frac{7}{4}} + \sum_{k=0}^{\log_{\eta_n}(n/\delta_n)-1} \pi\delta_n^2\left((1+\eta_n)^{2(k+1)} - (1+\eta_n)^{2k}\right) c\delta_n^{-\frac{1}{4}}(1+\eta_n)^{-\frac{k+1}{4}} \\
&= 1 + c\pi\delta_n^{\frac{7}{4}} + c\pi\delta_n^{\frac{7}{4}}\frac{\eta_n(2+\eta_n)}{(1+\eta_n)^{\frac{1}{4}}} \sum_{k=1}^{\log_{\eta_n}(n/\delta_n)-1} (1+\eta_n)^{\frac{7k}{4}} \\
&= 1 + c\pi\delta_n^{\frac{7}{4}} + c\pi\delta_n^{\frac{7}{4}}\frac{\eta_n(2+\eta_n)}{(1+\eta_n)^{\frac{1}{4}}}\left(\frac{n}{\delta_n}\right)^{\frac{7}{4}}\frac{1}{\eta_n} \\
&= 1 + c\pi\delta_n^{\frac{7}{4}} + c\pi\frac{(2+\eta_n)}{(1+\eta_n)^{\frac{1}{4}}}n^{\frac{7}{4}}.
\end{aligned}$$

We need to estimate the sum of covariances between the spin at $(0,0)$ with any other spin in the square $[-n,n]^2$. The estimation above, however, works only for those spins which are in the circle of radius n. To get around this problem, we shall take a large $K \in \mathbb{N}$, and approximate the square from below (and, in the sequel, from above) with the union of K pieces of circular sector of equal central angle of $2\pi/k$, but of changing radii. For a fixed central angular sector we choose its radius as the maximal radius such that the respective circular sector is inside the square. We shall denote by $r_{n,l}$ this radius for the lth sector, where $l \in [K]$. In turn, each annulus is cut into k sectors and thus, when summing up the annular sectors belonging to the same angular sector, we get just the same sum as above, with a constant factor of $1/K$ and with differing upper bounds of summation for every sector:

$$\begin{aligned} S_n &= \sum_{x \in \mathbb{Z}_n^2} \mathbb{E}[\sigma(0)\sigma(x)] \\ &\geq 1 + c\pi\delta_n^{\frac{7}{4}} + c\pi \frac{(2+\eta_n)}{(1+\eta_n)^{\frac{1}{4}}} \frac{1}{K} \sum_{l=1}^{K} \sum_{k=1}^{\log_{\eta_n}(r_{n,l}/\delta_n)-1} (1+\eta_n)^{\frac{7k}{4}} \\ &= 1 + c\pi\delta_n^{\frac{7}{4}} + c\pi \frac{(2+\eta_n)}{(1+\eta_n)^{\frac{1}{4}}} \frac{1}{K} \sum_{l=1}^{K} r_{n,l}^{\frac{7}{4}}. \end{aligned}$$

We turn to estimating the denominator. Observe that

$$\mathrm{Var}(M_n^{s_n}) = \frac{1}{k_n^2} \sum_{x \in H^{s_n}} \mathbb{E}[\sigma(0)\sigma(x)].$$

So we only need to deal with the "sparse susceptibility". We are going to apply the same strategy as above for S_n. In order to make sure that the estimation for the number of vertices of H^{s_n} in a particular annulus is accurate, we set the condition $s_n \ll \eta_n\delta_n \ll \delta_n$. Otherwise the only difference is that now we give an upper bound and therefore we are going to estimate $\mathbb{E}[\sigma(0)\sigma(x)]$ for all x in the kth annulus by $c\left(\delta_n(1+\eta_n)^k\right)^{-\frac{1}{4}}$. We also split the sum into K terms according to sectors as before. Again, the upper bound of summation is differ according to sector: For a fixed central angular sector we sum until the minimal radius such that the intersection of the square and the respective angular sector falls in to the circular sector. We denote this radius by $R_{n,l}$ for $l \in [K]$. So we get

$$\begin{aligned} &\sum_{x \in H^{s_n}} \mathbb{E}[\sigma(0)\sigma(x)] \\ &\leq 1 + c\pi\left(\frac{\delta_n}{s_n}\right)^2 + \frac{1}{K}\sum_{l=1}^{K} \sum_{k=0}^{\log_{\eta_n}(R_{n,l}/\delta_n)-1} \pi\left(\frac{\delta_n}{s_n}\right)^2 \left((1+\eta_n)^{2(k+1)} - (1+\eta_n)^{2k}\right) c\delta_n^{-\frac{1}{4}}(1+\eta_n)^{-\frac{k}{4}} \\ &= 1 + c\pi\delta_n^2 s_n^{-2} + c\pi\delta_n^{\frac{7}{4}} s_n^{-2}\eta_n(2+\eta_n)\frac{1}{K}\sum_{l=1}^{K} \sum_{k=1}^{\log_{\eta_n}(R_{n,l}/\delta_n)-1} (1+\eta_n)^{\frac{7k}{4}} \\ &= 1 + c\pi\delta_n^2 s_n^{-2} + c\pi(2+\eta_n)s_n^{-2}\frac{1}{K}\sum_{l=1}^{K} R_{n,l}^{\frac{7}{4}}. \end{aligned}$$

For every $\epsilon > 0$, one may choose K large enough so that $r_{n,l} \geq (1-\epsilon)R_{n,l}$ for every

$l \in [K]$. Choosing such a K and putting together the two estimates we obtain that

$$\begin{aligned}
&\mathrm{Corr}^2(M_n, M_n^{s_n}) \\
&\geq \frac{k_n^2}{n^2} \frac{1 + c\pi\delta_n^{\frac{7}{4}} + c\pi\frac{(2+\eta_n)}{(1+\eta_n)^{\frac{1}{4}}} n^{\frac{7}{4}}}{1 + c\pi\delta_n^2 s_n^{-2} + c\pi(2+\eta_n)s_n^{-2}n^{\frac{7}{4}}} \\
&\geq \frac{(k_n s_n)^2}{n^2} \frac{1 + c\pi\delta_n^{\frac{7}{4}} + c\pi\frac{(2+\eta_n)}{(1+\eta_n)^{\frac{1}{4}}}(1-\epsilon)^{\frac{7}{4}}\frac{1}{K}\sum_{l=1}^K R_{n,l}^{\frac{7}{4}}}{s_n^2 + c\pi\delta_n^2 + c\pi(2+\eta_n)\frac{1}{K}\sum_{l=1}^K R_{n,l}^{\frac{7}{4}}}.
\end{aligned}$$

Choose the sequence s_n in such a way that $s_n^2 = o(n^{\frac{7}{4}})$, that is, $s_n = o(n^{\frac{7}{8}})$ and let $\delta_n = o(n^{\frac{7}{4}})$ as well. We claim that with this choice the above correlation can be arbitrary close to $1-\epsilon$ for large n. Indeed, first note that $\frac{(k_n s_n)^2}{n^2} \geq 1$, so the first factor can be ignored. Observe that $n^{\frac{7}{4}} \leq \frac{1}{K}\sum_{l=1}^K R_{n,l}^{\frac{7}{4}} \leq (\sqrt{2}n)^{\frac{7}{4}}$ and therefore under these conditions in both the numerator and the denominator, the first two terms are negligible in comparison with the third one. Recalling that $\eta_n \to 0$, we get that for all large n

$$\mathrm{Corr}^2(M_n, M_n^{s_n}) \geq 1 - 2\epsilon.$$

This shows that the magnetization can be reconstructed from H^{s_n}, as we stated. In general, however, this does not imply that Maj_n can be reconstructed from H^{s_n}. In order to see this we show that the scaling limit of the magnetisation, which exists by [CGN15] and the scaling limit of the sparse magnetisation has the same distribution. This follows from the fact that the moments of M_n and $M_n^{s_n}$ are asymptotically the same by Proposition 3.5 in [CGN15] (proving exponential tails in the scaling limit). Indeed, we have just shown this for the second moment above. Using the n-point function established in [CHI15] and the technology from Section 3.3 in [CGN15] this equality can be extended for higher moments.

We also have to note that although the results we used are established for the square lattice, not the torus, by the classical unicity result (see [Y52]), stating that there is only one Ising measure at criticality on $\mathbb{Z}^2$, implies that these results are equally valid for the torus. □

The following is likely to be true, but still not within easy reach:

Conjecture 3.4.6. *The equivalent of Theorem 3.4.5 is true for the sequence $\mathbb{Z}_n^d$ for every $d > 2$ as well.*

An exciting area of research is to investigate a sequence tending to the + measure of a supercritical (low temperature model). As we mentioned in Section 3.1, if a transitive function was possible to reconstruct, then it would not be the magnetisation, as this measure has finite susceptibility. On the other hand, all the natural models we investigated has the property that whenever there is SR, magnetization can be reconstructed. Also, different proofs of Theorem 2.1.1 stating no SR for product measures highlight the extremal role played by magnetisation (in particular the first one and the one using Proposition 3.2.3): indeed, in case of product measures, for any sparse sequence of subsets it is the transitive function that achieves the highest clue possible.

Question 3.4.7. *Is there a sequence $\{\mu_n\}$ of Ising measures converging to the supercritical + measure on $\mathbb{Z}^d$ such that there is*

1. *SR for* $\{\mu_n\}$*?*
2. *RSR for* $\{\mu_n\}$ *?*

As it was mentioned in Section 3.1, the results of [BM02] suggest that the answer might be positive.

3.4.3 The Curie-Weiss model

In this section we shall return to the argument of Theorem 2.3.1 to show that there is no sparse reconstruction for the subcritical Curie-Weiss model, the Ising model on the complete graph K_n.

The Curie-Weiss measure is defined trough the following Hamiltonian:

$$H(\sigma) = -\frac{1}{n}\sum_{(x,y)} \sigma_x \sigma_y - h \sum_{x\in[n]} \sigma_x.$$

We have a normalisation term $\frac{1}{n}$, since in this sequence of graphs the vertex degree is growing linearly with n.

Theorem 3.4.8. *There is no sparse reconstruction for the subcritical Curie-Weiss model with* 0 *external magnetic field.*

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

Lemma 3.4.9. *Let* $\sigma[n]$ *be a sequence of spin systems and suppose that there is a* $C > 0$ *such that for every* n

$$H(\sigma[n]) \geq n - C,$$

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

Proof. The proof repeats that of Lemma 2.3.3 and Theorem 2.3.1. First observe that

$$\sum_j^L H(\sigma(S_j)) \leq \sum_j \sum_{i\in S_j} H(\sigma(i)) = k\sum_{i\in[n]} H(\sigma(i)) \leq k(H(\sigma[n]) + C), \tag{3.4.2}$$

were for the last inequality we used the condition of the Lemma. In turn, together with the Shearer inequality as in 2.3.4, we obtain that

$$\sum_j^L I(Z, \sigma(S_j)) \leq k(I(Z, \sigma[n]) + C).$$

Now we can use this inequality just as in the proof of Theorem 2.3.1 to get that

$$nI(Z, \sigma_U) \leq |U|(I(Z, \sigma[n]) + C).$$

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

Theorem 3.4.10 (Tail of subcritical Curie-Weiss)**.** *If* $\beta < \beta_c = 1$ *then*

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

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

For a proof of this result see [T11], Chapter 3, Theorem 11.

Lemma 3.4.11. *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 positive integer $0 < i \leq n$,*

$$\mathbb{P}[M_i > t\sqrt{n}] \leq \frac{e^{-Ct^2}}{1 - 4e^{-\frac{t^2}{4}}}, \tag{3.4.3}$$

for some positive constant C.

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

$$\mathbb{P}[M_n \leq \frac{t}{2}\sqrt{n} \mid M_i > t\sqrt{n}] \leq 4e^{-\frac{t^2}{4}}. \tag{3.4.4}$$

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 $S_k \, : \, k = 1, 2, \dots n - i$, 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:

$$\begin{aligned}
&\mathbb{P}[M_n \leq \frac{t}{2}\sqrt{n} \mid M_i > t\sqrt{n}] \\
&\leq \mathbb{P}[\min_k \{M_{i+k}\} \leq 0 \mid M_i > t\sqrt{n}] + \mathbb{P}[M_n \leq \frac{t}{2}\sqrt{n} \text{ and } \min_k \{M_{i+k}\} > 0 \mid M_i > t\sqrt{n}] \\
&\leq \mathbb{P}[\min S_1, S_2, \dots S_{n-i} \leq -t\sqrt{n}] + \mathbb{P}[S_{n-i} \leq -\frac{t}{2}\sqrt{n}] \\
&\leq 2\mathbb{P}[S_{n-i} > t\sqrt{n}] + \mathbb{P}[S_{n-i} = t\sqrt{n}] + \mathbb{P}[S_{n-i} \geq \frac{t}{2}\sqrt{n}] \leq 4e^{-\frac{t^2}{4}}.
\end{aligned}$$

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 $\mathbb{P}[\{\max S_1, S_2, \dots S_{n-i}\} \geq l] = 2\mathbb{P}[S_{n-i} > l] + \mathbb{P}[S_{n-i} = l]$. Finally in the last row we used the Gaussian estimate for the tail of a binomially distributed random variable.

After using the definition of conditional probability and rearranging (3.4.4)

$$\mathbb{P}[M_i > t\sqrt{n}] \leq \frac{\mathbb{P}[M_n > \frac{t}{2}\sqrt{n} \text{ and } M_i > t\sqrt{n}]}{1 - 4e^{-\frac{t^2}{4}}}.$$

Using that, by Theorem 3.4.10,

$$\mathbb{P}[M_n > \frac{t}{2}\sqrt{n} \text{ and } M_i > t\sqrt{n}] \leq \mathbb{P}[M_n > \frac{t}{2}\sqrt{n}] \leq e^{-Ct^2},$$

we obtain

$$\mathbb{P}[M_i > t\sqrt{n}] \leq \frac{e^{-Ct^2}}{1 - 4e^{-\frac{t^2}{4}}}.$$

□

Now we are ready to show that the condition of Lemma 3.4.9 is satisfied for the subcritical Curie-Weiss model.

Lemma 3.4.12. *The subcritical Curie-Weiss model with a fixed inverse temperature β and with $h = 0$ satisfies the conditions of Lemma 3.4.9, 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*

$$H(\sigma_\beta([n])) \geq n - C.$$

Proof. According to the chain rule of entropy,

$$H(\sigma_\beta[n]) = \sum_{k=0}^{n-1} H(\sigma(k+1) \mid \sigma[k]). \tag{3.4.5}$$

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:

$$H\left(\sigma(k+1) \mid \sigma([k])\right) = \sum_t \mathbb{P}\left[M_k = t\right] H\left(\sigma(k+1) \mid M_k = t\right), \tag{3.4.6}$$

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,

$$H\left(\sigma(k+1) \mid M_k = t\right) = h\left(\mathbb{E}[\sigma(k+1) = 1 \mid M_k = t]\right), \tag{3.4.7}$$

where

$$h(x) := \frac{1-x}{2} \log \frac{1-x}{2} + \frac{1+x}{2} \log \frac{1+x}{2} = 1 - \frac{1}{\ln 4} x^2 + O(x^4), \tag{3.4.8}$$

using the Taylor expansion of h around 0. Let us compute the Hamiltonian conditioned on the event that the sum of the first k spins is t:

$$\begin{aligned} H_{n,0}(\sigma \mid M_k = t) &= -\frac{1}{2n}\left(\sum_{i,j>k} \sigma(i)\sigma(j) - \sum_{i \leq k} \sigma(i) \sum_{l>k} \sigma(l)\right) - \frac{1}{2n} \sum_{i,j \leq k} \sigma(i)\sigma(j) \\ &= -\frac{t}{2n} \sum_{i>k} \sigma(i) - \frac{1}{2n} \sum_{i,j>k} \sigma(i)\sigma(j) - \frac{t^2}{2n}. \end{aligned}$$

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,

$$\mathbb{E}\left[\sigma(k+1) \mid M_k = t\right] = m_{n-k}\left(\beta, \frac{t}{2n}\right),$$

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

$$m_{n-k}(\beta, \frac{t}{2n}) = m_{n-k}(\beta, 0) + \frac{t}{2n}\frac{\partial m}{\partial h} + O\left(\frac{t^2}{n^2}\right).$$

Recall, that by (3.4.1), $\frac{\partial m_n}{\partial h} \leq \frac{\partial m}{\partial h} = \beta S$ where S denotes the limiting susceptibility at inverse temperature β . It is well known (see for example [Pet] Section 13.1) that the susceptibility is finite in the subcritical (high temperature) regime (or, which is the same the variance of the magnetisation is $O(n)$). Obviously $m_n(\beta, 0) = 0$, so the first order approximation says that for every $t \geq 0$

$$\mathbb{E}\left[\sigma(k+1) \mid M_k = t\right] = \beta S \frac{t}{2n} + O\left(\frac{t^2}{n^2}\right).$$

From Equation (3.4.7), taking into account the expansion of h as in (3.4.8) we obtain that

$$H\left(\sigma(k+1) \mid M_k = t\right) = 1 - C\frac{t^2}{n^2} + O\left(\frac{t^3}{n^3}\right). \tag{3.4.9}$$

We introduce the following notation:

$$\begin{aligned} f(t) &= \mathbb{P}\left[M_k = t\right], \\ F(t) &= \sum_{s=0}^{t} \mathbb{P}\left[M_k = s\right] = \mathbb{P}\left[0 \leq M_k \leq t\right], \\ h(t) &= 1 - H\left(\sigma(k) \mid M_k = t\right). \end{aligned}$$

Now we can rewrite (3.4.6)

$$\begin{aligned} &H\left(\sigma(k+1) \mid \sigma[k]\right) \\ &= \sum_t \mathbb{P}\left[M_k = t\right] H\left(\sigma(k+1) \mid M_k = t\right) \\ &= 1 - \sum_t \mathbb{P}\left[M_k = t\right]\left(1 - H\left(\sigma(k+1) \mid M_k = t\right)\right) \\ &= 1 - \sum_{t=-k}^{k} f(t)h(t). \end{aligned}$$

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:

$$\begin{aligned} \sum_{t=0}^{k} f(t)h(t) =& F(k)h(k) - F(0)h(0) + \sum_{t=0}^{k-1} F(t)(h(t+1) - h(t)) = \\ =& F(k)h(k) + \sum_{t=0}^{k-1} (F(k) - \mathbb{P}\left[M_k > t\right])(h(t+1) - h(t)) = \\ =& \sum_{t=0}^{k-1} \mathbb{P}\left[M_k > t\right]\left(h(t+1) - h(t)\right), \end{aligned}$$

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.

$$\sum_{t=0}^{k-1} \mathbb{P}\left[M_k > t\right]\left(h(t+1) - h(t)\right) = \sum_{t=0}^{L\sqrt{n}-1} (\dots) + \sum_{t=L\sqrt{n}}^{n^{\frac{3}{4}}-1} (\dots) + \sum_{t=n^{\frac{3}{4}}}^{k-1} (\dots).$$

Let us start with the first sum:

$$\sum_{t=0}^{L\sqrt{n}-1} \mathbb{P}\left[M_k > t\right] \left(h(t+1) - h(t)\right) \leq \sum_{t=0}^{L\sqrt{n}-1} h(t+1) - h(t) = h(L\sqrt{n}) - h(0).$$

Since $h(0) = 0$ 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 (3.4.9), we get

$$\sum_{t=0}^{L\sqrt{n}-1} \mathbb{P}\left[M_k > t\right] \left(h(t+1) - h(t)\right) \leq CL^2\frac{1}{n} + O\left(\frac{1}{n^{3/2}}\right).$$

Now we turn to the second sum. By Lemma 3.4.11

$$\sum_{t=L\sqrt{n}}^{n^{\frac{3}{4}}-1} \left(\mathbb{P}\left[M_k > t\right] \left(h(t+1) - h(t)\right) \leq \sum_{t=L\sqrt{n}}^{n^{\frac{3}{4}}-1} \left(h(t+1) - h(t)\right)\frac{e^{-\frac{Ct^2}{n}}}{1 - 4e^{-\frac{t^2}{4n}}}.$$

First note that $t = o(n)$, so it is valid to 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

$$e^{-\frac{Ct^2}{n}} \leq \left(\frac{t^2}{2n}\right)^{-2} \text{ and } \left(1 - 4e^{-\frac{t^2}{4n}}\right) \geq \frac{1}{2} \tag{3.4.10}$$

are satisfied whenever $t \geq L\sqrt{n}$. With such an L we have:

$$\sum_{t=L\sqrt{n}}^{n^{\frac{3}{4}}-1} \left(h(t+1) - h(t)\right)\frac{e^{-\frac{Ct^2}{n}}}{1 - 4e^{-\frac{t^2}{4n}}} \leq \frac{C}{n^2}\sum_{t=L\sqrt{n}}^{n^{\frac{3}{4}}-1} t\left(\frac{t}{2\sqrt{n}}\right)^{-4} = C'\sum_{t=L\sqrt{n}}^{n^{\frac{3}{4}}-1} t^{-3}.$$

Therefore, approximating the sum with the respective integral, we get that

$$\sum_{t=L\sqrt{n}}^{n^{\frac{3}{4}}-1} \mathbb{P}\left[M_k > t\right] \left(h(t+1) - h(t)\right) \leq C\text{"}\,(L\sqrt{n})^{-2} - n^{-\frac{3}{2}}) \leq C\text{"}\,\frac{1}{L^2 n}.$$

Finally, using the tail estimation of Lemma 3.4.11 and (3.4.10):

$$\sum_{t=n^{\frac{3}{4}}}^{k-1} \mathbb{P}\left[M_k > t\right] \left(h(t+1) - h(t)\right) \leq \frac{1}{2}\sum_{t=n^{\frac{3}{4}}}^{k-1} e^{-\frac{Ct^2}{2n}} = o\left(\frac{1}{n}\right),$$

where we used the trivial bound $h(t+1) - h(t) \leq 1$.

Now we can put everything together, using that $\sum_{t=1}^{k} f(t)h(t) = \sum_{t=-k}^{-1} f(t)h(t)$.

$$1 - H\left(\sigma(k)|\sigma([k-1])\right) \leq 2\sum_{t=0}^{k-1} \mathbb{P}\left[M_{k-1} > t\right] \left(h(t+1) - h(t)\right) \leq C\left(L^2 + \frac{1}{L^2}\right)\frac{1}{n} + o\left(\frac{1}{n}\right)$$

and therefore, substituting this estimate into the chain rule, we have that for some constant $K > 0$

$$H(\sigma_\beta([n])) = \sum_{k=1}^{n} H(\sigma(k)|\sigma([k-1])) \geq n\left(1 - K\frac{1}{n} + o\left(\frac{1}{n}\right)\right) = n - K + o\left(1\right).$$

□

Bibliography

[AGMT15] D. Ahlberg, S. Griffiths, R. Morris, V. Tassion. Quenched Voronoi percolation. *Advances in Mathematics.* (2015) `arXiv:1501.04075 [math.PR]`

[ADS13] M. Aizenman, H. Duminil-Copin, V. Sidoravicius, Vladas. Random Currents and Continuity of Ising Model's Spontaneous Magnetization. *Communications in Mathematical Physics.* **334** (2013).

[AF86] M. Aizenman, R. Fernandez. On the critical behavior of the magnetization in high-dimensional Ising models. *Journal of Statistical Physics.* **44** (1986). 393–454.

[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.

[BV93] E, Bernstein, U. Vazirani (1993). Quantum complexity theory. *Proceedings of the Twenty-Fifth Annual ACM Symposium on Theory of Computing (STOC 93).* (1993), 11-20

[Bo05] T. Bodineau. Slab percolation for the Ising model. *Probability Theory and Related Fields.* **132** (2005) 83–118.

[BM02] T. Bodineau, and F. Martinelli. Some New Results on the Kinetic Ising Model in a Pure Phase. *J. Statist. Phys.* **109** (2002) 207–235.

[BLM13] S. Boucheron, G. Lugosi and P. Massart. *Concentration Inequalities: A nonasymptotic theory of independence.* Oxford University Press, 2013.

[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.

[CGN15] F. Camia, C. Garban, C. M. Newman. Planar Ising magnetization field I. Uniqueness of the critical scaling limit. *Ann. Probab.* **43** (2015), 528–571.

[C14] S. Chatterjee. *Superconcentration and Related Topics.* Springer-Verlag Berlin Heidelberg, 2014.

[CHI15] D. Chelkak, C. Hongler and K. Izyurov, Conformal invariance of spin correlations in the planar Ising model. *Annals of Mathematics* **181**, 3, (2015), 1087–1138

[F18] M. P. Forsström. Denseness of volatile and nonvolatile sequences of functions. *Stochastic Processes and their Applications.* (2018), 128 (11): 3880–3896.

[F15] M. P. Forsström. Noise Sensitivity and Noise Stability for Markov Chains: Existence Results. `arXiv:1501.01824v1 [math.PR]`

[FK96] E. Friedgut and G. Kalai. Every monotone graph property has a sharp threshold. *Proc. Amer. Math. Soc.* **124** (1996), 2993–3002.

[Ga19+] P. Galicza. Pivotality versus noise stability for monotone transitive functions. *Electronic Communications in Probability*, accepted, (2019). `arXiv:1909.05375 [math.PR]`

[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]`.

[GLSS12] D. Gavinsky, S. Lovett, M. Saks, S. Srinivasan. A Tail Bound for Read-k Families of Functions. *Random Struct. Algorithms* **47** (2012): 99-108.

[GW95] M. X. Goemans and D. P. Williamson. *Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. J. ACM 42*, 6 (November 1995), 1115–1145.

[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)

[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]

[Le74] J. Lebowitz. GHS and other inequalities. *Commun. Math. Phys.* **35** (1974), 87–92.

[Lu94] A. Lubotzky. *Discrete groups, expanding graphs and invariant measures.* Progress in Mathematics, 125. Birkhäuser Verlag, Basel, 1994. With an appendix by J. D. Rogawski.

[LyP] R. Lyons and Y. Peres. Probability on Trees and Networks. Book in preparation, available at http://pages.iu.edu/˜rdlyons/

[LMN93] N. Linial, Y. Mansour, and N. Nisan. Constant depth circuits, Fourier transform, and learnability. *J. Assoc. Comput. Mach.* **40(3)** (1993), 607-620

[Mo10] E. Mossel. Gaussian bounds for noise correlation of functions. *Geom. Funct. Anal.* Vol. 19 (2010) 17131756

[MOO05] E. Mossel, R. O'Donnell, and K. Oleszkiewicz. Noise stability of functions with low influences invariance and optimality. *In Proceedings of the 46th Annual IEEE Symposium on Foundations of Computer Science (FOCS '05).* (2005) IEEE Computer Society, Washington, DC, USA, 21-30.

[OD14] Ryan O'Donnell. *Analysis of Boolean functions.* Cambridge University Press, 2014. http://get.analysisofbooleanfunctions.org

[OW87] D. Ornstein, B. Weiss. Entropy and isomorphism theorems for actions of amenable groups. *Journal d'Analyse Mathmatique.* **48(1)** (1987), 1–141.

[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.

[Pi96] A. Pisztora. Surface order large deviations for Ising, Potts and percolation models. *Probab. Theory Relat. Fields.* **104** (1996), 427-466

[RS18] M. Roberts and B. Şengül. Exceptional times of the critical dynamical Erdős-Rényi graph. *Ann. Appl. Probab.* **28** (2018), no. 4, 2275–2308.

[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]

[Si80] Barry Simon. Correlation inequalities and the decay of correlations in ferromagnets. *Commun. Math. Phys.* **77 (1980), 111–126**

[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

[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, (2009), pp. 145–174. `arXiv:0901.4760 [math.PR]`

[T11] B. Tóth. A statisztikus fizika matematikai módszerei. (2011) Lecture notes. `http://tankonyvtar.ttk.bme.hu/pdf/15.pdf`.

[Tao05] Terence Tao. Szemerédi's regularity lemma revisited, ***Contributions to Discrete Mathematics*** 125 (2003), 195–224

[Y52] C.N. Yang. The spontaneous magnetization of a two-dimensional Ising model. *Phys. Rev.* 85 (1952), no. 5, 808-816

[W09] Wendelin Werner. Lectures on two-dimensional critical percolation. In Statistical mechanics, IAS/Park City Math. Ser. 16 (2009), 297–360.