Superconcentration and Related Topics

Sourav Chatterjee

Transcribed by Mihai Nica

This is a draft of my notes from Professor Sourav Chatterjee's six lecture course on "Superconcentration and Related Topics" which were given in July 2012 at the Cornell Probability Summer School.

Contents

CHAPTER 1

Overview

We will look at some random variables X_n where we have the phenomenon that $\frac{\mathbf{Var}(X_n)}{n} \to 0$ as $n \to \infty$ This is in contrast to the usual phenomenon that $\mathbf{Var}(X_n) = O(n)$ that we are more routinely used to in our study of probability. This idea is sometimes referred to as "super concentration" and as we will see it has some connections to chaos and other interesting phenomenon that appear.

In this section we will introduce some of the models that give rise to the problems that are of interest to us and state some of the theorems we will work towards. The proof of the theorems are not given here, but are instead reserved for later sections where they will be filled in with more details.

1. First Passage Percolation

One place where this phenomenon appears is the First Passage Percolation (FPP) model.

DEFINITION 1.1. Consider the graph $G = \left(\mathbb{Z}^d, E\left(\mathbb{Z}^d\right)\right)$ with the usual nearest neighbor edges for the lattice. Consider a set of iid random non-negative weights called the *passage time* or *edge rates:*$(w_e)_{e \in E(\mathbb{Z}^d)}$. Define *the passage time for a path* $\pi : x \leftrightarrow y$ between two points $x, y \in \mathbb{Z}^d$ by:

$$P_\pi(x, y) = \sum_{e \in \pi} w_e$$

And define the *first passage time* between x, y to be the minimum over all such paths:

$$T(x, y) = \inf_{\pi : x \leftrightarrow y} P_\pi(x, y)$$

In this model, one is mostly interested in the asymptotics of the passage time in any direction. That is for fixed $x \in \mathbb{R}^d$ let:

$$T_n(x) = T\left(\vec{0}, [nx]\right)$$

Where $[nx]$ is the nearest lattice point to the vector $nx \in \mathbb{R}^d$. One is interested in the behavior of $T_n(x)$ as $n \to \infty$. We will sometimes write T_n to mean $T_n\left(\vec{e_1}\right)$.

FACT 1.2. *One can show that* $\lim_{n\to\infty} \frac{T_n(x)}{n}$ *exists and is deterministic. Moreover, it is positive when* $\mathbf{P}(w_e = 0) < p_c(d)$ *the critical probability for percolation in* $\mathbb{Z}^d$.

PROBLEM 1.3. What is the behavior as $n \to \infty$of $\mathbf{Var}(T_n(x))$?

Keston proved that $\mathbf{Var}(T_n) \leq Cn$ in any dimension. There are also exponential tail inequalities which control the probability that $\mathbf{Var}(T_n)$ is very large.

Benjamin-Kelai-Shram (2003) proved that if the edge weights only take two values, $w_e \in \{a, b\}$that $\mathbf{Var}(T_n) \leq \frac{Cn}{\log n}$. Notice here that $\mathbf{Var}(T_n) = o(n)$ which will somehow turn out to be qualitatively different than other bounds as we will see.

CONJECTURE 1.4. *Under some mild conditions on the weights w_e, some physicists believe, based on non-rigorous ideas, that:*

$$\mathbf{Var}(T_n(x)) = O\left(n^{\frac{2}{3}}\right)$$

This is an open problem to prove rigorously.

2. (1+1)dimensional Gaussian Random Polymer

DEFINITION 2.1. In $\mathbb{Z}^2$, consider the set $T_n = \{-1, 1\}^n = \{(0, 0), (1, a_1), (2, a_2), \ldots, (n, a_n)\}$ with $a_j = a_{j-1} \pm 1$ which we think of as the set of all possible NE-SE paths of length n. Let $(g_v)_{v\in\mathbb{Z}^2}$be iid standard Gaussian $g_v \sim N(0, 1)$ random variables. Define the energy of a path $p = \{(0, a_1), (1, a_2), \ldots (n, a_n)\} \in T_n$ by:

$$H(p) = -\sum_{i=1}^{n} g_{(i,a_i)}$$

The object of interest in this model is the minimum energy path $E_n = \inf_{p\in T_n} H(p)$.

THEOREM 2.2. *If E_n is the minimum energy of the (1+1)-dim'l Gaussian Random Polymer as described above, then:*

$$\mathbf{Var}(E_n) \leq \frac{Cn}{\log n}$$

REMARK. Compare this result to a more unrestricted polymer, one where we allow $T_n = \{(1, a_1), (2, a_2) \ldots (n, a_n)\}$ where we allow $a_i \in \{1, 2, \ldots n\}$ with no restriction on the size of each step. In this space, the maximum energy is plainly seen by maximizing in each column seperatly:

$$\begin{aligned} L_n &= \max_p \sum_{i=1}^{n} g_{(i,a_i)} \\ &= \sum_{i=1}^{n} \max_{j\in\mathbb{Z}} g_{ij} \end{aligned}$$

It is a not so hard fact to see that $\mathbf{Var}(L_n) \sim \frac{Cn}{\log n}$here by using the fact that if $Z_1, \ldots Z_n \sim N(0, 1)$ are iid then $\mathbf{Var}(\max_i Z_i) \sim \frac{C}{\log n}$. This is a classical result one can show without using the ideas of concentration.

3. Sherington-Kirkpatrick Model of Spin Glasses

DEFINITION 3.1. Consider n particles which are all either spin $+1$ or -1, and let $\sigma = (\sigma_1, \sigma_2, \ldots \sigma_n) \in \{+1, -1\}^n$ denote the spin configuration. Suppose we have an array of interaction weights $(g_{ij})_{1\leq i\leq j\leq n}$ which are iid random variables, usually taken to be distributed like a standard Gaussian, $g_{ij} \sim N(0, 1)$, Define the *energy of the spin configuration σ* by:

$$H_n(\sigma) = -\frac{1}{\sqrt{n}} \sum_{1\leq i\leq j\leq n} g_{ij}\sigma_i\sigma_j$$

This is an example of a *disordered system* and we sometimes refer to the weights $(g_{ij})_{1\leq i\leq j\leq n}$ as the *disorder* of the system. This energy induces a measure on the set of all such configurations, known as the *Gibbs measure* by:

$$\mu_n(\sigma) \propto \exp\left(-\beta H_n(\sigma)\right)$$

Where β is a parameter, sometimes interpreted as the inverse temperature of the system. The normilization constant $Z_n(\beta) = \sum_\sigma e^{-\beta H(\sigma)}$ is called the *partition function* and a number of interesting properties can be derived from it. One such parameter is the *free energy* of the system, given by:

$$F_n\left(\beta\right) = -\frac{1}{\beta}\log Z_n\left(\beta\right)$$

The *Parisi formula* tells us that $\lim_{n\to\infty}\frac{F_n(\beta)}{n}$ converges to a deterministic limit.

PROBLEM 3.2. What is the behavior of $\mathbf{Var}\left(F_n\left(\beta\right)\right)$?

One can show that this model has a phase transition at $\beta = 1$, with qualitatively different behavior for $\beta < 1$ and for $\beta > 1$. For $\beta < 1$ (the high temperature phase) one can show that as $n \to \infty$ that:

$$\mathbf{Var}\left(F_n(\beta)\right) \to C(\beta)$$

For $\beta > 1$, the best known classical upper bound was that $\mathbf{Var}\left(F_n\left(\beta\right)\right) \leq C(\beta)n$.

THEOREM 3.3. (Chatterjee 2009)
For all β:

$$\begin{aligned}\mathbf{Var}\left(F_n\left(\beta\right)\right) &\leq \frac{C(\beta)n}{\log n}\\ &= o(n)\end{aligned}$$

CONJECTURE 3.4. *It is conjectured that* $\mathbf{Var}\left(F_n\left(\beta\right)\right) = O(1)$. *Again, this is an open problem.*

4. Chaos

PROBLEM 4.1. In the setting of the (1+1)-Dimensional Gaussian Polymer, let $\hat{p}_n$be the optimal (i.e. minimum energy) path of length n.

Is $\hat{p}_n$chaotic? That is, how sensitive is $\hat{p}_n$ to small perturbations of the vertex weights g_{ij}?

To make a precise model, we can take a new set of iid $N(0,1)$ variables g'_{ij} which are related to the old vertex weights $g_{ij}$by the relation that $\left(g_{ij}, g'_{ij}\right)$ is a multidimensional Gaussian with marginals $N(0,1)$ and with correlation $1-\epsilon$. How similar is $\hat{p}'_n$ to $\hat{p}_n$ under this slight change of vertex weights?

THEOREM 4.2. (Chatterjee 2008)
In the above set up for the Gaussian Random Polymer, where $\left(g_{ij}, g'_{ij}\right)$ are weights that have standard Gaussian marginals with correlation $1-\epsilon$, and $\hat{p}_n$ is the optimal minimum energy path, we have that:

$$\mathbf{E}\left(|\hat{p}_n \cap \hat{p}'_n|\right) \leq \frac{Cn}{\epsilon \log n}$$

This is saying that the two paths have very few vertices in common. Notice that if $\epsilon \gg \log n^{-1}$, then this is $o(n)$. We will also show that:

$$\mathbf{Var}\left(E_n\right) = o(n) \iff \exists \epsilon_n \to 0 \text{ s.t. } \mathbf{E}\left(\left|\hat{p}_n \cap \hat{p}_n^{(\epsilon_n)}\right|\right) = o(n)$$

PROBLEM 4.3. In the setting of the spin glass model, we can ask what is the distance between two randomly chosen configurations. More precisely, if we let σ^1 and σ^2 be iid picks from the Gibbs measure $\mu(\sigma) = Z_n(\beta)^{-1}\exp(-\beta H(\sigma))$, (remember that the energy H here is governed by the random interaction weights g_{ij}) then we define the distance:

$$R_{1,2} = \frac{1}{n}\sum_{i=1}^{n}\sigma_i^1\sigma_i^2$$

It is known that for $\beta \leq 1$ that $\mathbf{E}\left(R_{1,2}^2\right) \to 0$ as $n \to \infty$. It is thought to be positive when $\beta > 1$. To investigate how the distance between two randomly chosen configurations, we again perturb our weights by choosing $\left(g_{ij}, g'_{ij}\right)$, which again have standard Gaussian marginals with correlation $1-\epsilon$ and we choose σ_1 from the Gibbs measure induced by the $g'_{ij}s$ and σ^2 from the Gibbs measure built from the g'_{ij}'s. In this case we define:

$$R_{1,2}(\epsilon) = \frac{1}{n}\sum_{i=1}^{n}\sigma_i^1\sigma_i^2$$

This will be some kind of measure for how far apart typical configurations from the Gibbs measure g_{ij} and g'_{ij} are. If $R_{1,2}$is small, the configurations σ^1 and σ^2 are almost orthogonal..i.e. very different.

THEOREM 4.4. (Chatterjee 2009)
For every β we will have that:

$$\mathbf{E}\left(R_{1,2}^2(\epsilon)\right) \leq \frac{C(\beta)}{\epsilon \log n}$$

In particular, we will show that this is related to the variance in the minimal energy for the problem, namely:

$$\mathbf{Var}\left(E_n\right) = o(n) \iff \exists \epsilon_n \to 0 \text{ s.t. } \mathbf{E}\left(R_{1,2}^2(\epsilon_n)\right) \to 0$$

CHAPTER 2

Multiple Valleys and the Poincare Inequality

1. Stability in Optimization Problems

An optimization problem is called *stable* if there exists a unique optimum solution and if any "near" optimum solution is "close" to the actual optimum solution. Of course, one must make precise what near and close mean here to make this a mathematical statement.

EXAMPLE 1.1. Let $(c_{ij})_{i,j\in[n]}$ be an iid collection of $Exp(1)$ random variables. The interpretation here is that we have n people and n tasks and if assigned to task j it takes person i c_{ij} units of energy to complete the task. Define:

$$C_n = \min_{\sigma\in S_n} \sum_{i=1}^{n} c_{i\sigma(i)}$$

This is the minimum energy the team of people need to complete the task. Aldous (2001) proved that:

$$C_n \to \frac{\pi^2}{6}$$

Moreover, he proved that that this problem has a property known as *Asymptotic Essential Uniqueness,* meaning that if $\frac{\sum_i c_{i\sigma(i)}}{C_n} \approx 1$ then $\sigma \approx \hat{\sigma}$ where $\hat{\sigma}$is the minimizer for C_n. Similar results hold for the minimum cost spanning tree on a uniformly weighted lattice.

Asymptotic essential uniqueness (AEU) is the type of stability we will look at here. We will formulate it precisely and prove some results, but first lets do a few more examples to build some intuition.

EXAMPLE 1.2. Let $g_1, \dots g_n \sim N(0,1)$ be iid and let $T_n = \{-1,1\}^n$ be the space of configurations $\sigma = \{\sigma_1, \dots \sigma_n\} \in T_n$. Define $f : T_n \to \mathbb{R}$ by:

$$f(\sigma) = \sum_{i=1}^{n} g_i\sigma_i$$

Clearly, to maximize f, we just take σ_i =sgn (g_i). This has the Asymptotic Essential Uniqueness property.

EXERCISE 1.3. Formulate precisely the asymptotic essential uniqueness property for the above example and then prove it.

EXAMPLE 1.4. In our spin glass model, we were optimizing over the space of configurations $\sigma = (\sigma_1, \dots, \sigma_n)$ with interaction weights $g_{ij} \sim N(0,1)$ iid. The energy we maximize is:

$$H_n(\sigma) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} g_{ij}\sigma_i\sigma_j$$

In this model there is *not AEU*, which we will prove later.

DEFINITION 1.5. *(Multiple Valley Property)*

Suppose we have a sequence of (usually finite) sets X_n and a sequence of random functions $f_n : \Omega \times X_n \to \mathbb{R}$ on a probability space Ω. Suppose also we have a non-negative *similarity measure* $s_n : X_n \times X_n \to \mathbb{R}^+$ which measures how similar to elements on X_n are. (This is like a distance function, but with two major differences: i) no triangle inequality nessasary and ii) s_n small means the elements are very *dissimilar*)

We will say that the sequence (f_n, X_n, s_n) has the *multiple valley property* if $\exists \epsilon_n, \delta_n, \gamma_n \to 0$ and $K_n \to \infty$ so that for every n:

$$\mathbf{P}\left\{\exists A \subset X_n \text{ s.t. } |A| > K_n, \text{ and } s_n(x,y) \leq \epsilon_n \ \forall x, y \in A \text{ and } \left|\frac{f_n(x)}{\min_{y \in X_n} f_n(y)} - 1\right| \leq \delta_n \ \forall x \in A\right\} \geq 1 - \gamma_n$$

The set A here is the set of "multiple valleys" for the functions f_n. The condition above is saying that with high probability (meaning $> 1 - \gamma_n$), that there is a set A with many elements (at least K_n), so that every element of A is close to the optimum value of f_n (within $1 - \delta_n$), and no two of them are very similar (in the sense of s_n and ϵ_n).

EXAMPLE 1.6. The spin glass model fits this setup with:

$$\begin{aligned} X_n &= \{-1, +1\}^n \\ f_n &= H_n \\ &= \frac{1}{\sqrt{n}} \sum_{i,j} g_{ij} \sigma_i \sigma_j \end{aligned}$$

Here the probability space is the space the $g'_{ij}s$ live on, this is where the randomness for this problem come from. We choose the similarity measure:

$$s_n(\sigma, \sigma') = \left(\frac{1}{n} \sum_{i=1}^{n} \sigma_i \sigma'_i\right)^2$$

In this way, two configurations are dissimilar if they are close to being orthogonal. We will see later that our choice of similarity measure is motivated by the model in question.

We will show later that this model, with this similarity measure does indeed have the multiple valley property.

EXAMPLE 1.7. The (1+1)-dim Guassian Polymers fit this setup with:

$$\begin{aligned} X_n &= \{(0,0), (1,a_1), (2,a_2), \ldots, (n,a_n)\} \equiv \{-1,+1\}^n \\ f_n &= H_n \\ &= \sum_{i=1}^{n} g_{(i,a_i)} \end{aligned}$$

Here the probability space is the space the $g'_v s$ live on. We choose the similarity measure to be the number of edges in common between two paths.:

$$s_n(p, p') = \frac{1}{n} |p \cap p'|$$

We will prove later that this model has the multiple valley property.

2. Markov Semi-Groups

In this section, we will review the theory of Markov semi groups. This will be important and useful to us because of results that concern the equilibrium distribution of the Markov semi group. By studying these Markov chains, we get a handle on the properties of functions of the equilibrium distribution.

Our models so far all have randomness coming from a collection of iid standard Gaussian variables, so we will want a Markov chain whose limit distribution is such a Gaussian. It will turn out that the Ornstein-Uhlenbeck process will ideal for this; the $n-$dim'l O-U process has a Gaussian equilibrium distribution and also flowing through time in the O-U process changes the coordinates in a predictable smooth way.

DEFINITION 2.1. Suppose that $(X_t)_{t\geq 0}$ is a continuous time Markov process on some state space S. This defines for us a *semi-group of operators* $(P_t)_{t\geq 0}$ indexed by time which act on the space $S^{\mathbb{R}} = \{f : S \to \mathbb{R}\}$. Each $P_t : S^{\mathbb{R}} \to S^{\mathbb{R}}$ by:

$$P_t(f)(x) = \mathbf{E}\left(f(X_t)|X_0 = x\right)$$

Here X_t is the Markov chain which we start at the point x. The Markov property of the process makes this a semi-group, as one can easily verify that:

$$P_{t+s} = P_t \circ P_s$$

The *generator* of the semi-group is the operator $L : S^{\mathbb{R}} \to S^{\mathbb{R}}$ defined by:

$$Lf = \lim_{t\to 0} \frac{P_t f - f}{t}$$

PROPOSITION 2.2. *The semi-group satisfies the "heat equation":*

$$\partial_t P_t = LP_t = P_t L$$

For this reason, we will sometimes write $P_t = \exp(tL)$.

PROOF. By the semi-group property:

$$\begin{aligned} \partial_t P_t &= \lim_{h\to 0} \frac{P_{t+h} - P_t}{h} \\ &= \lim_{h\to 0} \frac{P_h - Id}{h} \circ P_t \\ &= L \circ P_t \end{aligned}$$

The same idea holds to show this is equal to $P_t \circ L$ too.

□

DEFINITION 2.3. A Markov process is said to have an *equilibrium distribution* or *stationary distribution* μ when, if we start with the distribution $X_0 \overset{d}{\sim} \mu$, we have the same distribution for all time, namely $X_t \overset{d}{\sim} \mu$. In terms of our semi-group operators, another way of saying the same thing is that for all t and for all f:

$$\mathbf{E}\left((P_t f)(Z)\right) = \mathbf{E}\left(f(Z)\right)$$

Where $Z \overset{d}{=} \mu$ is a random variable with distribution μ. Sometimes we will write this in "integral form":

$$\int P_t f \, d\mu = \int f \, d\mu$$

Or sometimes:

$$\mathbf{E}_\mu(P_t f) = \mathbf{E}_\mu(f)$$

It is a fact about Markov chains that as time increases, the distribution of the Markov chain approaches the stationary distribution μ. In our semi-group formulation, this is expressed by the fact that $P_t f$ converges to a constant:

$$\lim_{t\to\infty} P_t f = \mathbf{E}_\mu(f)$$

From here onward, we will always assume that the Markov chain X_t has a stationary distribution μ. We will sometimes let Z be a random variable which has the distribution μ.

DEFINITION 2.4. The equilibrium distribution of the Markov chain defines an inner product space on the functions $f : S \to \mathbb{R}$,whhich we call $L^2(\mu)$ with inner product given by:

$$\begin{aligned} \langle f, g\rangle_{L^2(\mu)} &= \int fg\, d\mu \\ &= \mathbf{E}(f(Z)g(Z)) \\ &= \mathbf{E}_\mu(f \cdot g) \\ \|f\|_{L^2(\mu)} &= \langle f, f\rangle^{\frac{1}{2}}_{L^2(\mu)} \end{aligned}$$

This is an L^2 space, so it satisfies familair things like the Cauchy Shwarz inequality, $|\langle f, g\rangle|^2 \le \langle f, f\rangle \langle g, g\rangle$ and so on. We also have something related called the *Dirchelet form:*

$$\mathcal{E}(f, g) = -\langle f, Lg\rangle$$

When the Markov chain is reversible, the L operator is self-adjoint with respect to this inner product, and so the Dirchelet form is symmetric. (Recall that a Markov chain is reversible if it has a stationary distribution and if $\mathbf{P}(X_0 = a, X_t = b) = \mathbf{P}(X_0 = b, X_t = a)$ where we are understood to start at the stationary distribution, i.e. $\mathbf{P}(X_0 = a) = \mu(a)$)

$$\mathcal{E}(f, g) = \mathcal{E}(g, f)$$

From now on, we will assume that this is indeed the case for any Markov chain (X_t) we talk about.

PROPOSITION 2.5. *Show that P_t obeys the following nice inequality in the space $L^2(\mu)$:*

$$\|P_t(h)\|_{L^2(\mu)} \le \|h\|_{L^2(\mu)}$$

PROOF. We first claim that $\phi(P_t f) \le P(\phi(f))$ for convex functions ϕ. This follows Jensen's inequality, since P_t is an expectation. We have for any $x \in S$:

$$\begin{aligned} \phi(P_t f)(x) &= \phi(\mathbf{E}(f(X_t)|\, X_0 = x)) \\ &\le \mathbf{E}(\phi(f(X_t))|\, X_0 = x) \\ &= P(\phi(f)) \end{aligned}$$

Now, to prove the result of the exercise, use this with $\phi(\cdot) = (\cdot)^2$, along with the fact that $\mathbf{E}_\mu(P_t(g)) = \mathbf{E}_\mu(g)$:

$$\begin{aligned}\left\|P_t(h)\right\|^2{}_{L^2(\mu)} &= \mathbf{E}_\mu\left(\left(P_t h\right)^2\right)\\ &\leq \mathbf{E}_\mu\left(P_t\left(h^2\right)\right)\\ &= \mathbf{E}_\mu\left(h^2\right)\\ &= \|h\|^2_{L^2(\mu)}\end{aligned}$$

□

DEFINITION 2.6. The equilibrium distribution also gives us a natural covariance functional defined by:

$$\begin{aligned}\mathbf{Cov}_\mu(f,g) &= \left(\int fg\,d\mu\right) - \left(\int f\,d\mu\right)\left(\int g\,d\mu\right)\\ &= \mathbf{E}_\mu(f\cdot g) - \mathbf{E}_\mu(f)\,\mathbf{E}_\mu(g)\end{aligned}$$

LEMMA 2.7. (The Covariance Lemma)
The covariance is related to the Dirichlet form by:

$$\mathbf{Cov}_\mu(f,g) = \int_0^\infty \mathcal{E}(f, P_t g)\,dt$$

PROOF. We use the fact that $P_t f \to \mathbf{E}_\mu(f)$. Have:

$$\begin{aligned}\mathbf{Cov}_\mu(f,g) &= \mathbf{E}_\mu(f\cdot g) - \mathbf{E}_\mu(f)\,\mathbf{E}_\mu(g)\\ &= \mathbf{E}_\mu(f\cdot(P_0 g)) - \mathbf{E}_\mu(f)\lim_{t\to\infty} P_t g\\ &= \mathbf{E}_\mu\left(f\cdot\left(P_0 g - \lim_{t\to\infty} P_t g\right)\right)\\ &= -\mathbf{E}_\mu\left(f\cdot\left(\int_0^\infty \partial_t P_t g\,dt\right)\right)\\ &= -\mathbf{E}_\mu\left(f\cdot\int_0^\infty L P_t g\,dt\right)\\ &= -\int_0^\infty \mathbf{E}_\mu\left(f\cdot L P_t g\right)\,dt\\ &= \int_0^\infty \mathcal{E}(f, P_t g)\,dt\end{aligned}$$

The interchange of the expectation with the integral can be justified with mild conditions which we will not concern ourselves with here.

□

The Poincare inequality, whose definition follows, is one of the most important classical inequalities one can get from Markov chains.

DEFINITION 2.8. *(The Poincare Inequality)*
A Markov Process is with Dirichlet form $\mathcal{E}$ is said to satisfy a Poincare inequality with constant C if $\forall f \in L^2(\mu)$:

$$\mathbf{Var}_\mu(f) = \mathbf{Cov}_\mu(f,f) \leq C\mathcal{E}(f,f)$$

3. The Ornstein-Ulenbeck Semigroup

Consider the Markov process $(X_t)_{t\geq 0}$known as the Ornstein-Ulenbeck process, which is given by the stochastic differential equation:

$$dX_t = -X_t dt + \sqrt{2} dB_t$$

This is the continuous limit of the auto-regressive process $AR(1)$. The stationary measure for this process is the standard Gaussian $N(0,1)$. The process can be realized in terms of Brownian motion, B_s,by:

$$X_t = e^{-t} X_0 + e^{-t} B_{(e^{2t}-1)}$$

This is the most useful form of the O-U process. From this representation, one can see that the semi-group of operators are given by:

$$(P_t f)(x) = \mathbf{E}_Z \left[f\left(e^{-t} x + \sqrt{1-e^{-2t}} Z \right) \right]$$

Where $Z \sim N(0,1)$ is a standard Gaussian variable. In particular, this representation shows that $N(0,1)$ is the equilibrium measure. We can also calculate the generator:

$$\begin{aligned} (Lf)(x) &= \partial_t P_t f|_{t=0} \\ &= \mathbf{E}_Z \left[f'\left(e^{-t} x + \sqrt{1-e^{-2t}} Z \right) \left(-e^{-t} x + \frac{e^{-2t}}{\sqrt{1-e^{-2t}}} Z \right) \right] \Big|_{t=0} \end{aligned}$$

From here, one can use the Gaussian integration by parts formula that $\mathbf{E}_Z(Zg(Z)) = \mathbf{E}_Z(g'(Z))$ to get:

$$(Lf)(x) = f''(x) - x f'(x)$$

Similarly, one can compute:

$$\mathcal{E}(f,g) = \mathbf{E}_Z(f'(Z) g'(Z))$$

The $d-$dimensional O-U process is created by running d independent O-U processes $X_t^1, \ldots X_t^d$ and using these as coordinates for a process in $\mathbb{R}^d$, i.e. $X_t = \left(X_t^1, \ldots, X_t^d\right)$. From our results in the $1-D$ case, it is easy to see the semi-group of operators and the generator for this Markov process:

$$(Lf)(\vec{x}) = (\Delta f)(\vec{x}) - \vec{x} \cdot ((\nabla f)(\vec{x}))$$

The equilibrium distribution is γ^d, the standard d-dimensional Gaussian random variable. The semi-group of operators and Dirichlet form are (here $Z \overset{d}{\sim} \gamma^d$):

$$\begin{aligned} (P_t g)(\vec{x}) &= \mathbf{E}_Z \left[g\left(e^{-t}\vec{x} + \sqrt{1-e^{-2t}} Z \right) \right] \\ \mathcal{E}(f,g) &= \mathbf{E}_Z \left(\sum_{i=1}^n (\partial_i f)(Z) (\partial_i g)(Z) \right) \\ &= \mathbf{E}_{\gamma^d}(\nabla f \cdot \nabla g) \end{aligned}$$

THEOREM 3.1. (Poincare Inequality for the O-U process)
The multidimensional O-U process satisfies the Poincare inequality with constant $C = 1$. That is to say:

$$\begin{aligned} \mathbf{Var}_{\gamma^d}(f) &\leq \mathcal{E}(f,f) \\ &= \mathbf{E}_{\gamma^d}\left(|\nabla f|^2 \right) \end{aligned}$$

Moreover, the constant $C = 1$ is optimal. That is to say, the inequality is not true for $C < 1$.

REMARK. The statement of this inequality has nothing to do with the O-U process, rather it is a statement about the variance of functions $f : \mathbb{R}^d \to \mathbb{R}$ when we consider the random variable $f(Z)$ with $Z \sim N(0,1)^d$ the standard Gaussian variable in $\mathbb{R}^d$. This is why this theorem will be important to us.

Of course the *proof* of the theorem goes through the machinery of our reversible Markov chains for the O-U process and only works because the O-U process is a reversible Markov chain with the Gaussian measure γ^d as its stationary distribution.

PROOF. We will use the covariance lemma and the Cauchy Shwarz and Jenson inequalities to get the result. To simply the terms that appear in the covariance lemma, we use the fact that for the O-U process that $\partial_{x_i} (P_t g)(\vec{x}) = e^{-t} P_t (\partial_{x_i} g)(\vec{x})$. (This can be seen from the expression $(P_t g)(\vec{x}) = \mathbf{E}_Z \left[g\left(e^{-t}\vec{x} + \sqrt{1-e^{-2t}} Z \right) \right]$ we saw earlier). Since P_t is linear (just as expectation is), this means that $\nabla (P_t g) = e^{-t} (P_t (\nabla g))$. Now consider:

$$\begin{aligned}
\mathbf{Cov}_{\gamma^d}(f,g) &= \int_0^\infty \mathcal{E}(f, P_t g)\, dt \\
&= \int_0^\infty \left(\mathbf{E}_\mu \left[\nabla f \cdot \nabla (P_t g) \right] \right) dt \\
&= \int_0^\infty \left(\mathbf{E}_\mu \left[\nabla f \cdot e^{-t} (P_t (\nabla g)) \right] \right) dt \\
&= \int_0^\infty e^{-t} \left\langle \nabla f, P_t (\nabla g) \right\rangle_{L^2(\gamma^d)} dt \\
&\leq \int_0^\infty e^{-t} \|\nabla f\|_{L^2(\gamma^d)} \|P_t (\nabla g)\|_{L^2(\gamma^d)} dt
\end{aligned}$$

The last line follows by the Cauchy Schwarz inequality. Finally, we use the inequality that $\|P_t (\nabla g)\|_{L^2(\mu)} \leq \|\nabla g\|_{L^2(\mu)}$ (See **??**) Have:

$$\begin{aligned}
\mathbf{Cov}_{\gamma^d}(f,g) &\leq \int_0^\infty e^{-t} \|\nabla f\|_{L^2(\gamma^d)} \|\nabla g\|_{L^2(\gamma^d)} dt \\
&= \|\nabla f\|_{L^2(\gamma^d)} \|\nabla g\|_{L^2(\gamma^d)}
\end{aligned}$$

Taking $f = g$ gives:

$$\begin{aligned}
\mathbf{Var}_{\gamma^d}(f) &\leq \|\nabla f\|^2_{L^2(\gamma^d)} \\
&= \langle \nabla f, \nabla f \rangle_{L^2(\gamma^d)} \\
&= \mathbf{E}_{\gamma^d} \left(|\nabla f|^2 \right) \\
&= \mathcal{E}(f,f)
\end{aligned}$$

The function $f(x_1, \ldots x_n) = \sum_i x_i$ shows that this bound is optimal since in this case $\mathbf{Var}_{\gamma^d}(f) = d$ while $\mathbf{E}_{\gamma^d} \left(|\nabla f|^2 \right) = \mathbf{E}_{\gamma^d}(d) = d$, so the constant $C = 1$ is optimal for this Poincare inequality.

□

4. Applications

Recall in our (1+1)-dimensional Gaussian polymer model that we had for paths $P = \{(0,0),(1,a_1),(2,a_2)\ldots,(n,a_n)\} \in \{-1,1\}^n$ the ground state energy E_n given by:

$$E_n = -\max_{|P|=n} \sum_{i=1}^{n} g_{(i,a_i)}$$

We think of E_n here as a function $E_n : \mathbb{R}^{n^2} \to \mathbb{R}$ where we evaluate $E_n\left(\{g_v : v \in [n]\times[n]\}\right)$ at Gaussian random variables on the grid. A bit of thought reveals:

$$\partial_{g_v} E_n = -\mathbf{1}_{\{v \in \text{optimal path producing } E_n\}}$$

Hence $|\nabla E_n|^2 = n+1$ for any value of $g_v's$, since every path goes through exactly $n+1$ vertices. Hence the Poincare inequality above gives:

$$\begin{aligned} \mathbf{Var}_{\gamma^{n^2}}(E_n) &\leq \mathbf{E}_{\gamma^{n^2}}\left(|\nabla E_n|^2\right) \\ &= n+1 \end{aligned}$$

Exercise 4.1. Prove that if $(g_1,\ldots,g_n)$ are jointly Gaussian (not necessarily uncorrelated) that:

$$\mathbf{Var}\left(\max_{1\leq i\leq n} g_i\right) \leq \max_{1\leq i\leq n} \mathbf{Var}(g_i)$$

Exercise 4.2. If $F_n(\beta)$ is the free energy of the the Sherington-Kirkpatrick Model of Spin Glasses, then:

$$\mathbf{Var}\left(F_n(\beta)\right) \leq C(\beta)n$$

Recall here that $H_n(\sigma) = -\frac{1}{\sqrt{n}}\sum g_{ij}\sigma_i\sigma_j$ and $F_n(\beta) = -\frac{1}{\beta}\log\left(\sum_\sigma \exp\left(-\beta H_n(\sigma)\right)\right)$.

We will see in the next lecture that the Poincare inequality is not providing a very good bound for these problems. That is to say, the true variance is less than what is being bounded here, but the Poincare inequality fails to capture that. The idea that the Poincare inequality is suboptimal is related to the idea of superconcentration, which we will look at next lecture.

CHAPTER 3

Superconcentration and Chaos

1. Superconcentraion

We say last time that the Poincare inequality can be used to show that, for a Gaussian vector $(g_1, \dots g_n)$ that:

$$\mathbf{Var}\left(\max_{1\leq i\leq n} g_i\right) \leq \max_{1\leq i\leq n} \mathbf{Var}\left(g_i\right)$$

In the case of iid $N(0,1)$ variables, this is saying:

$$\mathbf{Var}\left(\max_{1\leq i\leq n} g_i\right) \leq 1$$

However, it is a fact that for $N(0,1)$ variables that:

$$\mathbf{Var}\left(\max_{1\leq i\leq n} g_i\right) \leq \frac{C}{\log n}$$

Which is a much better bound. The idea that the classical Poincare inequality is suboptimal is the idea for superconcentration.

Definition 1.1. *($\epsilon-$Superconcentration)*
Suppose we have a Markov process $X_t \in S$ with Dirchelet form $\mathcal{E}$ and equilibrium measure μ. Suppose also that this Markov process satisfies a Poincare inequality with constant C. That is $\mathbf{Var}_\mu\left(f\right) \leq C\mathcal{E}\left(f,f\right)$.
A function $f : S \to \mathbb{R}$ is called *ϵ-superconcentrated* when we have the inequality:

$$\mathbf{Var}_\mu\left(f\right) \leq \epsilon C\mathcal{E}\left(f,f\right)$$

Example 1.2. If we take our familiar example of the O-U process on $\mathbb{R}^d$ that we have been considering, we know that we have a Poincare inequality with $C = 1$. If we look at the max however, we have:

$$\mathbf{Var}_{\gamma^d}\left(\max_{1\leq i\leq n} g_i\right) \leq \frac{C}{\log n}$$

Which means that the function $f(x_1, \dots, x_n) = \max_{1\leq i\leq n} x_i$ is $\frac{C}{\log n}$-superconcentrated. Note that this really depends on the function. For example $\mathbf{Var}_{\gamma^d}\left(\sum_i g_i\right) = n$ shows that the function $f(x_1, \dots x_n) = \sum_i x_i$ is not superconcentrated (Technically speaking, it is "1-superconcentrated" but we really want ϵ small and preferably going to zero as some parameter $n \to \infty$)

Problem 1.3. Suppose that $(g_1, \dots, g_n)$ are jointly Gaussian, but not necessarily independent. We know that if they are independent, we have superconcentration. More generally, under what conditions is $\max_{1\leq i\leq n} g_i$ super concentrated? In other words, do small correlations ruin the superconcentration?

As a concrete example, let $(g_1, \ldots, g_n)$ be iid $N(0,1)$ variables and consider for configurations $\sigma \in \{-1,+1\}^n$ the function $H(\sigma) = \frac{1}{\sqrt{n}} \sum_i g_i \sigma_i$. The vector $(H(\sigma))_{\sigma \in \{-1,+1\}^n} \in \mathbb{R}^{2^n}$ is jointly Gaussian (each entry is a linear combination of independent Gaussian).

The correlations $\mathbf{Corr}\left(H(\sigma), H(\sigma')\right) = \frac{1}{n} \sum_i \sigma_i \sigma_i'$, which are small. ($\frac{1}{n} \sum_i \sigma_i \sigma_i' = \frac{1}{n}$ (# of agree coords - # of disagree coords), so if we choose two configurations uniformly at random from $\{-1,+1\}^n$, then each coordinate has a probability $\frac{1}{2}$ of agreement, and probability of $\frac{1}{2}$ of disagreeing. Thus $\frac{1}{n} \sum_i \sigma_i \sigma_i' = \frac{1}{n} \sum_i X_i$ where each $X_i = \pm 1$ with probability $\frac{1}{2}$ for each outcome. By the central limit theorem, we know that $\frac{1}{n} \sum_i \sigma_i \sigma_i' \approx N(0, \frac{1}{n})$ so this is really small on average)

However, even though the correlations are small, the variance of the maximum is still $O(1)$:

$$\begin{aligned} \max_{\sigma \in \{-1,+1\}^n} H(\sigma) &= \frac{1}{\sqrt{n}} \sum_{i=1}^n |g_i| \\ \mathbf{Var}\left(\max_{\sigma \in \{-1,+1\}^n} H(\sigma) \right) &= \frac{1}{n} \sum_{i=1}^n C \\ &= C \end{aligned}$$

Where C here is the variance of the absolute value of a $N(0,1)$ random variable. This example shows that even a small amount of correlation can (potentially) ruin superconcentration.

On the other hand, if we consider $H(\sigma) = \frac{1}{n} \sum g_{ij} \sigma_i \sigma_j$, this *is* (supposed to be) superconcentrated. This is an open problem.

2. Fourier Expansions

Recall our setup for studying Markov processes, we had:

$(X_t)_{t \geq 0}$	–	A time reversible Markov process
$(P_t)_{t \geq 0}$	–	Semigroup of operators for evolving the process in time
L	–	Generator of the semigroup of operators (self-adjoint)
μ	–	Stationary distribution for the process
$\mathcal{E}$	–	Dirchelet form (symmetric)

In this section we will do some more analysis by taking an eigenbasis for L. This is akin to doing Fourier analysis (where one takes the eigenbasis for the Laplacian).

Proposition 2.1. *The generator L is negative semi-definite.*

Proof. By the Cauchy-Schwarz inequality and the Jensen inequality $\|P_t f\|_{L^2(\mu)} \leq \|f\|_{L^2(\mu)}$we have:

$$\begin{aligned} \mathbf{E}_\mu \left(f \left(P_t f \right) \right) &= \langle f, P_t f \rangle_{L^2(\mu)} \\ &\leq \|f\|_{L^2(\mu)} \|P_t f\|_{L^2(\mu)} \\ &\leq \|f\|_{L^2(\mu)}^2 \\ &= \mathbf{E}_\mu \left(f^2 \right) \end{aligned}$$

Hence, rearranging and dividing by t gives:

$$\mathbf{E}_\mu\left(f\left(\frac{P_t f - f}{t}\right)\right) \leq 0 \,\forall t \geq 0$$

Taking the limit $t \to 0$ now gives the result:

$$\begin{aligned}\lim_{t\to 0}\mathbf{E}_\mu\left(f\left(\frac{P_t f - f}{t}\right)\right) &\leq 0\\ \mathbf{E}_\mu\left(f\left(Lf\right)\right) &\leq 0\\ \left\langle f, Lf\right\rangle_{L^2(\mu)} &\leq 0\end{aligned}$$

□

REMARK. Since L is a non-negative definite operator, we might expect L to have some nice eigenvalues and eigenfunction. Often, under some mild conditions, this is indeed the case and the eigenvalues of $-L$ form an increasing sequence:

$$0 = \lambda_0 \leq \lambda_1 \leq \dots$$

Notice that 0 is always the smallest eigenvalue since constant functions are in the null space of L. (Notice $P_t 1 = \mathbf{E}_\mu(1) = 1$, so $L(1) = \lim_{t\to 0} t^{-1}(P_t 1 - 1) = 0$) One can show that the eigenspace of $\lambda_0 = 0$ is simple because if $Lf = 0$ then $P_t f = e^{tL} f = f$ for every t. But since $\lim_{t\to\infty} P_t f = \mathbf{E}_\mu(f)$ then f must be constant. Hence $f = \mathbf{E}_\mu(f)$ is a constant.

THEOREM 2.2. *If the eigenvalue $\lambda_1 > 0$ we say that there is a "spectral gap" . In this case we can prove that the Poincare inequality holds with constant $C = \frac{1}{\lambda_1}$. That is to say, we have:*

$$\mathbf{Var}_\mu(f) \leq \frac{1}{\lambda_1}\mathcal{E}(f,f)$$

PROOF. In the case that L has eigenvalues, just as in ordinary Fourier theory, we can establish the Plancherel identity:

$$\|f\|^2_{L^2(\mu)} = \sum_{k=0}^{\infty}\langle u_k, f\rangle^2_{L^2(\mu)}$$

We know $u_0 = 1$ though so $\langle u_k, f\rangle_{L^2(\mu)} = \mathbf{E}_\mu(f)$ this is saying then that:

$$\|f\|^2_{L^2(\mu)} - \mathbf{E}_\mu(f)^2 = \sum_{k=1}^{\infty}\langle u_k, f\rangle^2_{L^2(\mu)}$$

We now recognize the left hand side as $\mathbf{Var}_\mu(f)$. Now on the other hand, we write $P_t f$ in its eigenbasis to see that:

$$\begin{aligned}\mathcal{E}(f,f) &= -\langle f, Lf\rangle_{L^2(\mu)}\\ &= \sum_{k=0}^{\infty}\lambda_k\langle u_k, f\rangle^2_{L^2(\mu)}\\ &= \sum_{k=1}^{\infty}\lambda_k\langle u_k, f\rangle^2_{L^2(\mu)} \text{ since } \lambda_0 = 0\end{aligned}$$

In particular, since the eigenvalues are increasing $\lambda_1 \leq \lambda_2 \leq \ldots$, we have finally:

$$\begin{aligned}\mathbf{Var}_\mu(f) &= \sum_{k=1}^{\infty} \langle u_k, f\rangle^2_{L^2(\mu)} \\ &\leq \frac{1}{\lambda_1}\sum_{k=1}^{\infty} \lambda_k \langle u_k, f\rangle^2_{L^2(\mu)} \\ &= \frac{1}{\lambda_1}\mathcal{E}(f,f)\end{aligned}$$

□

REMARK. In this framework of eigenvalues and eigendecompositions, we can see that the phenomenon of $\epsilon-$superconcentration just means that $\mathbf{Var}_\mu(f) \ll \frac{1}{\lambda_1}\mathcal{E}(f,f)$ is $\lambda_1 \sum_{k=1}^{\infty} \langle u_k, f\rangle^2_{L^2(\mu)} \ll \sum \lambda_k \langle u_k, f\rangle^2_{L^2(\mu)}$. Heuristically, this is saying that "most of the Fourier mass concentrates on large eigenvalues".

This appears in the literature under the name of "noise-sensitive" functions. However, only boolean functions are studied for the most part.

3. Chaos

DEFINITION 3.1. We say that a function f is $\epsilon - \delta-$*chaotic* if $\forall t \geq \delta$:

$$\mathcal{E}(f, P_t f) \leq \epsilon e^{-\lambda_1 t}\mathcal{E}(f,f)$$

REMARK. Notice the inequality is *always* true if the ϵ is removed. That is:

$$\mathcal{E}(f, P_t f) \leq e^{-\lambda_1 t}\mathcal{E}(f,f)$$

This can be seen from the Fourier analysis we did earlier, write $f = \sum_i \langle u_k, f\rangle_{L^2(\mu)} u_k$. Hence:

$$P_t f = e^{tL} f = \sum_{k=0}^{\infty} \lambda_k e^{-\lambda_k t} \langle u_k, f\rangle_{L^2(\mu)} u_k$$

Now using $\mathcal{E}(f,f) = \sum_{k=1}^{\infty} \lambda_k \langle u_k, f\rangle^2$, and $\lambda_1 < \lambda_2 < \ldots$ we have:

$$\begin{aligned}\mathcal{E}(f, P_t f) &= \sum_{k=0}^{\infty} \lambda_k e^{-\lambda_k t} \langle u_k, f\rangle_{L^2(\mu)} u_k \\ &\leq \sum_{k=0}^{\infty} \lambda_k e^{-\lambda_1 t} \langle u_k, f\rangle_{L^2(\mu)} u_k \\ &= e^{-\lambda_1 t}\sum_{k=0}^{\infty} \lambda_k \langle u_k, f\rangle_{L^2(\mu)} u_k \\ &= e^{-\lambda_1 t}\mathcal{E}(f,f)\end{aligned}$$

REMARK.

PROPOSITION 3.2. *We will give a condition on the polymer model that is sufficient to show that E_n is $\epsilon - \delta-$chaotic.*

In the Polymer model, consider the ground state energy E_n. This depends on the environment of iid Gaussian variables $(g_v)_{v\in\mathbb{Z}^2}$. Let $(h_v)_{v\in\mathbb{Z}^2}$ be another set of iid Gaussian random variables. Fix a time parameter t and define a new realization of the environment $(g_v^t)_{v\in\mathbb{Z}^2}$ by $g_v^t = e^{-t}g_v + \sqrt{1-e^{-2t}}h_v$. Let $\hat{p}$ be the optimal path

that achieved E_n in the environment $(g_v)_{v\in\mathbb{Z}^2}$ and let $\hat{p}^t$ be the optimal path in the environment $(g'_v)_{v\in\mathbb{Z}^2}$. Let $|\hat{p}\cap\hat{p}^t|$ be the number of vertices they have in common.

If for all $t>\delta$, $\mathbf{E}\left(|\hat{p}\cap\hat{p}^t|\right)\leq\epsilon(n+1)$, then E_n is $\epsilon-\delta-$chaotic.

PROOF. The proof use the O-U process. Recall that when we use the O-U process we have $\lambda_1=1$ (it is the size of the constant in the Poincare inequality.) We've seen already that the ground state energy E_n has:

$$\begin{aligned}\frac{\partial E_n}{\partial g_v} &= \mathbf{1}_{\{v\in\hat{p}\}}\\ \mathcal{E}(E_n,E_n) &= \mathbf{E}_{\gamma^d}\left(|\nabla E_n|^2\right)=n+1\end{aligned}$$

Now, consider that:

$$\begin{aligned}\mathcal{E}(E_n,P_tE_n) &= \mathbf{E}_{\gamma^d}\left(\nabla E_n\cdot\nabla\left(P_tE_n\right)\right)\\ &= e^{-t}\mathbf{E}_{\gamma^d}\left(\nabla E_n\cdot P_t\left(\nabla E_n\right)\right)\\ &= e^{-t}\sum_v\mathbf{E}_{\gamma^d}\left(\frac{\partial E_n}{\partial g_v}\cdot P_t\left(\frac{\partial E_n}{\partial g_v}\right)\right)\end{aligned}$$

But for a function f, we have

$$\begin{aligned}\mathbf{E}_{\gamma^d}\left(f\left(P_tf\right)\right) &= \mathbf{E}_Z\left(f(Z)\left(P_tf\right)(Z)\right)\\ &= \mathbf{E}_{Z,Z'}\left(f(Z)f\left(e^{-t}Z+\sqrt{1-e^{-2t}}Z'\right)\right)\end{aligned}$$

Where $Z,Z'$are iid $N(0,1)^d$ variables. This works because of the expression for the O-U process $X_t=e^{-t}X_0+e^{-t}B_{e^{2t}-1}\overset{d}{=}e^{-t}X_0+N\left(0,\sqrt{1-e^{-2t}}\right)$. Notice that if we identify $Z=g_v$ and $Z=h_v$, then g_v^t is precisly $e^{-t}Z+\sqrt{1-e^{-2t}}Z'$. That is:

$$\mathbf{E}_{\gamma^d}\left(f\left(P_tf\right)\right)=\mathbf{E}_{g_v,g'_v}\left(f(g_v)f\left(g_v^t\right)\right)$$

In our case, $\frac{\partial E_n}{\partial g_v}=\mathbf{1}_{\{v\in\hat{p}\}}$ so $\frac{\partial E_n}{\partial g_v}(g_v)\frac{\partial E_n}{\partial g_v}(g'_v)=\mathbf{1}_{\{v\in\hat{p}\}}\mathbf{1}_{\{v\in\hat{p}^t\}}=\mathbf{1}_{\{v\in\hat{p},v\in\hat{p}^t\}}$ So we have:

$$\begin{aligned}\mathcal{E}(E_n,P_tE_n) &= e^{-t}\sum_v\mathbf{E}_{\gamma^d}\left(\frac{\partial E_n}{\partial g_v}\cdot P_t\left(\frac{\partial E_n}{\partial g_v}\right)\right)\\ &= e^{-t}\sum_v\mathbf{P}_{g_v,g'_v}\left(v\in\hat{p}\text{ and }v\in\hat{p}^t\right)\\ &= e^{-t}\mathbf{E}\left(\left|\hat{p}\cap\hat{p}^t\right|\right)\end{aligned}$$

From this, the definition of $\epsilon-\delta-$chaotic, and the fact that $\mathcal{E}(f,f)=n+1$ it is clear that $\mathbf{E}\left(|\hat{p}\cap\hat{p}^t|\right)\leq n+1$ gives that E_n is $\epsilon-\delta-$chaotic.

□

EXERCISE 3.3. Show that in the S-K model for spin glasses, prove the following sufficient condition for the free energy $F_n(\beta)$ to be $\epsilon-\delta-$chaotic.

Define g_v^t as in the previous statement. Let $R_{1,2}(t)$ be the average overlap between configurations chosen from the measure induced by g_v and g_v^t. If $\forall t>\delta$, $\mathbf{E}\left(R_{1,2}^2(t)\right)\leq C(\beta)\epsilon$, then $F_n(\beta)$ is $\epsilon-\delta-$chaotic.

4. Equivalence of Chaos and Superconcentration

The ideas of $\epsilon - \delta-$chaotic and $\epsilon-$superconcentration are actually equivalent if one adjusts the constants a bit as the following two theorems make precise. The idea of the proof is to use the fact that $e^{\lambda_1 t}\mathcal{E}(f, P_t f)$ is decreasing (which is apparent from the Fourier analysis) and then use the covariance lemma $\mathbf{Var}_\mu(f) = \int_0^\infty \mathcal{E}(f, P_t f)\ dt$.

With this relation in hand, we can see heuristically the relationship. If we have superconcentration, then $\mathbf{Var}_\mu(f)$ is small, so $\mathcal{E}(f, P_t f)$ must be decreasing quickly, which is $\epsilon - \delta-$chaos. If we have $\epsilon - \delta-$chaos then $\mathcal{E}(f, P_t f)$ is decreasing quickly, so $\mathbf{Var}_\mu(f)$ must be small, which is superconcentration.

THEOREM 4.1. *If f is $\epsilon-$superconcentrated then $\forall \delta > 0$, f is $\epsilon' - \delta-$chaotic where $\epsilon' = \frac{\epsilon}{\lambda_1 \delta}$*

PROOF. Suppose that f is $\epsilon-$superconcentrated. Then we know that:

$$\mathbf{Var}_\mu(f) \leq \frac{\epsilon}{\lambda_1}\mathcal{E}(f, f)$$

Now, for any $\delta > 0$, observe that $\mathcal{E}(f, P_t f) = \sum_{k=1}^\infty \lambda_k e^{-\lambda_k t} \langle u_k, f\rangle^2_{L^2(\mu)}$ is a *decreasing* function of t because of the decaying exponential term. Moreover, since $\lambda_1 < \lambda_2 < \ldots$ we can do even better and say that $e^{\lambda_1 t}\mathcal{E}(f, P_t f) = \sum_{k=1}^\infty \lambda_k e^{-(\lambda_k - \lambda_1)t} \langle u_k, f\rangle^2_{L^2(\mu)}$ is a *decreasing* function of t. Its also apparent from this formulation that $\mathcal{E}(f, P_t f) \geq 0$ is *non-negative*. We use these facts along with the covariance lemma $\mathbf{Var}_\mu(f) = \int_0^\infty \mathcal{E}(f, P_t f)$ to see, that for any $\delta > 0$:

$$\begin{aligned}
\frac{\epsilon}{\lambda_1}\mathcal{E}(f, f) &\geq \mathbf{Var}_\mu(f) \\
&= \int_0^\infty \mathcal{E}(f, P_t f)\ dt \\
&\geq \int_0^\delta \mathcal{E}(f, P_t f)\ dt \text{ since } \mathcal{E}(f, P_t f) \geq 0 \\
&= \int_0^\delta e^{-\lambda_1 t} e^{\lambda_1 t}\mathcal{E}(f, P_t f)\ dt \\
&\geq \int_0^\delta e^{-\lambda_1 t}\left(e^{\lambda_1 \delta}\mathcal{E}(f, P_\delta f)\right)\ dt \text{ since } e^{\lambda_1 t}\mathcal{E}(f, P_t f) \text{ decreasing} \\
&= e^{\lambda_1 \delta}\mathcal{E}(f, P_\delta f)\int_0^\delta e^{-\lambda_1 t}\ dt \\
&= e^{\lambda_1 \delta}\mathcal{E}(f, P_\delta f)\frac{1 - e^{-\lambda_1 \delta}}{\lambda_1} \\
&= \mathcal{E}(f, P_\delta f)\frac{e^{\lambda_1 \delta} - 1}{\lambda_1} \\
&\geq \frac{e^{\lambda_1 \delta}}{\lambda_1}\mathcal{E}(f, P_\delta f)
\end{aligned}$$

Finally, since $e^{\lambda_1 t}\mathcal{E}(f, P_t f)$ is decreasing, we have that for all $t > \delta$ that:

$$\begin{aligned} e^{\lambda_1 t}\mathcal{E}(f, P_t f) &\leq e^{\delta t}\mathcal{E}(f, P_\delta f) \\ &\leq \frac{\epsilon}{\lambda_1 \delta}\mathcal{E}(f, f) \\ &= \epsilon'\mathcal{E}(f, f) \end{aligned}$$

Dividing through by $e^{\lambda_1 t}$ gives the desired result.

□

THEOREM 4.2. *If f is $\epsilon - \delta-$chaotic, then f is $\epsilon'-$superconcentrated, where $\epsilon' = \epsilon + \lambda_1 \delta$.*

PROOF. We again use the fact that $\mathcal{E}(f, P_t f)$ and $e^{\lambda_1 t}\mathcal{E}(f, P_t f)$ are decreasing. Also recall that $P_0 f = f$. Have:

$$\begin{aligned} \mathbf{Var}_\mu(f) &\leq \int_0^\infty \mathcal{E}(f, P_t f)\, dt \\ &= \int_0^\delta \mathcal{E}(f, P_t f)\, dt + \int_\delta^\infty e^{-\lambda_1 t}\left(e^{\lambda_1 t}\mathcal{E}(f, P_t f)\right) dt \\ &\leq \int_0^\delta \mathcal{E}(f, P_0 f)\, dt + \int_\delta^\infty e^{-\lambda_1 t}\left(e^{\lambda_1 \delta}\mathcal{E}(f, P_\delta f)\right) dt \\ &= \delta\mathcal{E}(f, f) + e^{\lambda_1 \delta}\mathcal{E}(f, P_\delta f)\int_\delta^\infty e^{-\lambda_1 t} dt \\ &= \delta\mathcal{E}(f, f) + \left(e^{\lambda_1 \delta}\mathcal{E}(f, P_\delta f)\right)\frac{e^{-\lambda_1 \delta}}{\lambda_1} \\ &= \delta\mathcal{E}(f, f) + \left(\epsilon\mathcal{E}(f, f)\right)\frac{e^{-\lambda_1 \delta}}{\lambda_1} \end{aligned}$$

On the last line we have used the definition of $\epsilon-\delta-$chaotic when $t = \delta$, namely $e^{\lambda_1 \delta}\mathcal{E}(f, P_\delta f) \leq \epsilon\mathcal{E}(f, f)$. Finally, we bound $e^{-\lambda_1 \delta} \leq 1$ to get:

$$\begin{aligned} \mathbf{Var}_\mu(f) &\leq \delta\mathcal{E}(f, f) + \left(\epsilon\mathcal{E}(f, f)\right)\frac{1}{\lambda_1} \\ &= \frac{1}{\lambda_1}\left(\epsilon + \delta\lambda_1\right)\mathcal{E}(f, f) \end{aligned}$$

Which is exactly $\epsilon'-$superconcentration, since $C = \frac{1}{\lambda_1}$ in the Poincare inequality.

□

EXERCISE 4.3. In the SK spin glass model, show the equivalence of superconcentration for $F_n(\beta)$ and chaos in the disorder.

CHAPTER 4

Multiple Valleys in the Polymer Model and Hypercontractivity

1. The Polymer Model

EXERCISE 1.1. In the polymer model, prove that

$$\begin{array}{ccc} \mathbf{Var}\left(E_n\right) & = & o(n) \\ & \Longleftrightarrow & \\ \exists \delta_n \to 0 \text{ s.t. } \forall t \geq \delta_n \quad \text{we have} & & \sup_{t \geq \delta_n} \mathbf{E}\left(|\hat{p} \cap \hat{p}_n^t|\right) = o(n) \end{array}$$

This is a corollary of the equivalence of superconcentration and chaos, and results we have proven earlier.

THEOREM 1.2. *Assume that there is a sequence $\epsilon_n \to 0$ and $\delta_n \to 0$ so that the ground state energy E_n in the Polymer model has the $\epsilon_n - \delta_n$-chaos property. Then E_n has the multiple valleys property.*

REMARK. Here is an outline of the proof. Remember that for the polymer model we have an energy functional $H(\sigma)$ and we are interested in the ground state energy E_n which is the minimal value of $H(\sigma)$. The idea of the proof can be broken into a few steps, which are outlined below. We assume here that E_n satisfies $\epsilon_n - \delta_n-$chaos here for some sequence $\epsilon_n \to 0$ and $\delta_n \to 0$ (This will be proven later.)

(1) The function $H(\sigma)$ depends only on the random environment (g_v). If we run an O-U process for some small time t we get a new environment (g_v^t) and new function $H^t(\sigma)$. Since we will run this for a small amount of time, g_v and g_v^t will be highly correlated.
(2) Let $\hat{p}$ and $\hat{p}^t$ be the optimal paths in the environment (g_v) and (g_v^t) respectively. We know from our previous work on the Polymer Model that $\epsilon_n - \delta_n-$chaos here implies that the perturbed path $\hat{p^t}$ and the original path $\hat{p}$ are almost disjoint, in the sense that $\mathbf{E}\left(|\hat{p} \cap \hat{p}^t|\right)$ is small.
(3) Now, since $H(\hat{p}^t) \approx H^t(\hat{p}^t)$ (since the measure (g_v) is highly correlated with (g_v^t)), and since $H^t(\hat{p}^t) \approx \mathbf{E}\left(E_n\right) \approx H(\hat{p})$ (since E_n is superconcentrated, so it has very low variance), we combine these approximations to get $H(\hat{p}^t) \approx H\left(\hat{p}\right)$.
(4) But now $\hat{p}^t$ is a path that is almost disjoint from $\hat{p}$ and yet is very close to the minimum energy $E_n = H\left(\hat{p}\right) \approx H(\hat{p}^t)$.
(5) By controlling the size of the approximation, and doing this for many independent perturbations $\left(g_v^{t_1}\right), \ldots, \left(g_v^{t_k}\right)$ we can find many paths that are near the ground state energy, and yet do not overlap very much.

PROOF. To follow in a later version...

□

2. Hypercontractivity

In general there are two ways to prove function are superconcentrated:

(1) Hypercontractivity
(2) Other methods

The idea of hypercontractivity is related to the inequality we proved earlier as a consequence of Jensen's inequality, namely:

$$\|P_t f\|_{L^2(\mu)} \leq \|f\|_{L^2(\mu)}$$

This is saying that P_t is contractive on $L^2(\mu)$. The hypercontractivity is an even stronger result, saying that P_t is extra contractive.

DEFINITION 2.1. For a Markov chain X_t with semigroup P_t and equilibrium distribution μ, we say that P_t is *hypercontractive* if $\forall p > 1$ and $t > 0$,$\exists q = q(t,p) > p$ s.t. $\forall f \in L^p(\mu)$ we have the inequality:

$$\|P_t f\|_{L^q(\mu)} \leq \|f\|_{L^p(\mu)}$$

REMARK. The hypercontractive inequality can be seen to be a *stronger* inequality than the Jensen inequality we had earlier, because on a probability space, for $p < q$ we have:

$$\|f\|_{L^p(\mu)} \leq \|f\|_{L^q(\mu)}$$

(This also comes from Jensen's inequality.) So the upper bound of the hypercontractive inequality is *smaller* than the usual upper bound: a stronger inequality.

THEOREM 2.2. (Hypercontractivity for the O-U semigroup)
Nelson (1973) proved that the O-U semigroup is hypercontractive in any dimension with:

$$q = 1 + (p-1)e^{2t}$$

This is a hard theorem and we will not prove it here. Instead we will assume it and use the result to examine superconcentration phenomenon. Below is an example of how the hypercontractive inequality can be useful.

PROPOSITION 2.3. *Let $g = (g_i)_{1 \leq i \leq d}$ be a standard $N(0,1)^d$ Gaussian. Let $h = (h_i)_{1 \leq i \leq d}$ be an independent $N(0,1)^d$ Gaussian. For some fixed $t > 0$, define $g^t = e^{-t} g + \sqrt{1 - e^{-2t}} h$ (this is flowing forward with the O-U process for a time t). Let $p = 1 + e^{-2t}$ and let A be any event. Then:*

$$\mathbf{P}\left(g^t \in A \middle| g \in A\right) \leq \mathbf{P}\left(g \in A\right)^{\frac{1}{2}\frac{1-e^{-2t}}{1+e^{-2t}}}$$

REMARK. Notice that for small t, we have $\frac{1}{2}\frac{1-e^{-2t}}{1+e^{-2t}} \approx \frac{1}{2}t + O(t^3)$. The is saying that if $\mathbf{P}\left(g \in A\right)$ is small, then $\mathbf{P}\left(g^t \in A | \, g \in A\right)$is small too. The interpretation is that if we are in a rare event A, even flowing forward in time by a small amount of time t is likely to make us exit the rare event.

PROOF. Let $f(x_1, \ldots, x_d) = \mathbf{1}_{\{x \in A\}}$. Let $p = 1 + e^{-2t}$ so that $q = 1 + (p-1)e^{2t} = 2$. Then by first using the Caucy shwarz inequality, and then using hypercontractivity inequality,$\|P_t f\|_{L^2} = \|P_t f\|_{L^q} \leq \|f\|_{L^p}$, we have:

$$\begin{aligned}\mathbf{P}\left(g^t \in A \cap g \in A\right) &= \mathbf{E}\left(f(g)f(g^t)\right)\\ &= \mathbf{E}_{Z,Z'}\left(f(Z)f(e^{-t}Z+\sqrt{1-e^{-2t}}Z'\right)\\ &= \mathbf{E}_{\gamma^d}\left(f\left(P_t f\right)\right)\\ &\leq \|f\|_{L^2}\|P_t f\|_{L^2}\\ &\leq \|f\|_{L^2}\|f\|_{L^p}\end{aligned}$$

Now, $\|f\|_{L^p}^p = \mathbf{E}_{\gamma^d}\left(\mathbf{1}_{\{g\in A\}}^p\right) = \mathbf{P}\left(g \in A\right)$. Hence $\|f\|_{L^2} = \mathbf{P}\left(g \in A\right)^{\frac{1}{2}}$, $\|f\|_{L^p} = \mathbf{P}\left(g \in A\right)^{\frac{1}{p}}$ so we have:

$$\begin{aligned}\mathbf{P}\left(g^t \in A \cap g \in A\right) &\leq \|f\|_{L^2}\|f\|_{L^p}\\ &= \mathbf{P}\left(g \in A\right)^{\frac{1}{2}+\frac{1}{p}}\end{aligned}$$

Dividing through by $\mathbf{P}\left(g \in A\right)$ and calculating $\frac{1}{2}+\frac{1}{p}-1 = \frac{1}{2}\frac{1-e^{-2t}}{1+e^{-2t}}$ gives:

$$\mathbf{P}\left(g^t \in A \middle| g \in A\right) \leq \mathbf{P}\left(g \in A\right)^{\frac{1}{2}\frac{1-e^{-2t}}{1+e^{-2t}}}$$

As desired.

□

CHAPTER 5

Hypercontractivity and the BKS trick

1. Talagrand's L1-L2 inequality

The Poincare inequality for the $n-$dimensional O-U process is:

$$\begin{aligned}\mathbf{Var}_{\gamma^n}(f) &\leq \mathbf{E}_{\gamma^n}\left(|\nabla f|^2\right)\\ &= \sum_{i=1}^n \|\partial_i f\|^2_{L^2(\gamma^n)}\end{aligned}$$

As we have seen this inequality is not always optimal. Part of the proof of the Poincare theorem is using the inequality $\|P_t f\|_{L^2} \leq \|f\|_{L^2}$, and we can improve the inequality a little bit by replacing this with the hypercontractive inequality $\|P_t f\|_{L^q} \leq \|f\|_{L^p}$. One way to do this will lead to *Talagrand's $L^1 - L^2$ inequality:*

$$\mathbf{Var}_{\gamma^n}(f) \leq C\sum_{i=1}^n \frac{\|\partial_i f\|^2_{L^2}}{1+\log\left(\frac{\|\partial_i f\|_{L^2}}{\|\partial_i f\|_{L^1}}\right)}$$

Here C is a universal constant that does not depend on n or on f. This inequality is not always an improvement over Poincare, but when $\|\partial_i f\|_{L^2} \gg \|\partial_i f\|_{L^1}$ the log term in the denominator is large, so we get substantial improvement.

EXAMPLE 1.1. For $f(x_1,\ldots,x_n) = \sum_i x_i, \partial_i f = 1$ so we have $\|\partial_i f\|_{L^1} = \|\partial_i f\|_{L^2} = 1$ so the log term vanishes and the inequality is $\mathbf{Var}_{\gamma^n}(f) \leq Cn$. We know in this case that $\mathbf{Var}_{\gamma^d}(f) = n$, so in this case Talagrands inequality is not improving the Poincare inequality.

EXAMPLE 1.2. We know that when $f(x_1,\ldots,x_n) = \max_i x_i$ that the Poincare inequality gives the suboptimal bound $\mathbf{Var}_{\gamma^d}(f) \leq 1$ (see **??**). Here we have that $\partial_i f = \mathbf{1}_{\{x_i \text{ is the max}\}}$ so $\|\partial_i f\|^2_{L^2} = \mathbf{P}_{\gamma^n}(g_i \text{ is the max}) = \frac{1}{n}$ by symmetry considerations. Hence $\|\partial_i f\|_{L^2} = \frac{1}{\sqrt{n}}$. Similarly, $\|\partial_i f\|_{L^1} = \mathbf{P}_{\gamma^n}(g_i \text{ is the max}) = \frac{1}{n}$ so $\frac{\|\partial_i f\|_{L^2}}{\|\partial_i f\|_{L^1}} = \sqrt{n}$ and Talagrands inequality gives:

$$\begin{aligned}\mathbf{Var}_{\gamma^n}(f) &\leq C\sum_{i=1}^n \frac{\frac{1}{n}}{1+\log n}\\ &\leq \frac{C}{\log n}\end{aligned}$$

Which is the correct order. Notice that this is a substancial improvement over the $O(1)$ bound that we got from the Poincare inequality alone.

EXERCISE 1.3. Consider a rooted binary tree of depth n. Suppose that each edge is given a weight $g_e \sim N(0,1)$ which are iid standard Gaussian variables.

Define the weight of each leaf l to be $w_l = \sum_{e \in \pi} g_e$, where π is the unique path from the root to the leaf l. Let $M = \max_l w_l$ be the maximum weight of a leaf.

This is the same as the position of the rightmost particle of a branching random walk with gaussian steps after n steps (at each step each particle splits into two new particles and each particle takes a step whose size is a Gaussian random variable.)

Prove that $\mathbf{E}(M) \sim Cn$ (easy)

Prove that $C = \sqrt{2 \log 2}$ (hard)

Use Talagrands $L^1 - L^2$ bound to prove that $\mathbf{Var}(M) \leq C \log n$. In this problem, it is know that $\mathbf{Var}(M) \leq K$ is bounded by a constant independent of n. There is no known simple proof of this fact however, it is really a very hard problem.

PROBLEM 1.4. Is the method of the $L^1 - L^2$ inequality necessary and sufficient to prove superconcentration? A recent result by Chatterjee shows that this is true for functions f which are monotone.

THEOREM 1.5. (Talagrands $L^1 - L^2$ Inequality)
For the O-U process semi-group we have that:

$$\mathbf{Var}_{\gamma^n}(f) \leq C \sum_{i=1}^{n} \frac{\|\partial_i f\|_{L^2}^2}{1 + \log\left(\frac{\|\partial_i f\|_{L^2}}{\|\partial_i f\|_{L^1}}\right)}$$

PROOF. As in the proof of the Poincare inequality, we use the covariance lemma to bound the variance (The beginning of this argument is identical to the proof of the Poincare inequality for the O-U process)

$$\begin{aligned}
\mathbf{Var}_{\gamma^d}(f) &= \int_0^\infty \mathcal{E}(f, P_t f)\, dt \\
&= \int_0^\infty \sum_{i=1}^{n} \left(\mathbf{E}_\mu\left[\partial_i f \cdot \partial_i (P_t f)\right]\right) dt \\
&= \int_0^\infty \left(\sum_{i=1}^{n} \mathbf{E}_\mu\left[\partial_i f \cdot e^{-t}(P_t(\partial_i f))\right]\right) dt \\
&= \int_0^\infty e^{-t} \sum_{i=1}^{n} \langle \partial_i f, P_t(\partial_i f)\rangle_{L^2(\gamma^d)}\, dt \\
&\leq \int_0^\infty e^{-t} \sum_{i=1}^{n} \|\partial_i f\|_{L^2(\gamma^d)} \|P_t(\partial_i f)\|_{L^2(\gamma^d)}\, dt
\end{aligned}$$

At this point, instead of using the Jensen inequality bound $\|P_t(\partial_i f)\|_{L^2(\gamma^d)} \leq \|\partial_i f\|_{L^2(\gamma^d)}$ we will use the hypercontractivity inequality bound for the O-U process that $\|P_t(\partial_i f)\|_{L^2(\gamma^d)} \leq \|\partial_i f\|_{L^p(\gamma^d)}$ where $p = 1 + e^{-2t}$. The form of p comes from the explicit formula we have for the hypercontractive inequality for the O-U process. Once we do this, since $1 < p < 2$ we have the opportunity to use Holder's inequality

$\|\partial_i f\|_{L^p} \leq \|\partial_i f\|_{L^2}^{2(p-1)/p} \|\partial_i f\|_{L^1}^{(2-p)/p}$. We have:

$$\begin{aligned}
\mathbf{Var}_{\gamma^d}(f) &\leq \int_0^\infty e^{-t} \sum_{i=1}^n \|\partial_i f\|_{L^2(\gamma^d)} \|\partial_i f\|_{L^p(\gamma^d)} \, dt \\
&\leq \int_0^\infty e^{-t} \sum_{i=1}^n \|\partial_i f\|_{L^2} \|\partial_i f\|_{L^2}^{2(p-1)/p} \|\partial_i f\|_{L^1}^{(2-p)/p} \, dt \\
&= \leq \sum_{i=1}^n \|\partial_i f\|_{L^2}^2 \left(\int_0^\infty e^{-t} \|\partial_i f\|_{L^2}^{(p-2)/p} \|\partial_i f\|_{L^1}^{(2-p)/p} \, dt \right)
\end{aligned}$$

To simplify this further, we write $a_i = \frac{\|\partial_i f\|_{L^2}}{\|\partial_i f\|_{L^1}} < 1$ and $\tanh(t) = \frac{e^t - e^{-t}}{e^t + e^{-t}} = \frac{p-2}{p}$ to get:

$$\mathbf{Var}_{\gamma^d}(f) \leq \sum_{i=1}^n \|\partial_i f\|_{L^2}^2 \left(\int_0^\infty e^{-t} a_i^{\tanh(t)} dt \right)$$

By the exercise below, $\int_0^\infty e^{-t} a_i^{\tanh(t)} dt \leq \frac{C}{1 - \log(a_i)}$, so we get:

$$\mathbf{Var}_{\gamma^n}(f) \leq C \sum_{i=1}^n \frac{\|\partial_i f\|_{L^2}^2}{1 + \log\left(\frac{\|\partial_i f\|_{L^2}}{\|\partial_i f\|_{L^1}} \right)}$$

As desired. □

EXERCISE 1.6. Prove that for a constant $a < 1$ that:

$$\int_0^\infty e^{-t} a^{\tanh(t)} dt \leq \frac{C}{1 - \log(a)}$$

2. The BKS Trick

Here we will use a trick due to Itai Benjamini, Gil Kalai, and Oded Schramm that allows one to show that the first passage time, T_n in the first passage percolation model is superconcentrated with $\mathbf{Var}(T_n) \leq Cn/\log n$. We will first introduce the problem and explain how the BKS trick gets around the difficulty.

PROPOSITION 2.1. *Consider the first passage percolation model with iid edge weights given by $w_e = |g_e|$ where $g_e \sim N(0,1)$ as usual. Let T_n be the first passage time to get from $(0,0)$ to $(n,0)$ and let $p_e = \mathbf{P}(e \in$ optimal path from $(0,0)$to $(n,0))$. Then:*

$$\mathbf{Var}(T_n) \leq C \sum_{e \in E(\mathbb{Z}^2)} \frac{p_e}{1 + \log\left(\frac{1}{p_e} \right)}$$

The result also holds when $w_e = F(g_e)$ for F a non-negative and differentiable Lipshitz function, $|F'| \leq C$.

PROOF. The proof uses the $L^1 - L^2$ inequality. We have, for any edge $e \in E(\mathbb{Z}^2)$ that:

$$\partial_e T_n = \mathbf{1}_{\{e \in \text{optimal path}\}}$$

So that $||\partial_e T_n||_{L^1} = p_e$ and $||\partial_e T_n||_{L^2} = \sqrt{p_e}$. Hence, by the $L^1 - L^2$ inequality, we have:

$$\begin{aligned}
\mathbf{Var}\left(T_n\right) &\leq C \sum_{e \in E(\mathbb{Z}^2)} \frac{\|\partial_e T_n\|_{L^2}^2}{1 + \log\left(\frac{\|\partial_e T_n\|_{L^2}}{\|\partial_e T_n\|_{L^1}}\right)} \\
&= C \sum_{e \in E(\mathbb{Z}^2)} \frac{p_e^2}{1 + \log\left(\frac{\sqrt{p_e}}{pe}\right)} \\
&\leq C \sum_{e \in E(\mathbb{Z}^2)} \frac{p_e}{1 + \log\left(\frac{1}{p_e}\right)}
\end{aligned}$$

Note that there is a little technical issue here to apply the $L^1 - L^2$ inequality to the infinite dimensional collection $(g_e)_{e \in E(\mathbb{Z}^2)}$, but this can be handled because we can show that the optimal path that achieves T_n can be controlled to stay in a finite (but large) box with high probability.

□

Remark. If we also knew that p_e was "small" for most e, this result would give us the result that $\mathbf{Var}(T_n) \sim Cn/\log(n)$. Unfortunately, there is no known elementary way to see this. The BKS trick lets us get around this difficulty by taking an average over many edges.

Theorem 2.2. (BKS inequality)
In the first passage percolation model as above, with $w_e = |g_e|$ we have:

$$\mathbf{Var}\left(T_n\right) \leq C \frac{n}{\log n}$$

Proof. (We will use $T = T_n$ for shorthand in this proof.) We will create a variable $\tilde{T}$ which approximates T. We will first show that the variance of T and $\tilde{T}$ are not too different, and then use the $L^1 - L^2$ theorem to prove the $\tilde{T}$ has the required variance.

Fix a constant k (to be chosen later) and let B be the $(2k+1) \times (2k+1)$ box centered around the origin. For every vertex $x \in B$, let T_x be the first passage time for the path between x and $x + (n, 0)$. Let $\tilde{T}$ be the average over all such paths:

$$\tilde{T} = \frac{1}{|B|} \sum_{x \in B} T_x$$

This average can be related to the passage time T, the first passage time between $\vec{0}$ and $(n, 0)$ by the inequalities:

$$\begin{aligned}
T_x &\leq T_{\{x \leftrightarrow \vec{0}\}} + T_{\{\vec{0} \leftrightarrow (n,0)\}} + T_{\{(n,0) \leftrightarrow (n,0)+x\}} \\
&= T_{\{x \leftrightarrow \vec{0}\}} + T + T_{\{(n,0) \leftrightarrow (n,0)+x\}} \\
T &\leq T_{\{\vec{0} \leftrightarrow x\}} + T_{\{x \leftrightarrow (n,0)+x\}} + T_{\{(n,0)+x \leftrightarrow (n,0)\}} \\
&= T_{\{\vec{0} \leftrightarrow x\}} + T_x + T_{\{(n,0)+x \leftrightarrow (n,0)\}}
\end{aligned}$$

Where for $y, z \in \mathbb{Z}^2$, $T_{\{y \leftrightarrow z\}}$ is the first passage time between y and z. These inequalities hold by a sort of triangle inequality for first passage percolation: the first passage time from a to c cannot be longer than the sum of the first passage time from a to b and the first passage time from b to c, for the path connecting a

to c is allowed to travel through b en route. These two inequalities together give us that for any $x \in B_k$:

$$\begin{aligned} |T_x - T| &\leq T_{\{x \leftrightarrow \vec{0}\}} + T_{\{(n,0) \leftrightarrow (n,0)+x\}} \\ &\leq \sum_{e \in \pi_{x \leftrightarrow \vec{0}}} w_e + \sum_{e \in \pi_{(n,0) \leftrightarrow (n,0)+x}} w_e \end{aligned}$$

Since x is inside the box B_k the direct paths from $\{x \leftrightarrow \vec{0}\}$ and $\{(n,0) \leftrightarrow (n,0)+x\}$ are no longer than $2k$ steps long, and the edges here are independent, we have:

$$\begin{aligned} \mathbf{E}\left(|T_x - T_n|^2\right) &\leq \mathbf{E}\left(\left(\sum_{e \in \pi_{x \leftrightarrow \vec{0}}} w_e + \sum_{e \in \pi_{(n,0) \leftrightarrow (n,0)+x}} w_e\right)^2\right) \\ &\leq 4k\mathbf{E}\left(w_e^2\right) + \binom{4k}{2}\mathbf{E}\left(w_e\right)\mathbf{E}\left(w_{e'}\right) \\ &\leq Ck^2 \end{aligned}$$

Hence:

$$\begin{aligned} \mathbf{E}\left(|\tilde{T} - T|^2\right) &\leq \frac{1}{|B|}\sum_{x \in B}\mathbf{E}\left(|T_x - T|^2\right) \\ &\leq Ck^2 \end{aligned}$$

Finally, since $\mathbf{E}\left(\tilde{T}\right) = \mathbf{E}\left(T_x\right) = \mathbf{E}\left(T\right)$ and for every x (since first passage percolation is transition invariant), we have with a little bit of algebra:

$$\mathbf{Var}(T) \leq 2\mathbf{Var}(\tilde{T}) + Ck^2$$

So it remains only to bound the variance of $\tilde{T}$. This is done with the $L^1 - L^2$ inequality as follows. Here we let $\hat{p}_x$ be the optimal path from x to $x + (n,0)$ that achieves T_x and $\hat{p}$ be the path from $\vec{0}$ to $(n,0)$ that achieves T. Have:

$$\partial_e \tilde{T} = \frac{1}{|B|}\sum_{x \in B}\mathbf{1}_{\{e \in \hat{p}_x\}}$$

So then:

$$\begin{aligned} ||\partial_e \tilde{T}||_{L^1} &= \frac{1}{|B|}\sum_{x \in B}\mathbf{P}\{e \in \hat{p}_x\} \\ &= \frac{1}{|B|}\sum_{x \in B}\mathbf{P}\{e - x \in \hat{p}\} \\ &= \frac{1}{|B|}\sum_{x \in B+e}\mathbf{P}\{x \in \hat{p}\} \\ &= \frac{1}{|B|}\mathbf{E}\left(\#\left\{\hat{p} \cap (B+e)\right\}\right) \end{aligned}$$

Now, we remark that the path $\hat{p}$ is a geodesic. Hence, moving along $\hat{p}$, if we let x_{in} be the first point that $\hat{p}$ intersects $B + e$ and x_{out} be the last point that $\hat{p}$ intersect $B + e$, then $\hat{p}$ is a geodesic between x_{in} and x_{out}. But both x_{in} and x_{out} are on the boundary of B, which is a box of size $2k + 1$. It is a fact in first passage percolation that the expectation of the length geodesic of two points which are

$O(k)$ apart is also $O(k)$ (This can be proven with large deviations, or by trying to bootstrap the result from the case where the edge weights are bounded $w_e \in (a,b)$ a.s.). Hence we have:

$$\begin{aligned} ||\partial_e \tilde{T}||_{L^1} &= \frac{1}{|B|}\mathbf{E}\left(\#\left\{\hat{p}\cap(B+e)\right\}\right) \\ &\leq \frac{1}{|B|}\mathbf{E}\left(|\hat{p}_{x_{in}\leftrightarrow x_{out}}|\right) \\ &= \frac{1}{|B|}O(k) \\ &= O\left(\frac{1}{k}\right) \end{aligned}$$

The last equality is because $|B| = \frac{1}{k^2}$. Now using the result which is like the $L^1 - L^2$ inequality, when $|\partial_i f| \leq C$ a.s. for every i we have that:

$$\mathbf{Var}_{\gamma^d}(f) \leq C' \sum_{i=1}^{n} \frac{||\partial_i f||_{L^1}}{1 - \log\left(||\partial_i f||_{L^1}\right)}$$

In our case, we have that $|\partial_e \tilde{T}| \leq 1$ a.s. and $||\partial_e \tilde{T}||_{L^1} = O(\frac{1}{k})$ so the bound is:

$$\begin{aligned} \mathbf{Var}\left(\tilde{T}\right) &\leq C' \sum_{e\in E(\mathbb{Z}^2)} \frac{||\partial_e \tilde{T}||_{L^1}}{1 - \log\left(O\left(\frac{1}{k}\right)\right)} \\ &\leq \frac{C'}{\log k}\sum_e ||\partial_e \tilde{T}||_{L^1} \\ &\leq \frac{C'}{\log k}\mathbf{E}\left(|\hat{p}|\right) \\ &\leq \frac{C'n}{\log k} \end{aligned}$$

Where we have again used the fact that the expected length of the first passage path between two points of distance $O(n)$ is $O(n)$ too. Finally, this gives:

$$\begin{aligned} \mathbf{Var}(T) &\leq 2\mathbf{Var}(\tilde{T}) + Ck^2 \\ &\leq \frac{C'n}{\log k} + Ck^2 \end{aligned}$$

Choosing $k = n^{\alpha}$for $\alpha < \frac{1}{2}$ gives the desired result:

$$\mathbf{Var}(T) \leq \frac{Cn}{\log n}$$

□

Exercise 2.3. Prove the result we used which is similar to the $L^1 - L^2$ inequality. Namely, if $|\partial_i f| \leq C$ for every i then:

$$\mathbf{Var}_{\gamma^d}(f) \leq C' \sum_{i=1}^{n} \frac{||\partial_i f||_{L^1}}{1 - \log\left(||\partial_i f||_{L^1}\right)}$$

CHAPTER 6

The Ising Model and the Improved Poincare Inequality

Recall that in the Ising model, we had $F_n(\beta) = -\frac{1}{\beta}\log\left(\sum_{\sigma\in\{-1,+1\}^n}\exp\left(-\beta H_n(\sigma)\right)\right)$ where $H_n(\sigma) = -\frac{1}{\sqrt{n}}\sum_{1\leq i,j\leq n} g_{ij}\sigma_i\sigma_j$. We aim here to show superconcentration in this model, namely that $F_n(\beta) = o(n)$ as $n\to\infty$. On the way to proving superconcentration, notice that:

$$\begin{aligned}\frac{\partial F_n}{\partial g_{ij}} &= -\frac{1}{\sqrt{n}}\frac{\sum_\sigma \sigma_i\sigma_j\exp(-\beta H_n(\sigma))}{\sum_\sigma\exp\left(-\beta H_n(\sigma)\right)}\\ &= -\frac{1}{\sqrt{n}}\left\langle\sigma_i,\sigma_j\right\rangle\end{aligned}$$

Where $\langle\sigma_i,\sigma_j\rangle = \mathbf{E}_{H_n}(\sigma_i\sigma_j)$ where the expectation is taken over the probability space given by the Gibbs measure $\mu_n(\sigma)\propto\exp\left(-\beta H_n(\sigma)\right)$. To further analyze this, we will need a bit more theory about the O-U semi-group.

1. Spectrum of the O-U semigroup

We first describe the spectrum and eigenfunction for the O-U semigroup in 1 dimension, and then we will generalize to n dimensions.

DEFINITION 1.1. The $k-$th Hermite polynomial is given by the expression:

$$H_k(x) = (-1)^k e^{\frac{x^2}{2}}\frac{d^k}{dx^k}\left(e^{-\frac{x^2}{2}}\right)$$

The first few are $H_0(x)=1$, $H_1(x)=x$, $H_2(x)=x^2-1$.

FACT 1.2. (1-D Spectrum of the O-U Process)
The sequence $(H_k)_{k=0}^\infty$ is an orthogonal basis for the Gaussian measure γ in the sense that:

$$\begin{aligned}||H_k||^2_{L^2(\gamma)} &= k!\\ \langle H_k,H_j\rangle_{L^2(\gamma)} &= 0 \text{ for } j\neq k\end{aligned}$$

Moreover, these are an eigenbasis for the O-U generator $-L$ with eigenvalue k:

$$\langle H_k, -LH_k\rangle = kH_k$$

FACT 1.3. (n-D spectrum of the O-U Process)
Let $Lf(x) = \Delta f(x) - x\cdot\nabla f(x)$ be the generator of the n dimensional O-U process. For $\vec{k} = (k_1,\dots k_n)$ the operator$-L$ has eigenfunctions $H_{\vec{k}}$ given by:

$$H_{\vec{k}} = \prod_{i=1}^n H_{k_i}(x_i)$$

Where H_k is the $k-$th Hermite polynomial defined above. The eigenvalue of $H_{\vec{k}}$ is $|\vec{k}| = \sum_i k_i$, and the norm of the eigenfunction is:

$$||H_{\vec{k}}||^2_{L^2(\gamma^n)} = \vec{k}! = k_1! \cdot k_2! \cdot \ldots \cdot k_n!$$

PROPOSITION 1.4. *For the 1-D Hermite polynomials, one can verify that $H_k' = kH_{k-1}$, and using this one can verify that:*

$$\langle f, H_k \rangle_{L^2(\gamma)} = \mathbf{E}_\gamma \left(\frac{d^k f}{dx^k} \right)$$

In n-D one can similarly show that:

$$\left\langle f, H_{\vec{k}} \right\rangle_{L^2(\gamma^n)} = \mathbf{E}_{\gamma^n} \left(\frac{\partial^{|\vec{k}|} f}{\partial x_1^{k_1} \partial x_2^{k_2} \ldots \partial x_n^{k_n}} \right)$$

PROOF. We use the explicit form of the Gaussian density $d\gamma = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$ and the definition of the Hermite polynomial $H_k(x) = (-1)^k e^{\frac{x^2}{2}} \frac{d^k}{dx^k} \left(e^{-\frac{x^2}{2}} \right)$ to see the result:

$$\begin{aligned} \langle f, H_k \rangle_{L^2(\gamma)} &= \int f H_k d\gamma \\ &= \int f \left((-1)^k e^{\frac{x^2}{2}} \frac{d^k}{dx^k} \left(e^{-\frac{x^2}{2}} \right) \right) \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \\ &= \int f (-1)^k \frac{d^k}{dx^k} \left(e^{-\frac{x^2}{2}} \right) \frac{1}{\sqrt{2\pi}} dx \end{aligned}$$

From here, one applies integration by parts to see that:

$$\begin{aligned} \langle f, H_k \rangle_{L^2(\gamma)} &= -\int f' (-1)^k \frac{d^{k-1}}{dx^{k-1}} \left(e^{-\frac{x^2}{2}} \right) \frac{1}{\sqrt{2\pi}} dx \\ &= \int f' (-1)^{k-1} \frac{d^{k-1}}{dx^{k-1}} \left(e^{-\frac{x^2}{2}} \right) \frac{1}{\sqrt{2\pi}} dx \\ &= \langle f', H_{k-1} \rangle_{L^2(\gamma)} \end{aligned}$$

Where we have recognized the integral from our above calculation. Repeating this $k-$times (or using a proof by induction to be more formal) one sees that:

$$\begin{aligned} \langle f, H_k \rangle_{L^2(\gamma)} &= \left\langle f^{(k)}, H_0 \right\rangle_{L^2(\gamma)} \\ &= \mathbf{E}_\gamma \left(\frac{d^k f}{dx^k} \right) \end{aligned}$$

The last equality holds since $H_0 \equiv 1$. The proof in $n-$D is very similar.

□

2. The Improved Poincare Inequality

We have seen before that we can write the variance of a function using through the Plancherel identity:

$$\mathbf{Var}_\mu (f) = \sum_{k=1}^{\infty} \langle u_k, f \rangle^2_{L^2(\mu)}$$

We have seen that using the fact that $\lambda_1 < \lambda_2 < \dots$ and this formula gives rise to the Poincare inequality, $\mathbf{Var}_\mu(f) \leq \frac{1}{\lambda_1}\sum_{k=1}^{\infty}\lambda_k \langle u_k, f\rangle^2_{L^2(\mu)}$. If we have better control over the eigenvalues however, we can get an even finer estimate. This is what we do in this section in the case of the O-U semigroup, where we know the eigenvalues from the analysis in the previous section.

PROPOSITION 2.1. *In the sett the $n-$dimensional O-U semigroup, define $\theta_m(f)$ by:*

$$\theta_m(f) = \sum_{1\leq i_1,i_2,\dots,i_m\leq n}\left(\frac{\partial^{|\vec{k}|}f}{\partial x_{i_1}\partial x_{i_2}\dots\partial x_{i_m}}\right)^2$$

Where here $\vec{k} = (k_1, k_2, \dots k_n)$ with $k_i \geq 1$ and $|\vec{k}| = \sum_i k_i$. Then:

$$\mathbf{Var}_{\gamma^n}(f) = \sum_{m=1}^{\infty}\frac{1}{m!}\theta_m(f)$$

PROOF. The result is a simple manipulation starting with the eigenfunctions $H_{\vec{k}}$ with eigenvalues $|\vec{k}|$ and the formula for the variance from Parseval's identity using the normalized H_k's as an eigenbasis:

$$\begin{aligned}\mathbf{Var}_{\gamma^n}(f) &= \sum_{\vec{k}\in(\mathbb{Z}^n)^+}\left\langle\frac{H_k}{\vec{k}!}, f\right\rangle^2_{L^2(\gamma^n)}\\ &= \sum_{\vec{k}\in(\mathbb{Z}^n)^+}\frac{1}{\vec{k}!}\langle H_k, f\rangle^2_{L^2(\gamma^n)}\end{aligned}$$

By our proposition 1.4, we know $\left\langle H_{\vec{k}}, f\right\rangle_{L^2(\gamma^n)} = \mathbf{E}_{\gamma^n}\left(\frac{\partial^{|\vec{k}|}f}{\partial x_1^{k_1}\partial x_2^{k_2}\dots\partial x_n^{k_n}}\right)$ so this sum is:

$$\mathbf{Var}_{\gamma^n}(f) = \sum_{\vec{k}\in(\mathbb{Z}^n)^+}\frac{1}{\vec{k}!}\mathbf{E}_{\gamma^n}\left(\frac{\partial^{|\vec{k}|}f}{\partial x_1^{k_1}\partial x_2^{k_2}\dots\partial x_n^{k_n}}\right)^2$$

If we now collect terms which have $\vec{k} = m$ and we multiply and divide my $m!$ we get:

$$\begin{aligned}\mathbf{Var}_{\gamma^n}(f) &= \sum_{m=1}^{\infty}\sum_{|\vec{k}|=m}^{\infty}\frac{1}{\vec{k}!}\mathbf{E}_{\gamma^n}\left(\frac{\partial^{|\vec{k}|}f}{\partial x_1^{k_1}\partial x_2^{k_2}\dots\partial x_n^{k_n}}\right)^2\\ &= \sum_{m=1}^{\infty}\frac{1}{m!}\sum_{|\vec{k}|=m}^{\infty}\frac{m!}{k_1!k_2!\dots k_n!}\mathbf{E}_{\gamma^n}\left(\frac{\partial^{|\vec{k}|}f}{\partial x_1^{k_1}\partial x_2^{k_2}\dots\partial x_n^{k_n}}\right)^2\end{aligned}$$

Finally, notice that $\frac{m!}{k_1!k_2!\dots k_n!}$ is the combinatorial expression for the number of ways to put m balls in n distiguished boxes so that the number of balls in box #1 is k_1, number of balls in box #2 is k_2 and so on. Another way to achieve an arrangement like this would be to assign each of the m balls an *index*, say the $k-$th ball gets the index $1 \leq i_k \leq n$, and put the ball k into the box $\#i_k$. If we think of each ball in box $\#j$ as the number of x_j derivatives to take, and since derivatives

commute, we see that the above expression is:

$$\mathbf{Var}_{\gamma^n}(f) = \sum_{m=1}^{\infty} \frac{1}{m!} \sum_{1\leq i_1,\ldots,i_m\leq n}^{\infty} \mathbf{E}_{\gamma^n}\left(\frac{\partial^m f}{\partial x_{i_1}\partial x_{i_2}\ldots\partial x_{i_n}}\right)^2$$

Which is exactly the expression written in terms of the θ_m which we desired. □

PROPOSITION 2.2. *For the case of the free energy of the Ising model, we can show that:*

$$\theta_m(F_n) \leq C_m(\beta)$$

Where $C_m(\beta)$ is a constant depending only on m and β (independent of n)

PROOF. *(Sketch)*

We will sketch the idea of the proof here by examining the cases $m = 1$ and $m = 2$. The same ideas can by applied to all values of m.

In the case $m = 1$ we have: (It is worth noting here, in order to avoid confusion, that the random Gaussian here are labeled $(g_{ij})_{1\leq i\leq j\leq n}$, so that the dimension of the space we are integrating over is not n but rather $d = \binom{n}{2} + n$.)

$$\begin{aligned}\theta_1(F_n) &= \sum_{1\leq i\leq j\leq n} \mathbf{E}_{\gamma^d}\left(\frac{\partial F_n}{\partial g_{ij}}\right)^2 \\ &= \sum_{1\leq i\leq j\leq n} \mathbf{E}_{\gamma^d}\left(-\frac{1}{\sqrt{n}}\langle\sigma_i,\sigma_j\rangle\right)^2 \\ &= \frac{1}{n}\sum_{1\leq i\leq j\leq n} \mathbf{E}_{\gamma^d}\left(\mathbf{E}_{H_n}(\sigma_i\sigma_j)\right)^2\end{aligned}$$

Where we have used $\frac{\partial F_n}{\partial g_{ij}} = -\frac{1}{\sqrt{n}}\langle\sigma_i,\sigma_j\rangle = -\frac{1}{\sqrt{n}}\mathbf{E}_{H_n}(\sigma_i\sigma_j)$ is the expectation of the product $\sigma_i\sigma_j$ over the Gibbs measure induced by H_n. The trick now is to use that $\mathbf{E}_{\gamma^d}\left(\mathbf{E}_{H_n}(\sigma_i\sigma_j)\right) = \mathbf{E}_U\left(\sigma_i\sigma_j\right)$ where U is the uniform distribution on the set $\{-1,+1\}^n$. This is can be seen because of the symmetry of the g_{ij}'s. With this trick we have:

$$\begin{aligned}\theta_1(F_n) &= \frac{1}{n}\sum_{1\leq i\leq j\leq n} \mathbf{E}_U\left(\sigma_i\sigma_j\right)^2 \\ &= 0\end{aligned}$$

The expectation is 0 in this case, since we are looking over the uniform distribution so $\sigma_i\sigma_j \in \{+1,-1\}$ with both equally likely. The same trick can be used for θ_2, but the terms become more complicated:

$$\begin{aligned}\frac{\partial^2 F_n}{\partial g_{ij}\partial g_{kl}} &= -\frac{1}{\sqrt{n}}\frac{\partial}{\partial g_{kl}}\langle\sigma_i\sigma_j\rangle \\ &= \frac{\beta}{n}\left(\langle\sigma_i\sigma_j\sigma_k\sigma_l\rangle - \langle\sigma_i\sigma_j\rangle\langle\sigma_k\sigma_l\rangle\right)\end{aligned}$$

And we will have

$$\begin{aligned}\theta_2(F_n) &= \sum_{\substack{1\leq i\leq j\leq n\\ 1\leq k\leq l\leq n}} \mathbf{E}_{\gamma^d}\left(\frac{\partial^2 F_n}{\partial g_{kl}\partial g_{ij}}\right)^2 \\ &= \frac{\beta^2}{n^2}\sum_{\substack{1\leq i\leq j\leq n\\ 1\leq k\leq l\leq n}} \mathbf{E}_{\gamma^d}\left(\langle\sigma_i\sigma_j\sigma_k\sigma_l\rangle - \langle\sigma_i\sigma_j\rangle\langle\sigma_k\sigma_l\rangle\right)^2\end{aligned}$$

Now we will use our uniform distribution over $\{-1,+1\}^n$ trick to kill some of the terms again. The term $\mathbf{E}_U(\sigma_i\sigma_j\sigma_k\sigma_l)$ will vanish just as it did in the computation of θ_1. We will remain with:

$$\begin{aligned}\mathbf{E}_{\gamma^d}\left(\langle\sigma_i\sigma_j\rangle\langle\sigma_k\sigma_l\rangle\right)^2 &= \mathbf{E}_{\gamma^d}\left(\mathbf{E}_{H_n}(\sigma_i\sigma_j)\mathbf{E}_{H_n}(\sigma_k\sigma_l)\right)^2 \\ &= \mathbf{E}_{\gamma^d}\left(\sigma_i^1\sigma_j^1\sigma_k^2\sigma_l^2\right)^2 \\ &= \mathbf{E}_{\gamma^d}\left(\sigma_i^1\sigma_j^1\sigma_k^2\sigma_l^2\tilde{\sigma}_i^1\tilde{\sigma}_j^1\tilde{\sigma}_k^2\tilde{\sigma}_l^2\right)\end{aligned}$$

Here we have used the trick $\mathbf{E}(X)^2 = \mathbf{E}(X^1X^2)$ with X^1, X^2 being iid variables having the distribution of X. (We did this with superscript numbers and with tildes). In our specific case, all these variables are chosen with respect to the Gibbs distribution. We have:

$$\begin{aligned}\theta_2(F_n) &= \frac{\beta^2}{n^2}\sum_{\substack{1\leq i\leq j\leq n\\ 1\leq k\leq l\leq n}} \mathbf{E}_{\gamma^d}\left(\sigma_i^1\sigma_j^1\sigma_k^2\sigma_l^2\tilde{\sigma}_i^1\tilde{\sigma}_j^1\tilde{\sigma}_k^2\tilde{\sigma}_l^2\right) \\ &\approx \frac{\beta^2}{n^2}\mathbf{E}_{\gamma^d}\left(\left(\sum_{1\leq i\leq n}\sigma_i^1\tilde{\sigma}_i^2\right)\left(\sum_{1\leq k\leq n}\sigma_k^1\tilde{\sigma}_k^2\right)\right)\end{aligned}$$

Here we have done a manipulation to bring the terms into a more manageable product (Some terms may vanish because of the uniform trick?). Finally, one uses the fact that $\left(\sum_{1\leq i\leq n}\sigma_i^1\tilde{\sigma}_i^2\right) = O(n)$ and the Cauchy-Shwarz inequality to get:

$$\begin{aligned}\theta_2(F_n) &\approx \frac{\beta^2}{n^2}O(n^2) \\ &\leq C_\beta\end{aligned}$$

The same type of idea works for general θ_m by using the Cauchy-Shwarz distribution on m terms.

□

Exercise 2.3. Show how the symmetry of the g_{ij}'s lead to the fact that $\mathbf{E}_{\gamma^d}\left(\mathbf{E}_{H_n}(f(\sigma))\right) = \mathbf{E}_U(f(\sigma))$ where U is the uniform distribution on $\{-1,+1\}^n$.

So we have a bound on the θ_m and so we have a bound on the variance of F_n:

$$\begin{aligned}\mathbf{Var}(F_n) &= \sum_{m=1}^{\infty}\frac{1}{m!}\theta_m(F_n) \\ &\leq \sum_{m=1}^{\infty}\frac{1}{m!}C_m(\beta)\end{aligned}$$

Now if it was the case that the sum $\sum \frac{1}{m!} C_m(\beta)$ converged, we would have that the variance is bounded by a *constant.* Unfortunately, the best estimate we can get for $C_m(\beta)$ from using the crude estimate of the Cauchy-Shwarz inequality is:

$$C_m(\beta) \leq \exp\left(c(\beta) n \log n\right)$$

For which $\sum \frac{1}{m!} C_m(\beta)$ diverges. However we can still get a bound of $o(n)$ by improving the Poincare inequality a little bit.

THEOREM 2.4. (Improved Poincare Inequality)

Consider a Markov process whose generator $-L$ has eigenfunctions $u_1, u_2, \ldots$ with eigenvalues $\lambda_1 < \lambda_2 < \ldots$. The variance of a random variable can be bounded by:

$$\mathbf{Var}_\mu(f) \leq \sum_{k=1}^{m-1} \langle u_k, f \rangle^2_{L^2(\mu)} + \frac{\mathcal{E}(f,f)}{\lambda_m}$$

COROLLARY 2.5. *For Gaussian random variables and the O-U process this translates to:*

$$\mathbf{Var}_{\gamma^n}(f) \leq \sum_{k=1}^{m-1} \frac{1}{k!} \theta_k(f) + \frac{\mathbf{E}_{\gamma^n}\left(|\nabla f|^2\right)}{m}$$

PROOF. As in the proof of the Poincare inequality, we start with the Plancheral identity and we use the fact that the sequence of eigenvalues λ_i are *increasing*:

$$\begin{aligned}
\mathbf{Var}_\mu(f) &= \sum_{k=1}^{\infty} \langle u_k, f \rangle^2_{L^2(\mu)} \\
&= \sum_{k=1}^{m-1} \langle u_k, f \rangle^2_{L^2(\mu)} + \sum_{k=m}^{\infty} \langle u_k, f \rangle^2_{L^2(\mu)} \\
&\leq \sum_{k=1}^{m-1} \langle u_k, f \rangle^2_{L^2(\mu)} + \frac{1}{\lambda_m} \sum_{k=m}^{\infty} \lambda_k \langle u_k, f \rangle^2_{L^2(\mu)} \\
&\leq \sum_{k=1}^{m-1} \langle u_k, f \rangle^2_{L^2(\mu)} + \frac{1}{\lambda_m} \mathcal{E}(f,f)
\end{aligned}$$

□

THEOREM 2.6. *For any $\beta > 0$, the free energy in the Ising model has superconcentration:*

$$\mathbf{Var}\left(F_n(\beta)\right) \leq \frac{c(\beta) n \log\left(\log n\right)}{\log n}$$

PROOF. If we use our improved Poincare inequality, along with the bound we had that $\theta_m(F_n) \leq e^{c(\beta) m \log m}$ we get:

$$\begin{aligned}
\mathbf{Var}_{\gamma^n}(f) &\leq \sum_{k=1}^{m-1} \frac{1}{k!} \theta_k(f) + \frac{\mathbf{E}_{\gamma^n}\left(|\nabla f|^2\right)}{m} \\
&\leq \sum_{k=1}^{m-1} \frac{1}{k!} e^{c(\beta) m \log m} + \frac{c'(\beta) n}{m} \\
&\approx \frac{1}{m!} e^{c(\beta) m \log m} + \frac{c'(\beta) n}{m}
\end{aligned}$$

If we choose $m = \frac{a(\beta)\log n}{\log(\log n)}$ we get the desired inequality.

□