MINICOURSE SPRING 2026: THE SUPERCONCENTRATION PHENOMENON

NICHOLAS A. COOK*

Contents

0. Notation 1
1. The concentration of measure phenomenon 2
2. Superconcentration: Examples 3
3. Markov semigroups 10
4. Gaussian Poincaré inequality and applications 13
5. Superconcentration and chaos 15
6. Superconcentration from hypercontractivity 18
7. Spectral approach to superconcentration 24
8. Superconcentration for the top eigenvalue of random matrices 30
9. Noise sensitivity and disorder chaos for last-passage percolation models 33
10. More recent developments 34
References 34

Abstract. These are daily lecture notes for a graduate probability minicourse (in progress in March 2026) on the superconcentration phenomenon. The primary source is Chatterjee's monograph [Cha14]. The notes below do not give proper citations for all results – one should check [Cha14] for missing references. References to more recent works will be collected here. In the final two lectures we discuss some heuristics and proof ideas from the more recent papers 1) [BLZ20] on disorder chaos for the top eigenvector of Wigner matrices, and 2) [GH24] on the transition from stability to chaos for dynamical Brownian last passage percolation.

0. Notation

Throughout the notes we use standard asymptotic notation $f = O(g)$ or $f \lesssim g$ to mean $|f| \leq Cg$ for a constant C, assumed to be universal unless otherwise stated. Further, $f \asymp g$ and $f = \Theta(g)$ mean $g \lesssim f \lesssim g$. When f is nonnegative then $g = \Omega(f)$ and $g \gtrsim f$ mean $f \lesssim g$. Dependence of the implicit constant on parameters is indicated with subscripts.

We often write $[n]$ for the discrete interval $\{1, \dots, n\}$.

With no risk of confusion, we use $|\cdot|$ for the length of a vector in a finite-dimensional Euclidean space as well as for the cardinality of a finite set.

Date: April 1, 2026.

We tend to use the dot product notation $x \cdot y$ for the inner product in a real finite dimensional vector space, reserving $\langle \cdot, \cdot \rangle$ for Hilbert spaces of functions over a probability space (we will often have both sorts of inner products in the same expression).

1. The concentration of measure phenomenon

Before we start on the superconcentration phenomenon, what is "regular" concentration? Here we give a quick and incomplete overview; for more background we refer to the texts [Led01, BLM13].

The most basic example from probability is a random walk $S_n = X_1 + \cdots + X_n$ for independent variables X_i. Let us assume the X_i are bounded a.s. While S_n ranges over a set of diameter $\Theta(n)$, it typically stays in a region of size $O(\sqrt{n})$ around its expectation, as shown by computing the variance. Moreover, the distribution of S_n decays rapidly away from this region, as quantified by Hoeffding's inequality:

$$\mathbb{P}(|S_n - \mathbb{E}S_n| \geq t\sqrt{n}) \leq 2\exp(-ct^2) \quad \forall t \geq 0. \tag{1.1}$$

Such *sub-Gaussian* tail bounds are common in the theory of concentration of measure. However, in this course we focus on the question of identifying the correct scale of fluctuations, as captured by the variance. The fact that a random walk has variance $O(n)$ and hence fluctuates at scale $O(\sqrt{n})$ is already a surprising form of "concentration of measure".

This *square-root cancellation* phenomenon arises more generally for martingales, as well as (conjecturally) pseudorandom sequences in number theory (the Riemann hypothesis is equivalent to square-root cancellation for partial sums of the Möbius function).

The square-root cancellation form of concentration of measure extends to more general functions of sequences $X_1, \ldots, X_n$ of independent variables. For instance, if $F : \mathbb{R}^n \to \mathbb{R}$ has the property that varying any single input while holding all others fixed can only change the output by an amount $O(1)$, then McDiarmid's inequality says

$$\mathbb{P}(|F(X) - \mathbb{E}F(X)| \geq t\sqrt{n}) \leq 2\exp(-ct^2) \quad \forall t \geq 0. \tag{1.2}$$

This generalizes (1.1) since $F(X) = X_1 + \cdots X_n$ has this regularity property (which one can view as a Lipschitz condition under the Hamming metric).

At the level of the variance, a bound on the scale of fluctuations for $F(X)$ is provided by the Effron–Stein inequality, which states that, with $X_1', \ldots, X_n'$ independent copies of $X_1, \ldots, X_n$, respectively, and

$$X^{(i)} := (X_1, \ldots, X_{i-1}, X_i', X_{i+1}, \ldots, X_n)$$

we have

$$\operatorname{Var} F(X) \leq \frac{1}{2}\sum_{i=1}^{n} \mathbb{E}(F(X) - F(X^{(i)}))^2. \tag{1.3}$$

The general concentration of measure phenomenon arises for functions F of high-dimensional random vectors X (not necessarily with indepenent coordiates), with F satisfying a Lipschitz condition under some metric. The above examples are for the Hamming metric, but there is also a powerful set of inequalities for the Euclidean

metric. For instance, if $G \in \mathbb{R}^n$ is a standard Gaussian vector and F is 1-Lipschitz with respect to the Euclidean norm $\|\cdot\|_2$, then

$$\mathbb{P}(|F(G) - \mathbb{E}F(G)| \geq t) \leq 2\exp(-ct^2) \quad \forall t \geq 0. \tag{1.4}$$

While this inequality does not extend to general vectors of independent variables with light tails, it does extend under some additional hypotheses:

Theorem 1.1 (Talagrand's inequality)**.** *Assume $X_1, \dots, X_n$ are independent random variables with $|X_i| \leq 1$ a.s. for all i, and let $F : \mathbb{R}^n \to \mathbb{R}$ be 1-Lipschitz and convex. Then*

$$\mathbb{P}(|F(X) - \mathbb{E}F(X)| \geq t) \leq 2\exp(-ct^2) \quad \forall t \geq 0. \tag{1.5}$$

(The result is usually stated for concentration about a median for $F(X)$, but it is not hard to show this is equivalent up to modification of the constant c in the exponent.)

As a quick application, we get a strong concentration inequality for the top eigenvalue of random matrices with independent entries, an example we will return to several times in the course.

Proposition 1.2. *Let X be an $m \times n$ matrix with independent entries X_{ij} bounded in absolute value by 1 a.s. Then*

$$\mathbb{P}(|\|X\|_{\mathrm{op}} - \mathbb{E}\|X\|_{\mathrm{op}}| \geq t) \leq 2\exp(-ct^2) \quad \forall t \geq 0. \tag{1.6}$$

In particular, $\|X\|_{\mathrm{op}}$ has variance $O(1)$. The same holds for the largest eigenvalue of a symmetric $n \times n$ random matrix with independent entries on and above the diagonal a.s. bounded by 1.

Exercise 1.1. Use Theorem 1.1 to prove Proposition 1.2 (including the statements about the variance and the symmetric case). Note that the Euclidean norm on the space $\mathbb{R}^{m\times n}$ of $m \times n$ matrices is the Frobenius (a.k.a. Hilbert–Schmdt) norm $\|M\|_F = (\sum_{i,j=1}^n M_{ij}^2)^{1/2} = (\operatorname{Tr} M^{\mathsf{T}} M)^{1/2}$.

That the scale of fluctuations of $\|X\|_{\mathrm{op}}$ is order 1 is surprising, given that the typical scale of $\|X\|_{\mathrm{op}}$ is $\max(m,n)^{1/2}$ (one sees this is a lower bound by considering the norm of a row or column). As we will see, when $m \asymp n$, the true scale of fluctuations is in fact smaller, of order $n^{-1/6}$! This is an instance of superconcentration.

2. Superconcentration: Examples

One of the main outcomes of Chatterjee's theory is to connect the phenomenon of superconcentration with two other phenomena called "disorder chaos" (a form of noise sensitivity) and "multiple valleys (or peaks)". Before giving formal definitions of these phenomena we illustrate them with examples.

2.1. First-passage percolation.

Let $E = E(\mathbb{Z}^d)$ be the set of nearest-neighbor edges for the integer lattice, and let $(\omega_e)_{e\in E}$ be iid non-negative random variables, called the *weights* or *edge passage times*. For a path p – by which we mean a finite set of edges that can be labeled $e_1, \dots, e_n$ so that e_i and e_{i+1} are connected for each $1 \leq i \leq n-1$ – the

passage time is defined as the sum of weights ω_{e_i}. For sites $x, y \in \mathbb{Z}^d$, the *first-passage time* $T(x,y)$ is

$$T(x,y) = \min_p \sum_{e \in p} \omega_e \tag{2.1}$$

where the minimum is taken over all paths p joining x to y.

For $x \in \mathbb{R}^d$ set $T_n(x) := T(0,[nx])$, where $[nx]$ is the nearest lattice point to nx. From the subadditive ergodic theorem one can show that for all x, $T_n(x)/n$ converges a.s. and in L^1 to a deterministic limit $\mu(x)$, which is positive under mild conditions on the weights. Hence, under mild conditions, $T_n(x)$ is typically of order n when n is large.

We refer to [Cha14] and the survey [ADH17] for additional background. Here we are only concerned with the scale of fluctuations, which is a wide open problem.

Consider $T_n := T_n(e_1)$, with $e_1 = (1,0,\dots,0)$ the first canonical basis vector in $\mathbb{R}^d$. Using a martingale argument, Kesten '93 showed

$$\operatorname{Var} T_n = O(n)$$

so that T_n exhibits at least a square-root cancellation amount of concentration. Note this is sharp for the case $d = 1$, in which case T_n is a random walk of the form considered in (1.1).

However, it is conjectured that for higher dimensions the scale of fluctations is significantly smaller than the standard square-root cancellation. For instance, for $d = 2$ it is conjectured that

$$\operatorname{Var} T_n \asymp n^{2/3}. \tag{2.2}$$

See [Cha07] for further discussion and references to physics literature and conjectures.

For $d = 2$ the state of the art at the time of this writing is

$$\log n \lesssim \operatorname{Var} T_n \lesssim n/\log n. \tag{2.3}$$

The lower bound was proved independently by Newman-Piza '97 and Pemantle–Peres '94. The upper bound holds for all $d \geq 2$, and is a seminal result of Benjamini–Kalai–Schramm '03 for this case of weights supported on 2 points, and the result has since been generalized to other distributions; see [ADH17]. While the upper bound is still far from the conjecture (2.2), it does beat the square-root cancellation implied by "off-the-shelf" martingale concentration, and we view it as an instance of superconcentration.

For a related model, oriented last-passage percolation (LPP), fluctuations at scale $n^{1/3}$ have been established. Here we maximize the passage time over up-right paths connecting the origin to (n,n). The scale of fluctuations and limiting distribution for the recentered and rescaled LPP was famously established by Johansson '00 for the case that weights are geometric or exponential. The limiting distribution is the same as that of the (appropriately recentered and rescaled) largest eigenvalue of a GUE matrix.

2.2. **Gaussian polymer.** Let P_n be the set of all 2^n integer walks of n steps starting at the origin, where each step is either distance 1 forwards or 1 backwards. The path of a walk is a set of vertices $p = \{(1,a_1),\dots,(n,a_n)\}$ in $\mathbb{Z}^2$, with a_i the position of the walker after i steps. Let $(g_v)_{v \in \mathbb{Z}^2}$ be iid standard Gaussians, which we view as a random environment for the polymer. (We will only be using $O(n^2)$ of these variables.) The

energy of a path is

$$H_n(p) = -\sum_{v \in p} g_v.$$

For now we consider the ground state energy

$$E_n := \min_{p \in P_n} H_n(p)$$

and let $\hat{p}_n$ denote the minimizer, which is a.s. unique since the weights have a density.

In the third lecture we'll use "classical" concentration techniques to show

$$\operatorname{Var} E_n = O(n). \tag{2.4}$$

As with FPP this would be sharp if our polymer were constrained to the horizontal axis (the "time" axis for the walk).

Later in the course we'll prove the following superconcentration result:

Theorem 2.1 (Chatterjee '08)**.** $\operatorname{Var}(E_n) = O(n/\log n)$.

In Graham '10 this result was extended to the $(1+d)$-dimensional Gaussian polymer model, where we consider walks from the origin in $\mathbb{Z}^d$ taking n nearest-neighbor steps (the "1" in "$1+d$" begin the time dimension for the walker, which is a spatial dimension in the polymer interpretation).

As with FPP it is conjectured that for the $(1+1)$-dimensional Gaussian polymer the fluctuations are in fact at scale $n^{1/3}$.

The argument for Theorem 2.1 will apply equally to the following modified model where the variance is indeed of order $n/\log n$. Consider now the set Q_n of all "paths" $\{(1, a_n), \ldots, (n, a_n)\}$ where we allow each a_i to take any value in $[n]$, independent of the value at other indices. Thus we consider all n^n graphs of functions $i \mapsto a_i$ from $[n]$ to $[n]$. Let

$$\widetilde{E}_n := \max_{q \in Q_n} \sum_{v \in q} g_v \,.$$

We have

$$\widetilde{E}_n = \max_{1 \le a_1, \ldots, a_n \le n} \sum_{k=1}^{n} g_{(k,a_k)} = \sum_{k=1}^{n} X_k$$

where $X_k := \max_{1 \le a \le n} g_{(k,a)}$. Now $X_1, \ldots, X_n$ are iid and it is a standard exercise to show they each have variance of order $1/\log n$, so $\widetilde{E}_n$ has variance of order $n/\log n$ as claimed.

Hence, a proof of the conjectured order $n^{2/3}$ for the variance, or any gain by a fractional power of n over the easy bound $O(n)$, will have to distinguish the sets P_n and Q_n. What is it about the geometry of P_n that should lead to such "super-super" concentration?

Remark 2.2 (Polymers with a temperature). In this short course we focus on the ground state energy for the polymer model, which can also be interpreted as a directed last-passage percolation model. More generally one can consider the polymer model at "positive temperature", sampling paths p from a Gibbs measure with Hamiltonian proportional to $\beta H_n(p)$ for an inverse-temperature parameter $\beta \in [0, +\infty]$. This nicely

interpolates between a random walk at $\beta = 0$ and last-passage percolation at $\beta = +\infty$. For more on this general setting we refer to the Saint Flour lectures of Comets [Com17].

2.3. **Maxima of general Gaussian fields.** We can pose the last question in a more general setup. Let $G = (g_v)_{v \in V}$ be a centered multivariate Gaussian vector indexed by a finite set V. (In most examples of interest where the field is defined on an infinite set, for questions about the maximum/minimum it is possible to pass to a finite set by a discretization argument.) Initially V has no special structure, but it gets a metric structure from the field: we define the distance

$$d(u,v) := (\mathbb{E}(g_u - g_v)^2)^{1/2} = \sqrt{R(u,u) + R(v,v) - 2R(u,v)}$$

where $R = (R(u,v))_{u,v \in V}$ is the covariance matrix for G.

In the Gaussian polymer model, $V = P_n$ and $g_p = \langle G, 1_p \rangle$, where 1_p is the vector $(1_{v \in p})_{v \in V}$. In this case $R(p,q) = \langle 1_p, 1_q \rangle = |p \cap q|$, and $d(p,q) = \sqrt{2(n - |p \cap q|)}$.

In the third lecture we should be in a position to prove the following inequality of Houdré '95:

$$\operatorname{Var} \max_{v \in V} g_v \leq \max_{v \in V} \operatorname{Var} g_v. \tag{2.5}$$

This is tight for $R(u,v) \equiv 1$, i.e. when the field is constant with standard normal distribution.

A general problem: What assumptions on R (or equivalently on the metric d) lead to improvements to (2.5)?

2.4. **Chaos for Gaussian polymers.** Now we show how the Gaussian polymer exhibits what in statistical physics is called "disorder chaos", i.e. a sensitivity to small changes to the disorder variables $G = (g_v)_{v \in \mathbb{Z}^2}$.

We perturb the disorder in the following way: let $G' = (g'_v)_{v \in \mathbb{Z}^2}$ be an independent copy of G, and for $\varepsilon \geq 0$ put

$$g^\varepsilon_v := e^{-\varepsilon} g_v + \sqrt{1 - e^{-2\varepsilon}} g'_v \tag{2.6}$$

so that $g^0_v = g_v$, and $\operatorname{Cov}(g_v, g^\varepsilon_v) = e^{-\varepsilon}$. When ε is small, $G^\varepsilon = (g^\varepsilon_v)_{v \in \mathbb{Z}^2}$ can be viewed as a small perturbation to the environment G. Let $\hat{p}^\varepsilon_n$ be the minimizer of the energy $H^\varepsilon_n(p) := -\sum_{v \in p} g^\varepsilon_v$.

Theorem 2.3 (Chatterjee '08)**.** *For the $(1+1)$-dimensional Gaussian polymer,*

$$\mathbb{E}|\hat{p}_n \cap \hat{p}^\varepsilon_n| \lesssim \frac{n}{\varepsilon \log n}$$

for all $\varepsilon \leq 1$.

Hence, after perturbing the disorder, as long as $\varepsilon = \omega(1/\log n)$, with high probability the ground state paths share a vanishing proportion of vertices.

In fact, superconcentration goes hand in hand with disorder chaos:

Theorem 2.4 (Chatterjee '08)**.** *For the $(d+1)$-dimensional Gaussian polymer,*

$$\operatorname{Var} E_n = o(n)$$

if and only if

$$\mathbb{E}|\hat{p}_n \cap \hat{p}^\varepsilon_n| = o(n)$$

for some sequence $\varepsilon = \varepsilon_n = o(1)$.

We will later see a quantitative version in a more general setup.

Since Graham '10 extended the superconcentration result of Theorem 2.1 to general dimension, it follows from the above that the Gaussian polymer of any dimension exhibits disorder chaos, at least at a qualitative level.

2.5. **Noise sensitivity of the top eigenvector.** Let $G = (g_{ij})_{i,j=1}^n$ be a GUE matrix. That is, G is Hermitian, and $(g_{ij})_{1\le i\le j\le n}$ are independent, with diagonal entries having standard real Gaussian distribution, and off-diagonal entries having standard complex Gaussian distribution (so the real and imaginary parts are independent with variance $\frac{1}{2}$). From G we get a Gaussian field on the complex unit sphere:

$$X_u = u^* G u\,, \qquad u \in S^{n-1}. \tag{2.7}$$

Let $\lambda_1 = \max_{u\in S^{n-1}} X_u$ and $u_1 = \arg\max_{u\in S^{n-1}} X_u$ be the top eigenvalue and associated eigenvector (u_1 is only determined up to multiplication by a phase but we ignore this issue).

Form $G^\varepsilon = (g_{ij}^\varepsilon)$ as in the previous subsection, so that $\mathbb{E}\overline{g_{ij}} g_{ij}^\varepsilon = e^{-\varepsilon}$ for all i, j. This gives a Gaussian field $(X_u^\varepsilon)_{u\in S^{n-1}}$, and we define $\lambda_1^\varepsilon, u_1^\varepsilon$ as before. Later in the course we'll show

Theorem 2.5 (Noise sensitivity for GUE ([Cha14])**.** $\mathbb{E}|\langle u_1, u_1^\varepsilon\rangle|^2 \lesssim \frac{1}{\varepsilon n^{1/3}}$ *for all* $\varepsilon \le 1$.

Hence, as soon as $\varepsilon = \omega(n^{-1/3})$, the top eigenvectors of G and G^ε are almost orthogonal with high probability.

Another natural way to perturb the disorder is to repeatedly resample entries chosen uniformly at random. We have the following more recent result:

Theorem 2.6 (Bordenave–Lugosi–Zhivotovskiy '20 [BLZ20])**.** *Let W be an $n \times n$ real Wigner matrix – i.e. a real symmetric matrix with standardized entries satisfying a tail condition, independent up to the symmetry constraint – let u_1 be its top eigenvector. Resample k entries chosen uniformly at random with replacement to form a matrix $W^{(k)}$, and denote its top eigenvector by $u_1^{(k)}$.*

If $k = \omega(n^{5/3})$ then $\mathbb{E}|\langle u_1, u_1^{(k)}\rangle| = o(1)$.

If $k \le n^{5/3} e^{-C(\log\log n)^2} = n^{5/3-o(1)}$, then

$$\mathbb{E} \min_{s=\pm 1} \|u_1 - s u_1^{(k)}\|_2 = o(1).$$

Note that with $k \asymp n^{5/3}$ we are resampling $\Theta(n^{2/3})$ entries on average per row, so that a typical row of $W^{(k)}$ is a random perturbation of the corresponding row of W by a vector of ℓ^2-norm $\Theta(n^{1/3})$. This is the same as for the Gaussian perturbation in Theorem 2.5, where G and G^ε differ by a complex Gaussian vector with typical norm of order $\sqrt{\varepsilon n}$, which is $\Theta(n^{1/3})$ at the threshold $\varepsilon \asymp n^{-1/3}$.

2.6. **Superconcentration and disorder chaos for the SK model.** The Sherrington–Kirkpatrick model is the most intensely studied mean-field model for a spin glass. It is a random probability measure on $\{-1, 1\}^n$, whose elements $\sigma = (\sigma_1, \dots, \sigma_n)$ are called spin configurations. Let $(g_{ij})_{1\le i<j\le n}$ be iid standard Gaussian variables a define the

Hamiltonian

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

The sign is a convention. The normalization by $\sqrt{n}$ makes it so that $H_n(\sigma)$ is typically of the same order n as the entropy $\log |\{-1,1\}^n| = n \log 2$ of the configuration space. For an inverse temperature parameter $\beta \ge 0$ we define a random Gibbs measure $\nu_{n,\beta}$ on the hypercube with

$$\nu_{n,\beta}(\{\sigma\}) = \frac{1}{Z_n(\beta)} \exp(-\beta H_n(\sigma)) \tag{2.8}$$

where

$$Z_n(\beta) = \sum_{\sigma \in \{-1,1\}^n} \exp(-\beta H_n(\sigma)) \tag{2.9}$$

is the normalizing constant, also called the partition function. With $\beta = 0$ ("infinite temperature") the Gibbs measure is the uniform measure, while as β increases the measure becomes increasingly concentrated on configurations of low energy $H_n(\sigma)$; at $\beta = +\infty$ (zero temperature) the measure concentrates on the *ground state* – the global minimizer of H_n.

Unlike we did for the polymer model, here we don't consider the ground state or its energy. Rather, we consider softened versions of these at positive temperature. The free energy is defined

$$F_n(\beta) := -\frac{1}{\beta} \log Z_n(\beta) \tag{2.10}$$

Note that $F_n(\beta) \to \min_\sigma H_n(\sigma)$ as $\beta \to \infty$, so that $F_n(\beta)$ can indeed be viewed as a sort of soft minimizer. At positive temperature the free energy is determined by a balance between the energy level and the number of spin configurations at that level (this is where the scaling of order n for the Hamiltonian is important).

A seminal result of Talagrand '06, building on an interpolation strategy of Guerra '03, established Parisi's formula for the a.s. deterministic limit of $\frac{1}{n}F_n(\beta)$.

How about fluctuations? Aizenman–Lebowitz–Ruelle '87 proved a central limit theorem for $F_n(\beta)$ when $\beta < 1$, holding without rescaling; in particular, the variance of $F_n(\beta)$ is bounded in this range.

Chen–Lam have shown the variance is $O(\log^2 n)$ at and near the critical temperature $\beta = 1$, where the model undergoes a phase transition. For general β Talagrand showed the variance is $O_\beta(n)$ using what in this course we would call "classical" concentration tools. Chatterjee '09 established superconcentration:

$$\operatorname{Var} F_n(\beta) \lesssim_\beta n/\log n \qquad \forall \beta \ge 0. \tag{2.11}$$

We turn to the question of disorder-chaos for the SK model. Since we are not considering the ground state, what could we mean by this? At positive temperature we consider samples fro the Gibbs measure. Sample σ^1 from the Gibbs measure $\nu_{\beta,n}$, and independently sample σ^2 from the Gibbs measure $\nu^{(\varepsilon)}_{\beta,n}$ formed in the same way as before

with correlated disorder variables $g_{ij}^{(\varepsilon)}$. We consider the *normalized overlap*

$$R_{1,2}(\varepsilon) := \frac{1}{n}\sum_{i=1}^{n} \sigma_i^1 \sigma_i^2. \tag{2.12}$$

It has been shown $\mathbb{E}R_{1,2}(0)^2 \to 0$ for $\beta \le 1$, while the limit is conjectured to be positive for all $\beta > 1$. Chatterjee '09 showed

$$\mathbb{E}R_{1,2}(\varepsilon)^2 \lesssim_\beta n^{-c(\beta)\varepsilon} \qquad \forall \beta \in [0,\infty), \varepsilon \in [0,1]\,. \tag{2.13}$$

2.7. **Multiple peaks/valleys.** We'll show that superconcentration for the maximum of a Gaussian field occurs when, typically, there are many well-separated near-maximizers. Thus, maximizers are highly unstable; one can see on an intuitive level how this would entail disorder chaos.

Before giving a more precise statement we review a well known example from asymptotic probabilistic combinatorial optimization where the optimizer is *stable*.

In the **random assignment problem** we have an $n \times n$ array (c_{ij}) of iid variables uniform in $[0,1]$, called *costs*, and we want to find the minimal cost matching and its associated cost:

$$\hat{c}_n := \min_\pi \sum_{i=1}^{n} c_{i,\pi(i)}\,, \qquad \hat{\pi}_n := \arg\min_\pi \sum_{i=1}^{n} c_{i,\pi(i)}\,.$$

Here π ranges over the set S_n of all permutations of $[n]$. Aldous '01 proved a conjecture of Parisi that

$$\hat{c}_n \to \zeta(2) = \pi^2/6 \quad a.s.$$

as $n \to \infty$. In fact, he proved a strong stability form of this statement: for every $\delta > 0$ there exists $\varepsilon > 0$ such that if $\pi_n \in S_n$ is any sequence (random or deterministic) satisfying

$$\liminf_{n\to\infty} \mathbb{E}\frac{1}{n}|\{i \in [n] : \pi_n(i) \ne \hat{\pi}_n(i)\}| \ge \delta$$

then

$$\liminf_{n\to\infty} \mathbb{E}\sum_{i=1}^{n} c_{i,\pi_n(i)} \ge \zeta(2) + \varepsilon\,.$$

This is an instance of what he termed the Asymptotic Essential Uniqueness (AEU) property.

Remark 2.7. A question that came up in class was whether one could quantify distance to the optimizer with other metrics. Above we're using the Hamming metric (or edit distance) but what about e.g. the Kendall tau distance, or transposition distance? This is a natural question from a computational point of view.

Exercise 2.1 ([Cha14, Exercise 1.14])**.** For $G = (g_1, \dots, g_n) \sim N(0, I_n)$, define a random function f on the hypercube $\{-1,1\}^n$ by

$$f(\sigma) := \langle G, \sigma\rangle = \sum_{i=1}^{n} g_i \sigma_i\,.$$

Let $\hat{\sigma} \in \{-1,1\}^n$ be the (a.s. unique) maximizer for f. Show that for every $\varepsilon > 0$ there exists $\delta > 0$ such that with high probability,

$$f(\sigma) \geq (1-\delta) f(\hat{\sigma}) \quad \Longrightarrow \quad |\{i \leq n : \sigma_i \neq \hat{\sigma}_i\}| \leq \varepsilon n \,.$$

Better: give an explicit dependence of δ on ε and a quantitative bound on the failure probability for this stability property.

The Multiple Valleys property is a strong negation of the above stability statements. Here we give a qualitative formulation of the property. In the general setup, we have a sequence $\mathcal{X}_n$ of sets of growing size equipped with a "similarity" function $s_n : \mathcal{X}_n \times \mathcal{X}_n \to [0,1]$, with $s_n(x,y) \approx 1$ (resp. 0) if x, y are highly aligned/similar (resp. almost orthogonal / dissimilar). A random function $f_n : \mathcal{X}_n \to \mathbb{R}$ has the Multiple Valleys property if it has many dissimilar near-optimizers with high probability. That is, with probability $1 - o(1)$, there is $A \subset \mathcal{X}_n$ with $|A| = \omega(1)$ such that

- $s_n(x,y) = o(1)$ for all distinct $x, y \in A$, and
- $f_n(x) \leq (1 + o(1)) \min_{z \in X_n} f_n(z)$ for all $x \in A$.

One defines the Multiple Peaks property analogously.

In 2008–09 Chatterjee established forms of the MV property for the (1+1)-dimensional Gaussian polymer and the SK model. Since then there have been important lines of work establishing successively improved quantitative forms of the Multiple Valleys property for different models. I hope to cover some of these developments in this course.

Chatterjee also showed that for general Gaussian fields, the Multiple Valleys property follows from Superconcentration:

Theorem 2.8 (Chatterjee '08)**.** *In the setup of Section 2.3, suppose $R(u,u) = 1$ for all $u \in V$ and $R(u,v) \geq 0$ for all $u, v \in V$. If $\operatorname{Var} \max_v g_v \leq \varepsilon$, then for some $\delta = O(1/\sqrt{\log(1/\varepsilon)})$ we have that with probability at least $1 - \delta$, there exists $A \subset V$ with $|A| \geq 1/\delta$ such that*

- *$g_u \geq (1-\delta) \max_v g_v$ for all $u \in A$, and*
- *$R(u,v) \leq \delta$ for all distinct $u, v \in A$.*

3. Markov semigroups

Chatterjee formally defines superconcentration and disorder chaos in the general setting of a Markov semigroup. Here we outline this general setting at a similar level of detail as in [Cha14], and specialize to the two examples we'll consider in this course, namely, the Ornstein–Uhlenbeck (OU) semigroup and the independent flips semigroup. For more detail and rigorous statements about Markov semigroups see for instance [BGL14].

The general theme is that to understand properties of a probability measure μ on a measurable space $\mathcal{X}$, and the distribution of functions on $(\mathcal{X}, \mu)$, it is useful to consider a Markov process $(X_t)_{t \geq 0}$ in $\mathcal{X}$ with μ as its unique stationary distribution.

3.1. **General setup.** So, let $(X_t)_{t \geq 0}$ be a continuous-time Markov process in an abstract measurable state space $\mathcal{X}$. From this we get a semigroup of operators P_t acting on bounded measurable functions $f : \mathcal{X} \to \mathbb{R}$ by

$$P_t f(x) := \mathbb{E}(f(X_t) | X_0 = x) = \mathbb{E}_x f(X_t)$$

the last expression introducing a convenient notation. The collection $(P_t)_{t\geq 0}$ is called an operator semigroup because $P_{t+s} = P_t \circ P_s$ for all $s, t \geq 0$.

We assume (X_t) has a unique stationary distribution μ and

$$P_t f(x) = \mathbb{E}_x f(X_t) \to \mathbb{E}_\mu f \quad \text{as } t \to \infty \quad \forall x \in \mathcal{X} \tag{3.1}$$

where $\mathbb{E}_\mu f$ is another way of writing $\int_{\mathcal{X}} f d\mu$. We denote the $L^2(\mathcal{X}, \mu)$-inner product by

$$\langle f, g \rangle := \int_{\mathcal{X}} f g d\mu = \mathbb{E}_\mu f g \,.$$

We denote the covariance

$$\mathrm{Cov}(f, g) = \mathrm{Cov}_\mu(f, g) := \int f g d\mu - \int f d\mu \int g d\mu.$$

The *generator* of the semigroup is an operator L on (nice enough) functions $f : \mathcal{X} \to \mathbb{R}$ defined as the derivative of P_t at $t = 0$:

$$Lf := \lim_{t\to 0} \frac{P_t f - f}{t} = \partial_t P_t|_{t=0} \tag{3.2}$$

for nice enough f. It is commonly the case (and is the case for the OU semigroup) that L is a differential operator and hence is only defined on a domain $\mathcal{D}(L)$ that is dense in $L^2(\mathcal{X}, \mu)$. Below we tacitly assume any function we apply L to is in its domain.

By applying (3.2) with $P_s f$ in place of f we find

$$\partial_t P_t = L P_t \tag{3.3}$$

for all $t \geq 0$.

We define the *Dirichlet form*

$$\mathcal{E}(f, g) := -\langle f, Lg \rangle = -\int_{cX} f L g d\mu \tag{3.4}$$

defined for nice enough f, g.

In this general setting we have the following useful lemma:

Lemma 3.1 (Covariance lemma)**.** *We have*

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

for all $f, g \in L^2(\mathcal{X}, \mu)$ for which the right hand side makes sense.

Proof (sketch). Assume f, g are (at least) bounded and measurable. Since $P_t g \to \int g d\mu$ pointwise (by (3.1)), from the bounded convergence theorem we can express

$$\begin{aligned}
\mathrm{Cov}_\mu(f,g) &= \langle f, g\rangle - \langle f, \lim_{t\to\infty} P_t g\rangle \\
&= \langle f, g\rangle - \lim_{t\to\infty} \langle f, P_t g\rangle \\
&= -\int_0^\infty \partial_t \langle f, P_t g\rangle dt \\
&= -\int_0^\infty \langle f, \partial_t P_t g\rangle dt \\
&= -\int_0^\infty \langle f, L P_t g\rangle dt \\
&= \int_0^\infty \mathcal{E}(f, P_t g) dt
\end{aligned}$$

as desired, where we applied (3.3) in the penultimate line. □

Exercise 3.1. Rigorously justify all steps in the above proof in the setting of the OU and independent flips semigroups (defined below).

3.2. **OU semigroup.** Here $\mathcal{X} = \mathbb{R}^d$ and $\mu = \gamma^d$, the standard Gaussian measure. Consider first the case $d = 1$. The OU process $(X_t)_{t\geq 0}$ satisfies the SDE

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

for $B = (B_t)$ a standard Brownian motion. An explicit solution is $X_t = e^{-t} X_0 + e^{-t} W_{e^{2t}-1}$ for W a different SBM.

We only need the semigroup, which is

$$P_t f(x) = \mathbb{E} f(e^{-t} x + \sqrt{1 - e^{-2t}} Z) \tag{3.5}$$

for Z standard Gaussian. Thus,

$$P_t f(x) \to f(x) \quad \text{as } t \downarrow 0\,, \qquad P_t f(x) \to \mathbb{E} f(Z) = \int_{\mathbb{R}} f d\gamma \quad \text{as } t \uparrow \infty.$$

Using the Gaussian integration by parts idendity

$$\mathbb{E} Z g(Z) = \mathbb{E} g'(Z) \tag{3.6}$$

one verifies

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

for f that is C^2 and not growing too fast at $\pm\infty$. Indeed, differentiating under the expectation and using the chain rule,

$$\partial_t P_t f(x) = -e^{-t} x \mathbb{E} f'(e^{-t} x + \sqrt{1 - e^{-2t}} Z) + e^{-2t} (1 - e^{-2t})^{-1/2} \mathbb{E} Z f'(e^{-t} x + \sqrt{1 - e^{-2t}} Z).$$

For the second term apply (3.6). Then take $t \to 0$ to obtain $L f(x)$.

From (3.6) one also gets the Dirichlet form

$$\mathcal{E}(f, g) = \mathbb{E} f'(Z) g'(Z). \tag{3.7}$$

The OU process in $\mathbb{R}^d$ is a random vector with entries that are independent 1-dimensional OU processes. One verifies (3.5) is still the formula for the semigroup, only with $Z \sim N(0, I_d)$. The generator and Dirichlet form are

$$Lf(x) = \Delta f(x) - x \cdot \nabla f(x)\,, \qquad \mathcal{E}(f,g) = \mathbb{E}_{\gamma^n} \nabla f \cdot \nabla g\,. \tag{3.8}$$

3.3. **Independent flips semigroup.**

....

(We'll come back to this)

4. Gaussian Poincaré inequality and applications

Here we state and prove a useful classical concentration inequality for Gaussian measure and give some applications.

In the setting of Section 3.1, a Markov semigroup $(P_t)_{t\geq 0}$ is said to satisfy the Poincaré inequality with constant K, or briefly, have property PI(K), if

$$\operatorname{Var}_\mu f \leq K\mathcal{E}(f,f) \tag{4.1}$$

for all $f \in L^2(\mathcal{X}, \mu)$ for which the right hand side makes sense. The smallest K for which this holds is called the Poincaré constant for the semigroup.

Theorem 4.1 (Gaussian Poincaré inequality)**.** *The OU semigroup on $(\mathbb{R}^d, \gamma^d)$ satisfies the Poincaré inequality with constant 1. That is, for $Z \sim N(0, I_d)$,*

$$\operatorname{Var} f(Z) \leq \mathbb{E}|\nabla f(Z)|^2 \tag{4.2}$$

for all smooth f.

Proof. By differentiating under the integral in (3.5) we see that

$$\nabla P_t f = e^{-t} P_t \nabla f. \tag{4.3}$$

Combining this with Lemma 3.1 we have

$$\operatorname{Var} f = \int_0^\infty \mathcal{E}(f, P_t f) dt = \int_0^\infty \mathbb{E}\nabla f \cdot \nabla P_t f dt = \int_0^\infty e^{-t}\mathbb{E}\nabla f \cdot P_t \nabla f dt \tag{4.4}$$

where the variance and expectation are taken with respect to the Gaussian measure γ^n. Applying Cauchy–Schwarz twice (first to for dot product on $\mathbb{R}^n$, then for the $L^2(\gamma^n)$ inner product), we have

$$\mathbb{E}\nabla f \cdot P_t \nabla f \leq \mathbb{E}|\nabla f||P_t f| \leq (\mathbb{E}|\nabla f|^2 \mathbb{E}|P_t \nabla f|^2)^{1/2} \tag{4.5}$$

where $|\cdot|$ is the Euclidean norm on $\mathbb{R}^n$. Now note that for any $h : \mathbb{R}^n \to \mathbb{R}^n$ with components $h_i \in L^2(\gamma^n)$, from Jensen's inequality

$$|P_t h(x)|^2 = |\mathbb{E}_Z h(e^{-t}x + \sqrt{1-e^{-2t}}Z)|^2 \leq \mathbb{E}_Z |h(e^{-t}x + \sqrt{1-e^{-2t}}Z)|^2$$

so

$$\mathbb{E}|P_t \nabla f|^2 = \mathbb{E}_G |P_t h(G)|^2 \leq \mathbb{E}_G \mathbb{E}_Z |h(e^{-t}G + \sqrt{1-e^{-2t}}Z)|^2$$

where G, Z are independent standard normal vectors. Noting that $e^{-t}G + \sqrt{1-e^{-2t}}Z$ is also standard normal, we conclude

$$\mathbb{E}|P_t h|^2 \leq \mathbb{E}|h|^2. \tag{4.6}$$

Applying this with $h = \nabla f$ in (4.5), we conclude

$$\mathbb{E}\nabla f \cdot P_t \nabla f \leq \mathbb{E}|\nabla f|^2$$

for all $t \geq 0$. Substituteing this bound in (4.4) and integrating concludes the proof. □

Example 4.2 (Sum of iid Gaussians)**.** Note that the Gaussian Poincaré inequality is sharp for $f(x) = x_1 + \cdots + x_n$, as both sides are equal to n.

Example 4.3 (Max of iid Gaussians)**.** Let $f(x) = \max_{1 \leq v \leq n} x_v$. Then f is absolutely continuous with partial derivatives $\partial_v f(x) = 1_{x_v \geq x_u \, \forall u} = 1_{\hat{v}(x)=v}$ a.e., where $\hat{v}(x)$ is the maximizing coordinate, uniquely defined for a.e. $x \in \mathbb{R}^n$. Thus

$$|\nabla f(x)|^2 = \sum_{v \in [n]} 1_{\hat{v}(x)=v} = 1 \quad a.e.$$

For iid Gaussian inputs g_v we thus have $\mathbb{E}|\nabla f(g)|^2 = 1$, so Theorem 4.1 gives $\operatorname{Var} f(g) \leq 1$. However, a standard exercise shows $\operatorname{Var} f(g) \asymp 1/\log n$, so the Poincaré inequality does not capture the correct scale of fluctuations for this function.

Example 4.4 (Gaussian polymer)**.** Now we use Theorem 4.1 to prove the bound (2.4) for the ground state energy of the Gaussian polymer model. We regard E_n as a function on Euclidean space, with a coordinate direction for each lattice site that the path can use. Note that this space is finite dimensional: the walk can only pass through the sites

$$V := \{(x, y) \in \mathbb{Z}^2 : |y| \leq x \leq n\}$$

which has size of order n^2. As a function on $\mathbb{R}^V$ the ground state energy is

$$E_n = \min_{p \in P_n} -\sum_{v \in p} x_v = -\sum_{v \in \hat{p}_n(x)} x_v$$

where $\hat{p}_n(x)$ is that a.e. uniquely defined minimizing path. While E_n is not differentiable everywhere, it's a minimum of linear functionals and hence absolutely continuous, with a.e. defined partial derivatives

$$\frac{\partial E_n}{\partial x_u}(x) = -1_{u \in \hat{p}_n(x)} \qquad a.e.$$

Hence, the gradient $\nabla E_n : \mathbb{R}^V \to \mathbb{R}^V$ is

$$\nabla E_n(x) = -\mathbf{1}_{\hat{p}_n(x)} \qquad a.e.$$

where $\mathbf{1}_U \in \mathbb{R}^V$ is the indicator of $U \subset V$. Since all paths $p \in P_n$ are of size n,

$$|\nabla E_n(x)|^2 = n \qquad a.e.$$

so Theorem 4.1 gives

$$\operatorname{Var}_{\gamma^n} E_n \leq n\,.$$

as desired.

Example 4.5 (Branching random walk)**.** Let T_n be the set of 2^n edges in the binary tree of depth n. Let $(g_e)_{e\in T_n}$ be iid standard Gaussians. Let V_n be the set of leaf vertices at level n in T_n. For each $v \in V_n$ let $X_v = \sum_e g_e$, where the sum is over the n edges along the path from the root of the tree to v. Thus, $(X_v)_{v\in V_n}$ is a collection of 2^n Gaussian random walks of length n, with a branching correlation structure. Let

$$M_n := \max_{v\in V_n} X_v\,. \tag{4.7}$$

This is the position of the right-most particle at time n in a binary Gaussian branching random walk process. By similar reasoning as in Example 4.4, the gradient of M_n, as a function of the disorder variables (g_e), is the indicator of the path $\hat{p} \subset T_n$ from the root to the maximizer $\hat{v}$ in (4.7). By the same reasoning we get $\operatorname{Var} M_n \le n$.

Note: The true order of the variance is $O(1)$. See [Cha14] for discussion and references.

Exercise 4.1 (Exercise 2.3 in [Cha14])**.** For the SK model (see Section 2.6), use Theorem 4.1 to show the free energy $F_n(\beta)$ has variance $O(n)$, where the implicit constant is universal.

5. Superconcentration and chaos

Here we define superconcentration and chaos and show they are equivalent.

Definition 5.1 (Superconcentration)**.** Let $(P_t)_{t\ge 0}$ be a Markov semigroup on a measurable space $\mathcal{X}$ with stationary distribution μ, as in the setup of Section 3.1, and suppose the semigroup satisfies a Poincaré inequality as in (4.1). We say a function $f : \mathcal{X} \to \mathbb{R}$ is *ε-superconcentrated* if

$$\operatorname{Var}_\mu f \le \varepsilon K\mathcal{E}(f, f). \tag{5.1}$$

Thus, for a semigroup satisfying a Poincaré inequality, any function is 1-superconcentrated. We informally say a function is superconcentrated if it is ε-superconcentrated for small ε (usually of size $o(1)$ with respect to an asymptotic parameter $n \to \infty$).

Example 5.2. From the computations in Example 4.3 we see that for $\mathbb{R}^n$ equipped with the standard Gaussian measure γ^n, the function $f(x) = \max_{1\le v\le n} x_v$ is $\Theta(1/\log n)$-superconcentrated. By contrast, the function $f(x) = x_1 + \cdots + x_n$ considered in Example 4.2 is only 1-superconcentrated, i.e. is not superconcentrated.

We turn to define (disorder) chaos. In the examples of Section 2, the chaos phenomenon was informally illustrated as a sensitivity of optimizers to small perturbations of the underlying Gaussian random variables. In the abstract Markov semigroup setup we interpret this dynamically in terms of the semigroup.

In our applications (and often) it holds that for the generator L for the Markov process, $-L$ is self-adjoint, non-negative definite, with a countable sequence of eigenvalues $0 = \lambda_0 \le \lambda_1 \le \cdots$, and a corresponding sequence of eigenfunctions $u_0, u_1, \ldots$, where $u_0 \equiv 1$, forming a complete orthonormal basis for $L^2(\mathcal{X}, \mu)$. We can expand functions, the semigroup and the Dirichlet form in the spectral basis: for $f \in L^2(\mathcal{X}, \mu)$ we have

$$f = \sum_{k=0}^{\infty} \langle u_k, f\rangle u_k$$

and $\langle u_0, f\rangle = \int f d\mu$, so

$$\mathrm{Var}_\mu f = \|f\|_{L^2}^2 - (\int f d\mu)^2 = \sum_{k=1}^{\infty} \langle u_k, f\rangle^2.$$

Also,

$$\mathcal{E}(f,f) = -\int fLf d\mu = \sum_{k=1}^{\infty} \lambda_k \langle u_k, f\rangle^2. \tag{5.2}$$

From the previous two displays we see:

Proposition 5.3. *A semigroup satisfies a Poincaré inequality if and only if $\lambda_1 > 0$, in which case the Poincaré constant is $1/\lambda_1$.*

Proof. If $\lambda_1 > 0$ then

$$\mathrm{Var}_\mu f = \sum_{k=1}^{\infty} \langle u_k, f\rangle^2 \le \frac{1}{\lambda_1} \sum_{k=1}^{\infty} \lambda_k \langle u_k, f\rangle^2 = \frac{1}{\lambda_1} \mathcal{E}(f,f)$$

with equality for $f = u_1$. On the other hand, if $\lambda_1 = 0$ then for $f = u_1$ we have $\mathrm{Var}_\mu f = 1$ while $\mathcal{E}(f,f) = 0$, so that a Poincaré inequality cannot hold. □

From the heat equation $\partial_t P_t = LP_t$ we have $P_t = e^{tL}$, so P_t is also diagonalized by the eigenbasis u_i, and

$$\mathcal{E}(f, P_t f) = -\int fLe^{tL} f d\mu = \sum_{k=1}^{\infty} \lambda_k e^{-\lambda_k t} \langle u_k, f\rangle^2. \tag{5.3}$$

When the semigroup satisfies a Poincaré inequality we can bound

$$\mathcal{E}(f, P_t f) \le \sum_{k=1}^{\infty} \lambda_k e^{-\lambda_1 t} \langle u_k, f\rangle^2 = e^{-\lambda_1 t} \mathcal{E}(f,f).$$

We say f is chaotic if we can improve this bound:

Definition 5.4 (Chaos)**.** In the above setting, we say a function $f \in L^2(\mu)$ is (ε, δ)-chaotic if

$$\mathcal{E}(f, P_t f) \le \varepsilon e^{-\lambda_1 t} \mathcal{E}(f,f) \quad \forall t \ge \delta\,.$$

Remark 5.5 (Disorder vs. temperature chaos). We have defined what is more specifically sometimes called *disorder chaos*, which encapsulates the sensitivity of the model to small perturbations to the underlying random disorder. One can also consider sensitivity with respect to variation of other model parameters. A common alternative for models of statistical physics is *temperature chaos* (which is easier to define and implement in an experimental setup).

As we will see, in many examples where f takes the form of a maximum of linear functionals of a Gaussian vector, the Dirichlet form $\mathcal{E}(f, P_t f)$ can be related to a measure of the "overlap" between optimizing functionals for two different Gaussian inputs with covariance e^{-t}. Moreover, it can be shown $\mathcal{E}(f, P_t f)$ is non-increasing in t, which via Lemma 3.1 implies a connection between superconcentration and chaos.

Let's see what this definition looks like in a particular setup generalizing the Gaussian polymer of Section 2.4. Suppose for a finite set V and a family $P \subset 2^V$ of subsets ("paths") we set

$$f : \mathbb{R}^V \to \mathbb{R}, \qquad f(x) := \max\{x \cdot \mathbf{1}_p : p \in P\} = \max_{p \in P} \sum_{v \in p} x_v.$$

Let $\hat{p}(x) \in P$ denote the optimizer, which is uniquely defined for a.e. $x \in \mathbb{R}^V$. Then the gradient at a point x is a.e. given by the indicator of the optimizing path:

$$\nabla f(x) = \mathbf{1}_{\hat{p}(x)} \in \mathbb{R}^V .$$

Equip $\mathbb{R}^V$ with the standard Gaussian measure $\gamma = \gamma^V$, and let P_t be the OU semigroup, with its Dirichlet form $\mathcal{E}(h,k) = \int \nabla h \cdot \nabla k d\gamma$. Then for $g \sim \gamma$ and f as above,

$$\mathcal{E}(f,f) = \mathbb{E}|\nabla f(g)|^2 = \mathbb{E}|\hat{p}(g)|$$

where $|\hat{p}(g)|$ is the cardinality of $\hat{p}(g)$. On the other hand, writing

$$x^t := e^{-t}x + \sqrt{1 - e^{-2t}}Z$$

for $x \in \mathbb{R}^V$ and $Z \sim \gamma$ independent of g, recalling (4.3)–(4.4) we have

$$\mathcal{E}(f, P_t f) = e^{-t}\mathbb{E}\nabla f(g) \cdot \nabla f(g^t) = e^{-t}\mathbb{E}|\hat{p}(g) \cap \hat{p}(g^t)|. \tag{5.4}$$

Thus, f is chaotic precisely when the overlap $|\hat{p}(g) \cap \hat{p}(g^t)|$ between the optimizing paths for g and for the perturbed vector g^t is small in expectation compared with $\mathbb{E}|\hat{p}(g)|$, for all t exceeding a small threshold.

The following result of Chatterjee establishes equivalence of superconcentration and chaos (with some dependence of parameters):

Theorem 5.6. *In the above setup, if f is ε-superconcentrated, then for any $\delta > 0$, f is (ε', δ)-superconcentrated, for $\varepsilon' = \varepsilon e^{\lambda_1 \delta}/(\lambda_1 \delta)$. Conversely, if f is (ε, δ) chaotic, then f is $(\varepsilon + \lambda_1 \delta)$-superconcentrated.*

Proof. Suppose f is ε-superconcentrated. Then by Lemma 3.1,

$$\frac{\varepsilon}{\lambda_1}\mathcal{E}(f,f) \geq \mathrm{Var}_\mu f = \int_0^\infty \mathcal{E}(f, P_t f)dt.$$

Since $P_t = e^{tL}$,

$$\mathcal{E}(f, P_t f) = \sum_{k=1}^{\infty} \lambda_k e^{-\lambda_k t}\langle u_k, f\rangle^2 \tag{5.5}$$

from which we see that $\mathcal{E}(f, P_t f)$ is non-negative and decreasing in t. Thus for any $\delta > 0$,

$$\int_0^\infty \mathcal{E}(f, P_t f)dt \geq \int_0^\delta \mathcal{E}(f, P_t f)dt \geq \delta \mathcal{E}(f, P_\delta f). \tag{5.6}$$

Moreover, from (5.5) we see that $e^{\lambda_1 t}\mathcal{E}(f, P_t f)$ is non-increasing, so for $t \geq \delta$,

$$e^{\lambda_1 t}\mathcal{E}(f, P_t f) \leq e^{\lambda_1 \delta}\mathcal{E}(f, P_\delta f) \leq \varepsilon \frac{e^{\lambda_1 \delta}}{\lambda_1 \delta}\mathcal{E}(f,f) \tag{5.7}$$

so f is (ε', δ)-chaotic.

Now assume f is (ε, δ)-chaotic. Then since $t \mapsto \mathcal{E}(f, P_t f)$ is non-increasing,

$$\operatorname{Var} f = \int_0^\infty \mathcal{E}(f, P_t f)dt \leq \int_0^\delta \mathcal{E}(f, f)dt + \int_\delta^\infty \varepsilon e^{-\lambda_1 t}\mathcal{E}(f, f)dt \leq (\delta + \frac{\varepsilon}{\lambda_1})\mathcal{E}(f, f) \quad (5.8)$$

as desired. □

For later use we record the following lemma, which follows from similar lines as in the above proof:

Lemma 5.7. *For any $t > 0$ we have*

$$\operatorname{Var} f \geq \frac{1}{\lambda_1}(1 - e^{-\lambda_1 t})e^{\lambda_1 t}\mathcal{E}(f, P_t f).$$

Proof. We noted in the proof of Theorem 5.6 that $e^{\lambda_1 t}\mathcal{E}(f, P_t f)$ is non-increasing in t, so from Lemma 3.1,

$$\operatorname{Var} f = \int_0^\infty e^{-\lambda_1 s}e^{\lambda_1 s}\mathcal{E}(f, P_s f)ds \geq e^{\lambda_1 t}\mathcal{E}(f, P_t f)\int_0^t e^{-\lambda_1 s}ds$$

and the claim follows. □

6. Superconcentration from hypercontractivity

6.1. **Hypercontractivity.** We continue the Markov semigroup setup from Section 3. From Jensen's inequality we see that a semigroup P_t is always an L^2-contraction:

$$\|P_t f\|_{L^2(\mu)} \leq \|f\|_{L^2(\mu)} \quad \forall t \geq 0. \quad (6.1)$$

Indeed, the square of the left hand side is $\int |\mathbb{E}_x f(X_t)|^2 d\mu(x) \leq \int \mathbb{E}_x |f(X_t)|^2 d\mu(x)$, and for $x \sim \mu$ we have $X_t \sim \mu$ for all t.

A semigroup (P_t) is said to be *hypercontractive* if the above can be strengthened to

$$\forall p > 1, t > 0 \quad \exists q(t, p) > p : \qquad \|P_t f\|_{L^q(\mu)} \leq \|f\|_{L^p(\mu)} \quad (6.2)$$

for all measurable f. Such an inequality quantifies the smoothing nature of the semigroup: for larger q, the L^q norm is more sensitive to "spikes" of f, so the above family of inequalities expresses that as t increases, P_t softens the spikes of f (and of course as $t \to \infty$, $P_t f$ converges to a constant, which isn't spiky at all).

We only consider two sorts of semigroups in this class, and both are hypercontractive (with the same $q(t, p)$).

Theorem 6.1 (Nelson '73)**.** *The OU semigroup is hypercontractive, with $q(t, p) = 1 + (p - 1)e^{2t}$.*

Theorem 6.2 (Bonami '70, Beckner '75)**.** *The independent flips semigroup is hypercontractive, with $q(t, p) = 1 + (p - 1)e^{2t}$.*

In fact it was shown by Gross in 1975 that any semigroup satisfying a log-Sobolev inequality is hypercontractive. See the appendix of [Cha14].

Here is a way to see that hypercontractivity is connected with disorder chaos (and hence superconcentration): Let g be a standard Gaussian vector in $\mathbb{R}^n$, and let (g^t) be

the OU process with $g^0 = 0$. Fix a Borel set $A \subset \mathbb{R}^n$. Let $p = p(t) \in (1,2)$ satisfy $2 = q(t,p)$. Then with $f = 1_A$ we have

$$\begin{aligned}\mathbb{P}(g \in A, g^t \in A) &= \mathbb{E} f(g) f(g^t) \\ &= \int f P_t f \, d\gamma \le \|f\|_{L^2} \|P_t f\|_{L^2} \le \|f\|_{L^2} \|f\|_{L^p} = \mathbb{P}(g \in A)^{\frac{1}{2}+\frac{1}{p}}\end{aligned}$$

where all L^p norms are with respect to γ. Thus,

$$\mathbb{P}(g^t \in A | g \in A) \le \mathbb{P}(g \in A)^{\frac{1}{2}+\frac{1}{p}-1}.$$

Since $p < 2$, the exponent on the right hand side is strictly positive. Thus, the OU process quickly exits any set of small Gaussian measure.

Remark 6.3 (t is small!). One should note that in all of the applications of hypercontractivity that follow we take t to be small (or, when we integrate a bound resulting from hypercontractivity over all t, the main contribution will come from small t).

6.2. **Talagrand's L^1-L^2 bound.** The following is a Gaussian version of a result of Talagrand '94 for the Bernoulli cube:

Theorem 6.4. *For any $f \in L^2(\mathbb{R}^n, \gamma^n)$,*

$$\mathrm{Var}_{\gamma^n} f \lesssim \sum_{i=1}^n \frac{\|\partial_i f\|^2_{L^2(\gamma^n)}}{1 + \log \frac{\|\partial_i f\|_{L^2(\gamma^n)}}{\|\partial_i f\|_{L^1(\gamma^n)}}} \tag{6.3}$$

where the implicit constant is universal.

Remark 6.5. Note that the right hand side in (6.3) increases with $\|\partial_i f\|_{L^1(\gamma^n)}$ as well as with the ratio $\|\partial_i f\|_{L^2(\gamma^n)}/\|\partial_i f\|_{L^1(\gamma^n)}$ (since $x \mapsto x^2/(1+\log x)$ is monotone on $[1,\infty)$). Hence to bound the variance it suffices to find *upper* bounds for the L^1 an L^2 norms of partial derivatives of f.

Before proving the theorem we consider a few simple examples.

Example 6.6 (Sum of iid Gaussians)**.** For $f(x) = x_1 + \cdots + x_n$, we have $\partial_i f(x) \equiv 1$ for each i, so $\|\partial_i f\|_{L^q(\gamma^n)} = 1$ for all q. Theorem 6.4 thus gives $\mathrm{Var}_{\gamma^n} f \lesssim n$, which as was noted in Example 4.2 is sharp.

Example 6.7 (Max of iid Gaussians)**.** Let $f(x) = \max_{v \in [n]} x_v$. In Example 4.3 we saw $\partial_v f(x) = 1_{v = \hat{v}(x)}$ a.e., where $\hat{v}(x) \in [n]$ is the a.e.-uniquely defined maximizing index. Thus,

$$\|\partial_v f\|_{L^q(\gamma^n)} = \mathbb{P}(\hat{v}(g) = v)^{1/q} = n^{-1/q}$$

for all $q > 0$, and Theorem 6.4 obtains the sharp bound $O(1/\log n)$ for the variance.

Example 6.8 (Branching random walk)**.** (Returning to Example 4.5). Viewing M_n as a function $F(g)$ of the Gaussian weights $g = (g_e)_{e \in T_n}$, with $\partial_e F$ the partial derivative of F with respect to the weight on edge e, we have

$$\partial_e F(g) = 1_{e \in \hat{p}(g)}$$

where $\hat{p}(g)$ is the path of length n from the root to the (a.s. unique) optimizing vertex $\hat{v}(g) = \arg\max_{v \in V_n} X_v$. Say an edge $e \in T_n$ is at level k if its endpoints are at distance $k-1$ and k from the root. There are 2^k edges at level k. By symmetry, for any $q > 0$,

$$\|\partial_e F\|_{L^q(\gamma)} = \mathbb{P}(e \in \hat{p}(g))^{1/q} = 2^{-k/q}$$

(writing γ for the standard Gaussian measure on $\mathbb{R}^{T_n}$). From Theorem 6.4 we can thus bound

$$\operatorname{Var} M_n \lesssim \sum_{k=1}^{n} \sum_{e \text{ at level } k} \frac{2^{-k}}{1 + \log(2^{k/2})} = \sum_{k=1}^{n} \frac{2^k 2^{-k}}{1 + \frac{k}{2}\log 2} \asymp \sum_{k=1}^{n} \frac{1}{k} \asymp \log n$$

which drastically improves on the bound n obtained in Example 4.5 using the Poincaré inequality.

6.3. **Proof of Theorem 6.4.** From the Covariance Lemma (Lemma 3.1) and identity (4.4) for the OU semigroup, for $g \sim \gamma^n$ we have

$$\operatorname{Var} f(g) = \int_0^\infty e^{-t} \mathbb{E} \nabla f(g) \cdot P_t \nabla f(g) dt = \sum_{i=1}^{n} \int_0^\infty e^{-t} \mathbb{E} \partial_i f(g) P_t \partial_i f(g)\, dt\,. \tag{6.4}$$

Consider arbitrary fixed $t > 0$ and $h \in L^2(\gamma^n)$. Set

$$p := 1 + e^{-2t} \in (1,2) \tag{6.5}$$

so that with $q(t,p) = 1 + e^{2t}(p-1)$ as in Theorem 6.1 we have $q(t,p) = 2$. From Cauchy–Schwarz and Theorem 6.1,

$$\mathbb{E} h(g) P_t h(g) \le \|h\|_{L^2(\gamma^n)} \|P_t h\|_{L^2(\gamma^n)} \le \|h\|_{L^2(\gamma^n)} \|h\|_{L^p(\gamma^n)}\,.$$

Next we use interpolation to bound the L^p norm in terms of L^1 and L^2 norms (recall $p \in (1,2)$). By Hölder's inequality,

$$\|h\|^p_{L^p(\gamma^n)} = \| |h|^{2(p-1)} |h|^{2-p} \|_{L^1(\gamma^n)} \le \|h^2\|^{p-1}_{L^1(\gamma^n)} \|h\|^{2-p}_{L^1(\gamma^n)} = \|h\|^{2p-2}_{L^2(\gamma^n)} \|h\|^{2-p}_{L^1(\gamma^n)}$$

so

$$\|h\|_{L^p(\gamma^n)} \le \|h\|^{2-\frac{2}{p}}_{L^2(\gamma^n)} \|h\|^{\frac{2}{p}-1}_{L^1(\gamma^n)}.$$

Hence, if $A < \infty$, $a \le 1$ are constants such that

$$\|h\|_{L^2(\gamma^n)} \le A\,, \qquad \frac{\|h\|_{L^1(\gamma^n)}}{\|h\|_{L^2(\gamma^n)}} \le a$$

then combining the previous displays we get

$$\mathbb{E} h(g) P_t h(g) \le \|h\|^2_{L^2(\gamma^n)} \left(\frac{\|h\|_{L^1(\gamma^n)}}{\|h\|_{L^2(\gamma^n)}} \right)^{\frac{2}{p}-1} \le A^2 a^{\frac{2}{p}-1} \tag{6.6}$$

(note $\frac{2}{p} - 1 \in (0,1)$ since $p \in (1,2)$). In terms of t we have

$$\frac{2}{p} - 1 = \tanh t \ge 1 - e^{-t} \tag{6.7}$$

so

$$\mathbb{E} h(g) P_t h(g) \le A^2 a^{1-e^{-t}}. \tag{6.8}$$

Returning to (6.4), we apply (6.8) with $h = \partial_i f$ to get if $\|\partial_i f\|_{L^2(\gamma^n)} \le A_i$ and $\|\partial_i f\|_{L^1(\gamma^n)} \le a_i \le 1$, then

$$\operatorname{Var} f(g) \le \sum_{i=1}^n A_i^2 \int_0^\infty e^{-t} a_i^{1-e^{-t}} dt = \sum_{i=1}^n A_i^2 \int_0^1 a_i^u du \lesssim \sum_{i=1}^n \frac{A_i^2}{1+\log(1/a_i)}. \tag{6.9}$$

The claim follows. □

The bound (6.7) and change of variable in (6.9) is slick, but one checks a more bare-hands approach using e.g.

$$\frac{2}{p} - 1 \ge \frac{1}{10}\min(1,t) \tag{6.10}$$

leads to the same bound (up to modification of the implicit constant):

Exercise 6.1. Give an alternative derivation of the final bound in (6.9) using (6.10) in place of (6.7).

The point is that when integrating the bound (6.6) over t (with $p = p(t)$) most of the gain comes from the linear growth of $\frac{2}{p} - 1$ for small t (as was forecast in Remark 6.3).

6.4. **Application to first-passage percolation.** We use Theorem 6.4 to prove the following variant of the seminal result of Benjamini–Kalai–Schramm establishing superconcentration for first-passage percolation (see the upper bound in (2.3)). (The original work of BKS covered weights with distribution supported on two points.)

Theorem 6.9. *In the setting of Section 2.1, suppose the weights ω_e take the form $\omega_e = \phi(g_e)$ with g_e standard Gaussian, and $\phi : \mathbb{R} \to [a,b]$ a fixed Lipschitz function into a compact interval $[a,b] \subset (0,\infty)$. Then*

$$\operatorname{Var} T_n \lesssim_{\phi,d} n/\log n. \tag{6.11}$$

The assumptions on the distribution of weights covers, for instance, the uniform distribution on $[a,b]$.

Before turning to the proof we discuss the idea. The proof is not a direct application of Theorem 6.4 as it was for the the branching random walk.

First note that T_n only depends on the weights $(g_e)_{e\in E'}$ for the finite set E' of edges within an origin-centered ball of radius $O_{a,b,d}(n)$. Indeed, by considering the path from the origin to ne_1 along the e_1-axis we see that $T_n \le bn$ a.s., while any path that exits an origin-centered ball of radius r (in the ℓ^1-metric, say) has weight at least ar, so the optimal path will not exit such a ball if $r > bn/a$.

For a direct application of Theorem 6.4 we would regard T_n as a function $T_n = F(g)$ of the Gaussian variables $(g_e)_{e\in E'}$, with

$$F : \mathbb{R}^{E'} \to \mathbb{R}\,, \qquad F(x) = \min_p \sum_{e\in P} \phi(x_e)\,. \tag{6.12}$$

Writing ∂_e for partial differentiation with respect to the variable x_e, we have

$$\partial_e F(x) = \phi'(x_e) 1_{e\in \hat p(x)} \tag{6.13}$$

for Lebesgue-a.e. x. So for $q = 1, 2$ and γ the standard Gaussian measure on $\mathbb{R}^{E'}$, we have

$$\|\partial_e F\|_{L^q(\gamma)} \leq \|\phi'\|_{L^\infty} \mathbb{P}(e \in \hat{p}(g))^{1/q}.$$

Thus, if we could show that for most $e \in E'$,

$$\mathbb{P}(e \in \hat{p}(g)) \lesssim n^{-c} \tag{6.14}$$

for some constant $c > 0$, then (6.11) would quickly follow from an application of Theorem 6.4, since the expected length of $\hat{p}(g)$ is $O(n)$. While (6.14) certainly seems likely to hold, it is not clear how to show this directly. To get around this issue, Benjamini et al use a nice averaging trick to pass to another random variable $\widetilde{T}_n$ for which L^1 and L^2 norms of partial derivatives can be controlled using basic considerations of the geometry of paths.

While a direct application of Theorem 6.4 via (6.14) does not seem possible for LPP on $\mathbb{Z}^d$, this approach does work in settings with more symmetry:

Exercise 6.2 (FPP on the torus)**.** Let E be the edge set of the nearest-neighbor graph on the d-dimensional discrete torus $(\mathbb{Z}/n\mathbb{Z})^d$. With weights $\omega_e = \phi(g_e)$ as in Theorem 6.9, let T_n denote the minimum passage time for a closed path p that "wraps around" the torus in the e_1 direction – that is, p contains at least one edge from each of the n slices $S_k = \{e = (u, v) : u_1 = k\}$, $k = 0, \ldots, n-1$. Show that $\operatorname{Var} T_n \lesssim_{\phi,d} n/\log n$.

Proof of Theorem 6.9. By scaling ϕ we may assume without loss of generality that $\|\phi'\|_{L^\infty} \leq 1$. We will show

$$\operatorname{Var} T_n \lesssim_{a,b,d} n/\log n. \tag{6.15}$$

Let $s = s(n)$ be a slowly growing function of n to be chosen later, and let B denote the origin-centered lattice box of sidelength s. Set

$$\widetilde{T}_n := \frac{1}{|B|} \sum_{v \in B} T_n^v \tag{6.16}$$

where T_n^v denotes the passage time from vertex v to vertex $v + ne_1$ (thus $T_n = T_n^0$, where 0 denotes the origin in $\mathbb{Z}^d$). Note that

$$T_n^u \leq T_n^v + O_d(s) \tag{6.17}$$

for any $u, v \in B$, since we get a path from u to $u + ne_1$ of weight $T_n^v + O_d(s)$ by taking the union of the optimizing path from v to ne_1 with paths of length $O_d(s)$ from u to v and $u + ne_1$ to $v + ne_1$. By interchanging u, v we get the reverse bound, and hence

$$T_n = \widetilde{T}_n + O_d(s). \tag{6.18}$$

From this one sees that

$$\operatorname{Var} T_n \leq 2 \operatorname{Var} \widetilde{T}_n + O_d(s^2). \tag{6.19}$$

To prove (6.15) it thus suffices to show

$$\operatorname{Var} \widetilde{T}_n \lesssim_{a,b,d} n/\log n \tag{6.20}$$

for some choice of s of size at most $O_{a,b,d}(\sqrt{n/\log n})$.

As in (6.12) we may view $\widetilde{T}_n$ as a function $\widetilde{F}(g)$ of a standard Gaussian vector $g \in \mathbb{R}^{E'}$, where E' is the set of edges in an origin-centered ball of radius $O_{a,b,d}(n)$ (assuming $s \lesssim n$). Now we have

$$\partial_e \widetilde{F}(g) = \frac{\phi'(g_e)}{|B|} \sum_{v \in B} 1_{e \in \hat{p}_v(g)}. \tag{6.21}$$

Hence,

$$\|\partial_e \widetilde{F}\|_{L^1(\gamma)} \le \|\phi'\|_{L^\infty} A_e \tag{6.22}$$

where

$$A_e := \frac{1}{|B|} \sum_{v \in B} \mathbb{P}(e \in \hat{p}_v(g)). \tag{6.23}$$

To bound this quantity, we note that by stationarity of the Gaussian field $(g_e)_{e \in \mathbb{Z}^d}$ under translations of the lattice, we have

$$\mathbb{P}(e \in \hat{p}_v(g)) = \mathbb{P}(e - v \in \hat{p}_v(g)) \tag{6.24}$$

where $e - v$ denotes the edge obtained by translating its endpoints by the lattice vector v. Thus,

$$\sum_{v \in B} \mathbb{P}(e \in \hat{p}_v(g)) = \mathbb{E}|\hat{p}_0(g) \cap (e - B)| \tag{6.25}$$

where $e - B$ is the set of edges parallel to e whose endpoint closest to the origin is in the box of sidelength s centered at the endpoint of e closest to the origin.

Now for the punchline: we must have

$$|\hat{p}_0(g) \cap (e - B)| \lesssim_{a,b,d} s \qquad a.s. \tag{6.26}$$

Indeed, we can always join the edges of first and last entry of $\hat{p}_0(g)$ into the box $e - B$ by a path of length at most ds, which has passage time at most bds, whereas any path of length ℓ in the box has passage time at least ℓa, so the optimizing path $\hat{p}_0(g)$ can have at most bds/a edges in B, from which (6.26) follows. (In fact we get a bound independent of d since $e - B$ only contains edges in one of the coordinate directions.) Since B has $O(s)^d$ vertices, we conclude

$$A_e = \frac{1}{|B|} \mathbb{E}|\hat{p}_0(g) \cap (e - B)| \lesssim_{a,b,d} s^{1-d}. \tag{6.27}$$

For the L^2 norms, we can bound

$$\|\partial_e F\|^2_{L^2(\gamma)} \le \mathbb{E}\left(\frac{1}{|B|} \sum_{v \in B} 1_{e \in \hat{p}_v(g)} \right)^2 \le \mathbb{E} \frac{1}{|B|} \sum_{v \in B} 1_{e \in \hat{p}_v(g)} = A_e \tag{6.28}$$

where we applied Cauchy–Schwarz to the average over B. By Theorem 6.4,

$$\operatorname{Var} \widetilde{T}_n \lesssim \sum_{e \in E'} \frac{A_e}{1 + \log(1/\sqrt{A_e})} \lesssim_{a,b,d} \frac{1}{\log s} \sum_{e \in E'} A_e \,.$$

Now we can bound

$$\sum_{e \in E'} A_e = \frac{1}{|B|} \sum_{v \in B} \sum_{e \in E'} \mathbb{P}(e \in \hat{p}_v(g)) = \frac{1}{|B|} \sum_{v \in B} \mathbb{E}|\hat{p}_v(g)| \lesssim_{a,b} n$$

so

$$\operatorname{Var} \widetilde{T}_n \lesssim_{a,b,d} n/\log s.$$

Taking $s = n^c$ for any fixed constant $c \in (0, \frac{1}{2})$, we get (6.20) to conclude the proof. □

Exercise 6.3. By tracking the dependence of bounds on the dimension in the proof of Theorem 6.9, obtain a version of the bound (6.11) with explicit dependence on d.

7. Spectral approach to superconcentration

7.1. **Expansion in Hermite polynomials.** In Section 5 we considered bounds for the variance and Dirichlet form obtained by diagonalizing the generator. Here we consider the case of the OU semigroup on $\mathbb{R}^d$:

$$P_t f(x) = \mathbb{E}\Big(e^{-t}x + \sqrt{1-e^{-2t}}Z\Big), \qquad Z \sim N(0, I_d)$$

with generator

$$Lf(x) = \Delta f(x) - x \cdot \nabla f(x).$$

The (probabilist's) Hermite polynomials are

$$H_k(x) := (-1)^k e^{x^2/2} \frac{d^k}{dx^k} e^{-x^2/2}, \qquad x \in \mathbb{R},\ k \geq 0. \tag{7.1}$$

Each H_k is a polynomial of degree k with degree-k coefficient equal to 1. The first three are

$$H_0 \equiv 1, \quad H_1(x) = x, \quad H_2(x) = x^2 - 1.$$

By direct differentiation one verifies that the H_k are eigenfucntions of the generator L for the 1-dimensional OU semigroup, with

$$LH_k = -kH_k. \tag{7.2}$$

For f a Schwartz function on $\mathbb{R}$, from integration by parts we get

$$\langle f, H_k \rangle_{L^2(\gamma)} = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} f(x)(-1)^k \frac{d^k}{dx^k} e^{-x^2/2} dx = \langle f', H_{k-1} \rangle_{L^2(\gamma)}. \tag{7.3}$$

Continuing, we get

$$\langle f, H_k \rangle_{L^2(\gamma)} = \langle f^{(k)}, H_0 \rangle_{L^2(\gamma)} = \mathbb{E} f^{(k)}(Z) \tag{7.4}$$

for $Z \sim N(0,1)$. In particular, taking $f = H_\ell$ for $\ell < k$ we see that $\langle H_\ell, H_k \rangle_{L^2(\gamma)} = 0$ since H_ℓ is of degree ℓ. Since H_k has leading coefficient one we see that $H_k^{(k)} \equiv k!$, so $\langle H_k, H_k \rangle_{L^2(\gamma)} = k!$. Hence, for any integers $k, \ell \geq 0$ we have

$$\langle H_\ell, H_k \rangle_{L^2(\gamma)} = k! 1_{k=\ell}. \tag{7.5}$$

Since the linear span of the Hermite polynomials contains all polynomials with real coefficients, one can show (using Weierstrass approximation and a truncation argument, say) that the sequence $(\frac{1}{\sqrt{k!}} H_k)_{k \geq 0}$ is a complete orthonormal basis for $L^2(\mathbb{R}, \gamma)$.

Note that one also obtains the Hermite polynomials by applying the Gram–Schmidt procedure to the monomials $1, x, x^2, \ldots$ under the inner product $\langle \cdot, \cdot \rangle_{L^2(\gamma)}$.

We get an orthonormal basis for $L^2(\mathbb{R}^d, \gamma^d)$ by tensorizing. For $k = (k_1, \dots, k_d) \in \mathbb{Z}^d_{\geq 0}$ let

$$H_k : \mathbb{R}^d \to \mathbb{R}\,, \quad H_k(x) := \prod_{i=1}^{d} H_{k_i}(x_i). \tag{7.6}$$

The H_k are eigenfunctions of L:

$$-LH_k = (k_1 + \cdots + k_d)H_k\,. \tag{7.7}$$

Again with (Stone–)Weierstrass one can show $(\frac{1}{\sqrt{k!}}H_k)_{k\in\mathbb{Z}^d_{\geq 0}}$ is a complete orthonormal basis for $L^2(\mathbb{R}^d, \gamma^d)$, where we write $k! := k_1! \cdots k_d!$.

For nice enough f (say Schwartz) we have the Plancherel identity:

$$\|f\|^2_{L^2(\gamma^d)} = \sum_{k\in\mathbb{Z}\geq 0^d} \frac{1}{k!}\langle f, H_k\rangle^2_{L^2(\gamma^d)} = \sum_{k\in\mathbb{Z}^d_{\geq 0}} \frac{1}{k!}(\mathbb{E}f^{(k)}(Z))^2 \tag{7.8}$$

where $Z \sim N(0, I_d)$ and we denote the mixed partial derivatives

$$f^{(k)}(x) := \frac{\partial^{k_1+\cdots+k_d} f}{\partial x_1^{k_1} \cdots \partial x_d^{k_d}}\,.$$

We can reorganize the sum as:

$$\|f\|^2_{L^2(\gamma^d)} = \sum_{m=0}^{\infty} \frac{1}{m!} \sum_{\substack{k\in\mathbb{Z}^d_{\geq 0} \\ k_1+\cdots+k_d=m}} \frac{m!}{k!}(\mathbb{E}f^{(k)}(Z))^2 =: \sum_{m=0}^{\infty} \frac{\theta_m(f)}{m!} \tag{7.9}$$

where

$$\begin{aligned}\theta_m(f) &:= \sum_{k_1,\dots,k_d=0}^{m} 1_{k_1+\cdots+k_d=m} \frac{m!}{k_1! \cdots k_d!}(\mathbb{E}f^{(k)}(Z))^2 \\ &= \sum_{i_1,\dots,i_m=1}^{d} \left(\mathbb{E}\frac{\partial^m f}{\partial x_{i_1} \cdots \partial x_{i_m}}(Z)\right)^2 \end{aligned} \tag{7.10}$$

is the squared norm of the projection of f to the eigenspace associated to eigenvalue m of $-L$. (We used the equality of mixed partial derivatives to reorganize the sum over the index set $[0, m] \times [1, d]$.) Hence,

$$\mathrm{Var}_{\gamma^d} f = \sum_{m=1}^{\infty} \frac{\theta_m(f)}{m!}\,. \tag{7.11}$$

We have the following truncated version of the above expansion:

Lemma 7.1. *For any integer $r \geq 1$ and $f \in C^r(\mathbb{R}^d)$ having derivatives of polynomial growth,*

$$\mathrm{Var}_{\gamma^d} f \leq \sum_{m=1}^{r-1} \frac{\theta_m(f)}{m!} + \frac{1}{r}\mathbb{E}|\nabla f(Z)|^2\,.$$

Note that the case $r = 1$ is the Poincaré inequality (Theorem 4.1). Hence, we can improve over Poincaré if we can control $\theta_m(f)$ for large m. Note that, crucially, the expectation is under the square in (7.10).

Note the statement and proof of the lemma easily carry over to general semigroups whose generator has a complete orthonormal basis of eigenvectors.

Proof. From (7.8), (7.10) and (7.11) it suffices to show

$$\sum_{\substack{k\in\mathbb{Z}^d_{\geq 0}\\ k_1+\cdots+k_d\geq r}} \frac{1}{k!}\langle f, H_k\rangle^2_{L^2(\gamma^d)} \leq \frac{1}{r}\mathbb{E}|\nabla f(Z)|^2\,. \tag{7.12}$$

This follows from a bound in the general semigroup setting: If (u_k) is a complete orthonormal basis of $L^2(\mu)$ that are eigenfunctions for the generator L associated to a Markov semigroup with stationary measure μ, and $-\lambda_k$ are the associated ordered eigenvalues (possibly having multiplicity, as in our setting), then for any $f \in L^2(\mu)$ and any $r > 0$,

$$\sum_{k:\lambda_k\geq r}\langle f, u_k\rangle^2_{L^2(\mu)} \leq \frac{1}{r}\sum_{k:\lambda_k\geq r}\lambda_k\langle f, u_k\rangle^2_{L^2(\mu)} \leq \frac{1}{r}\sum_{k\geq 0}\lambda_k\langle f, u_k\rangle^2_{L^2(\mu)} = \frac{1}{r}\mathcal{E}(f, f)$$

where the last equality is (5.2). The claim follows. □

7.2. **Application to the SK model.** Recall the setup from Section 2.6. Here we apply Lemma 7.1 to prove the following theorem of Chatterjee, showing the free energy of the SK model is superconcentrated at all temperatures.

Theorem 7.2 (Superconcentration of the free energy for the SK model)**.** *We have*

$$\operatorname{Var} F_n(\beta) \lesssim \frac{n}{\log n}(\log\log n + \log(1+\beta)).$$

Proof. We follow the physics notation by writing $\langle\cdot\rangle$ for expectation under the Gibbs measure, conditional on a realization of the disorder G. That is, for $f : \{-1, 1\}^n \to \mathbb{R}$,

$$\langle f\rangle = \frac{1}{Z}\sum_{\sigma\in\{-1,1\}^n} f(\sigma)e^{-H(\sigma)}.$$

We write $\mathbb{E}(\cdot) := \mathbb{E}_G\langle\cdot\rangle$ for total expectation over the Gibbs measure and the disorder.

To lighten notation we denote generic elements of the triangular array $A = \{(i, j) : 1 \leq i < j \leq n\}$ by $a = (i, j)$. Viewing quantities as functions of the Gaussian variables $g_a, a \in A$, we write ∂_a for partial differentiation with respect to the variable indexed by a. For $\sigma \in \{-1, 1\}^n$ and $a = (i, j) \in A$ we abbreviate the product

$$\sigma_a := \sigma_i\sigma_j\,.$$

In this notation we have

$$H(\sigma) = -\frac{\beta}{\sqrt{n}} \sum_{a \in A} \sigma_a g_a$$
$$Z = \sum_{\sigma} e^{-H(\sigma)}$$
$$F = -\frac{1}{\beta} \log Z.$$

All of these quantities are functions of $G = (g_a)_{a \in A}$. To apply Lemma 7.1 we need to bound the partial derivatives of F. For derivatives of first order,

$$\partial_a F = -\frac{1}{\beta} \frac{\partial_a Z}{Z} = -\frac{1}{\sqrt{n}} \frac{1}{Z} \sum_{\sigma} \sigma_a e^{-H(\sigma)} = -\frac{1}{\sqrt{n}} \langle \sigma_a \rangle.$$

Since σ is in the hypercube, $\sigma_a \in \{-1, 1\}$, so $|\langle \sigma_a \rangle| \le 1$ a.s. Thus, all partial derivatives of F of first order are a.s. bounded by $1/\sqrt{n}$ in magnitude, so

$$|\nabla F|^2 \le \frac{1}{n} \sum_{a \in A} |\langle \sigma_a \rangle|^2 \le n \quad a.s. \tag{7.13}$$

Note that at this point we can apply the Poincaré inequality to get

$$\operatorname{Var} F \le n. \tag{7.14}$$

Since $\mathbb{E}\sigma_a = \mathbb{E}\sigma_i \sigma_j = 0$ by symmetry, for the first coefficient $\theta_1(F)$ in the Hermite expansion we get

$$\theta_1(F) = \sum_{a \in A} (\mathbb{E}\partial_a F)^2 = 0. \tag{7.15}$$

This is a first indication that Lemma 7.1 might lead to and improvement on the Poincaré inequality.

Now we consider higher-order coefficients. We go through the second-order case in detail and then show how higher-order derivatives can be bounded similarly.

For $a, b \in A$, the corresponding partial derivative of the free energy yields the scared covariance of σ_a, σ_b under the Gibbs measure:

$$\partial_a \partial_b F = -\frac{1}{\beta} \partial_b \frac{\partial_a Z}{Z} = -\frac{1}{\beta} \left(\frac{\partial_a \partial_b Z}{Z} - \frac{\partial_a Z}{Z} \frac{\partial_b Z}{Z} \right) = -\frac{\beta}{n} \big(\langle \sigma_a \sigma_b \rangle - \langle \sigma_a \rangle \langle \sigma_b \rangle \big).$$

We use a trick to put everything under a single Gibbs expectation: Let σ^1, σ^2 be two independent samples from the Gibbs measure, conditional on the disorder G. (Thus we are technically considering a Gibbs measure on the enlarged space $\{-1, 1\}^n \times \{-1, 1\}^n$ with partition function Z^2.) Then we can reexpress

$$\partial_a \partial_b F = -\frac{\beta}{n} \big(\langle \sigma_a^1 \sigma_b^1 \rangle - \langle \sigma_a^1 \sigma_b^2 \rangle \big).$$

Thus,

$$\theta_2(F) = \sum_{a,b\in A} (\mathbb{E}\partial_a\partial_b F)^2 = \frac{\beta^2}{n^2}\sum_{a,b\in A}\left(\mathbb{E}\sigma_a^1\sigma_b^1 - \mathbb{E}\sigma_a^1\sigma_b^2\right)^2$$
$$= \frac{\beta^2}{n^2}\sum_{a,b\in A}(\mathbb{E}\sigma_a^1\sigma_b^1)^2 + (\mathbb{E}\sigma_a^1\sigma_b^2)^2 - 2\mathbb{E}\sigma_a^1\sigma_b^1\mathbb{E}\sigma_a^1\sigma_b^2\,. \qquad (7.16)$$

Now we use the 'replicas' trick again, this time on the disorder: letting $\widetilde{G}$ be an independent sample of the disorder, and $\tilde\sigma^1, \tilde\sigma^2$ independent samples from the corresponding Gibbs measure, we can reexpress

$$(\mathbb{E}\sigma_a^1\sigma_b^1)^2 + (\mathbb{E}\sigma_a^1\sigma_b^2)^2 - 2\mathbb{E}\sigma_a^1\sigma_b^1\mathbb{E}\sigma_a^1\sigma_b^2 = \mathbb{E}\sigma_a^1\sigma_b^1\tilde\sigma_a^1\tilde\sigma_b^1 + \mathbb{E}\sigma_a^1\sigma_b^2\tilde\sigma_a^1\tilde\sigma_b^2 - 2\mathbb{E}\sigma_a^1\sigma_b^1\tilde\sigma_a^1\tilde\sigma_b^2\,.$$

Next we take the sum over a, b under the expectations. Note that for any two spin vectors $\sigma, \tilde\sigma \in \{-1,1\}^n$,

$$\sum_{a\in A}\sigma_a\tilde\sigma_a = \sum_{1\le i<j\le n}\sigma_i\tilde\sigma_i\sigma_j\tilde\sigma_j = \frac{1}{2}\Big((\sigma\cdot\tilde\sigma)^2 - \sum_{i=1}^n \sigma_i^2\tilde\sigma_i^2\Big) = \frac{1}{2}\big((\sigma\cdot\tilde\sigma)^2 - n\big). \qquad (7.17)$$

Applying this in (7.16) we get

$$\theta_2(F) = \frac{1}{4}\frac{\beta^2}{n^2}\Big(\mathbb{E}((\sigma^1\cdot\tilde\sigma^1)^2 - n)^2 + \mathbb{E}((\sigma^1\cdot\tilde\sigma^1)^2 - n)((\sigma^2\cdot\tilde\sigma^2)^2 - n)$$
$$- 2\mathbb{E}((\sigma^1\cdot\tilde\sigma^1)^2 - n)((\sigma^1\cdot\tilde\sigma^2)^2 - n)\Big)$$
$$\le \frac{\beta^2}{n^2}\mathbb{E}((\sigma^1\cdot\tilde\sigma^1)^2 - n)^2$$

where in the last line we applied Cauchy–Schwarz to the last two expectations from the previous line and used that $\sigma^i\cdot\tilde\sigma^j$ are identically distributed for all $(i,j)\in\{1,2\}^2$. Now note that by symmetry, the vector

$$\big(\sigma_1^1\tilde\sigma_1^1, \dots, \sigma_n^1\tilde\sigma_n^1\big)$$

is uniform in $\{-1,1\}^n$, and hence

$$\sigma^1\cdot\tilde\sigma^1 = \sum_{j=1}^n \sigma_j^1\tilde\sigma_j^1$$

has the distribution of a simple random walk S of length n. One verifies the variance of S^2, that is, $\mathbb{E}(S^2-n)^2$, is of order n^2 (in fact it is $3n^2 + O(n)$) so altogether we have the n-independent bound

$$\theta_2(F) \lesssim \beta^2. \qquad (7.18)$$

The argument for higher-order Hermite coefficients $\theta_k(F)$ follows similar lines. For general mixed partial derivatives we have the following:

Claim 7.3. For any $a_1, \dots, a_k \in A$,

$$\partial_{a_1}\cdots\partial_{a_k}F = \frac{\beta^{k-1}}{n^{k/2}}\sum_{\ell_1,\dots,\ell_k=1}^k c_k(\ell_1,\dots,\ell_k)\langle\sigma_{a_1}^{\ell_1}\cdots\sigma_{a_k}^{\ell_k}\rangle$$

where $\sigma^1, \dots, \sigma^k$ are independent samples from the Gibbs measure conditional on G, and $c_k(\ell_1, \dots, \ell_k)$ are constants bounded in absolute value by $(k-1)!$.

Exercise 7.1. Prove the claim.

From (7.10) and Lemma 7.1 we see we need to bound the square of the (total) expectation of partial derivatives. We have

$$(\mathbb{E}\partial_{a_1} \cdots \partial_{a_k} F)^2 = \frac{\beta^{2k-2}}{n^k} \sum_{\boldsymbol{\ell}, \boldsymbol{\ell}' \in [k]^k} c_k(\boldsymbol{\ell}) c_k(\boldsymbol{\ell}') \mathbb{E}\sigma_{a_1}^{\ell_1} \cdots \sigma_{a_k}^{\ell_k} \mathbb{E}\sigma_{a_1}^{\ell_1'} \cdots \sigma_{a_k}^{\ell_k'}.$$

Similarly as in the case $k = 2$, we let $\widetilde{G}$ be an independent sample of the disorder and take independent samples $\tilde{\sigma}^1, \dots, \tilde{\sigma}^k$ from the corresponding Gibbs measure, with which we can reexpress

$$\mathbb{E}\sigma_{a_1}^{\ell_1} \cdots \sigma_{a_k}^{\ell_k} \mathbb{E}\sigma_{a_1}^{\ell_1'} \cdots \sigma_{a_k}^{\ell_k'} = \mathbb{E}\sigma_{a_1}^{\ell_1} \tilde{\sigma}_{a_1}^{\ell_1'} \cdots \sigma_{a_k}^{\ell_k} \tilde{\sigma}_{a_k}^{\ell_k'}.$$

Summing over $a_1, \dots, a_k$ and using the observation (7.17), we find

$$\begin{aligned} \theta_k(F) &= \sum_{a_1, \dots, a_k \in A} (\mathbb{E}\partial_{a_1} \cdots \partial_{a_k} F)^2 \\ &= 2^{-k} \frac{\beta^{2k-2}}{n^k} \sum_{\boldsymbol{\ell}, \boldsymbol{\ell}' \in [k]^k} c_k(\boldsymbol{\ell}) c_k(\boldsymbol{\ell}') \mathbb{E}[(\sigma^{\ell_1} \cdot \tilde{\sigma}^{\ell_1'})^2 - n] \cdots [(\sigma^{\ell_k} \cdot \tilde{\sigma}^{\ell_k'})^2 - n]. \end{aligned}$$

Applying Hölder's inequality to the expectation:

$$\mathbb{E}[(\sigma^{\ell_1} \cdot \tilde{\sigma}^{\ell_1'})^2 - n] \cdots [(\sigma^{\ell_k} \cdot \tilde{\sigma}^{\ell_k'})^2 - n] \le \mathbb{E}|(\sigma^{\ell_1} \cdot \tilde{\sigma}^{\ell_1'})^2 - n|^k$$

since the variables $\sigma^{\ell_i} \cdot \tilde{\sigma}^{\ell_i'}$ are identically distributed. Now just as for the case $k = 2$ we see that by symmetry,

$$\sigma^{\ell_1} \cdot \tilde{\sigma}^{\ell_1'} = \sum_{j=1}^{n} \sigma_j^{\ell_1} \tilde{\sigma}_j^{\ell_1'}$$

has the distribution of a simple random walk. Hence (by Hoeffding's or Bernstein's inequality) it is $O(\sqrt{n})$-sub-Gaussian, so its square is $O(n)$-sub-exponential, as is the centered variable $(\sigma^{\ell_1} \cdot \tilde{\sigma}^{\ell_1'})^2 - n$ (see [Ver18, Section 2.7]). From this it follows that the kth moment of $|(\sigma^{\ell_1} \cdot \tilde{\sigma}^{\ell_1'})^2 - n|$ is $O(kn)^k$ for every $k \in \mathbb{N}$ (again see [Ver18, Section 2.7]). Substituting this bound above, and using the bounds $c_k(\boldsymbol{\ell}) \le (k-1)! \le k^k$, we conclude that

$$\theta_k(F) = O(\beta k)^{O(k)} \tag{7.19}$$

for all k. Applying Lemma 7.1 along with the above bound and (7.13), we get that for any $r \in \mathbb{N}$,

$$\operatorname{Var} F \le \sum_{k=1}^{r-1} \frac{1}{k!} O(\beta k)^{O(k)} + \frac{n}{r} \le O(\beta r)^{O(r)} + \frac{n}{r}.$$

The claim follows by taking r of size $c \log n / (\log(1+\beta) + \log\log n)$ for a sufficiently small constant $c > 0$. □

8. Superconcentration for the top eigenvalue of random matrices

(While the titles of this and the previous section might suggest they are related, they have nothing to do with each other.)

Recall the definition of an $n \times n$ GUE random matrix G from Section 2.5. In this section we prove that the top eigenvalue $\lambda_1 = \lambda_1(G)$ is superconcentrated. We then deduce Theorem 2.5 establshing noise sensitivity for the top eigenvector, via the connection between superconcentration and disorder chaos that was explored in Section 5.

8.1. Superconcentration for λ_1.

It is well known that $\lambda_1/\sqrt{n}$ converges in probability to 2. Moreover, from the variational formula

$$\lambda_1 = \sup_{u \in S^{n-1}} u^* G u$$

one sees that λ_1 is a $O(1)$-Lipschitz (with respect to the Euclidean norm) function of the n^2 real Gaussian variables for the real and imaginary parts of entries of G. From the Gaussian Poincaré inequality (Theorem 4.1) we deduce

$$\operatorname{Var} \lambda_1 = O(1).$$

From the Gaussian concentration inequality (1.4) we further deduce that $\lambda_1 - 2\sqrt{n}$ has sub-Gaussian tails (with sub-Gaussian constant $O(1)$).

However, as is also well known, λ_1 in fact asymptotically has fluctuations of order $n^{-1/6}$. Indeed,

$$n^{1/6}(\lambda_1 - 2\sqrt{n})$$

converges in law to the TW_2 distribution. Moreover, the upper tail has non-sub-Gaussian shape $\exp(-\Theta(x^{3/2}))$. See [AGZ10].

From the convergence in law it does not follow that $\operatorname{Var} \lambda_1 = O(n^{-1/3})$ for finite n. We deduce this from the following results of Ledoux [Led03, Led07] giving non-asymptotic tail bounds.

Theorem 8.1. *We have*

$$\mathbb{P}(\lambda_1 \geq 2(1+\delta)\sqrt{n}) \lesssim \frac{1}{\sqrt{\delta}} \exp(-c\delta^{3/2} n) \qquad \forall\, \delta \in (0,1) \tag{8.1}$$

and

$$\mathbb{P}(\lambda_1 \leq 2(1-\delta)\sqrt{n}) \lesssim \exp(-c\delta^{3/2} n) \qquad \forall\, \delta \in (0,1). \tag{8.2}$$

By integrating these tail bounds we obtain:

Corollary 8.2. $\operatorname{Var} \lambda_1 = O(n^{-1/3})$.

Remark 8.3.

- The proof gives the rate $\frac{8\sqrt{2}}{3}(\delta^{3/2} + O(\delta^2))$ for the upper tail, though this constant is not asymptotically sharp.
- The upper tail (8.1) was extended to more general classes of "structured" random matrices with independent but not-necessarily identically distributed (in particular non-Gaussian) entries by Brailovskaya–van Handel in [BvH].

- While sufficient to obtain the sharp variance bound, the lower tail bound (8.2) is not sharp. The true order is $\exp(-\Theta(\delta^3 n^2)$. It is intuitive that lower tail deviations of λ_1 should be more rare than upper tail deviations, as a deviation to the left pushes several other eigenvalues along with it.

We only give the proof of the upper bound (8.1), which uses hypercontractivity (Theorem 6.1) applied to the Hermite polynomials introduced in the previous section. The proof of the lower bound involves ideas unrelated to this course.

Denote the normalized Hermite polynomials

$$h_k := \frac{1}{\sqrt{k!}} H_k$$

which are orthonormal in $L^2(\gamma)$.

Lemma 8.4. *Let $k \geq 0$, $t \geq 0$ and $q = 1 + e^{2t}$. Then*

$$\|h_k\|_{L^q(\gamma)} \leq e^{kt}.$$

Proof. $\|h_k\|_{L^q(\gamma)} = e^{kt}\|P_t h_k\|_{L^q(\gamma)} \leq e^{kt}\|h_k\|_{L^2(\gamma)} = e^{kt}$. □

We need the following lemma relating Hermite polynomials to the spectral distribution of the GUE. It says the one-point intensity function (with respect to Lebesgue measure) of the eigenvalue point process is $\sum_{k=0}^{n-1} h_k(\lambda)^2$ times the standard Gaussian density.

Lemma 8.5. *Let $\lambda_1 \geq \cdots \geq \lambda_n$ be the eigenvalues of a GUE matrix G. Then for any measureable function $f : \mathbb{R} \to \mathbb{R}$,*

$$\mathbb{E}\sum_{j=1}^{n} f(\lambda_j) = \int_{\mathbb{R}} f(\lambda) \sum_{k=0}^{n-1} h_k(\lambda)^2 d\gamma(\lambda).$$

This can be proved using Vandermonde determinants; see [Led07].

Proof of (8.1) in Theorem 8.1. For brevity write $y = 2(1+\delta)\sqrt{n}$. From Lemma 8.5 we have

$$\mathbb{P}(\lambda_1 \geq y) = \mathbb{E}1_{\lambda_1 \geq y} \leq \mathbb{E}\sum_{j=1}^{n} 1_{\lambda_j \geq y} = \int_y^\infty \sum_{k=0}^{n-1} h_k(\lambda)^2 d\gamma(\lambda). \tag{8.3}$$

Let $r > 1$ to be optimized later (we will take $r - 1 \asymp \sqrt{\delta}$) and let t be such that $2r = 1 + e^{2t}$. Applying Hölder's inequality and Lemma 8.4 to each summand, we get

$$\int_y^\infty h_k^2 d\gamma \leq \gamma([y,\infty))^{1-\frac{1}{r}} \|h_k^2\|_{L^r(\gamma)} = \gamma([y,\infty))^{1-\frac{1}{r}} \|h_k\|_{L^{2r}(\gamma)}^2 \leq 2e^{2kt} \exp(-\tfrac{1}{2}y^2(1-\tfrac{1}{r})).$$

The rest is plugging in and tidying up. Inserting this estimate in the previous line and summing over k yields

$$\mathbb{P}(\lambda_1 \geq y) \leq \frac{2}{e^{2t}-1} \exp\left(n(2t - 2(1+\delta)(1-\tfrac{1}{r}))\right) \tag{8.4}$$

With $r = 1 + \rho$ for small ρ we have

$$2t = \log(2r-1) = \log(1+2\rho) = 2\rho - 2\rho^2 + \frac{8}{3}\rho^3 + O(\rho^4)$$

and

$$2(1-\frac{1}{r}) = \frac{2\rho}{1+\rho} = 2\rho - 2\rho^2 + 2\rho^3 + O(\rho^4).$$

Substituting these estimates in the previous line gives

$$\mathbb{P}(\lambda_1 \geq y) \lesssim \frac{1}{\rho} \exp\Big(n\big(\frac{2}{3}\rho^3 - 4\delta\rho + O(\delta\rho^2 + \delta^2\rho + \rho^4)\big)\Big).$$

Finally, taking $\rho = c\sqrt{\delta}$ (with $c = \sqrt{2}$ optimizing the leading order terms) yields the claim. □

8.2. **Disorder chaos for u_1.** Sensitivity of the top eigenvector u_1 for Wigner matrices has been considered in two setups. In [Cha14, §3.6] Chatterjee considers the OU flow for a GUE matrix and shows that the top eigenvector u_1^t at time t goes from being nearly perfectly aligned with u_1^0 to nearly perfectly orthogonal from u_1^0 around time $t = \Theta(n^{-1/3})$.

In [BLZ20] a discrete analogue of this is shown, where k entries of a Wigner matrix are sampled one at a time uniformly without replacement and replaced with independent copies, with the transition happening at $k = \Theta(n^{5/3})$.

For both arguments, the sharp variance bound $O(n^{-1/3})$ for λ_1 is a key input.

Lemma 8.6. *For a finite set $\mathcal{X} \subset \mathbb{R}^n$ let $f : \mathbb{R}^n \to \mathbb{R}$ be given by $f(y) := \max_{x\in\mathcal{X}} x \cdot y$. Let $\hat{x}(y)$ denote a maximizer, which is unique for a.e. y. Then for a standard Gaussian vector $g \in \mathbb{R}^n$ we have*

$$\operatorname{Var} f(g) = \int_0^\infty e^{-t}\mathbb{E}[\hat{x}(g)\cdot\hat{x}(g^t)]dt$$

where as usual $(g^t)_{t\geq 0}$ is the OU process in $\mathbb{R}^n$ with $g^0 = g$. Moreover,

$$t \mapsto \mathbb{E}[\hat{x}(g)\cdot\hat{x}(g^t)] \tag{8.5}$$

is non-increasing, and we have

$$\operatorname{Var} f(g) \geq (1-e^{-t})\mathbb{E}[\hat{x}(g)\cdot\hat{x}(g^t)]. \tag{8.6}$$

Proof. From Lemma 3.1 and (4.4) we have

$$\operatorname{Var} f(g) = \int_0^\infty \mathcal{E}(f, P_t f)dt$$

and

$$\begin{aligned}\mathcal{E}(f, P_t f) &= \mathbb{E}\nabla f(g)\cdot\nabla P_t f(g)\\ &= e^{-t}\mathbb{E}\nabla f(g)\cdot P_t\nabla f(g)\\ &= e^{-t}\mathbb{E}\nabla f(g)\cdot\nabla f(g^t).\end{aligned}$$

Note that the function f is absolutely continuous, given by a maximum of a finite collection of linear functionals. Its gradient thus satisfies

$$\nabla f(y) = \hat{x}(y) \qquad a.e.\ y \in \mathbb{R}^n.$$

The variance formula follows.

For the second claim, we saw in the proof of Theorem 5.6 that $e^t\mathcal{E}(f, P_t f)$ is non-increasing in t (recall that $\lambda_1 = 1$ for the OU semigroup). The final variance bound (8.6) is a consequence of Lemma 5.7. □

By a limiting argument one can show the following:

Corollary 8.7. *As in the previous subsection, let λ_1, u_1 denote the top eigenvalue and associated eigenvector for an $n \times n$ GUE matrix G, and let u_1^t be the top eigenvector for the matrix G^t obtained by running the OU process on the space of Hermitian matrices, started from $G^0 = G$. (Here we view G^t as a Gaussian vector in a real vector space of dimension n^2, with $\binom{n+1}{2}$ coordinates for the real parts the entries on and above the diagonal and $\binom{n}{2}$ for the imaginary parts of entries above the diagonal.) Then*

$$\operatorname{Var} \lambda_1 = \int_0^\infty e^{-s}\mathbb{E}|u \cdot u^s|^2 ds \geq (1 - e^{-t})\mathbb{E}|u \cdot u^t|^2$$

for any $t \geq 0$.

Proof. $\lambda_1 = \max_{u \in S^{n-1}} u^* G u = \max_{u \in S^{n-1}} \langle uu^*, G \rangle_F$ (using the Frobenius inner product) has the form of a maximum of linear functionals of a Gaussian vector as in Lemma 8.6, with the maximum taken over the set $\mathcal{X}$ of rank-1 Hermitian matrices with nonzero eigenvalue equal to 1. While $\mathcal{X}$ is not finite, a discretization argument allows one to extend the lemma to this setting. The details are left to the reader. □

Combining Corollary 8.7 with Corollary 8.2, we get

$$\mathbb{E}|u \cdot u^t|^2 \lesssim \frac{1}{(1 - e^{-t})n^{1/3}} \tag{8.7}$$

and hence $|u \cdot u^t| = o(1)$ with high probability for $t = \omega(n^{-1/3})$.

9. Noise sensitivity and disorder chaos for last-passage percolation models

In this last lecture we give an overview of what's known on superconcentration, noise sensitivity and disorder chaos for last-passage percolation (LPP) models. We'll see heuristic arguments that nicely parallel what we've seen for the top eigenvalue/eigenvector of Wigner matrices, and indeed 2-dimensional LPP is expected to be in the same universality class. We end by stating results and describing proof ideas from the formidable recent work [GH24] on the Brownian LPP model (see also the companion work [GH23]). Much of the literature review and heuristics that follow are taken from that paper. (In particular we refer to [GH24, §§2–3] for an excellent discussion of the ideas that includes figures.)

....

10. More recent developments

The main reference for the course was published about 12 years ago, and it inspired a lot of subsequent work by many authors. I will try to summarize some highlights here. (I'm not an expert on these topics so I will be updating this section as I learn more during this minicourse. Please let me know of any important omissions!)

10.1. **Open problems.** Here is an incomplete list of open problems stated in [Cha14] (with numbering taken from the text), with notes on more recent progress.

Open Problem 1.2. *Improve the upper bound $O(n/\log n)$ in (2.3) for the variance of FPP. Preferably, find the correct order of the variance.*

This appears to still be wide open.

Open Problem 1.4. *Improve the upper bound $O(n/\log n)$ in Theorem 2.1 for the variance of the ground state energy for the Gaussian polymer. Preferably, find the correct order of the variance.*

Open Problem 1.5. *Analyze the fluctuations of the free energy $F_n(\beta)$ for the SK model at the critical temperature $\beta = 1$.*

- It has been conjectured that the variance is of order $O(\log n)$. Wei-Kuo Chen and Wai-Kit Lam proved the variance is $O(\log^2 n)$ in [CL19]. See [CL24] and references therein.

Open Problem 1.7. *Find the correct order of* $\operatorname{Var} F_n(\beta)$. *In particular, try to prove it is $O(1)$ for all β.*

- The open part is for the low-temperature range $\beta > 1$. The previous problem asks for $\beta = 1$ and the case $\beta < 1$ was already proved in the 80s. Again see [CL24] for a review of the literature.

Open Problem 1.8. *(For the setting of Section 2.3) give a general and easily verifiable condition on the covariance matrix R that is equivalent to superconcentration of the maximum.*

Open Problem 3.2. *Prove that the ground state energy of the SK model is superconcentrated.*

References

[ADH17] Antonio Auffinger, Michael Damron, and Jack Hanson. *50 years of first-passage percolation*, volume 68 of *University Lecture Series*. American Mathematical Society, Providence, RI, 2017. doi:10.1090/ulect/068. (Cited on page 4.)

[AGZ10] Greg W. Anderson, Alice Guionnet, and Ofer Zeitouni. *An introduction to random matrices*, volume 118 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2010. (Cited on page 30.)

[BGL14] Dominique Bakry, Ivan Gentil, and Michel Ledoux. *Analysis and geometry of Markov diffusion operators*, volume 348 of *Grundlehren der mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer, Cham, 2014. doi:10.1007/978-3-319-00227-9. (Cited on page 10.)

[BLM13] Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. *Concentration inequalities*. Oxford University Press, Oxford, 2013. A nonasymptotic theory of independence, With a foreword by Michel Ledoux. doi:10.1093/acprof:oso/9780199535255.001.0001. (Cited on page 2.)

[BLZ20] Charles Bordenave, Gábor Lugosi, and Nikita Zhivotovskiy. Noise sensitivity of the top eigenvector of a Wigner matrix. *Probab. Theory Related Fields*, 177(3-4):1103–1135, 2020. doi:10.1007/s00440-020-00970-1. (Cited on pages 1, 7, and 32.)

[BvH] Tatiana Brailovskaya and Ramon van Handel. Extremal random matrices with independent entries and matrix superconcentration inequalities. arXiv:2401.06284. (Cited on page 30.)

[Cha07] Sourav Chatterjee. Stein's method for concentration inequalities. *Probab. Theory Related Fields*, 138(1-2):305–321, 2007. URL: http://dx.doi.org/10.1007/s00440-006-0029-y, doi:10.1007/s00440-006-0029-y. (Cited on page 4.)

[Cha14] Sourav Chatterjee. *Superconcentration and related topics*. Springer Monographs in Mathematics. Springer, Cham, 2014. doi:10.1007/978-3-319-03886-5. (Cited on pages 1, 4, 7, 9, 10, 15, 18, 32, and 34.)

[CL19] Wei-Kuo Chen and Wai-Kit Lam. Order of fluctuations of the free energy in the SK model at critical temperature. *ALEA Lat. Am. J. Probab. Math. Stat.*, 16(1):809–816, 2019. doi:10.30757/alea.v16-29. (Cited on page 34.)

[CL24] Wei-Kuo Chen and Wai-Kit Lam. Universality of superconcentration in the Sherrington-Kirkpatrick model. *Random Structures Algorithms*, 64(2):267–286, 2024. doi:10.1002/rsa.21183. (Cited on page 34.)

[Com17] Francis Comets. *Directed polymers in random environments*, volume 2175 of *Lecture Notes in Mathematics*. Springer, Cham, 2017. Lecture notes from the 46th Probability Summer School held in Saint-Flour, 2016. doi:10.1007/978-3-319-50487-2. (Cited on page 6.)

[GH23] Shirshendu Ganguly and Alan Hammond. The geometry of near ground states in Gaussian polymer models. *Electron. J. Probab.*, 28:Paper No. 60, 80, 2023. doi:10.1214/22-ejp895. (Cited on page 33.)

[GH24] Shirshendu Ganguly and Alan Hammond. Stability and chaos in dynamical last passage percolation. *Commun. Am. Math. Soc.*, 4:387–479, 2024. doi:10.1090/cams/35. (Cited on pages 1 and 33.)

[Led01] Michel Ledoux. *The concentration of measure phenomenon*, volume 89 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2001. (Cited on page 2.)

[Led03] Michel Ledoux. A remark on hypercontractivity and tail inequalities for the largest eigenvalues of random matrices. In *Séminaire de Probabilités XXXVII*, volume 1832 of *Lecture Notes in Math.*, pages 360–369. Springer, Berlin, 2003. URL: https://doi.org/10.1007/978-3-540-40004-2_14, doi:10.1007/978-3-540-40004-2_14. (Cited on page 30.)

[Led07] M. Ledoux. Deviation inequalities on largest eigenvalues. In *Geometric aspects of functional analysis*, volume 1910 of *Lecture Notes in Math.*, pages 167–219. Springer, Berlin, 2007. URL: https://doi.org/10.1007/978-3-540-72053-9_10, doi:10.1007/978-3-540-72053-9_10. (Cited on pages 30 and 31.)

[Ver18] Roman Vershynin. *High-dimensional probability*, volume 47 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge, 2018. An introduction with applications in data science, With a foreword by Sara van de Geer. doi:10.1017/9781108231596. (Cited on page 29.)

Department of Mathematics, Duke University, 120 Science Dr, Durham, NC 27710
Email address: nicholas.cook@duke.edu