Topics in concentration of measure: Lecture II

Sourav Chatterjee

Courant Institute, NYU

St. Petersburg Summer School, June 2012
Lecture II: Superconcentration and related topics
An optimization problem is called ‘stable’ if any near-optimum is close to the global optimum in some appropriate metric.

For example, the function $f(x) = x^2$ is minimized at 0, and the minimum is stable, since any near-minimal point must be close to zero.

On the other hand, the function $f(x) = x^2 e^{-x^2}$ is also minimized at 0, but there are points arbitrarily far away from 0 that are as close to being minimum as we want.
Asymptotic Essential Uniqueness

- Mathematics abounds with stability theorems for optimization problems.
- In probability theory, David Aldous introduced a notion of stability that he called ‘Asymptotic Essential Uniqueness’.
- A famous result of Aldous is that the random assignment problem has the AEU property.
- In the random assignment problem, we have an $n \times n$ table of i.i.d. non-negative random variables ($c_{ij}$) with mean 1, and the objective is to minimize the ‘cost’ $\sum c_{i\pi(i)}$ over all permutations $\pi$ of $\{1, \ldots, n\}$.
- $c_{ij}$ is interpreted as the cost of allocating task $i$ to agent $j$.
- Aldous (2001) proved (among many other things) that under mild conditions on the matrix entries, the minimizing permutation has the AEU property, in the sense that when $n$ is large, any permutation that nearly minimizes the cost must be close to the global minimizer.
A simple example

Aldous’s theorem is a highly non-trivial result. A much simpler example is provided by the following: Let $g_1, \ldots, g_n$ be i.i.d. random variables with a continuous density. Maximize $\sum g_i \sigma_i$ over all $\sigma_1, \ldots, \sigma_n \in \{-1, 1\}$.

Clearly, the maximum is attained when $\sigma_i = \text{sign}(g_i)$, $i = 1, \ldots, n$.

Moreover, it is quite easy to prove that the optimizer has the AEU property, since

$$\sum g_i \sigma_i = \sum |g_i| - \sum_{i: \sigma_i \neq \text{sign}(g_i)} 2|g_i|.$$
A question

In the previous problem, what happens if we replace the linear form by a quadratic form, i.e. try to maximize

\[ \sum g_{ij} \sigma_i \sigma_j, \]

where \( g_{ij} \) are i.i.d. random variables, does the AEU property still hold?

Guess: No! Why?

- If \( (\sigma_1, \ldots, \sigma_n) \) was allowed to take values on a sphere instead of the hypercube, the maximization problem is exactly the problem of finding the largest eigenvalue. (Assume that the matrix is symmetric, for simplicity.)
- We know that for symmetric matrices with i.i.d. entries, the large eigenvalues cluster at the top. (Not an easy theorem!)
- The corresponding eigenvectors are therefore all near-optimal solutions, yet mutually orthogonal to each other.

However, none of the random matrix tools are available for the hypercube, so it is not clear how to prove ‘no AEU’ here.
The quadratic form $\sum g_{ij} \sigma_i \sigma_j$ occurs as a multiple of the energy of a spin configuration $\sigma = (\sigma_1, \ldots, \sigma_n)$ in the famous Sherrington-Kirkpatrick (SK) model of spin glasses. (Long and distinguished history; see Talagrand’s books.)

To be completely precise, let $g_{ij}$ be a collection of i.i.d. $N(0, 1)$ random variables, with $g_{ij} = g_{ji}$. For any $\sigma \in \{-1, 1\}^n$, the SK model defines the energy of $\sigma$ as

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

Physicists have long claimed that the energy landscape of the SK model has ‘multiple valleys’. Although the precise meaning is never specified, a very simplistic interpretation may be that there are many spin configurations with near-minimal energy that are all nearly orthogonal to each other. In other words, the total opposite of AEU.
Multiple valleys

▶ Suppose that we have a sequence of sets $X_n$.

▶ Let $s_n$ be a ‘similarity measure’ on $X_n$, that is, $s_n$ is a function from $X_n \times X_n$ into $[0, \infty)$ such that $s_n(x, y)$ denotes the ‘degree of similarity between $x$ and $y$’.

▶ Let $f_n : X_n \rightarrow \mathbb{R}$ be a random function.

▶ **Definition:** We will say that the sequence $(f_n, X_n, s_n)$ exhibits ‘multiple valleys’ if there exists $\epsilon_n$, $\delta_n$ and $\gamma_n$ tending to zero and $K_n$ tending to $\infty$, such that for each $n$, with probability $\geq 1 - \gamma_n$, there exists a set $A \subseteq X_n$ of size $\geq K_n$, such that $s_n(x, y) \leq \epsilon_n$ for all $x, y \in A$, $x \neq y$, and for all $x \in A$,

$$\left| \frac{f_n(x)}{\min_{z \in X_n} f_n(z)} - 1 \right| \leq \delta_n.$$
Multiple valleys in the SK model

- Let $X_n = \{-1, 1\}^n$.
- Let $s_n$ be the similarity measure on $X_n$ defined as
  \[ s_n(\sigma^1, \sigma^2) := \left( \frac{1}{n} \sum_{i=1}^{n} \sigma^1_i \sigma^2_i \right)^2. \]
- Let $(g_{ij})_{1 \leq i < j \leq n}$ be i.i.d. $N(0, 1)$ random variables.
- Let $H_n : X_n \to \mathbb{R}$ be the random function
  \[ H_n(\sigma) := -\frac{1}{\sqrt{n}} \sum_{1 \leq i < j \leq n} g_{ij} \sigma_i \sigma_j. \]

**Theorem (C., 2009)**

The sequence $(H_n, X_n, s_n)$ exhibits multiple valleys.

(In other words, with probability $\approx 1$ there is a large number of points in $X_n$ that are nearly orthogonal to each other, where $H_n$ is near-minimal.)
The random matrix example indicates that there cannot be an ‘easy’ way to prove such a theorem. (At least not that I know of!)

I will outline a general plan of attack for proving such multiple valley theorems.

The first step is to prove superconcentration of a relevant random quantity. Superconcentration means, roughly, that the quantity has smaller fluctuations than predicted by classical theory.

The next step is to show that superconcentration is equivalent to chaos, that is, high sensitivity to small perturbations.

Lastly, prove that chaos implies multiple valleys.

The route superconcentration $\rightarrow$ chaos $\rightarrow$ multiple valleys can be used to establish multiple valleys in a variety of problems, but we will stick to this one example to keep things in focus.
This talk is based primarily on ideas from two papers I wrote in 2008 and 2009:

1. **Chaos, concentration, and multiple valleys.**
   arXiv:0810.4221v2 [The properties of superconcentration, chaos and multiple valleys are defined here, and shown to be interrelated, along with a number of examples.]

2. **Disorder chaos and multiple valleys in spin glasses.**
   arXiv:0907.3381v1 [The Sherrington-Kirkpatrick model is shown to have all three properties.]

However, the proofs I will show today will be somewhat different and (hopefully) more enlightening than those in the above manuscripts.
Let \((X_t)_{t \geq 0}\) be a Markov process on a state space \(\mathcal{X}\).

Assume that there exists an invariant probability measure \(\mu\) for this process, such that irrespective of the starting state, the limiting distribution of \(X_t\) as \(t \to \infty\) is \(\mu\). In other words, \(\mu\) is the equilibrium probability measure.

The Markov process \(X_t\) defines a semigroup of operators \((P_t)_{t \geq 0}\) acting on integrable functions:

\[
P_t f(x) = \mathbb{E}(f(X_t) \mid X_0 = x).
\]

This is a semigroup since \(P_{t+s} = P_t P_s\).

Note that \(\lim_{t \to \infty} P_t f(x) = \mathbb{E}_\mu(f)\), where \(\mathbb{E}_\mu\) denotes integration with respect to \(\mu\) on \(\mathcal{X}\).
The generator $L$ of a Markov semigroup $P_t$ is the operator

$$Lf := \lim_{t \to 0} \frac{P_t f - f}{t}.$$
Given a Markov semigroup $P_t$ with generator $L$, it is easy to prove the heat equation:

$$\partial_t P_t = \lim_{s \to 0} \frac{P_{t+s} - P_t}{s} = \lim_{s \to 0} \frac{(P_s - I)P_t}{s} = LP_t.$$ 

This implies the formula $P_t = e^{tL}$. 
The equilibrium measure $\mu$ defines the natural inner product on $L^2(\mu)$ as $(f, g) := \mathbb{E}_\mu(fg)$.

The Dirichlet form $\mathcal{E}$ of a Markov semigroup with generator $L$ and equilibrium measure $\mu$ is a bilinear form:

$$\mathcal{E}(f, g) := -\mathbb{E}_\mu(fLg) = -(f, Lg).$$

The Markov process is said to be reversible if $L$ is self-adjoint with respect to the inner product. This happens if and only if $\mathcal{E}(f, g) = \mathcal{E}(g, f)$ for all $f, g$. 
Poincaré inequality

- Given two functions $f$ and $g$, their covariance under $\mu$ is defined as

$$\text{Cov}_\mu(f, g) := \mathbb{E}_\mu(fg) - \mathbb{E}_\mu(f)\mathbb{E}_\mu(g).$$

- Similarly, the variance of $f$ is defined as

$$\text{Var}_\mu(f) = \text{Cov}_\mu(f, f).$$

- The Markov process $X_t$ is said to satisfy a Poincaré inequality with constant $C$ if for all $f \in L^2(\mu)$,

$$\text{Var}_\mu(f) \leq C \mathcal{E}(f, f).$$

- **Question:** When does a Markov process satisfy a Poincaré inequality?
Generator is negative semidefinite

- Let $L$ be the generator of a reversible Markov semigroup.

- Jensen's inequality implies that $P_t$ is an $L^2$ contraction, that is,

\[ \| P_t f \|_{L^2(\mu)}^2 \leq \| f \|_{L^2(\mu)}^2. \]

- Therefore by Cauchy-Schwarz,

\[ (f, P_t f) \leq \| f \|_{L^2(\mu)} \| P_t f \|_{L^2(\mu)} \leq \| f \|_{L^2(\mu)}^2 = (f, f). \]

- Since $Lf = \lim_{t \to 0} (P_t f - f)/t$, this shows that $(f, Lf) \leq 0$.

- Thus, $L$ is negative semidefinite.

- Note that $Lf \equiv 0$ for any constant function $f$. So 0 is always an eigenvalue.

- Since $P_t = e^{tL}$ and $P_t f \to \mathbb{E}_\mu(f)$, constant functions are the only functions such that $Lf \equiv 0$. 

Under mild conditions, the eigenvalues of $-L$ can be ordered as a countable sequence $0 = \lambda_0 \leq \lambda_1 \leq \lambda_2 \leq \cdots$.

If $\lambda_1 > 0$, we say that the Markov process has spectral gap $\lambda_1$.

Let $u_0, u_1, \ldots$ be a corresponding sequence of mutually orthogonal eigenvectors. Here $u_0$ is the constant function $1$.

Since $Lf \equiv 0 \implies f$ constant, therefore $(u_k)_{k \geq 0}$ is an orthogonal basis of $L^2(\mu)$.

Therefore, any $f$ may be written as

$$f = \sum_{k=0}^{\infty} (u_k, f) u_k.$$

Consequently,

$$\mathcal{E}(f, g) = -(f, Lg) = \sum_{k=0}^{\infty} \lambda_k (u_k, f)(u_k, g).$$
Fact: A Markov process satisfies a Poincaré inequality if and only if it has a spectral gap. Moreover, the optimal constant in the Poincaré inequality is $1/\lambda_1$.

To see this, first note that since $(u_k)_{k \geq 0}$ is an orthogonal basis of $L^2(\mu)$, we have the Plancherel identity:

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

Observe that $(u_0, f) = \mathbb{E}_\mu(f)$. Therefore, the Plancherel identity may be rewritten as $\text{Var}_\mu(f) = \sum_{k=1}^{\infty} (u_k, f)^2$.

Consequently, if $\lambda_1 > 0$,

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

and equality is achieved for the function $f = u_1$.

If $\lambda_1 = 0$, then the function $u_1$ violates Poincaré inequality.
The standard Ornstein-Uhlenbeck process $X_t$ is a Markov process on $\mathbb{R}$ that satisfies the stochastic differential equation

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

where $B_t$ is standard Brownian motion.

Even if you are unfamiliar with stochastic calculus, no worries! The OU process has an alternate description as a time-changed scaling of Brownian motion:

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

where again, $(B_s)_{s \geq 0}$ is standard Brownian motion.
The alternative representation shows that, given $X_0 = x$, $X_t$ has the same distribution as

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

where $Z \sim N(0, 1)$.

Therefore, the Markov semigroup is given by

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

Clearly, the standard Gaussian measure $\gamma$ is the equilibrium measure for $P_t$. 
Multidimensional OU process

- An \( n \)-dimensional standard OU process \( X_t \) is a Markov process in \( \mathbb{R}^n \) whose coordinates are independent one-dimensional OU processes.

- The semigroup is simply

\[
P_t f(x) = \mathbb{E}(e^{-t}x + \sqrt{1 - e^{-2t}}Z),
\]

where \( Z \) is an \( n \)-dimensional standard Gaussian random vector. We will denote the law of \( Z \) by \( \gamma^n \).

- The generator of this process is

\[
L f(x) = \Delta f(x) - x \cdot \nabla f(x),
\]

where \( \Delta \) is the Laplacian operator in \( \mathbb{R}^n \), \( \nabla f \) is the gradient of \( f \), and \( \cdot \) denotes the usual inner product.
Spectral decomposition of the $n$-dimensional OU process

- The eigenfunctions are indexed by elements of $\mathbb{Z}_+^n$, where $\mathbb{Z}_+$ is the set of nonnegative integers.
- For $k = (k_1, \ldots, k_n) \in \mathbb{Z}_+^n$, the eigenfunction $H_k$ is simply

$$H_k(x) = \prod_{i=1}^{n} H_{k_i}(x_i),$$

where $H_{k_i}$ is the $k_i$th univariate Hermite polynomial. The corresponding eigenvalue is $k_1 + \cdots + k_n$.
- Consequently, this process has spectral gap 1.
- The Dirichlet form is $\mathcal{E}(f, g) = \mathbb{E}_{\gamma^n}(\nabla f \cdot \nabla g)$.
- Therefore, we have the Gaussian Poincaré inequality:

$$\text{Var}_{\gamma^n}(f) \leq \mathbb{E}_{\gamma^n} |\nabla f|^2,$$

where $|x|$ denotes Euclidean norm of a vector $x$. 
Recall: The SK model defines a Hamiltonian $H_n$ on $\{-1, 1\}^n$ as

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

where $g_{ij}$ are i.i.d. $N(0, 1)$ random variables.

The **Gibbs measure at inverse temperature $\beta$** induced by this Hamiltonian is the probability measure on $\{-1, 1\}^n$ that assigns mass

$$Z_n(\beta)^{-1} e^{-\beta H_n(\sigma)}$$

to a configuration $\sigma$. Here $Z_n(\beta)$ is the (random) normalizing constant.
The free energy is the SK model at inverse temperature $\beta$ is defined as

$$F_n(\beta) := \frac{1}{\beta} \log Z_n(\beta).$$

Let $\sigma^1$ and $\sigma^2$ be two configurations drawn independently from the Gibbs measure at inverse temperature $\beta$. The overlap between $\sigma^1$ and $\sigma^2$ is defined as

$$R_{1,2} := \frac{1}{n} \sum_{i=1}^{n} \sigma^1_i \sigma^2_i.$$
The fluctuations of $F_n(\beta)$ will be particularly important for us. It is conjectured that $\text{Var}(F_n(\beta)) \leq C(\beta)$ for some constant depending only on $\beta$.

Aizenman, Lebowitz and Ruelle (1987) proved this when $\beta < 1$.

A direct application of the Gaussian Poincaré inequality shows that $\text{Var}(F_n(\beta)) \leq C(\beta)n$.

Indeed, this is the best that one can get by the Poincaré inequality when $\beta > 1$, and this was the best known upper bound until 2009.

**Theorem (C., 2009)**

*For any $\beta$, $\text{Var}(F_n(\beta)) \leq C(\beta)n/\log n$.*

In the next few slides, we will see how to prove this.
Superconcentration

- Suppose that we have a Markov process with equilibrium measure $\mu$ and Dirichlet form $\mathcal{E}$ that satisfies a Poincaré inequality with optimal constant $C$.

- We will say that a function $f$ is $\epsilon$-superconcentrated if

$$\text{Var}_\mu(f) \leq C \epsilon \mathcal{E}(f, f).$$

- $\epsilon$-superconcentration with small $\epsilon$ means that the Poincaré inequality is suboptimal for upper bounding the variance of $f$.

- Let $0 = \lambda_0 < \lambda_1 \leq \lambda_2 \leq \cdots$ be the eigenvalues of the generator of our process, with eigenfunctions $u_0, u_1, \ldots$.

- Since $\text{Var}_\mu(f) = \sum_{k=1}^{\infty} (u_k, f)^2$ and $\mathcal{E}(f, f) = \sum_{k=1}^{\infty} \lambda_k (u_k, f)^2$, superconcentration occurs if and only if most of the “Fourier mass” concentrates on the higher end of the spectrum. Similar to noise-sensitivity.
An improved Poincaré inequality

Proposition

For any $m \geq 1$ and any $f$,

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

Proof. Simply note that $\mathcal{E}(f, f) = \sum_{k=1}^{\infty} \lambda_k (u_k, f)^2$ and

$$\text{Var}_\mu(f) = \sum_{k=1}^{m-1} (u_k, f)^2 + \sum_{k=m}^{\infty} (u_k, f)^2$$

$$\leq \sum_{k=1}^{m-1} (u_k, f)^2 + \frac{1}{\lambda_m} \sum_{k=m}^{\infty} \lambda_k (u_k, f)^2.$$

How to use: Get crude bounds on $(u_k, f)^2$. Then optimize over $m$. (Note: $m = 1$ gives the usual Poincaré inequality.)
One can use the Hermite polynomial basis of the \( \binom{n}{2} \) dimensional standard Gaussian measure, together with the improved Poincaré inequality, to prove that

\[
\text{Var}(F_n(\beta)) \leq \frac{C(\beta)n \log \log n}{\log n},
\]

where \( F_n(\beta) \) is the free energy of the SK model at inverse temperature \( \beta \). (Will not go into the computational details.)

This improves the bound \( C(\beta)n \) given by the usual Poincaré inequality, thereby proving superconcentration of the free energy.

To improve it to \( C(\beta)n/\log n \), a different line of attack is needed. Will not discuss here.

**Remark:** Often, hypercontractive methods are used to prove superconcentration (Talagrand, Benjamini, Kalai, Schramm, ...). Hypercontractive methods do not seem to work in spin glasses.
As usual, Markov process $X_t$ with Dirichlet form $\mathcal{E}$, semigroup $P_t$, eigenvalues $(\lambda_k)_{k \geq 0}$, eigenfunctions $(u_k)_{k \geq 0}$, spectral gap $\lambda_1$, equilibrium measure $\mu$, etc. It is easy to show that for all $t$ and $f$,

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

We will say that a function $f \in L^2(\mu)$ is $(\epsilon, \delta)$-chaotic if for all $t \geq \delta$,

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

In other words, if $f$ is $(\epsilon, \delta)$-chaotic for small $\epsilon$ and $\delta$, then $\mathcal{E}(f, P_t f)$ decays to zero much faster than ‘usual’.

We will see in the next few slides why this may be called chaos.
Why chaos?

Recall the free energy in the SK model:

\[ F_n(\beta) = \frac{1}{\beta} \log \sum_{\sigma \in \{-1,1\}^n} \exp \left( \frac{\beta}{\sqrt{n}} \sum_{1 \leq i < j \leq n} g_{ij} \sigma_i \sigma_j \right) . \]

The relevant semigroup is the \( (\frac{n}{2}) \)-dimensional OU semigroup \( P_t \), with spectral gap 1.

Let \( (g_{ij}^t)_{1 \leq i < j \leq n, t \geq 0} \) be an \( (\frac{n}{2}) \)-dimensional OU process, with \( g_{ij}^0 = g_{ij} \).

Just like \( g_{ij} \), the collection \( (g_{ij}^t)_{1 \leq i < j \leq n} \) defines a Gibbs measure at inverse temperature \( \beta \). We will call this the Gibbs measure at time \( t \).
Fix $t$ and $\beta$. Let $\sigma^1$ be a configuration drawn from the Gibbs measure at time 0 and $\sigma^2$ be a configuration drawn from the Gibbs measure at time $t$.

Let $R_{1,2}(t) = \frac{1}{n} \sum_{i=1}^{n} \sigma^1_i \sigma^2_i$ be the overlap between $\sigma^1$ and $\sigma^2$.

An easy computation shows that

$$e^t \mathcal{E}(F_n(\beta), P_t F_n(\beta)) = n \mathbb{E}(R_{1,2}^2(t)).$$

Thus, if $F_n(\beta)$ is $(\epsilon, \delta)$ chaotic, then for all $t \geq \delta$,

$$\mathbb{E}(R_{1,2}^2(t)) \leq \epsilon \mathbb{E}(R_{1,2}^2(0)) \leq \epsilon.$$
Why chaos? — Another example

- Let \( (g_e)_{e \in E(\mathbb{Z}^2)} \) be i.i.d. \( N(0, 1) \) random variables, where \( E(\mathbb{Z}^2) \) is the edge set of \( \mathbb{Z}^2 \). Call these ‘edge weights’.
- Take any \( n \), and consider the set of all ‘up-right’ paths from \((0, 0)\) to \((n, n)\).
- The ‘weight’ of a path is the sum of edge-weights along the path.
- Let \( L_n \) be the weight of the heaviest up-right path from \((0, 0)\) to \((n, n)\). (Oriented last-passage percolation.)
- Let \( (g_e^t)_{e \in E(\mathbb{Z}^2), t \geq 0} \) be an infinite-dimensional OU process, with \( g_e^0 = g_e \).
- Let \( p(t) \) be the optimal path ‘at time \( t \)’.
- An easy computation gives \( e^t \mathcal{E}(L_n, P_t L_n) = \mathbb{E}|p(0) \cap p(t)| \).
- Thus if \( L_n \) is \((\epsilon, \delta)\)-chaotic, then for all \( t \geq \delta \),
  \[
  \mathbb{E}|p(0) \cap p(t)| \leq \epsilon \mathbb{E}|p(0) \cap p(0)| = 2\epsilon n.
  \]
Theorem (Essentially in C., 2008)

If \( f \) is \((\epsilon, \delta)\)-chaotic, then it is \( \epsilon'\)-superconcentrated, where \( \epsilon' = \epsilon + \lambda_1 \delta \). Conversely, if \( f \) is \( \epsilon \)-superconcentrated, then for any \( \delta \), \( f \) is \((\epsilon', \delta)\)-chaotic, where \( \epsilon' = \epsilon / \lambda_1 \delta \).

Remark: If \( f \) is \( \epsilon \)-superconcentrated for some small \( \epsilon \), then choosing \( \delta = \sqrt{\epsilon} \), we get \( \epsilon' = \sqrt{\epsilon} / \lambda_1 \), we see that \( f \) is \((\epsilon', \delta)\)-chaotic where both \( \epsilon' \) and \( \delta \) are small.
Proof sketch

- Since $P_t = e^{tL}$, spectral decomposition of $L$ gives

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

- Then by the Plancherel identity and the spectral formula shown above,

$$\text{Var}_\mu(f) = \sum_{k=1}^{\infty} (u_k, f)^2 = \int_0^{\infty} \mathcal{E}(f, P_t f) \, dt.$$  

- Thus, if $\mathcal{E}(f, P_t f)$ decays to zero ‘unusually rapidly’, then $\text{Var}_\mu(f)$ is ‘unusually small’.

- On the other hand, from the spectral formula we see that $\mathcal{E}(f, P_t f)$ is a decreasing, non-negative function of $t$.

- Consequently, if $\text{Var}_\mu(f)$ is ‘unusually small’, then $\mathcal{E}(f, P_t f)$ must decay to zero ‘unusually fast’.

- Easy to fill in the details.
Chaos in the SK model

As a consequence of the superconcentration of the free energy, and the equivalence of superconcentration and chaos, the following theorem is proved.

**Theorem (C., 2009)**

Let $R_{1,2}(t)$ be the overlap between a configuration drawn at time 0 and another drawn at time $t$. Then for all $t \geq \sqrt{1/\log n}$,

$$\mathbb{E}(R_{1,2}^2(t)) \leq C(\beta) \sqrt{\frac{1}{\log n}}.$$

- This is known as ‘chaos in disorder’ in the SK model. Was an open problem for a long time. Further recent progress by Chen and Panchenko.
- Actually, the theorem in (C., 2009) gives a better bound.
Chaos implies multiple valleys

- Gibbs measure at inverse temperature $\beta$ concentrates on near-minimal energy states if $\beta$ is large.
- Let $H^t(\sigma)$ denote the energy of $\sigma$ at time $t$. Let $H = H^0$.
- Fix a large $\beta$. If $\sigma^1$ and $\sigma^2$ are picked from the Gibbs measures at time 0 and time $t$, then $H(\sigma^1)$ and $H^t(\sigma^2)$ are close to $\min H(\sigma)$ and $\min H^t(\sigma)$.
- But if $t$ is small, then $H^t(\sigma) \approx H(\sigma)$ for all $\sigma$. In particular, $\min H^t(\sigma) \approx \min H(\sigma)$. So, if $\beta$ is large and $t$ is small, then $\sigma^2$ is a near-minimal energy state for $H$ also!
- Thus, if $\beta$ is large and $t$ is small, and are calibrated depending on $n$ such that $\mathbb{E}(R_{1,2}^2(t))$ is small, then we have found two nearly orthogonal states that are both of near-minimal energy at time 0. Repeating, we get many such states.
- This proves the multiple valley theorem for the SK model. The same idea goes through in many other examples.
Two research problems

- Rigorously compute the optimal order of variance of superconcentrated quantities. In almost all problems (except for a few exactly solvable ones), we cannot even get close to the optimal order.

- Define a notion of multiple valleys that is equivalent to chaos and superconcentration instead of being weaker.

*End of Lecture II.*