1 Overview

In the context of an optimization problem over a domain of random noise, the following three phenomena are roughly equivalent:

(1) **Superconcentration:** the optimal value concentrates more tighly than one can prove using classical concentration inequalities.

(2) **Chaos:** the location of the optimizer is highly sensitive to small changes in the noise field.

(3) **Multiple Valleys:** there are many near-optimal points.

Intuitively, all three of these phenomena indicate nonlinearity of the optimization problem in terms of the underlying noise. This is currently a quite vague statement, and the theorems that describe these equivalences in general are a bit too technical to state right at the beginning. So, let's illustrate this with a few examples.

2 First Example: Gaussian Polymers

Suppose that we have a "medium" or "environment" consisting of some amount of stuff at each vertex of $\mathbb{Z}^2$. More precisely, we have $(g_v)_{v\in\mathbb{Z}^2}$, a collection of i.i.d. standard Gaussian random variables. A 'polymer' through the medium is simply the graph of a path that could be taken by a simple random walk on $\mathbb{Z}$. In other words, it is a sequence $\{(0,0),(1,a_1),\ldots,(n,a_n)\}$ where $|a_{i+1}-a_i|=1$ for all i. The 'energy' of a polymer p of length n is

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

Using this as the (random) Hamiltonian in a (random) Gibbs measure, we get a distribution which favors polymers that pick up more stuff. Notice that the g_v for odd v are completely irrelevant.

For now, we will only consider the zero-temperature regime. This means that we want to look at the minimum energy path $\hat{p}_n$, which has energy E_n (this is the negative of the maximum possible sum of Gaussians along a polymer). These are both random variables which depend on the underlying noise (g_v). This is exactly the same as the "last-passage percolation" model from a point to a line, by the way.

2.1 Superconcentration

It can be proved that E_n has order n (using, for example, Kingman's subadditive ergodic theorem). It can also be proved using classical techniques (i.e. the Poincaré inequality) that $\mathrm{Var}(E_n) \le Cn$. This variance bound, however, is intuitively suboptimal; if we simply had a sum of n independent Gaussians, then the variance would be of order exactly n. However, here we are optimizing over all paths, so if a particular Gaussian is too low, the optimal polymer will just avoid it and therefore the variance should be less than that of a sum of Gaussians with the same size.

Indeed, if we replaced the Gaussians with exponential random variables, then results from integrable probability tell us that the variance of E_n would be of order $n^{2/3}$, and this is what is expected to be true in the Gaussian case as well. However, this is unproven. In fact, the best known result (I think) is

Theorem 1. *If E_n is the ground state energy in the Gaussian random polymer model, then*

$$\mathrm{Var}(E_n) \le \frac{Cn}{\log n},$$

where C does not depend on n.

This is what we call Superconcentration—it is an improvement on the classical variance bound of order n. This improvement might not seem very good since it is still far away from the conjectured order of $n^{2/3}$. However, this seemingly mundane result turns out to be equivalent to some other interesting properties, namely chaos and multiple valleys.

2.2 Chaos

First let's discuss a bit how we will typically perturb the noise field. For a standard Gaussian random variable g, we will write g^t to mean the evolution of g under the Ornstein-Uhlenbeck process for time t. This means that for any given fixed t,

$$g^t \stackrel{d}{=} e^{-t}g + \sqrt{1-e^{-2t}}g',$$

where g' is an independent standard Gaussian. Notice that the correlation between g and g^t is e^{-t}, so when t is small the two are close.

Now let's start with an environment $(g_v)_{v\in\mathbb{Z}^2}$ and perturb it to obtain $(g_v^t)_{v\in\mathbb{Z}^2}$ (each vertex should be perturbed independently). Let $\hat{p}_n$ denote the ground state polymer in the original environment, and let $\hat{p}_n^t$ denote the ground state polymer in the t-perturbed environment. Let $|\hat{p}_n \cap \hat{p}_n^t|$ denote the number of vertices in the intersection of the two polymers. Then

Theorem 2. *We have*

$$\mathbb{E}[|\hat{p}_n \cap \hat{p}_n^t|] \leq \frac{Cn}{(1-e^{-t})\log n},$$

where C does not depend on n or t.

Using the approximation $1-e^{-t} \approx t$ for small t, this shows that the two paths are 'almost disjoint' relative to their lengths as soon as $t \gg \frac{1}{\log n}$. This is an example of what we call Chaos—by slightly perturbing the noise field, we get a very different optimizer.

One of the main points of the book we are reading is that superconcentration and chaos are essentially equivalent. More precisely, we have the following theorem which we'll get to in the third lecture.

Theorem 3. *For the Gaussian random polymer model,*

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

if and only if there exists $t_n \to 0$ such that

$$\mathbb{E}[|\hat{p}_n \cap \hat{p}_n^{t_n}|] = o(n).$$

2.3 Multiple Valleys

Roughly speaking, a random optimization problem has multiple valleys if there are many vastly dissimilar near-optimal solutions. Here is a rather technical but precise definition. Suppose we have a sequence of functions $f_n : X_n \to \mathbb{R}$ on sets X_n with a 'similarity measure' $s_n : X_n^2 \to [0,\infty)$. The similarity measure should not be though of like a distance function but rather like the reciprocal of a distance function, where more similar inputs get a higher output.

Definition 1. We say that (f_n, X_n, s_n) (or simply f_n) has the Multiple Valley (MV) property if there are $\epsilon_n, \delta_n, \gamma_n \to 0$ and $K_n \to \infty$ such that, with probability $\geq 1-\gamma_n$, there exists a set $A \subseteq X_n$ of cardinality $\geq K_n$ such that $s_n(x,y) \leq \epsilon_n$ for all distinct $x, y \in A$, and we have

$$\left|\frac{f_n(x)}{\min_{y\in X_n} f_n(y)} - 1\right| \leq \delta_n$$

for all $x \in A$.

In the case of Gaussian polymers, we will take X_n to be the space of all random walk paths of length n, with similarity measure

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

The function we'll consider is the energy function H_n. In the fourth lecture or so, we will see how to prove that this (H_n, X_n, s_n) has multiple valleys; it will follow from the superconcentration of the minimum energy.

3 Another Example: Sherrington-Kirpatrick Model

Now that we've seen all three phenomena in one model, let's see how they manifest in another model. This is the SK model of spin glasses, a mean-field model sort of like the Curie-Weiss model of a ferromagnet, but with random interaction weights between vertices. Specifically, let $(g_{ij})_{1\le i<j\le n}$ be a collection of i.i.d. standard Gaussian random variables, called the 'disorder'. Given this, the energy of a spin configuration $\sigma \in \{\pm 1\}^n$ is

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

Again, this is a *random* Hamiltonian. We can define a Gibbs measure at inverse temperature β via

$$\pi_n^\beta(\sigma) \propto e^{-\beta H_n(\sigma)}.$$

We need to normalize by

$$Z_n(\beta) = \sum_{\sigma\in\{\pm 1\}^n} e^{-\beta H_n(\sigma)},$$

the partition function at inverse temperature β. The partition function encodes a lot of information about the distribution, since it is essentially a generating function for the energy levels. It also gives rise to the 'free energy' which is a popular quantity in statistical physics:

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

Notice that as $\beta\to\infty$ (i.e. as temperature approaches zero),

$$F_n(\beta) \to E_n,$$

the minimum energy of a spin configuration. It is known that the free energy has order n, and in fact

$$\lim_{n\to\infty} \frac{F_n(\beta)}{n}$$

exists and is a deterministic function of β.

3.1 Superconcentration

As for the fluctuations of $F_n(\beta)$, it is known that when $\beta < 1$ (i.e. high temperature),

$$\lim_{n\to\infty} \mathrm{Var}(F_n(\beta)) < \infty.$$

I tried to find an intuitive reason for why this should be true, but I couldn't come up with anything. For low temperatures $\beta > 1$, the classical result is $\mathrm{Var}(F_n(\beta)) \le C(\beta)n$. The superconcentration result is

Theorem 4.

$$\mathrm{Var}(F_n(\beta)) \le \frac{C(\beta)n}{\log n}.$$

It is believed that $\mathrm{Var}(F_n(\beta))$ should be bounded for all temperatures, but this is far from being proved. As a warning, we won't get to this until chapter 10 of the monograph (or chapter 6 for a slightly worse bound).

3.2 Chaos

The overlap between two spin configurations σ and σ' is defined as

$$R(\sigma,\sigma') := \frac{1}{n}\sum_{i=1}^n \sigma_i\sigma_i'.$$

(this is called $R_{1,2}$ in the monograph). Again, we have two different results for $\beta < 1$ and $\beta > 1$. For $\beta < 1$, it is known that if σ, σ' are drawn independently from the same Gibbs distribution, then

$$\lim_{n\to\infty} \mathbb{E}[R(\sigma, \sigma')^2] = 0.$$

This makes sense because in high temperatures, the spin configuration has essentially independent ± 1 entries, and the law of large numbers kicks in. On the other hand, for low temperatures $\beta > 1$, the above limit is positive, which means that for low temperatures, two random samples from the Gibbs distribution will be macroscopically correlated.

But what if instead of sampling two configurations using the same disorder, we slightly perturb the disorder before sampling another configuration. Specifically, let (g_{ij}^t) be a perturbed disorder from (g_{ij}). Sample σ from $\pi_n^\beta(g_{ij})$, and sample σ^t from $\pi_n^\beta(g_{ij}^t)$. Define

$$R(t) = R(\sigma, \sigma^t)$$

We have the following equivalence between chaos of the spin configurations and superconcentration of the free energy:

Theorem 5. *In the SK model,*

$$\mathrm{Var}(F_n(\beta)) = o(n)$$

if and only if there exists $t_n \to 0$ such that

$$\mathbb{E}[R(t_n)^2] = o(1).$$

This equivalence will be shown in the third lecture, and once we prove the superconcentration result will imply chaos for the low temperature SK spin glass model.

3.3 Multiple Valleys

Later on we will show that the Hamiltonian function in the SK model has multiple valleys with respect to the similarity measure

$$s_n(\sigma, \sigma') = R(\sigma, \sigma')^2.$$

This will follow from an equivalence result between superconcentration and the multiple valleys property, but we might also show how it follows directly from chaos.

4 Yet Another Example: Maxima of Gaussian Fields

Consider any centered n-dimensional Gaussian $g = (g_1, \ldots, g_n)$, meaning that each g_i has mean zero. The distribution of g is completely characterized by its covariance matrix (C_{ij}), where $C(i,j) = \mathrm{Cov}(g_i, g_j)$ (note that this is called $R(i,j)$ in the monograph). We will consider the maximum $\max_i g_i$.

4.1 Superconcentration

It is known that

$$\mathrm{Var}(\max_i g_i) \leq \max_i \mathrm{Var}(g_i).$$

This is tight in general, as can be seen by taking $C(i,j) = 1$ for all i, j, meaning that all g_i are the same standard normal random variable. However, when the g_i's are independent and standard, then $\max_i g_i$ has variance $\frac{1}{\log n}$, whereas $\max_i \mathrm{Var}(g_i) = 1$.

When we define superconcentration in general in the third lecture, we will see that $\max_i g_i$ is superconcentrated exactly when the above inequality is suboptimal.

4.2 Chaos

We can think of g as giving a geometry to the index set $\{1, \ldots, n\}$ via

$$d(i,j) = \sqrt{\mathbb{E}[(g_i - g_j)^2]}.$$

Assume for now that all $C(i,i) = 1$, so that

$$d(i,j) = \sqrt{2 - 2C(i,j)},$$

so we can say that the points i and j are further apart when $C(i,j)$ is smaller. Let's also assume that $C(i,j) \geq 0$ so that all the variables are positively correlated.

Assume that the Gaussian is full-dimensional, in the sense that its covariance matrix has full rank. This means that there will always be a unique index at which g attains its maximum. Let I denote this unique index, and let I^t denote the corresponding index for $g^t = (g_1^t, \ldots, g_n^t)$, a perturbed version of g.

We have the following equivalence between superconcentration and chaos (which, in this context manifests as $\mathbb{E}[C(I, I^t)]$ being close to zero for all t above a small threshold).

Theorem 6. *For each $t \geq 0$ we have*

$$0 \leq \mathbb{E}[C(I, I^t)] \leq \frac{\mathrm{Var}(\max_i g_i)}{1 - e^{-t}}, \qquad \textit{and}$$
$$\mathrm{Var}(\max_i g_i) \leq 1 - e^{-t} + \mathbb{E}[C(I, I^t)]e^{-t}.$$

The first inequality says that if we have superconcentration, then we have chaos. The second inequality is the converse statement.

4.3 Multiple Peaks

Here is a theorem which says that superconcentration implies multiple peaks in Gaussian fields.

Theorem 7. *Let $\epsilon = \mathrm{Var}(\max_i g_i)$. Then there is a universal constant D such that if*

$$\delta := \frac{D}{\sqrt{\log(1/\epsilon)}},$$

then with probability at least $1 - \delta$, there are at least $\frac{1}{\delta}$ points i satisfying

$$g_i \geq (1 - \delta) \max_k g_k,$$

such that for any two of these points i, j, we have $C(i,j) \leq \delta$.

If ϵ is small (superconcentration), then δ is also small above, which gives multiple peaks.

5 Conclusion

I think we should be able to finish this monograph this semester. There are twelve chapters in the book and twelve weeks of regularly scheduled seminars before the RRR week. Some of the chapters are more difficult/long than others, but I think after chapter 4 most of them are relatively independent from one another. We can also break up some of the more lengthy chapters if we decide to meet during the RRR or finals weeks.

Either way we will have to skip a few details here and there. That's why I think it will be important for people to follow along with the reading even if they are not giving a talk in a particular week. Even a quick skim should suffice before the lecture, since seeing the material twice will greatly aid in the understanding.

Finally, thanks to everyone for voting for this topic.