Chapter 1
Introduction

This monograph is about a certain curious feature of random objects that I call 'superconcentration', and two related topics, 'chaos' and 'multiple valleys'. Although superconcentration has been a recognized feature in a number of areas of probability theory in the last twenty years (under a variety of names), its connections with chaos and multiple valleys were discovered and explored for the first time in Chatterjee (2008b). This introductory chapter sketches the basic ideas behind these three concepts through some examples. Precise definitions are given in later chapters.

1 Superconcentration

The theory of concentration of measure gives probability theory a range of tools to compute upper bounds on the orders of fluctuations of complicated random variables. Usually, the general techniques of measure concentration are required when more direct problem-specific approaches do not work. The essential techniques of this theory are adequately summarized in the classic monograph of Ledoux (2001) and the recent book by Boucheron et al. (2013). Roughly speaking, superconcentration happens when the classical measure concentration techniques give suboptimal bounds on the order of fluctuations. The techniques available for proving superconcentration are rather inadequate at this point of time. They give only small improvements on the upper bounds from classical theory. One may wonder what is so interesting about such minor improvements. The main point of this monograph is to demonstrate how *any* improvement over a classical upper bound is equivalent to a number of strange and interesting phenomena, such as chaos under small perturbations and the emergence of multiple valleys.

A formal definition of superconcentration will be given later. Right now, the situation is best explained through examples.

S. Chatterjee, *Superconcentration and Related Topics*,
Springer Monographs in Mathematics, DOI 10.1007/978-3-319-03886-5_1,

1.1 First-Passage Percolation

Let $\mathbb{Z}^d$ be the integer lattice and let $E(\mathbb{Z}^d)$ be the set of edges of $\mathbb{Z}^d$. Let $(\omega_e)_{e\in E(\mathbb{Z}^d)}$ be a collection of i.i.d. non-negative random variables, called 'edge weights' or 'passage times'. The edge-weight of e models the 'time it takes for a certain fluid to pass through the edge e'. For a path p of connected edges in $\mathbb{Z}^d$, the 'passage time' of p is defined as the sum of passage times of edges along the path. The first-passage time $T(x, y)$ from a vertex x to a vertex y is defined as the minimum over all passage times of paths connecting x to y. Clearly, it suffices to consider self-avoiding paths only. This is the standard model of first-passage percolation, introduced by Hammersley and Welsh (1965) and subsequently studied by numerous authors.

Given $x \in \mathbb{R}^d$ and an integer n, let $T_n(x)$ denote the first-passage time $T(0, [nx])$, where 0 is the origin and $[nx]$ is the lattice point closest to nx. One of the fundamental results of first-passage percolation is that for all x, the limit

$$\lim_{n\to\infty} \frac{T_n(x)}{n}$$

exists and is a deterministic function of x. Moreover, the limit is positive if the probability that an edge weight is zero is less than the critical percolation probability in dimension d. For more on this and other properties of first-passage percolation, see Kesten (1986, 1993) and Grimmett and Kesten (2012).

Therefore under mild conditions (e.g. continuous edge-weights) the first-passage time from 0 to a point at distance n scales like n. For us, the question of interest is: What is the behavior of $\mathrm{Var}(T_n(x))$?

For notational simplicity, let $x = e_1 = (1, 0, 0, \ldots, 0)$, and denote $T_n(x)$ simply by T_n. It was proved by Kesten (1993) using a martingale argument that $\mathrm{Var}(T_n) \le Cn$, where C is a constant that depends only on the distribution of the edge weights and on the dimension. Kesten moreover proved exponential tail bounds for the fluctuations, which were strengthened by Talagrand (1995) using his famous concentration inequalities for product spaces.

However, none of these bounds improved the bound on the order of fluctuations, which remained at $\sqrt{n}$, until Benjamini et al. (2003) proved that for binary edge weights in dimension $d \ge 2$,

$$\mathrm{Var}(T_n) \le \frac{Cn}{\log n}.$$

Here again, C is a constant depending only the distribution of the edge-weights and the dimension. This result was extended to a more general class of edge weights by Benaïm and Rossignol (2008).

The Benjamini-Kalai-Schramm (BKS) theorem is an example of what I call 'superconcentration', which means, roughly, that the order of fluctuations is less than the upper bound given by classical theory. A more precise definition will be given later.

The physicists conjecture that when $d = 2$, $\mathrm{Var}(T_n)$ should scale like $n^{2/3}$. Besides numerical evidence, there is also some indirect mathematical evidence to support this conjecture: The closely related model of two-dimensional oriented last-passage percolation is 'exactly solvable' in the case of exponential or geometric edge-weights, famously proved by Johansson (2000), who derived the $n^{1/3}$ order of fluctuations as a consequence of the exact solution. See for example Borodin et al. (2012) for more details and a survey of exciting recent developments.

In higher dimensions, there is no consensus about the behavior of the variance of T_n. Some experts think that $\mathrm{Var}(T_n)$ remains bounded as $n \to \infty$ in sufficiently high dimension, while others disagree.

It is not hard to prove that $\mathrm{Var}(T_n)$ is bounded below by a positive constant depending on edge weights and the dimension. A non-trivial lower bound was proved by Newman and Piza (1995) and simultaneously by Pemantle and Peres (1994), who showed that in $d = 2$, $\mathrm{Var}(T_n)$ must grow at least as fast as a multiple of $\log n$ under mild conditions.

In Chap. 5, we will see a proof of the following variant of the BKS theorem.

Theorem 1.1 (Variant of Theorem 1 in Benjamini et al. 2003) *Consider first-passage percolation on $\mathbb{Z}^d$, $d \geq 2$. If the edge-weight distribution can be realized as the probability distribution of a non-negative Lipschitz function of a Gaussian random variable that is uniformly bounded away from zero and infinity, then for all n,*

$$\mathrm{Var}(T_n) \leq \frac{Cn}{\log n},$$

where C is a constant that depends only on the distribution of the edge-weights and the dimension.

The class of edge-weight distributions covered by the above theorem includes, for instance, uniform distributions on bounded intervals.

Open Problem 1.2 Improve the upper bound on the variance in Theorem 1.1. Preferably, find the correct order of the variance.

1.2 Gaussian Random Polymers

Let n be a positive integer, and consider the set of all 1-dimensional random walk paths of length n, starting at 0. In other words, each path is a sequence like $\{(0, a_0), (1, a_1), \ldots, (n, a_n)\}$, where $a_0 = 0$, and $|a_{i+1} - a_i| = 1$ for each i, so that there are 2^n such paths. Any such path is a possible shape of a $(1 + 1)$-dimensional polymer. Here $(1 + 1)$ means 1 space dimension and 1 time dimension.

Let $(g_v)_{v \in \mathbb{Z}^2}$ be a collection of i.i.d. standard Gaussian random variables, called the 'environment' or the 'medium'. Given a random walk path p of length n, define

the 'energy' of p as

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

This is the so-called 'Gaussian random polymer'—or more correctly, the '$(1+1)$-dimensional polymer in Gaussian random media'—originally defined in the mathematical physics literature (see Imbrie and Spencer 1988). The model generalizes easily in higher dimensions, leading to the so-called $(d+1)$-dimensional polymer.

An object of interest is the minimum energy of a path of length n. This is called the 'ground state energy', which will be denoted by E_n. Another object of interest is the minimum energy path, which will be denoted by $\hat{p}_n$. Incidentally, the $(1+1)$-dimensional random polymer model is exactly the same as the model of two-dimensional oriented last-passage percolation.

Along the lines of the BKS theorem (but with some additional technical difficulties), the following result was proved in Chatterjee (2008b).

Theorem 1.3 (Theorem 8.1 in Chatterjee 2008b) *If E_n is the ground state energy in the $(1+1)$-dimensional Gaussian random polymer model, then*

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

where C does not depend on n.

We will see a proof of this theorem in Chap. 5. The result was extended to $(d+1)$-dimensional polymers by Graham (2010).

Open Problem 1.4 Improve the upper bound on the variance in Theorem 1.3. Preferably, find the correct order of the variance.

As in first-passage percolation, classical techniques give a bound of order n instead of $n/\log n$. Theorem 1.3 is another instance of superconcentration.

Incidentally, the $\log n$ improvement over the classical bounds may be optimal in certain examples, although not in first-passage percolation or directed polymers. This is demonstrated by the following example. Let $(g_v)_{v \in \mathbb{Z}^2}$ be i.i.d. standard Gaussian random variables. Consider all 'paths' of the form

$$\{(1, a_1), (2, a_2), \ldots, (n, a_n)\},$$

where now the a_i's can take any value in $\{1, \ldots, n\}$, with no restrictions. Then clearly the maximum possible value of the sum of vertex weights along such a path is precisely

$$\sum_{i=1}^{n} \max_{1 \le j \le n} g_{(i,j)},$$

which has variance of order $n/\log n$ since the maximum of n i.i.d. standard Gaussian random variables has variance of order $1/\log n$ (see Appendix A).

1.3 The Sherrington-Kirkpatrick Model of Spin Glasses

In this model there are n particles, each carrying a spin of $+1$ or -1. A spin configuration $\sigma = (\sigma_1, \ldots, \sigma_n)$ is an element of $\{-1, 1\}^n$. Let $(g_{ij})_{1 \le i < j \le n}$ be a collection of i.i.d. standard Gaussian random variables, called the 'disorder'. Given the disorder, define the energy of a spin configuration σ as

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

This defines the so-called Sherrington-Kirkpatrick (SK) model of spin glasses introduced by Sherrington and Kirkpatrick (1975).

Having so defined the energy, the 'free energy' at 'inverse temperature' $\beta \ge 0$ is defined as

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

where

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

It is known that

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

exists and is a deterministic function of β. The existence of the limit was proved using an ingenious subadditive argument by Guerra and Toninelli (2002). The expression for the limit is given by the remarkable 'Parisi formula', proved rigorously in a famous paper of Talagrand (2006), who built upon an interpolation scheme introduced by Guerra (2003). Extensive details about this and other models of spin glasses, as well as references to the literature, are available in Talagrand (2011, 2012) and Panchenko (2013a,b).

We are interested in the order of fluctuations of $F_n(\beta)$. When $\beta < 1$, it was proved by Aizenman et al. (1987) that

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

exists and is finite. They moreover gave a formula for this limit and proved a central limit theorem for $F_n(\beta)$. Possibly, this result can be extended to $\beta = 1$ to prove an optimal bound for the variance.

Open Problem 1.5 Analyze the fluctuations of $F_n(\beta)$ when $\beta = 1$.

When $\beta > 1$, the best known upper bound for a long time was the classical bound $\mathrm{Var}(F_n(\beta)) \le Cn$. (See Talagrand 2003, Corollary 2.2.5, for example.) Superconcentration of the free energy was proved for the first time in Chatterjee (2009).

Theorem 1.6 (Theorem 1.5 in Chatterjee 2009) *If $F_n(\beta)$ is the free energy of the SK model at inverse temperature β, then*

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

where $C(\beta)$ depends only on β. This result holds for all $\beta \ge 0$.

Interestingly, though the result looks similar to the BKS theorem, the method of proof is very different. The hypercontractive tools that lie at the heart of the BKS proof do not seem to work for spin glasses.

I have heard it said that $\mathrm{Var}(F_n(\beta))$ should be $O(1)$ for all β, or at least for $\beta \ne 1$.

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

See Parisi and Rizzo (2009) for some physics conjectures and predictions related to the above problem.

In Chap. 10, we will see a proof of Theorem 1.6. A slightly worse bound with an easier proof will be presented in Chap. 6.

1.4 Maxima of Gaussian Fields

Consider a Gaussian field $g = (g_1, \dots, g_n)$ on the finite index set $\{1, \dots, n\}$. This is just an n-dimensional Gaussian random vector, but I use the term 'field' because there is a geometry involved, namely the geometry induced by the L^2 metric

$$d(i, j) := \big(\mathbb{E}(g_i - g_j)^2\big)^{1/2}.$$

Assume that g is centered, i.e. each g_i has mean zero. Let

$$R(i, j) := \mathrm{Cov}(g_i, g_j).$$

The question of interest is the following: What is the variance of $\max_i g_i$? It is a well-known result that

$$\mathrm{Var}\Big(\max_{1\le i\le n} g_i\Big) \le \max_{1\le i\le n} \mathrm{Var}(g_i). \tag{1.1}$$

This inequality was proved by Houdré (1995), although the method of proof seems to be implicit in the much earlier work of Nash (1958), and the works of Chernoff (1981), Chen (1982), and Houdré and Kagan (1995). In Chap. 2, we will see how to prove this inequality.

The inequality (1.1) is easily seen to be tight by taking $R(i, j) = 1$ for all i, j. However it can often be suboptimal, e.g. when the g_i's are independent. In fact, the ground state energy of Gaussian random polymers is a special case of this general problem. When we formally define superconcentration in Chap. 3, we will see that the maximum of a Gaussian field is superconcentrated if and only if inequality (1.1) is suboptimal.

Open Problem 1.8 Give a general and easily verifiable condition on the covariance matrix R that is equivalent to superconcentration of the maximum.

2 Chaos

Traditionally, chaos is defined as 'high sensitivity to initial conditions' in the context of dynamical systems. In the statistical physics of disordered systems (such as spin glasses), chaos takes on a slightly different but related meaning: a system in statistical physics is called chaotic if it is highly sensitive to small changes in parameters such as temperature, disorder, etc. While this is not a precise definition, its meaning may be easily clarified through examples.

2.1 Chaos in Gaussian Polymers

While studying random polymer models, physicists (e.g. Huse et al. 1985, Zhang 1987, Mézard 1990) have studied whether polymers are chaotic, that is, whether they are sensitive to small changes in the environment. A specific object of interest is the polymer with minimum energy, i.e. the ground state polymer. Recall that we introduced this model in Sect. 1.2, and denoted the ground state by $\hat{p}_n$.

The physicists have a standard way of perturbing a Gaussian environment. Consider the $(1 + 1)$-dimensional Gaussian polymer, for instance. Take a $t > 0$ and for each vertex $v \in \mathbb{Z}^2$, let (g_v, g_v^t) be a pair of jointly Gaussian random variables with mean 0, variance 1 and correlation e^{-t}. When t is close to zero, $g_v \approx g_v^t$ with high probability. Let $\{(g_v, g_v^t) : v \in \mathbb{Z}^2\}$ be a collection of i.i.d. pairs. Then if $(g_v)_{v\in\mathbb{Z}^2}$ is our original environment, we may call $(g_v^t)_{v\in\mathbb{Z}^2}$ the perturbed environment, or more precisely, the t-perturbed environment.

The following theorem was proved in Chatterjee (2008b). It may be proved using a consequence of Theorem 1.6 from Sect. 1.3 of this chapter and Theorem 3.5 from Chap. 3.

Theorem 1.9 (Theorem 8.1 in Chatterjee 2008b) *Consider the* $(1+1)$*-dimensional Gaussian random polymer of length* n*. 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 polymer paths. Then*

$$\mathbb{E}\left|\hat{p}_n \cap \hat{p}_n^t\right| \le \frac{Cn}{(1-e^{-t})\log n},$$

where C *is a constant that does not depend on* n *or* t.

The above result shows that the paths are 'almost disjoint' relative to their lengths when $t \gg 1/\log n$. This result is related to the formula

$$\mathrm{Var}(E_n) = \int_0^\infty e^{-t}\mathbb{E}\left|\hat{p}_n \cap \hat{p}_n^t\right| dt, \tag{1.2}$$

first proved in Chatterjee (2008b). This is a special case of the so-called 'dynamical formulas of variance' that are widely used in the literature on Gaussian fields (see e.g. Deuschel et al. 2000, Proposition 2.2 or Adler and Taylor 2007, Chap. 2). We will see a proof of this identity and its relation to Theorem 1.9 later in the monograph.

For the physics connection to Theorem 1.9, see for example da Silveira and Bouchaud (2004).

One of the main thrusts of this monograph is that a superconcentration result such that Theorem 1.3 and a chaos result such as Theorem 1.9 are essentially equivalent. Indeed, we will see how to prove the following theorem in Chap. 3.

Theorem 1.10 *Consider the Gaussian random polymer model in any dimension. If* E_n *is the ground state energy, then*

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

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

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

Combining this with the superconcentration result of Graham (2010), it follows that the Gaussian random polymer is chaotic in any dimension.

2.2 Chaos in the SK Model

Given an inverse temperature β and a disorder $(g_{ij})_{1\le i<j\le n}$, the SK model defines a 'Gibbs measure' on $\{-1, 1\}^n$ by putting mass

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

at each configuration σ, where Z_n and H_n are defined in Sect. 1.3. The Gibbs measure is a random probability measure on $\{-1, 1\}^n$.

An important quantity associated with the SK model is the 'overlap': Having defined the Gibbs measure, let σ^1 and σ^2 be two spin configurations drawn independently from the Gibbs measure. The overlap between σ^1 and σ^2 is defined as

$$R_{1,2} := \frac{1}{n}\sum_{i=1}^{n} \sigma_i^1 \sigma_i^2. \tag{1.3}$$

It is known from the work of Aizenman et al. (1987) that when $\beta < 1$,

$$\lim_{n\to\infty} \mathbb{E}\big(R_{1,2}^2\big) = 0.$$

It was shown later by Guerra (1995) that the same result holds when $\beta = 1$. However, it is believed that the system undergoes a phase transition at $\beta = 1$, one of the effects of which is that the above limit is positive when $\beta > 1$. (For a proof of this, see Example 1 in Panchenko 2008.)

The chaos question in the SK model is similar to the one in polymers: Is the Gibbs measure chaotic under small perturbations of the disorder? (This is known as chaos in disorder; there is also a similar 'chaos in temperature'.)

The question is precisely formulated as follows. As in the polymer model, let $(g_{ij}^t)_{1\le i<j\le n}$ be a t-perturbed disorder; this means, for each (i, j), (g_{ij}, g_{ij}^t) is a pair of Gaussian random variables with mean zero, variance 1, and correlation e^{-t}.

Fix $\beta \ge 0$. Let σ^1 be a configuration picked from the original Gibbs measure at inverse temperature β, and let σ^2 be a configuration picked from the Gibbs measure defined by the t-perturbed disorder. Given the two disorders, σ^1 and σ^2 are independent. Define

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

Physicists (e.g. Bray and Moore 1987, Fisher and Huse 1986, Krzakała and Bouchaud 2005) define chaos in disorder to be the phenomenon that $R_{1,2}(t)$ is close to zero for some small positive t. (For an up-to-date survey of the physics literature on this, see Rizzo 2009.) This is easily proved to be true when $\beta < 1$ following the Aizenman-Lebowitz-Ruelle argument; the point is that this is supposed to be true for all β. This was proved rigorously in Chatterjee (2009).

Theorem 1.11 (Theorem 1.3 in Chatterjee 2009) *Let $R_{1,2}(t)$ be defined as above. Then for any integer $k \ge 1$,*

$$\mathbb{E}\big(R_{1,2}^{2k}(t)\big) \le \big(C_1(\beta)k\big)^k n^{-C_2(\beta)k\min\{1,t\}},$$

where $C_1(\beta)$ and $C_2(\beta)$ are positive constants that depend only on β.

A non-trivial extension of the above theorem to the SK model in the presence of an external field was recently achieved by Chen (2011). Chaos in mixed p-spin models was proved by Chen and Panchenko (2012).

The above theorem will be proved in Chap. 10. Moreover, we will show in Chap. 3 that this theorem is equivalent to the superconcentration of the free energy:

Theorem 1.12 *In the SK model,*

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

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

$$\mathbb{E}\big(R_{1,2}^2(t_n)\big) = o(1).$$

2.3 Chaos in Gaussian Fields

Let $g = (g_1, \ldots, g_n)$ be a centered Gaussian field with covariance matrix R, as in Sect. 1.4. Let g' be an independent copy of g. For each $t \geq 0$, let

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

Under the mild condition that $R(i, j) \neq 1$ for all $i \neq j$, there is almost surely a unique (random) index $I \in \{1, \ldots, n\}$ where the field g attains its maximum. Similarly, for each $t \geq 0$ let I^t be the unique index at which g^t attains its maximum. Note that $g^0 = g$ and $I^0 = I$.

Suppose that the field g is 'positively correlated', i.e. $R(i, j) \geq 0$ for all i, j. Note that in the L^2 geometry on $\{1, \ldots, n\}$ induced by the field g, the distance $d(i, j)$ may be represented as

$$d(i, j) = \sqrt{R(i, i) + R(j, j) - 2R(i, j)}.$$

In other words, the smaller the value of $R(i, j)$, the further apart are the points i and j. Thus one may roughly say that the location of the maximum of g is 'chaotic under small perturbations' if $\mathbb{E}(R(I^0, I^t))$ is close to zero for all values of t above a small threshold.

The following generalization of the identity (1.2) was first proved in Chatterjee (2008b) and simultaneously in Nourdin and Viens (2009):

$$\mathrm{Var}\Big(\max_{1 \leq i \leq n} g_i\Big) = \int_0^\infty e^{-t} \mathbb{E}\big(R\big(I^0, I^t\big)\big)\, dt. \tag{1.4}$$

The important observation made in Chatterjee (2008b) is that the integrand on the right is always a decreasing function of t. From this the following theorem was derived. It shows that the maximum of a Gaussian field is superconcentrated if and only if the location of the maximum is chaotic under small perturbations. We will see how to prove this theorem in Chap. 3.

Theorem 1.13 (Theorem 3.2 in Chatterjee 2008b) *Let g, R and I^t be as above. Let $v = \mathrm{Var}(\max_{1\le i\le n} g_i)$ and $w := \max_{1\le i\le n} \mathrm{Var}(g_i)$. Then for each $t \ge 0$ we have*

$$0 \le \mathbb{E}\big(R\big(I^0, I^t\big)\big) \le \frac{v}{1-e^{-t}}, \quad \textit{and}$$

$$v \le w\big(1-e^{-t}\big) + \mathbb{E}\big(R\big(I^0, I^t\big)\big)e^{-t}.$$

To understand this result, suppose that the coordinates are normalized to satisfy $\mathrm{Var}(g_i) = 1$ for each i. Then $w = 1$, and the maximum is superconcentrated if and only if v is close to zero. If $v \approx 0$, then the first inequality shows that $\mathbb{E}(R(I^0, I^t)) \approx 0$ for all t above a small threshold, namely, for $t \gg v$. On the other hand, if $\mathbb{E}(R(I^0, I^t)) \approx 0$ for all t above a small threshold, let's say $t \ge \delta$, then the second inequality shows that $v \ll w$, e.g. by taking $t = \delta$. Thus, the first inequality proves that superconcentration of the maximum implies that the location of the maximum is chaotic under small perturbations, and the second inequality proves the converse statement.

Note that the condition that $R(i, j) \ge 0$ for all i, j is not required for the above theorem to hold, but the result does not carry a lot of meaning without this condition.

3 Multiple Valleys

This section is an introduction to the third property related to superconcentration, namely, multiple valleys. To motivate the definition of multiple valleys, we will first discuss a related concept that is in some sense the opposite of multiple valleys.

3.1 Asymptotic Essential Uniqueness

An optimization problem is called 'stable' if any near-optimum is close to the optimum in some appropriate metric. Mathematics abounds with stability questions for optimization problems. In probability theory, we usually have a sequence of optimization problems instead of a single problem. Aldous (2001) defined a notion of stability that makes sense in the probabilistic setting. This is known as Asymptotic Essential Uniqueness (AEU).

The first example where AEU was rigorously established is the so-called random assignment problem (see Aldous 2001). There are n tasks and n individuals. Each individual can be assigned exactly one task. The cost of assigning task j to individual i is c_{ij}. The assignment problem seeks to minimize $\sum_{i=1}^n c_{i\pi(i)}$ over all permutations π of $\{1, \ldots, n\}$. When the c_{ij}'s are random variables, probability enters into the picture. In the simplest scenario, the c_{ij}'s are i.i.d. non-negative random variables. Assume that the c_{ij}'s have a density with respect to Lebesgue measure and that the value of the density at 0 is 1. (For example, the c_{ij}'s may be i.i.d. exponential random variables with mean 1.) Aldous (2001) proved a famous conjecture of

Parisi, who claimed that as $n \to \infty$, the minimum cost of the random assignment problem approaches $\zeta(2)$, where ζ is the Riemann zeta function.

In the same paper, Aldous introduced and proved the AEU property for the random assignment problem. It was shown that any assignment (permutation) that nearly minimizes the cost must be almost the same as the optimal assignment. The precise result is as follows: For all $0 < \delta < 1$, there exists an $\epsilon(\delta) > 0$ such that if μ_n are (random) permutations depending on $(c_{ij})_{1 \le i,j \le n}$ such that for all n,

$$\mathbb{E}\left(\frac{1}{n}\left|\left\{i : \mu_n(i) \neq \pi_n(i)\right\}\right|\right) \geq \delta,$$

then

$$\liminf_{n\to\infty} \mathbb{E}\left(\sum_{i=1}^{n} c_{i\mu(i)}\right) \geq \zeta(2) + \epsilon.$$

Similar results for the minimal spanning tree on graphs with randomly weighted edges and a few other models were later proved by Aldous et al. (2008, 2009). Other problems, such as the random Euclidean traveling salesman problem, are thought to have the AEU property but with no proofs (see Aldous 2001, Sect. 7).

A simple example where AEU can be directly verified is as follows. Let $g_1, \ldots, g_n$ be i.i.d. standard Gaussian random variables. Define a random function $f_n : \{-1, 1\}^n \to \mathbb{R}$ as

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

The problem is to maximize this function over all possible values of σ. Clearly, the maximum is attained at $\hat{\sigma}$, where $\hat{\sigma}_i = \text{sign}(g_i)$. It is simple to prove that this optimization problem has the AEU property. More precisely, one can try to prove the following exercise.

Exercise 1.14 Show that for any given $\epsilon > 0$ there is a $\delta > 0$ small enough, such that with probability tending to 1 as $n \to \infty$, all σ such that $f_n(\sigma) \geq (1 - \delta) f_n(\hat{\sigma})$ satisfy

$$\frac{1}{n}\left|\{i : \sigma_i \neq \hat{\sigma}_i\}\right| \leq \epsilon.$$

3.2 Multiple Valleys and Peaks

The notion of multiple valleys (or multiple peaks) is sort of an opposite of the AEU property. This concept was introduced in Chatterjee (2008b) and is somewhat different than the notion of multiple valleys in the physics folklore.

Roughly speaking, we will say that a random optimization problem has multiple valleys if there are many vastly dissimilar near-optimal solutions. A precise definition may be given as follows. Suppose we have a sequence of sets X_n, and for each n, $f_n : X_n \to \mathbb{R}$ is a random function. Let s_n be a 'similarity measure' on X_n: for each $x, y \in X_n$, $s_n(x, y)$ is a non-negative real number that measures the degree of similarity between x and y. We impose no conditions on s_n besides non-negativity (and measurability, if the situation demands).

Definition 1.15 A sequence (f_n, X_n, s_n) (or simply f_n) is said to have the Multiple Valley (MV) property if there exist ϵ_n, δ_n and γ_n tending to zero and K_n tending to infinity, such that for each n, 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 $x, y \in A$, $x \neq y$, and for all $x \in A$,

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

If the min is replaced by max, we shall call it the Multiple Peaks property.

Let us now go back to the Sherrington-Kirkpatrick model. Take X_n to be the space $\{-1, 1\}^n$, f_n to be the Hamiltonian function

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

and s_n to be the similarity measure

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

The following result was first proved in Chatterjee (2009).

Theorem 1.16 (Chatterjee 2009) *For the similarity measure defined above, the Hamiltonian of the Sherrington-Kirkpatrick model has the multiple valley property. In other words, when n is large, then with high probability there are many states with near-minimal energy that are all nearly mutually orthogonal to each other.*

The connection with our previous discussion of chaos and superconcentration is that we will derive the multiple valley result as a consequence of the chaos property of the SK model.

We will prove a similar result for $(1 + 1)$-dimensional polymers. Here the space X_n is the space of all random walk paths of length n. The similarity between two paths p and p' is

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

where $|p \cap p'|$ is the number of vertices in the intersection of the two paths. The function f_n in this case is the Hamiltonian (energy) function H_n, that is, the sum of vertex weights along a given path, together with a negative sign in front. The following theorem will be proved as a consequence of the superconcentration of the ground state energy. This was originally proved in Chatterjee (2008b).

Theorem 1.17 (Chatterjee 2008b) *The energy function in the* $(1+1)$*-dimensional Gaussian polymer model has multiple valleys, if the similarity measure is defined as above. In other words, when* n *is large, then with high probability there are many paths that all have nearly minimal energy and are all nearly disjoint from each other.*

The result extends to the $(d+1)$-dimensional polymer for any $d \geq 2$ using the superconcentration result of Graham (2010).

In Chap. 4, we will see how the existence of multiple valleys follows from superconcentration. Combining this with the superconcentration of the free energy (in the case of the SK model) and the ground state energy (in the case of polymers), the proofs of Theorems 1.17 and 1.16 will follow.

For general Gaussian fields, it was proved in Chatterjee (2008b) that superconcentration (or equivalently, chaos) implies the existence of multiple peaks. The following theorem is a cleaned-up version of the result from Chatterjee (2008b).

Theorem 1.18 (Variant of Theorem 3.7 in Chatterjee 2008b) *Let* $g = (g_1, \ldots, g_n)$ *be a centered Gaussian field with covariance matrix* R*. Suppose that* $R(i,i) = 1$ *and* $R(i,j) \geq 0$ *for all* i, j*, and let*

$$\epsilon := \mathrm{Var}\Big(\max_{1 \leq i \leq n} g_i\Big).$$

Then there is a universal constant C *such that if*

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

then with probability at least $1-\delta$*, there are at least* $1/\delta$ *points* i *satisfying*

$$g_i \geq (1-\delta) \max_{1 \leq j \leq n} g_j,$$

such that for any two of these points i *and* j*,* $R(i,j) \leq \delta$*.*

In other words, if ϵ is close to zero (i.e. the maximum is superconcentrated), then with high probability there are a large number of indices that are mutually 'distant' from each other, at which g is near-maximal.

http://www.springer.com/978-3-319-03885-8

Superconcentration and Related Topics

Chatterjee, S.

2014, IX, 156 p., Hardcover
ISBN: 978-3-319-03885-8