5.1 Random energy model.1 (Disordered systems) ③

The nightmare of every optimization algorithm is a random landscape; if every new configuration has an energy uncorrelated with the previous ones, no search method is better than systematically examining every configuration. Finding ground states of disordered systems like spin glasses and random-field models, or equilibrating them at non-zero temperatures, is challenging because the energy landscape has many features that are quite random. The random energy model (REM) is a caricature of these disordered systems, where the correlations are completely ignored. While optimization of a single REM becomes hopeless, we shall see that the study of the ensemble of REM problems is quite fruitful and interesting.

The REM has \( M = 2^N \) states for a system with \( N \) ‘particles’ (like an Ising spin glass with \( N \) spins), each state with a randomly chosen energy. It describes systems in limit when the interactions are so strong and complicated that flipping the state of a single particle completely randomizes the energy. The states of the individual particles then need not be distinguished; we label the states of the entire system by \( j \in \{1, \ldots, 2^N\} \). The energies of these states \( E_j \) are assumed independent, uncorrelated variables with a Gaussian probability distribution

\[
P(E) = \frac{1}{\sqrt{\pi N}} e^{-E^2/N} \quad (1)
\]

of standard deviation \( \sqrt{N/2} \).

Microcanonical ensemble. Consider the states in a small range \( E < E_j < E + \delta E \). Let the number of such states in this range be \( \Omega(E) \delta E \).

---

1This exercise draws heavily from Mézard and Montanari, chapter 7.
(a) Calculate the average
\[ \langle \Omega(N\epsilon) \rangle_{\text{REM}} \]
over the ensemble of REM systems, in terms of the energy per particle \( \epsilon \). For energies near zero, show that this average density of states grows exponentially as the system size \( N \) grows. In contrast, show that \( \langle \Omega(N\epsilon) \rangle_{\text{REM}} \) decreases exponentially for \( E < -N\epsilon_* \) and for \( E > N\epsilon_* \), where the limiting energy per particle
\[ \epsilon_* = \sqrt{\log 2} . \]
(Hint: The total number of states \( 2^N \) either grows faster or more slowly than the probability density per state \( P(E) \) shrinks.)

What does an exponentially growing number of states mean? Let the entropy per particle be \( s(\epsilon) = S(N\epsilon)/N \). Then (setting \( k_B = 1 \) for notational convenience) \( \Omega(E) = \exp(S(E)) = \exp(Ns(\epsilon)) \) grows exponentially whenever the entropy per particle is positive.

What does an exponentially decaying number of states for \( \epsilon < -\epsilon_* \) mean? It means that, for any particular REM, the likelihood of having any states with energy per particle near \( \epsilon \) vanishes rapidly as the number of particles \( N \) grows large.

How do we calculate the entropy per particle \( s(\epsilon) \) of a typical REM? Can we just use the annealed average
\[ s_{\text{annealed}}(\epsilon) = \lim_{N \to \infty} (1/N) \log \langle \Omega(E) \rangle_{\text{REM}} \]
computed by averaging over the entire ensemble of REMs?

(b) Show that \( s_{\text{annealed}}(\epsilon) = \log 2 - \epsilon^2 \).

If the energy per particle is above \(-\epsilon_*\) (and below \( \epsilon_* \)), the expected number of states \( \Omega(E) \delta E \) grows exponentially with system size, so the fractional fluctuations become unimportant as \( N \to \infty \). The typical entropy will become the annealed entropy. On the other hand, if the energy per particle is below \(-\epsilon_*\), the number of states in the energy range \((E, E + \delta E)\) rapidly goes to zero, so the typical entropy \( s(\epsilon) \) goes to minus infinity. (The annealed entropy is not minus infinity because it gets a contribution from exponentially rare REMs that happen to have an energy level far into the tail of the probability distribution.) Hence
\[ s(\epsilon) = s_{\text{annealed}}(\epsilon) = \log 2 - \epsilon^2 \quad |\epsilon| < \epsilon_* \]
\[ s(\epsilon) = -\infty \quad |\epsilon| > \epsilon_* . \]

Notice why these arguments are subtle. Each REM model in principle has a different entropy. For large systems as \( N \to \infty \), the entropies of different REMs look more and more similar to one another\(^3\) (the entropy is self-averaging) whether \(|\epsilon| < \epsilon_* \) or \(|\epsilon| > \epsilon_* \). However, \( \Omega(E) \) is not self-averaging for \(|\epsilon| > \epsilon_* \), so the typical entropy is not given by the ‘annealed’ logarithm \( \langle \Omega(E) \rangle_{\text{REM}} \).

This sharp cutoff in the energy distribution leads to a phase transition as a function of temperature.

---

\(^2\)Annealing a disordered system (like an alloy or a disordered metal with frozen-in defects) is done by heating it to allow the defects and disordered regions to reach equilibrium. By averaging \( \Omega(E) \) not only over levels within one REM but also over all REMs, we are computing the result of equilibrating over the disorder—an annealed average.

\(^3\)Mathematically, the entropies per particle of REM models with \( N \) particles approach that given by equation 5 with probability one (Mézard and Montanari, eqn 5.10).
1. Plot $s(\epsilon)$ versus $\epsilon$, and illustrate graphically the relation $1/T = \partial S/\partial E = \partial s/\partial \epsilon$ as a tangent line to the curve, using an energy in the range $-\epsilon_* < \epsilon < 0$. What is the critical temperature $T_c$? What happens to the tangent line as the temperature continues to decrease below $T_c$?

When the energy reaches $\epsilon_*$, it stops changing as the temperature continues to decrease (because there are no states below $\epsilon_*$).

2. Solve for the free energy per particle $f(T) = \epsilon - Ts$, both in the high-temperature phase and the low temperature phase. (Your formula for $f$ should not depend upon $\epsilon$.) What is the entropy in the low temperature phase? (Warning: The microcanonical entropy is discontinuous at $\epsilon_*$. You’ll need to reason out which limit to take to get the right canonical entropy below $T_c$.)

The REM has a glass transition at $T_c$. Above $T_c$ the entropy is extensive and the REM acts much like an equilibrium system. Below $T_c$ one can show (Mézard and Montanari, eqn 5.25) that the REM thermal population condenses onto a finite number of states (i.e., a number that does not grow as the size of the system increases), which goes to zero linearly as $T \to 0$.

The mathematical structure of the REM also arises in other, quite different contexts, such as combinatorial optimization (Exercise 2) and random error correcting codes (Mézard and Montanari, ch 6).

---

4. The distribution of ground-state energies for the REM is an extremal statistics problem, which for large $N$ has a Gumbel distribution (see the Extreme Value Statistics exercise).

5. This exercise draws heavily from Mézard and Montanari, chapter 7.
Parts of the exercise are to be done on the computer; hints can be found on the computer exercises portion of the book Web site.

Let’s start with the brute-force approach.

(a) See if you can determine whether a fair split exists for the following sets (from Mézard and Montanari):

\[ S_1 = [10, 13, 23, 6, 20], \]
\[ S_2 = [6, 4, 9, 14, 12, 3, 15, 15], \]
\[ S_3 = [93, 58, 141, 209, 179, 48, 225, 228], \]
\[ S_4 = [2474, 1129, 1388, 3752, 821, 2082, 201, 739]. \]

Hint: \( S_1 \) has a balanced partition, and \( S_4 \) has a minimum cost of 48. Don’t spend an inordinate amount of time searching – cleverness doesn’t help much in this problem.

(b) In Figure 1, does it appear that there is a phase transition for large systems where fair partitions go from probable to unlikely? What value of \( \kappa_c \) would you estimate as the critical point?

Should we be calling this a phase transition? It emerges for large systems; only in the ‘thermodynamic limit’ where \( N \) gets large is the transition sharp. It separates two regions with qualitatively different behavior. The problem is much like a spin glass, with two kinds of random variables: the skill levels of each player \( a_j \) are fixed, ‘quenched’ random variables for a given random instance of the problem, and the assignment to teams can be viewed as spins \( s_j = \pm 1 \) that can be varied (‘annealed’ random variables) to minimize the cost \( C = |\sum_j a_j s_j| \).

---

6Quenched random variables are fixed terms in the definition of the system, representing dirt or disorder that was frozen in as the system was formed (say, by quenching the hot liquid material into cold water, freezing it into a disordered configuration). Annealed random variables are the degrees of freedom that the system can vary to explore different configurations and minimize its energy or free energy.
(c) Show that the square of the cost $C^2$ is of the same form as the Hamiltonian for a spin glass, $H = \sum_{i,j} J_{ij} s_i s_j$. What is $J_{ij}$?

The putative phase transition in the optimization problem (part (b)) is precisely a zero-temperature phase transition for this spin-glass Hamiltonian, separating a phase with zero ground-state energy from one with non-zero energy in the thermodynamic limit.

We can understand both the value $\kappa_c$ of the phase transition and the form of $p_{\text{perf}}(N, M)$ by studying the distribution of possible ‘signed’ costs $E_s = \sum_j a_j s_j$. These energies are distributed over a maximum total range of $E_{\text{max}} - E_{\text{min}} = 2 \sum_{j=1}^N a_j \leq 2N 2^M$ (all players playing on the plus team, through all on the minus team). For the bulk of the possible team choices $\{s_j\}$, though, there will be some cancellation in this sum. The probability distribution $P(E)$ of these energies for a particular NPP problem $\{a_j\}$ is not simple, but the average probability distribution $\langle P(E) \rangle$ over the ensemble of NPP problems can be estimated using the central limit theorem. (Remember that the central limit theorem states that the sum of $N$ random variables with mean zero and standard deviation $\sigma$ converges rapidly to a normal (Gaussian) distribution of standard deviation $\sqrt{N}\sigma$.)

(d) Estimate the mean and variance of a single term $s_j a_j$ in the sum, averaging over both the spin configurations $s_j$ and the different NPP problem realizations $a_j \in [1, \ldots, 2^M]$, keeping only the most important term for large $M$. (Hint: Approximate the sum as an integral, or use the explicit formula $\sum_{j=1}^K k^2 = K^3/3 + K^2/2 + K/6$ and keep only the most important term.) Using the central limit theorem, what is the ensemble-averaged probability distribution $P(E)$ for a team with $N$ players? Hint: Here $P(E)$ is non-zero only for even integers $E$, so for large $N$ $P(E) \approx (2/\sqrt{2\pi}\sigma) \exp(-E^2/2\sigma^2)$; the normalization is doubled.

Your answer to part (d) should tell you that the possible energies are mostly distributed among integers in a range of size $\sim 2^M$ around zero, up to a factor that goes as a power of $N$. The total number of states explored by a given system is $2^N$. So, the expected number of zero-energy states should be large if $N \gg M$, and go to zero rapidly if $N \ll M$. Let us make this more precise.

(e) Assuming that the energies for a specific system are randomly selected from the ensemble average $P(E)$, calculate the expected number of zero-energy states as a function of $M$ and $N$ for large $N$. What value of $\kappa = M/N$ should form the phase boundary separating likely from unlikely fair partitions? Does that agree well with your numerical estimate from part (b)?

The assumption we made in part (e) ignores the correlations between the different energies due to the fact that they all share the same step sizes $a_j$ in their random walks. Ignoring these correlations turns out to be a remarkably good ap-

---

7More precisely, we ignore correlations between the energies of different teams $s = \{s_i\}$, except for swapping the two teams $s \to -s$. This leads to the $N - 1$ in the exponent of the exponent for $p_{\text{perf}}$ in part (f). Notice that in this approximation, NPP is a form of the random energy model (REM, exercise 1), except that we are interested in states of energy near $E = 0$, rather than minimum energy states.
proximation. We can use the random-energy approximation to estimate \( p_{\text{perf}} \) that you plotted in part (b).

(f) In the random-energy approximation, argue that
\[
p_{\text{perf}} = 1 - (1 - P(0))^{2N-1}.
\]
Approximating \((1 - A/L)^L \approx \exp(-A)\) for large \( L \), show that
\[
p_{\text{perf}}(\kappa, N) \approx 1 - \exp \left[ -\sqrt{\frac{3}{2\pi N}} 2^{-N(\kappa - \kappa_c)} \right].
\]
(6)

(g) Does the random-energy approximation explain the data well?

Rigorous results show that this random-energy approximation gives the correct value of \( \kappa_c \). The entropy of zero-cost states below \( \kappa_c \), the probability distribution of minimum costs above \( \kappa_c \) (of the Weibull form, (see the Extreme Value Statistics exercise), and the probability distribution of the \( k \) lowest cost states are also correctly predicted by the random-energy approximation. It has also been shown that the correlations between the energies of different partitions vanish in the large \((N, M)\) limit so long as the energies are not far into the tails of the distribution, perhaps explaining the successes of ignoring the correlations.

What does this random-energy approximation imply about the computational difficulty of NPP? If the energies of different spin configurations (arrangements of kids on teams) were completely random and independent, there would be no better way of finding zero-energy states (fair partitions) than an exhaustive search of all states. This perhaps explains why the best algorithms for NPP are not much better than the exhaustive search you implemented in part (a); even among \( \text{NP}^{-}\text{complete} \) problems, NPP is unusually unyielding to clever methods.\(^8\) It also lends credibility to the conjecture in the computer science community that \( \text{P} \neq \text{NP}^{-}\text{complete} \); any polynomial-time algorithm for NPP would have to ingenuously make use of the seemingly unimportant correlations between energy levels.

\(^8\) The computational cost does peak near \( \kappa = \kappa_c \). For small \( \kappa \ll \kappa_c \) it’s relatively easy to find a good solution, but this is mainly because there are so many solutions; even random search only needs to sample until it finds one of them. For \( \kappa > \kappa_c \) showing that there is no fair partition becomes slightly easier as \( \kappa \) grows (Mézard and Montanari, fig 7.3)