4 The quenched free energy and replicas

In the previous lecture we saw that in disordered systems, averaging the partition function over the disorder results in a solution with undesirable properties. In particular, this annealed average represents behavior of a system with temperature-dependent and atypical values of the couplings between spins. We also argued that an alternative way of taking the average, over the free energy rather than the partition function, repairs these problems and gives an answer that describes the behavior of typical samples of the system and its couplings.

This is more easily said than done. We must calculate

$$F = -\frac{1}{\beta}\overline{\log Z_J(\beta)} \tag{1}$$

with all the Gaussian random variables making up the disordered inside a nasty nonlinear function. In general, we do not know how to take averages over such objects. Therefore, we are going to have to make use of some tricks to temporarily reduce the problem to one which *is* an ordinary exponential function of Gaussian variables, where we know how to take the average, and then later convert back. The trick relies on the following trivial calculus identity:

$$\frac{\partial}{\partial n}x^n = x^n \log x \tag{2}$$

This implies that we can write $\log x$ as

$$\log x \lim_{n\to 0}\frac{\partial}{\partial n}x^n \tag{3}$$

It seems simple enough, but how can it help with our quenched free energy? We would write

$$\log Z_J(\beta) = \lim_{n\to 0}\frac{\partial}{\partial n}Z_J(\beta)^n \tag{4}$$

Now we will make a step that mathematicians tend not to appreciate, which is to pretend for a while that n is actually an integer and not a real number. Then we can write

$$\begin{aligned}\log Z_J(\beta) &= \lim_{n\to 0}\frac{\partial}{\partial n}Z_J(\beta)^n = \lim_{n\to 0}\frac{\partial}{\partial n}\underbrace{Z_J(\beta)\cdots Z_J(\beta)}_{n} \\ &= \lim_{n\to 0}\frac{\partial}{\partial n}\left(\int ds_1\, e^{-\beta H(s_1)}\right)\cdots\left(\int ds_n\, e^{-\beta H(s_n)}\right) \\ &= \lim_{n\to 0}\frac{\partial}{\partial n}\int\left(\prod_{a=1}^{n} ds_a\right)\, e^{-\beta\sum_{a=1}^{n}H(s_a)}\end{aligned} \tag{5}$$

The result appears to be a single partition function for a system whose configuration space is the product of n copies of the original configuration space, and whose energy is the sum of the energies due to each of the individual configuration spaces. We say that we have *replicated* the original system n times, which gives this method for treating the logarithm the *replica method*. Note that, as desired, the dependence of our logarithm on the Gaussian random variables is in the form of a linear exponential, so that we can take averages.

The replicated system may seem trivial at first glance: if the energy is indeed the sum of energies of the replica subsystems, then there are no interactions between the replicas and we should get an answer that is identical to that of each of the subsystems independently. However, there is a deep subtly here that will quickly emerge as we begin to do calculations: each of the replicated subsystems has identical disordered couplings. The effect of this is similar to the effect that would arise from interactions: certain spin configurations will be favored at low temperature among each subsystem, and this will cause the subsystems to be correlated with one another despite the lack of interaction. When we average away the disorder, this correlation will be expressed through emergent interactions among the replicas.

Recall that the system we are studying, the spherical spin glasses, have a configuration space given by the sphere $\|\mathbf{s}\|^2 = N$ and a Hamiltonian given by

$$H_J(\mathbf{s}) = \frac{1}{\sqrt{2p!N^{p-1}}} \sum_{i_1,\ldots,i_p} J_{i_1,\ldots,i_p} s_{i_1} \cdots s_{i_p} \tag{6}$$

Recall further that because this Hamiltonian is a linear combination of Gaussian random variables, it is also a Gaussian random variable, and therefore completely specified by its mean $\overline{H_J(\mathbf{s})} = 0$ and its covariance

$$\overline{H_J(\mathbf{s})H_J(\mathbf{s}')} = Nf\left(\frac{\mathbf{s}\cdot\mathbf{s}'}{N}\right) \tag{7}$$

where $f(q) = \frac{1}{2}q^p$ for the p-spin models we are studying. Because H_J is a Gaussian random variable, it is simple to take the average over disorder in the replicated partition function. Using the usual properties of Gaussian random variables, we have

$$\begin{aligned}
\overline{\log Z_J(\beta)} &= \lim_{n\to 0} \frac{\partial}{\partial n} \int \left(\prod_{a=1}^{n} d\mathbf{s}_a\right) \overline{e^{-\beta \sum_{a=1}^{n} H_J(\mathbf{s}_a)}} \qquad (8)\\
&= \lim_{n\to 0} \frac{\partial}{\partial n} \int \left(\prod_{a=1}^{n} d\mathbf{s}_a\right) e^{\frac{1}{2}\beta^2 \sum_{a=1}^{n}\sum_{b=1}^{n} \overline{H_J(\mathbf{s}_a)H_J(\mathbf{s}_b)}}\\
&= \lim_{n\to 0} \frac{\partial}{\partial n} \int \left(\prod_{a=1}^{n} d\mathbf{s}_a\right) e^{\frac{1}{2}N\beta^2 \sum_{ab}^{n} f\left(\frac{\mathbf{s}_a\cdot\mathbf{s}_b}{N}\right)}
\end{aligned}$$

As promised, averaging over the disorder has produced a new effective energy for the configurations that couples configurations between different subsystems by their overlap. Notice that the annealed calculation we did in the previous lecture is recovered here if we take $n = 1$.

Now the averaged integrand still depends on the configurations $\mathbf{s}_a$ in a non-trivial way, but only through the $n \times n$ *matrix of overlaps* $Q_{ab} = \frac{1}{N}\mathbf{s}_a \cdot \mathbf{s}_b$. Because of the way this matrix appears in the averaged replicated partition function, it is a natural candidate for the order parameter of our model. We will discuss the physical significance of this matrix later, but for now we can allow ourselves to be led by the suggestion that it makes an appropriate order parameter. To use it, we must introduce δ functions that fix its value, with

$$\begin{aligned}\overline{\log Z_J(\beta)} &= \lim_{n\to 0}\frac{\partial}{\partial n}\int\left(\prod_{ab} dQ_{ab}\,\delta(NQ_{ab} - \mathbf{s}_a\cdot\mathbf{s}_b)\right)\left(\prod_{a=1}^{n} d\mathbf{s}_a\right)\, e^{\frac{1}{2}N\beta^2\sum_{ab}^{n} f(Q_{ab})} \\ &= \lim_{n\to 0}\frac{\partial}{\partial n}\int dQ\,d\tilde{Q}\left(\prod_{a=1}^{n} d\mathbf{s}_a\right)\, e^{\frac{1}{2}N\beta^2\sum_{ab}^{n} f(Q_{ab})+\frac{1}{2}\sum_{ab}\tilde{Q}_{ab}(NQ_{ab}-\mathbf{s}_a\cdot\mathbf{s}_b)} \\ &= \lim_{n\to 0}\frac{\partial}{\partial n}\int dQ\,d\tilde{Q}\, e^{\frac{1}{2}N\beta^2\sum_{ab}^{n} f(Q_{ab})+\frac{1}{2}N\sum_{ab}\tilde{Q}_{ab}Q_{ab}}\int\left(\prod_{a=1}^{n} d\mathbf{s}_a\right) e^{-\frac{1}{2}\sum_{ab}\tilde{Q}_{ab}\mathbf{s}_a\cdot\mathbf{s}_b}\end{aligned}$$

where because of the spherical constraint we must understand that the diagonal $Q_{aa} = 1$. As in the spherical Curie–Weiss model, introduction of an appropriate order parameter has separated the integral depending on the microscopic configurations from that due to the energy. The resulting integrals over $\mathbf{s}_a$ are N identical n-dimensional Gaussian integrals, one for each of the N dimensions, and performing those Gaussian integrals gives

$$\begin{aligned}\overline{\log Z_J(\beta)} &= \lim_{n\to 0}\frac{\partial}{\partial n}\int dQ\,d\tilde{Q}\, e^{\frac{1}{2}N\beta^2\sum_{ab}^{n} f(Q_{ab})+\frac{1}{2}N\sum_{ab}\tilde{Q}_{ab}Q_{ab}}\left(\sqrt{\frac{(2\pi)^n}{\det\tilde{Q}}}\right)^N \\ &= \lim_{n\to 0}\frac{\partial}{\partial n}\int dQ\,d\tilde{Q}\, e^{\frac{1}{2}N\beta^2\sum_{ab}^{n} f(Q_{ab})+\frac{1}{2}N\sum_{ab}\tilde{Q}_{ab}Q_{ab}-\frac{1}{2}N\log\det\tilde{Q}+\frac{1}{2}nN\log(2\pi)}\end{aligned}$$

We are clearly nearing a point where we can evaluate the remaining integrals by a saddle point approximation at large N. We begin with the matrix $\tilde{Q}$, which has a unique saddle point. The derivative of the effective active with respect to $\tilde{Q}$ is

$$0 = \frac{\partial}{\partial\tilde{Q}_{ab}}\left(\frac{1}{2}\sum_{cd}\tilde{Q}_{cd}Q_{cd} - \frac{1}{2}\log\det\tilde{Q}\right) = \frac{1}{2}Q_{ab} - \frac{1}{2}\tilde{Q}^{-1}_{ab} \tag{9}$$

This is a matrix equation that has the solution $\tilde{Q} = Q^{-1}$. We therefore have

$$\begin{aligned}\overline{\log Z_J(\beta)} &= \lim_{n\to 0}\frac{\partial}{\partial n}\int dQ\, e^{\frac{1}{2}N\beta^2\sum_{ab}^{n} f(Q_{ab})+\frac{1}{2}N\sum_{ab}Q^{-1}_{ab}Q_{ab}-\frac{1}{2}N\log\det Q^{-1}+\frac{1}{2}nN\log(2\pi)} \\ &= \lim_{n\to 0}\frac{\partial}{\partial n}\int dQ\, e^{\frac{1}{2}N\beta^2\sum_{ab}^{n} f(Q_{ab})+\frac{1}{2}N\log\det Q+\frac{1}{2}nN(1+\log(2\pi))}\end{aligned}$$

We have reduced the problem of the quenched free energy to a saddle point in the off-diagonal elements of the overlap matrix Q with effective action

$$\mathcal{S}_n(Q) = \frac{1}{n}\left[\frac{1}{2}\beta^2\sum_{ab}^{n} f(Q_{ab}) + \frac{1}{2}\log\det Q\right] + \frac{1}{2}(1+\log(2\pi)) \tag{10}$$

If we found a saddle point Q^*, we would find

$$\begin{aligned}\overline{\log Z_J(\beta)} &= \lim_{n\to 0}\frac{\partial}{\partial n}e^{nN\mathcal{S}_n(Q^*)} \\ &= N\lim_{n\to 0}(\mathcal{S}_n(Q^*) + n\tfrac{\partial}{\partial n}\mathcal{S}_n(Q^*))e^{nN\mathcal{S}_n(Q^*)} = N\lim_{n\to 0}\mathcal{S}_n(Q^*)\end{aligned} \tag{11}$$

This is a substantial challenge: we must find solutions to a nonlinear matrix optimization problem that can be rendered sensible in the limit of the size of the matrix going to zero.

Before we look at solutions to this problem, let's address the physical significance of Q, since this will help us reason about its form. It turns out to have a deep connection with the probability distribution of finding that two configurations drawn from the Boltzmann distribution of a given sample are separated by a given overlap. We can show this by calculating the moments of the distribution of overlaps. The mth moment is given by

$$\begin{aligned}\overline{\langle q^m\rangle} &= \overline{\int ds\, ds'\, p_J(s)\, p_J(s')}\left(\frac{s\cdot s'}{N}\right)^m \\ &= \overline{\int ds\, ds'\, \frac{1}{Z}e^{-\beta H_J(s)}\frac{1}{Z}e^{-\beta H_J(s')}}\left(\frac{s\cdot s'}{N}\right)^m\end{aligned}$$

We can treat the factors of $\frac{1}{Z}$ with replicas just as they can be used to treat the logarithm. Now we use the even more trivial identity $\lim_{n\to 0} Z^{n-2} = Z^{-2}$, and write

$$\overline{\langle q^m\rangle} = \lim_{n\to 0}\overline{\int\left(\prod_a^n ds_a\right)\, e^{-\beta\sum_a^n H_J(s_a)}\left(\frac{s_1\cdot s_2}{N}\right)^m}$$

where we have written $s = s_1$ and $s' = s_2$ for a total of n replicas. Following exactly the same steps we just saw, this reduces to

$$\overline{\langle q^m\rangle} = \lim_{n\to 0}\int dQ\, e^{N\mathcal{S}(Q)}Q_{12}^m$$

Since the effective action is invariant under permutations of the replica indices, we can equivalently write Q_{ab} for any $a \neq b$ in place of Q_{12}. Averaging over every possible option, this gives

$$\overline{\langle q^m\rangle} = \lim_{n\to 0}\int dQ\, e^{N\mathcal{S}(Q)}\frac{1}{n(n-1)}\sum_{a\neq b}Q_{ab}^m = \lim_{n\to 0}\frac{1}{n(n-1)}\sum_{a\neq b}(Q_{ab}^*)^m$$

The right hand side is the mth moment of the off-diagonal elements of the overlap matrix Q^* evaluated at the saddle point. Since for every moment m, the moment of the overlap q of two configurations in the Boltzmann distribution is equal to that of the off-diagonal of Q, we conclude that the two probability distributions are the same. Therefore, the distribution of off-diagonal elements of Q, $P(q)$, is equal to the probability distribution of overlaps from two configurations drawn from the Boltzmann distribution!

What does this look like in practice? Being in high dimensions strongly shapes the results. Consider the trivial case, where the energy is the same for every configuration (or infinite temperature), and every configuration is just as likely as any other. On the high-dimensional sphere, given some arbitrary axis, most of the volume is within a tiny vicinity of the equator. Therefore, drawing two random vectors, it is overwhelmingly likely that the are nearly orthogonal to each other. It therefore follows that

$$P(q) = \delta(q) \tag{12}$$

that is, the probability distribution of overlaps drawn at large N approaches a δ function at zero overlap. In fact, such a configuration corresponds to our annealed average in the previous lecture: when $P(q) = \delta(q)$, all off-diagonal elements of Q are zero and $Q = I$, and therefore $\det Q = 1$ and

$$\mathcal{S}(Q = I) = \frac{1}{2}\beta^2 f(1) + \frac{1}{2}(1 + \log(2\pi)) \tag{13}$$

exactly the annealed result. Therefore, in order to see a situation different from the annealed one, we must have some probability of seeing configurations at an overlap other than zero. In some sense this makes sense: in our models we will have a unique ground state, and therefore in the zero temperature limit the probability of overlaps drawn from the Boltzmann distribution must approach $P(q) = \delta(q - 1)$, e.g., all configurations are identical.

The next simplest possibility follows this line of thinking: that elements drawn from the Boltzmann distribution still only have one overlap with each other, but that this overlap is no longer zero. This guess is called *replica symmetry*, and corresponds to a matrix $Q_{ab} = \delta_{ab} + (1 - \delta_{ab})q$. In order to evaluate the free energy that would result from a replica symmetric ansatz, we need to know the eigenvalues of the matrix Q so we can compute its determinant. One eigenvector is $1 = [1, \dots, 1]$, with

$$Q1 = \sum_b (\delta_{ab} + (1 - \delta_{ab})q) = 1 + (n - 1)q \tag{14}$$

The other $n - 1$ eigenvectors are those orthogonal to 1, and are such that $\sum_b v_b = 0$. Therefore

$$Q\boldsymbol{v} = \sum_b (\delta_{ab} + (1 - \delta_{ab})q)v_b = (1 - q)v_a \tag{15}$$

We therefore have $\det Q = (1+(n-1)q)(1-q)^{n-1}$. Since the matrix has n elements 1 and $n^2 - n$ elements q, we have

$$\mathcal{S}(Q) = \frac{1}{2}\beta^2(f(1)+(n-1)f(q)) + \frac{1}{n}\left[\frac{1}{2}\log(1+(n-1)q) + \frac{1}{2}(n-1)\log(1-q)\right] + \frac{1}{2}(1+\log(2\pi)) \tag{16}$$

with the limit of $n \to 0$ giving

$$\lim_{n\to 0}\mathcal{S}(Q) = \frac{1}{2}\beta^2(f(1)-f(q)) + \frac{1}{2}\frac{q}{1-q} + \frac{1}{2}\log(1-q) + \frac{1}{2}(1+\log(2\pi)) \tag{17}$$

where we have used l'Hopital's rule to evaluate the limit involving the logarithms. Extremizing this with respect to q gives

$$0 = -\frac{1}{2}\beta^2 f'(q) + \frac{1}{2}\frac{q}{(1-q)^2} = -\frac{p}{4}\beta^2 q^{p-1} + \frac{1}{2}\frac{q}{(1-q)^2} \tag{18}$$

This equation gives a messy polynomial root for $p > 2$, and in fact the result is a discontinuous phase transition for all such p. We can see this by writing

$$q^{p-2}(1-q)^2 = \frac{2}{p}T^2 \tag{19}$$

and plotting the result. The formula on the left has zeros at $q = 0$ and $q = 1$ and a maximum for some intermediate q, whereas the right is a constant that decreases with decreasing T. Therefore for some $T = T_0$, the maximum coincides with the righthand side and we have a phase transition.

How does this solution do? For the $p = 2$ case, which we know does not behave like a spin glass but we understand that its ground state has energy density -1, this can be explicitly solved by

$$q^* = \frac{\beta - 1}{\beta} \tag{20}$$

a solution that is valid when $\beta > 1$, signaling a phase transition from zero to nonzero q^* at $\beta = 1$. The free energy under that condition would be

$$\begin{aligned} F &= -\frac{1}{\beta}\overline{\log Z_J(\beta)} = -N\frac{1}{\beta}\lim_{n\to 0}\mathcal{S}(Q^*) \\ &= -N\frac{1}{\beta}\left[\frac{1}{2}\beta^2(f(1)-f(q^*)) + \frac{1}{2}\frac{q^*}{1-q^*} + \frac{1}{2}\log(1-q^*) + \frac{1}{2}(1+\log(2\pi))\right] \\ &= -N\frac{1}{\beta}\left[-\frac{3}{4} + \beta - \frac{1}{2}\log\beta + \frac{1}{2}(1+\log(2\pi))\right] \end{aligned}$$

And therefore the energy density is

$$E = \frac{\partial}{\partial\beta}\beta F = N\left[\frac{1}{2}\frac{1}{\beta} - 1\right] \tag{21}$$

which indeed approaches $-N$ as $\beta \to \infty$.

It seems that the replica symmetric approach works for $p = 2$, but what about the richer models we actually care about? Next lecture we will look at the stability of this solution and see when it is not valid.