5 Stability of replica symmetry & replica symmetry breaking

Last lecture, we saw that for the spherical spin glasses the average logarithm of the partition function is given by

$$\overline{\log Z_J(\beta)} = N \lim_{n\to 0} \mathcal{S}(Q^*) \tag{1}$$

where the effective action $\mathcal{S}$ is a function of $n \times n$ matrices Q and is given by

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

and Q^* is a saddle point of $\mathcal{S}$ that maximizes the value of the average logarithm. We argued that in general, the probability distribution of off-diagonal elements of Q^* corresponds to the probability distribution of overlaps between pairs of configurations independently drawn from the Boltzmann distribution. We saw that one candidate for the saddle point is the so-called *replica symmetric* overlap matrix $Q_{ab} = \delta_{ab} + (1-\delta_{ab})q$, depending on the single overlap q.

$$\mathcal{S}(q) = \lim_{n\to 0}\mathcal{S}(Q) = \frac{1}{2}\beta^2(f(1)-f(q)) + \frac{1}{2}\frac{q}{1-q} + \frac{1}{2}\log(1-q) + \frac{1}{2}(1+\log(2\pi)) \tag{3}$$

This solution with $q^{p-2}(1-q)^2 = \frac{2}{p}T$ correctly reproduced the random matrix theory derived ground state energy of the $p=2$ spin model. It also produces a description of a phase transition in the $p>2$ spherical spin glosses. But are these descriptions correct?

A necessary but not sufficient condition for a saddle point solution to a problem to be correct is that it is stable. Usually this means that the point is a maximum of the real part of the effective action. But if we look at the (successful) 2-spin replica symmetric description, we see that the second derivative with respect to q at $q^* = (\beta-1)/\beta$ is

$$\frac{\partial^2\mathcal{S}}{\partial q^2} = -\frac{1}{2}f''(q^*) + \frac{1}{2}\frac{1}{(1-q^*)^2} + \frac{q^*}{(1-q^*)^3} = \beta^2(\beta-1) \tag{4}$$

which is *positive* for the whole replica symmetric phase of $\beta > 1$. This implies that the replica symmetric saddle point is a *minimum* of the action, not a maximum. What is going on here? It helps to look instead at the saddle point resulting from waiting to take the limit of $n \to 0$ until after the saddle point in N. For finite n, the effective action is

$$\mathcal{S}(Q) = \frac{1}{2}\beta^2(f(1)+(n-1)f(q)) + \frac{1}{n}\left[\frac{1}{2}\log(1+(n-1)q) + \frac{1}{2}(n-1)\log(1-q)\right] + \frac{1}{2}(1+\log(2\pi)) \tag{5}$$

The equivalent saddle point for nonzero n is

$$q = \frac{(n-2)\beta + \sqrt{4 + n(n\beta^2 - 4)}}{2(n-1)\beta} \tag{6}$$

while the second derivative evaluated at this saddle point is

$$\frac{\partial^2 \mathcal{S}}{\partial q^2} = -\frac{1}{4}\beta^2 \left(\sqrt{4 + n(n\beta^2 - 4)} + (n-2)\beta\right)\sqrt{4 + n(n\beta^2 - 4)} \tag{7}$$

The factor in parenthesis changes sign at $n = 1$ for all values of β, and therefore the stability flips at this point. We should be making this saddle point evaluation for the *matrix* Q, not its constituent parameters, which entails making it before the limit of n to zero. Therefore, we should look to the stability for $n > 1$, not the stability at $n = 0$. Fortunately, we don't really need to do this in practice: the flip of sign is extremely generic in replica calculations. The result is that we need to look for saddle points of the effective action that minimize it, not maximize it, when the parameter involved parameterizes a replica matrix.

The same analysis indicates that the saddle point is also a minimum with respect to variation of q for $p > 2$. But, we were less worried about this than whether the replica symmetric ansatz is unstable to other kinds of instabilities. To examine this, we need to look at the second derivative of $\mathcal{S}(Q)$ with respect to arbitrary matrices Q. This is given by

$$\begin{aligned}\frac{\partial^2 \mathcal{S}(Q)}{\partial Q_{ab}\partial Q_{cd}} &= \frac{1}{n}\frac{\partial}{\partial Q_{cd}}\left(\frac{1}{2}\beta^2 f'(Q_{ab}) + \frac{1}{2}Q^{-1}_{ab}\right) \\ &= \frac{1}{2}\frac{1}{n}\left(\delta_{ac}\delta_{bd}\beta^2 f''(Q_{ab}) - Q^{-1}_{ac}Q^{-1}_{bd}\right)\end{aligned} \tag{8}$$

Eigenvectors of this operator are constructed from matrices δQ whose rows and columns are eigenvectors of Q, *and* which are nonzero only in blocks where Q is constant. We can prove this term by term. For the first term, if δQ is constant only for a part of the matrix where the entries have value q, then $f''(Q)\odot\delta Q = f''(q)\delta Q$, making it an eigenmatrix with eigenvalue $f''(q)$. For the second term, the eigenvectors of Q are also eigenvectors of Q^{-1}, with eigenvalue λ^{-1}. Therefore, if the rows and columns of δQ are all eigenvectors of Q with the same eigenvalue, then

$$\begin{aligned}Q^{-1}\delta Q(Q^{-1})^T &= \begin{bmatrix} Q^{-1}\boldsymbol{v}_1 & \cdots & Q^{-1}\boldsymbol{v}_n\end{bmatrix}(Q^{-1})^{-1} \\ &= \begin{bmatrix}\lambda^{-1}\boldsymbol{v}_1 & \cdots & \lambda^{-1}\boldsymbol{v}_n\end{bmatrix}(Q^{-1})^T \\ &= \lambda^{-1}\begin{bmatrix}\boldsymbol{u}_1^T \\ \vdots \\ \boldsymbol{u}_n^T\end{bmatrix}(Q^{-1})^T \\ &= \lambda^{-1}\left(Q^{-1}\begin{bmatrix}\boldsymbol{u}_1 & \cdots & \boldsymbol{u}_n\end{bmatrix}\right)^T = \lambda^{-2}\delta Q\end{aligned}$$

which makes it an eigenvector with eigenvalue λ^{-2}. Therefore, for such a matrix δQ we would have an eigenvalue $\beta^2 f''(q) - \lambda^{-2}$, where λ is the eigenvalue associated with the given eigenvector and q is the element of Q which is nonzero on the support of δQ. One of the eigenvectors of a replica symmetric Q is the constant vector, to satisfy the second term we would need to construct the constant matrix, which does not satisfy the first term. In fact perturbations like the constant vector correspond precisely to changing the value of q, which is what we have already calculated the stability of. The other set of eigenvectors of Q are such that $\sum_i v_i = 0$. We can easily construct a matrix where every row and column is an eigenvector belonging to this subspace *and* the only nonzero elements correspond to elements with the same value in Q: for $n = 6$ we would have

$$\delta Q_0 = \begin{bmatrix} 0 & 1 & 1 & 1 & -1.5 & -1.5 \\ 1 & 0 & 1 & -1.5 & 1 & -1.5 \\ 1 & 1 & 0 & -1.5 & -1.5 & 1 \\ 1 & -1.5 & -1.5 & 0 & 1 & 1 \\ -1.5 & 1 & -1.5 & 1 & 0 & 1 \\ -1.5 & -1.5 & 1 & 1 & 1 & 0 \end{bmatrix} \tag{9}$$

We see that every row and column of δQ adds to zero, and so is an eigenvector of Q, and that the only nonzero elements correspond with the off-diagonal. Therefore this is a perturbation with eigenvalue

$$\lambda_R = \beta^2 f''(q) - \frac{1}{\lambda_0^2} = \beta^2 f''(q) - \frac{1}{(1-q)^2} \tag{10}$$

which is known in the literature as the *replicon eigenvalue*. For the replica symmetric saddle point of the 2-spin model, this eigenvalue is exactly zero, which means it does not necessarily indicate an instability. For general p, recall that solutions satisfy

$$0 = -\frac{1}{2}\beta^2 f'(q) + \frac{1}{2}\frac{q}{(1-q)^2} \tag{11}$$

or

$$\frac{1}{(1-q)^2} = \frac{\beta^2}{q} f'(q) \tag{12}$$

Substituting this into λ_R, we see an instability if

$$0 < \lambda_R = \beta^2 \left(f''(q) - \frac{1}{q} f'(q) \right) = \frac{1}{2}\beta^2 (p(p-1)q^{p-2} - pq^{p-2}) \tag{13}$$
$$= \frac{1}{2}\beta^2 p q^{p-2}(p-2)$$

Since the eigenvalue is positive for all $p > 2$, this represents an instability of the replica symmetric saddle point!

So, we need to consider more general forms of the matrix Q which break the symmetry among replicas. This is a very challenging problem, since there are very many ways to perturb a matrix. Early attempts considered splitting the replicas into two groups, where in each group there is a different mutual overlap and another overlap between groups. However, this was incorrect. The answer that works, found by Giorgio Parisi, is one that only slightly breaks the symmetry among individual replicas. The idea is that the statistics of no individual replica is different from any other. However, among *pairs* of replicas, you can find neighbors with different possible overlaps. For example, suppose you sample one state, and then look at its overlap with others. The others can take multiple possible values. However, if you sample another state, the statistics of its overlaps with others has the exact same form as that of the first state. The resulting matrix of overlaps is called *heirarchical*, because it implies a geometry of basins and subbasins organized in a fractal structure.

The form of the matrix implied by this has every row with the same set of values, but permuted. In the simple case where the set of overlap values is discrete, e.g., comes in $k+1$ possible values, one can visualize the matrix by a hierarchical block structure. Starting from the furthest from the diagonal, the width of the block is $m_0 = n$ and the overlap value inside it is q_0. Stepping inward, the kth interior block has width $m_k - m_{k-1}$ and overlap value q_0. Finally, there is the diagonal, with width $m_{k+1} = 1$. The number of elements in the matrix can be found simply by permuting all the rows so that they sorted: then the matrix is formed by rectangles, with one side n and for the kth block the other side $m_k - m_{k+1}$, so the number of elements in the kth block is $n(m_k - m_{k+1})$.

Recall that we showed the distribution of off-diagonal elements of Q corresponds to the probability distribution of overlaps. Therefore, in this k-RSB scenario, $P(q)$ consists of $k+1$ δ functions located at each of the q_i and with weight

$$\frac{n(m_i - m_{i+1)}}{n(n-1)} = \frac{m_i - m_{i+1}}{n-1} \tag{14}$$

This may seem to pose a problem: in the limit of $n \to 0$, this appears to result in negative probabilities if the size of the size of the blocks decreases as one moves towards the diagonal. However, the limit of $n \to 0$ must also influences the block sizes m, and in very unintuitive ways: just as in the limit $n \to 0$ with $m_i > 0$ means that the size of the block becomes larger than the size of the matrix containing it, *the whole hierarchy of block sizes inverts as well*, so that once $n \to 0$ is taken the 'block sizes' have the property $0 = n = m_0 < m_1 < \cdots < m_k < m_{k+1} = 1$. So, very strange, but so far self-consistent.

In order to analyze such a matrix in our model, we need to understand its determinant, and therefore its eigenvalue spectrum. For simplicity (and because it is going to be the final answer for the p-spin models), we stick only to 1RSB, which consists of just one interior block of size m_1 and value q_1. One eigenvector is the constant vector, which now has eigenvalue $1 + (m-1)q_1 + (n-m)q_0$, the sum over one of the rows. The next set of eigenvectors is orthogonal to this one, but constant inside one of the n/m interior blocks, and has the structure

$$[\underbrace{1,\ldots,1}_{m},\underbrace{-m/(n-m),\ldots,-m/(n-m)}_{n-m}] \tag{15}$$

We can check it is an eigenvector. There are two possibilities for the result: that from a row of Q with $a \leqslant m$, or that for a row with $a > m$. In the former case,

$$\begin{aligned}\sum_{b=1}^{m} Q_{ab} - \frac{m}{n-m}\sum_{b=m+1}^{n} Q_{ab} &= 1 + (m-1)q_1 - \frac{m}{n-m}(n-m)q_0 \\ &= 1 + (m-1)q_1 - mq_0\end{aligned} \tag{16}$$

while in the latter,

$$\begin{aligned}\sum_{b=1}^{m} Q_{ab} - \frac{m}{n-m}\sum_{b=m+1}^{n} Q_{ab} &= mq_0 - \frac{m}{n-m}[(n-2m)q_0 + (m-1)q_1 + 1] \\ &= -\frac{m}{n-m}[1 + (m-1)q_1 - mq_0]\end{aligned}$$

The eigenvector therefore has eigenvalue $\lambda_0 = 1 + (m-1)q_1 - mq_0$ and multiplicity $n/m - 1$ (because all n/m of them are linearly related to the constant vector). Finally, there is a set of eigenvectors that are orthogonal to both of these: one on the diagonal, negative and constant inside an interior block, and zero outside. These vectors have the form

$$[1,\underbrace{-1/(m-1),\ldots,-1/(m-1)}_{m-1},\underbrace{0,\ldots,0}_{n-m}] \tag{17}$$

with eigenvalue

$$\lambda_1 = \frac{1}{\nu_1}\sum_{b=1}^{n} Q_{1b}\nu_b = Q_{11} - \frac{1}{m-1}\sum_{a=2}^{m} Q_{1a} = 1 - q_1 \tag{18}$$

with multiplicity $n - n/m$, for the n places to put the diagonal, minus one per block to prevent linear combinations in each. Checking this is an eigenvector by testing the other two possibilities is left for an exercise. We therefore have

$$\sum_{ab}^{n} f(Q_{ab}) = nf(1) + n(m-1)f(q_1) + n(n-m)f(q_0) \tag{19}$$

$$\log\det Q = \log(1+(m-1)q_1+(n-m)q_0)+\left(\frac{n}{m}-1\right)\log(1+(m-1)q_1-mq_0)+n\left(1-\frac{1}{m}\right)\log(1-q_1) \tag{20}$$

In the limit of $n\to 0$, we have

$$\lim_{n\to 0}\frac{1}{n}\sum_{ab}^{n} f(Q_{ab}) = f(1)+(m-1)f(q_1)-mf(q_0) \tag{21}$$

$$\begin{aligned}\lim_{n\to 0}\frac{1}{n}\log\det Q &= \left(1-\frac{1}{m}\right)\log(1-q_1)+\frac{q_0}{1+(m-1)q_1-mq_0}+\frac{1}{m}\log(1+(m-1)q_1-mq_0)\\ &= \left(1-\frac{1}{m}\right)\log\lambda_1+\frac{q_0}{\lambda_0}+\frac{1}{m}\log(\lambda_0)\\ &= \log\lambda_1+\frac{q_0}{\lambda_0}+\frac{1}{m}\log\frac{\lambda_0}{\lambda_1}\end{aligned}$$

Before we get into examining saddle-point solutions of this, let's look at the stability conditions of those solutions. Recall that the stability is set by the operator

$$\frac{1}{2}\frac{1}{n}\left(\delta_{ac}\delta_{bd}\beta^2 f''(Q_{ab})-Q^{-1}_{ac}Q^{-1}_{bd}\right) \tag{22}$$

Repeating the argument we make before about the eigenmatrices of this operator, we can construct them from matrices whose rows and columns are all eigenvectors of Q with the same eigenvalue. This results in the stability conditions

$$0\geqslant\beta^2 f''(q_0)-\lambda_0^{-2}\qquad 0\geqslant\beta^2 f''(q_1)-\lambda_1^{-2} \tag{23}$$

The sum the two terms above gives the effective action in q_1, q_0, and m. First, consider the saddle-point condition for q_0, which is (using $\frac{\partial}{\partial q_0}\lambda_0=-m$)

$$0=\frac{\partial\mathcal{S}}{\partial q_0}=-m\beta^2 f'(q_0)+\frac{mq_0}{\lambda_0^2} \tag{24}$$

Combined with the stability condition, we find $0\geqslant\beta^2 f''(q_0)-\beta^2\frac{1}{q_0}f'(q_0)=\frac{1}{2}\beta^2(p-2)pq_0^{p-2}$, which always violated for $p\geqslant 2$ if $q_0>0$. Therefore, we must conclude $q_0=0$. Applying this, the rest of the effective action is

$$\mathcal{S}(m,q_1)=\frac{1}{2}\left[\beta^2(f(1)-(1-m)f(q_1))+\log(1-q_1)+\frac{1}{m}\log\frac{1-(1-m)q_1}{1-q_1}\right] \tag{25}$$

which gives conditions on the order parameters differentiating with respect to q_1 of

$$0=-\beta^2(1-m)f'(q_1)+\frac{(1-m)q_1}{(1-q_1)(1-(1-m)q_1)} \tag{26}$$

and with respect to m of

$$0 = \beta^2 f(q_1) + \frac{1}{m}\frac{q_1}{1-(1-m)q_1} - \frac{1}{m^2}\log\frac{1-(1-m)q_1}{1-q_1} \tag{27}$$

There are two interesting temperatures to look at: one where the 1RSB state with $m = 1$ becomes locally stable, and one where it becomes a valid thermodynamic state for $m < 1$. But when is it the stable state? For this, we need both equations to be satisfied. Once again looking for continuous transitions with $m = 1$,

$$0 = -\beta^2 f'(q_1) + \frac{q_1}{1-q_1} \qquad 0 = \beta^2 f(q_1) + q_1 + \log(1-q_1) \tag{28}$$

Dividing these equations gives

$$\frac{q_1}{p} = \frac{f(q_1)}{f'(q_1)} = -(q_1 + \log(1-q_1))\frac{1-q_1}{q_1} \tag{29}$$

or

$$-(1-q_1)(q_1+\log(1-q_1)) = \frac{1}{p}q_1^2 \tag{30}$$

The expression on the left is zero for $q_1 = 0$ and $q_1 = 1$ and positive otherwise, with zero derivative and curvature 1 at $q_1 = 0$. We can therefore graphically solve it for $p > 2$, where the left and right-hand curves have an intersection at nonzero q_1. Using again the equation for q_1 with $m - 1$ divided out, this implies

$$T_K = \sqrt{\frac{1-q_1}{q_1}f'(q_1)} = \sqrt{\frac{1}{2}p(1-q_1)q_1^{p-2}} \tag{31}$$

So, is the 1RSB solution stable when it arises? For our pure p-spin models, we just have to check that the temperature at which it becomes locally stable is less than the transition temperature. The conditions for q_1 being a saddle point and for the stability condition when $m = 1$ are

$$0 = -\beta^2 f'(q_1) + \frac{q_1}{1-q_1} \qquad 0 \geqslant \beta^2 f''(q_1) - \frac{1}{(1-q_1)^2} \tag{32}$$

Dividing these two equations gives a condition on q_1:

$$q_1(1-q_1) \leqslant \frac{f'(q_1)}{f''(q_1)}\frac{pq_1^{p-1}}{p(p-1)q_1^{p-2}} = \frac{1}{p-1}q_1 \tag{33}$$

which is solved by $q_1 \geqslant \frac{p-2}{p-1}$. The corresponding temperature is given by

$$T_d = \sqrt{\frac{1-q_1}{q_1}f'(q_1)} \leqslant \sqrt{\frac{p(p-2)^{p-2}}{2(p-1)^{p-1}}} \tag{34}$$

This is the temperature where a 1RSB state appears as a metastable state. On the table below, we see that for $p > 2$ this is always true. Therefore, the 1RSB solution is the correct description of the low temperature phase in these models!

p	2	3	4	5
q_d	0	0.5	0.667	0.75
T_d	1	0.612	0.544	0.513
q_K	0	0.645	0.805	0.871
T_K	1	0.586	0.503	0.461

How can we interpret this phase? Recall the correspondence we established in the previous lecture: the probability distribution of off-diagonal matrix elements corresponds to the probability distribution of overlaps between states drawn independently from the Boltzmann distribution. The 1RSB ansatz corresponds to two possible values of overlap: 0 or q_1. The picture one should have of the distribution of states in space is that under the transition temperature, states start belonging to one of many clusters. Two states drawn from different clusters will have overlap zero, while two states drawn from the same cluster will have overlap one.

Higher RSB corresponds to a similar picture, but with more levels of clustering: clusters can have subclusters, which themselves have subclusters. In these models, playing with $f(q)$ which is not homogeneous can result in systems described by arbitrary kinds of RSB, including so-called *full* RSB, where there is a continuous probability distribution of overlaps.

At the phase transition, we always have m decreasing continuously from $m = 1$. This indicates that pairs of clustered spins begin to emerge with an overlap q_1, but the proportion of those pairs relative to the proportion of spins that are uncorrelated grows continuously from zero at the phase transition.