Chapter 1

Disordered systems

There are systems in nature, science, technology, society, etc. in which some properties are heterogeneous and do not change quickly in time. Think, for example, of a binary mixture of beads (or atoms) with different sizes as those sketched in Fig. 1. In this ensemble, the formation of a crystal is prevented either by the slow dynamics or by the incommensurable sizes of the beads. Over the time scale of human beings, such system might retain its disordered *structure and its physical properties may depend on the presence of such disorder and even on its specific realization. Structural glasses are indeed formed in a way resembling the example of Fig. 1. By disorder, thus, we mean a form of randomness that varies slowly and that is on top of the stochastic variability to be taken into account in statistical physics.*

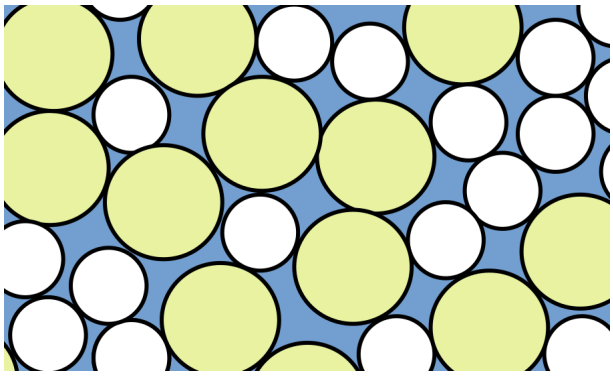

Figure 1. Sketch of a disordered binary mixture.

One of the standard theoretical models of condensed matter is the Ising model for ferromagnets, which is the paradigm of a system displaying a phase transition by varying the temperature. In this chapter, we use spin glass variants of this model to show how the free energy should be evaluated in systems with some quenched disorder. To emphasize the similarities and the differences between non-disordered and disordered systems, we start by recalling the solution of the mean field Ising model. With the Hopfield model we also discuss how spin glasses can be used to store *patterns, connecting this framework to computer applications. In fact, neural networks in machine learning are in a way systems evolving their disorder to learn a task. For this reason, the study of phase transitions in disordered systems sheds some light on the different regimes in which neural networks may happen to perform their tasks.*

1.1. A simple non-disordered system: the mean field Ising Model

For systems in equilibrium at a temperature T (and $\beta = 1/T$, with Boltzmann constant $k_B = 1$), average quantities in the canonical ensemble are obtained by running weighted sums over all possible configurations S,

$$\langle X \rangle = \frac{\sum_S X[S] e^{-\beta H[S]}}{\sum_S e^{-\beta H[S]}} \tag{1.1}$$

The denominator is the partition function

$$Z = \sum_S e^{-\beta H[S]} \tag{1.2}$$

and $H[S]$ is the Hamiltonian, or energy function of the system. The free energy is $F = -T \log Z$. If we have X coupled to a parameter α in $H' = H + \alpha X$, we see from the structure of (1.1) that $\langle X \rangle = \partial F / \partial \alpha|_{\alpha=0}$ derives from a suitable derivative of the free energy, which acts as a generating function. Hence, the knowledge of F allows to predict the mean behavior of the system. We are going to see that this picture is complicated by the presence of disorder.

A paradigm of interacting system is the Ising model, with spins $S_i \in \{-1, 1\}$ for $i = 1...N$ and Hamiltonian

$$H = -\frac{J}{N} \sum_{i \neq j} S_i S_j - h \sum_i S_i \,. \tag{1.3}$$

It is in a "mean-field" version if the sum runs over all possible pairs of $1 \leq i \leq N$ and $1 \leq j \leq N$ with $i \neq j$ (in the next section also the (i, i) pair will be included for simplifying the calculations; such constant energy shift is irrelevant thermodynamically). The system is ferromagnetic if $J > 0$ and hence $-J < 0$ favors the alignment of the spins. The external field h is the same for all spins.

We are interested in the thermodynamic limit $N \to \infty$. From the point of view of a given spin j, a myriad of other spins should contribute with an average effect due to the central limit theorem. The average magnetization

$$m = \frac{1}{N} \sum_i \langle S_i \rangle \tag{1.4}$$

should thus play a relevant role. Indeed, it enters in the calculation of the typical energetic contribution from j,

$$H_j = S_j \left[-\frac{2J}{N} \sum_i S_i - h \right] \tag{1.5}$$

$$\simeq S_j \left[-\frac{2J}{N} \sum_i \langle S_i \rangle - h \right] \qquad \text{(for large } N\text{)} \tag{1.6}$$

$$\equiv -h_m S_j \qquad \text{with average field} \quad h_m = 2Jm + h \tag{1.7}$$

(we would have had J and not $2J$ if pairs $i < j$ were considered).

The mean field approximation yields a non-interacting system of a single spin S_j in a field h_m. Its two possible states at inverse temperature $\beta = 1/T$ thus appear with probability given by Boltzmann weights normalized by the partition function Z,

$$P(S_j) = \frac{e^{-\beta H_j}}{Z} = \frac{e^{\beta h_m S_j}}{e^{\beta h_m} + e^{-\beta h_m}} \tag{1.8}$$

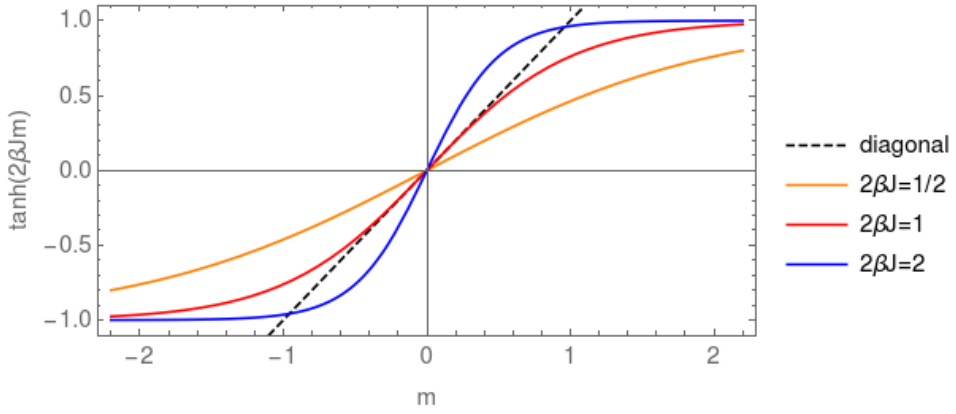

Figure 2. The graphical solution of (1.12) is obtained by finding the intersections of the $\tanh(2\beta Jm)$ function with the function m. The three curves are for low, critical, and large β.

There is still a self-consistency condition to impose on h_m and thus on the magnetization m:

$$m = \sum_{S_j=\pm 1} P(S_j)S_j \tag{1.9}$$

$$= \frac{e^{\beta h_m} - e^{\beta h_m}}{e^{\beta h_m} + e^{-\beta h_m}} \tag{1.10}$$

$$= \tanh(\beta h_m) \tag{1.11}$$

and by recalling what is h_m, the self-consistent equation becomes

$$m = \tanh(\beta 2Jm + \beta h) \tag{1.12}$$

Even for the simplest case $h = 0$ without external field, one should find the solution graphically as shown in Figure 2. At low β (high temperature) the function $\tanh(\beta 2Jm)$ is quite flat and stays below the function m, hence there is only one solution at $m = 0$. At a critical β_c, three solutions merge at $m = 0$ and they split above β_c into $m = -m^*, 0, +m^*$ because $\tanh(\beta Jm)$ is steep enough to cross the diagonal m three times. The critical β_c is found by requiring that the derivative of $\tanh(\beta 2Jm)$ is equal to 1 at $m = 0$, which yields $\beta_c = 1/(2J)$.

1.2. Random Field Ising Model (RFIM)

We aim at understanding what changes from the standard Ising model if the field h_i is now randomly assigned to each site i. To stress that this random field is a fixed feature of each system, we call it *quenched disorder*. We may expect that this disorder added to the thermal randomness is a crucial factor when much stronger than the total ferromagnetic coupling with the other spins. In this sense, we could have a paramagnetic phase also if the temperature is very low because the spins prefer to follow their own local h_i rather than the global trend given by the magnetization.

We thus study a Random Field Ising Model (RFIM): the ferromagnetic coupling between spins $S_i \in \{-1, 1\}$ (in a configuration denoted by $S = (S_1, \ldots, S_N)$) is as in the standard Ising model and the interaction is again not limited to nearest neighbor but runs over all pairs i, j,

$$H_h[S] = -\frac{J}{N}\sum_{i,j} S_i S_j - \sum_i h_i S_i \tag{1.13}$$

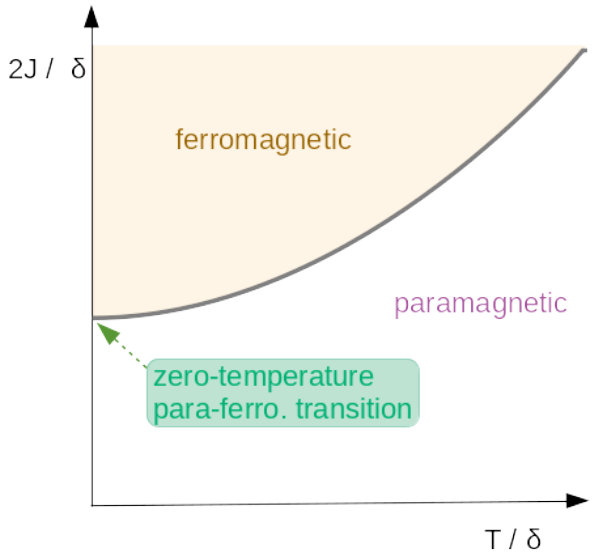

Figure 3. Phase diagram of the RFIM.

including the i, i interaction, for later convenience. Here, as a novelty, the disorder is realized by picking each local field h_i from the same Gaussian distribution with variance δ^2,

$$p(h_i) = \frac{1}{\sqrt{2\pi\delta^2}} e^{-h_i^2/2\delta^2} \qquad \forall i \leq N \tag{1.14}$$

and fixing it. Thus, every system is characterized by a given quenched disorder $h = \{h_i | i = 1, \ldots, N\}$ (of i.i.d. random variables) with full probability

$$p(h) = \prod_{i=1}^{N} p(h_i) \tag{1.15}$$

We would like to prove that the phase diagram of the RFIM is as that shown in Figure 3, where two phases (ferromagnetic with a macroscopic magnetization, and paramagnetic) appear. The diagram is as a function of the ratios T/δ and J/δ. The value T/δ quantifies how thermal energy is relevant with respect to the disorder. The value J/δ quantifies the relevance of the ferromagnetic coupling with respect to the disorder, and obviously the ferromagnetic phase appears where J/δ is sufficiently large, where "sufficiently" is quantified by the dense line separating the phases (note its monotonic increase with T: why?). The phase diagram also shows that, for given values of J and T, there is always a value of δ that can randomize the system enough to make it paramagnetic. Even for $T = 0$ we may see a para-ferromagnetic phase transition by varying δ.

The new issue is to find the typical behavior of a system by averaging its behavior over the realizations of the disorder. As discussed later, this centers around averaging the free energy

$$F_h = -T \log Z_h \tag{1.16}$$

over the disorder, rather than averaging the partition function

$$Z_h = \sum_S e^{-\beta H_h[s]} \tag{1.17}$$

An average over the disorder is denoted by an overline in the following. For example,

$$\overline{F} = -T\overline{\log Z_h} = -T\int \prod_i dh_i\, p(h) \log Z_h \tag{1.18}$$

The average of a nonlinear function as the log is problematic and it turns out to be simpler to average the n-th power $(Z_h)^n$. Because of this, it is useful to follow the *replica trick*. In its various forms, it reads

$$\overline{\log Z_h} = \lim_{n\to 0} \frac{\overline{Z^n} - 1}{n}\,, \tag{1.19}$$

$$\overline{\log Z_h} = \lim_{n\to 0} \frac{1}{n} \log \overline{Z^n}\,, \tag{1.20}$$

$$\overline{\log Z_h} = \frac{\partial}{\partial n} \overline{Z^n}\bigg|_{n=0} \tag{1.21}$$

This is a useful mathematical step that, however, comes at the price of performing a weird limit of $n \to 0$ replicas. Each replica is one out of n copies of the system, all sharing the same disorder h.

We label each replica by an index $a = 1, \dots, n$ let it understood that $Z = Z_h$ and $H_h[S] = H[S]$ both depend on a quenched disorder h. The n-th power of the partition function is then

$$\begin{aligned} \overline{Z^n} &= \overline{\sum_{\{S^a\}} \exp\left(\frac{\beta J}{N} \sum_a \sum_{ij} S_i^a S_j^a\right) \exp\left(\beta \sum_i \sum_a S_i^a h_i\right)} \\ &= \sum_{\{S^a\}} \exp\left(\frac{\beta J}{N} \sum_a \sum_{ij} S_i^a S_j^a\right) \underbrace{\overline{\exp\left(\beta \sum_i \sum_a S_i^a h_i\right)}}_{\equiv \overline{e^{\sum_i \lambda_i h_i}}} \end{aligned} \tag{1.22}$$

where $\sum_{\{S^a\}}$ means sum over all possible configurations of all replicas and

$$\lambda_i = \beta \sum_a S_i^a \tag{1.23}$$

Since the average over disorder is limited to the last term with h_i's, and since each term yields its own average via a Gauss integral,

$$\overline{e^{\lambda_i h_i}} = \int dh_i p(h_i) e^{\lambda_i h_i} = e^{\delta^2 \lambda_i^2/2}, \tag{1.24}$$

we may rewrite

$$\overline{Z^n} = \sum_{\{S^a\}} \exp\left[\frac{\beta J}{N} \sum_a \sum_{ij} S_i^a S_j^a + \frac{\beta^2 \delta^2}{2} \sum_i \left(\sum_a S_i^a\right)^2\right] \tag{1.25}$$

$$= \sum_{\{S^a\}} \exp\left[\frac{\beta J}{N} \sum_a \left(\sum_i S_i^a\right)^2 + \frac{\beta^2 \delta^2}{2} \sum_i \left(\sum_a S_i^a\right)^2\right] \tag{1.26}$$

where we noted that the first term in the exponential is just the square of $\sum_i S_i^a$. This is possible thanks to the choice of running the interactions also over the (i, i) pairs.

By inspecting the structure of (1.26) we note that we arrived at a system with interacting replicas! At the same time, the disorder has disappeared from the formulas. This trade of complications will finally lead to a solution of the RFIM.

Next we use the Hubbard-Stratonovich (HS) transformation,

$$e^{\frac{C}{2}z^2} = \frac{1}{\sqrt{2\pi C}} \int dx\, e^{-\frac{x^2}{2C} \pm zx} \tag{1.27}$$

(for negative exponent it becomes $e^{-\frac{C}{2}z^2} = \frac{1}{\sqrt{2\pi C}} \int dx e^{-\frac{x^2}{2C} \pm izx}$) which is useful for transforming squares in exponentials. Mathematically we might say that it corresponds to the inverse direction of solving a Gauss integral. In physical terms, this is translated to a replacement of interactions between degrees of freedom (z^2) by interactions with a mediating field x (the term zx) which follows a Gaussian statistics (x^2). The left-hand side of the HS formula can be seen in (1.26) if we identify

$$z_a = \sqrt{2J\beta} \sum_i S_i^a \tag{1.28}$$

$$C = \frac{1}{N} \tag{1.29}$$

$$e^{\frac{C}{2}z_a^2} = \frac{1}{\sqrt{2\pi C}} \int dx_a e^{-\frac{x_a^2}{2C} + z_a x_a} \tag{1.30}$$

By performing the HS transformation we get a version of $\overline{Z^n}$ in which spins S_i appear decoupled from the others,

$$\overline{Z^n} = \left(\frac{N}{2\pi}\right)^{n/2} \sum_{\{S^a\}} \int \prod_a dx_a \exp\Bigg[-\frac{N}{2} \sum_a x_a^2 + \underbrace{\sqrt{2J\beta} \sum_i \sum_a S_i^a x_a + \frac{\beta^2\delta^2}{2} \sum_i \left(\sum_a S_i^a\right)^2}_{\sum_i \ldots \text{ gives } N \text{ times the same object } \log Z_1} \Bigg] \tag{1.31}$$

$$= \left(\frac{N}{2\pi}\right)^{n/2} \int \prod_a dx_a \exp\left[N \left(-\frac{1}{2} \sum_a x_a^2 + \log Z_1(x_a) \right) \right] \tag{1.32}$$

with

$$Z_1(x_a) = \sum_{\{S^a = \pm 1\}} \exp\left(\sqrt{2\beta J} \sum_a x_a S^a + \frac{\beta^2\delta^2}{2} \left(\sum_a S^a \right)^2 \right) \tag{1.33}$$

where we set $S_i^a \to S^a$ due to the independence of Z_1 on the index i.

The exponent $\sim N$ in (1.32) shows that we can now use the saddle point approximation for large N. In doing this, we also assume that all replicas share the same $x_a = x$ (like in a replica symmetric solution), hence $\sum_a x_a = nx$, and $\sum_a x_a^2 = nx^2$. The saddle point, denoted by x_m, solves the equation

$$\frac{\partial}{\partial x} \left[-\frac{1}{2} n x^2 + \log Z_1(x) \right] = 0 \quad \to \quad nx = \frac{\partial}{\partial x} \log Z_1(x) \tag{1.34}$$

hence

$$n x_m = \sqrt{2\beta J} \frac{\sum_{\{S^a = \pm 1\}} \left(\sum_a S^a\right) e^{A[S, x_m]}}{\sum_{\{S^a = \pm 1\}} e^{A[S, x_m]}} \tag{1.35}$$

where

$$A[S, x] = \sqrt{2\beta J} x \sum_a S^a + \frac{\beta^2\delta^2}{2} \left(\sum_a S^a \right)^2 \tag{1.36}$$

The structure of (1.35) reveals that x_m is proportional to the average over the replicas of the spins, i.e. the magnetization m, in an ensemble where the Boltzmann weight e^A determines averages $\langle \dots \rangle$,

$$\frac{x_m}{\sqrt{2\beta J}} = \left\langle \frac{1}{n} \sum_a S^a \right\rangle \equiv m \tag{1.37}$$

We can thus rewrite everything by using $m = x_m/\sqrt{2\beta J}$:

$$\overline{Z^n} \propto e^{N\left[-n\beta J m^2 + \log Z_1(m)\right]} \tag{1.38}$$

$$Z_1(m) = \sum_{\{S^a=\pm 1\}} e^{A[S,m]} \tag{1.39}$$

$$A[S,m] = 2\beta J m \sum_a S^a + \frac{\beta^2\delta^2}{2} \left(\sum_a S^a \right)^2 \tag{1.40}$$

$$m = \frac{1}{Z_1(m)} \sum_{\{S^a=\pm 1\}} \left(\frac{1}{n} \sum_a S^a \right) e^{A[S,m]} \tag{1.41}$$

where $A[S,m]$ still couples the statistics of the replicas.

We recall that we are looking for a self-consistent equation for the magnetization, in analogy to the solution of the standard mean field Ising model. The square $(\sum_a S^a)^2$ that resisted so far in the exponent is removed by means of another HS transformation,

$$e^{A[S,m]} = \int \frac{d\nu}{\sqrt{2\pi}} e^{-\frac{1}{2}\nu^2 + (2\beta J m + \beta\delta\nu)\sum_a S^a} \tag{1.42}$$

This brings the advantage of decoupling the replicas. With this HS transformation, $Z_1(m)$ becomes

$$\begin{aligned} Z_1(m) &= \sum_{\{S^a=\pm 1\}} e^{A[S,m]} \\ &= \int \frac{d\nu}{\sqrt{2\pi}} e^{-\frac{1}{2}\nu^2} \prod_a \sum_{S^a=\pm 1} e^{(2\beta J m + \beta\delta\nu) S^a} \qquad (n \text{ decoupled replicas}) \\ &= \int \frac{d\nu}{\sqrt{2\pi}} e^{-\frac{1}{2}\nu^2} \left[2\cosh(2\beta J m + \beta\delta\nu)\right]^n \\ &= \int \frac{d\nu}{\sqrt{2\pi}} e^{-\frac{1}{2}\nu^2 + n\log[2\cosh(2\beta J m + \beta\delta\nu)]} \end{aligned} \tag{1.43}$$

We are now able to perform the limit $n \to 0$ dictated by the replica trick, for which $Z_1 \to 1$.

In analogy, we can prove that also the formula for m can be rewritten without any explicit reference to each replica but with just the number n of replicas appearing (exercise). It turns out that

$$m = \frac{1}{Z_1(m)} \int \frac{d\nu}{\sqrt{2\pi}} e^{-\frac{1}{2}\nu^2 + n\log[2\cosh(2\beta J m + \beta\delta\nu)]} \tanh(2\beta J m + \beta\delta\nu) \tag{1.44}$$

which, for $n \to 0$, gives

$$m = \int \frac{d\nu}{\sqrt{2\pi}} e^{-\frac{1}{2}\nu^2} \tanh\left(2\beta J m + \beta\delta\nu\right) \tag{1.45}$$

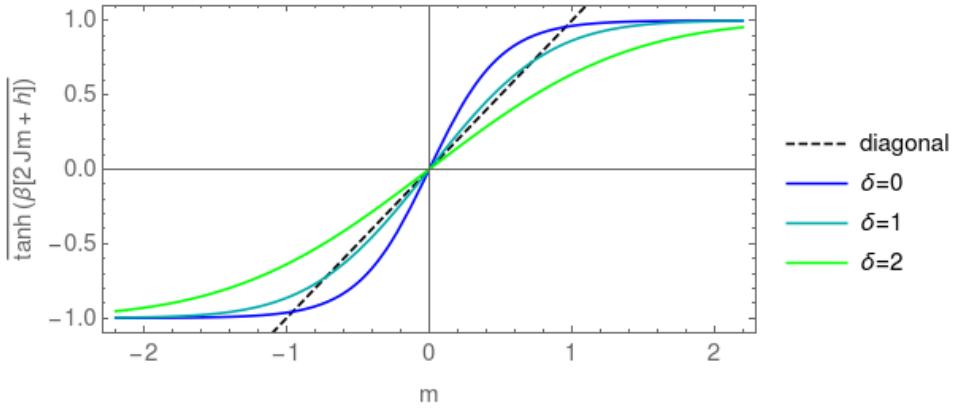

Figure 4. Plot of the self-consistent magnetization (1.46) vs m, for $\beta = J = 1$. The three curves are for different noise strengths: $\delta = 0$ (standard Ising), $\delta = 1$ and $\delta = 2$. Note that the system without disorder is ferromagnetic for these parameters but becomes paramagnetic at sufficiently high δ.

This further appearance of a Gaussian distribution for ν (unit variance) is welcome as one can convert it to a Gaussian distribution for $h = \delta\nu$ and translate the formula to an average over disorder [see (1.15)],

$$\begin{aligned} m &= \int \frac{dh}{\sqrt{2\pi\delta^2}} e^{-\frac{h^2}{2\delta^2}} \tanh\left(2\beta J m + \beta h\right) \\ &= \overline{\tanh\left(\beta\left(2Jm + h\right)\right)} \end{aligned} \tag{1.46}$$

This self-consistent equation for $m = m_{sc}(m)$ with $m_{sc}(m)$ given by the right-hand side of (1.46) is solved graphically, as shown for the Ising model. The critical line in the phase diagram of Figure 3 corresponds to the points where $\partial m_{sc}/\partial m = 1$, that the values of (T, δ, J) for which the curve is tangent to the diagonal line $m = m$. One can prove (exercise) that this condition turns into the equation

$$2\beta J \int dh\, p(h) \frac{1}{[\cosh(\beta h)]^2} = 1 \tag{1.47}$$

which can be recast in several forms; for example, by using reduced variables $J' = J/\delta$, $\beta' = \beta\delta$, $\tilde{h} = \beta h$ related to those in the axis of the phase diagram of Figure 3, we get

$$2\beta' J' \int \frac{d\tilde{h}}{\sqrt{2\pi}} e^{-\frac{\tilde{h}^2}{2\beta'^2}} \frac{1}{\left[\cosh \tilde{h}\right]^2} = 1 \tag{1.48}$$

Using this condition, one can show that even for zero temperature one has a para-ferromagnetic transition by varying the ratio $2J/\delta = 2J'$. The transition takes place (exercise) at $2J/\delta = \sqrt{\pi/2}$. Note that the self-consistent equation of the mean field Ising model is recovered from (1.46) for $\delta \to 0$. In Figure 4 there is an example of ferromagnetic phase disappearing by increasing δ.

The free energy averaged over the disorder in the end is

$$\begin{aligned}
\overline{F} &= -T\overline{\log Z} \\
&= -T\left.\frac{\partial}{\partial n}\overline{Z^n}\right|_{n=0} \\
&\simeq -T\frac{\partial}{\partial n}\left[e^{N(-n\beta Jm^2+\log Z_1)}\right]_{n=0} \\
&= -TN\left[-\beta Jm^2+\frac{\partial}{\partial n}\log Z_1\right]_{n=0} \\
&= N\left[Jm^2-\frac{T}{Z_1}\frac{\partial}{\partial n}Z_1\right]_{n=0} \\
&= N\left[Jm^2-T\int\frac{d\nu}{\sqrt{2\pi}}e^{-\frac{1}{2}\nu^2}\log[2\cosh(2\beta Jm+\beta\delta\nu)]\right] \\
&= N\left[Jm^2-T\int\frac{dh}{\sqrt{2\pi\delta^2}}e^{-\frac{h^2}{2\delta^2}}\log[2\cosh(\beta(2Jm+h))]\right] \qquad (1.49)
\end{aligned}$$

To wrap up, after deciding that the correct quantity to average is the free energy, one uses the replica trick to convert the computation to that of a system of interacting replicas without disorder. By some massage including two Hubbard-Stratonovich steps (to get rid of quadratic forms in the exponent till we get to the right quadratic form, i.e. the disorder average) and by the identification of some quantities with others having physical meaning (magnetization, average over disorder), one finds a self-consistent equation for the magnetization that represents the generalization to a system with quenched disorder of its version for the Ising model. This magnetization enters in the solution for the free energy. The phase diagram of the RFIM follows from these equations, with boundary between phases given by the points where the self-consistent function of the magnetization has derivative 1.

As a final point, let us highlight that the limits $\lim_{N\to\infty}\lim_{n\to 0}$ have been inverted in the above calculations, becoming $\lim_{n\to 0}\lim_{N\to\infty}$. Performing the thermodynamic limit before the limit to zero replicas is fine for the RFIM.

1.3. Neural networks and Hopfield model

The RFIM does not have a very interesting low temperature phase. In the ferromagnetic state it displays the two possible magnetizations and even in the paramagnetic case it stays around a single characteristic state which converges, for $T \to 0$, to a single state with spins fully aligned with local fields. This is not the case, in general, for disordered systems, which usually display a complex low temperature phase, or phases, with many basins of the free energy.

Our first example of disordered spin system with many nontrivial minima is a neural network where patterns are intentionally generated by an external agent, by encoding them in the coupling J_{ij} between neurons, which are biologically realized by axons. In this case each J_{ij} represents a *synaptic efficiency*, i.e. the kind of transmission of the axon from neuron j to neuron i. We will map neurons to spins and J_{ij} to their coupling, thus translating patterns into energetic minima.

Biologically, a neuron is activated when the incoming electrical signal overcomes a threshold. We define neuron states as

$$S_i = +1 \quad \text{(excited), and} \qquad S_i = -1 \quad \text{(at rest)} \qquad (1.50)$$

and a local field collecting all other impulses as

$$h_i = \sum_{j=1}^{N} J_{ij}(S_j + 1) \tag{1.51}$$

where also J's take ± 1 values,

$$\begin{aligned} J_{ij} &= +1 \quad \text{(excitatory synapse)} \\ J_{ij} &= -1 \quad \text{(inhibitory synapse)} \end{aligned} \tag{1.52}$$

These J's are quenched. The dynamical variables are the S_i, evolving by defining a new value of each S_i given the local excitation h_i determined by all values of other S_j's.

The dynamical rule, defined in discrete time $t = 1, 2, 3, \ldots$, is the following:

$$S_i(t+1) = \mathrm{sgn}(h_i(t) - \theta_i^*) \qquad \text{with} \qquad h_i(t) = \sum_{j=1}^{N} J_{ij}(S_j(t) + 1) \tag{1.53}$$

and local threshold θ_i^*. A simplifying hypothesis

$$\theta_i^* = \sum_{j=1}^{N} J_{ij} \tag{1.54}$$

leads to

$$S_i(t+1) = \mathrm{sgn}\left[h_i(t) - \sum_{j=1}^{N} J_{ij}\right] = \mathrm{sgn}\left[\sum_{j=1}^{N} J_{ij} S_j(t)\right] \tag{1.55}$$

We have all $J_{ij} \neq 0$ and thus the system is densely connected. However, we impose the $J_{ii} = 0$ (Hebb rule).

The purpose of this neural network is to store P patterns. By indexing the patterns with $\mu = 1, \ldots, P$, we have that each pattern

$$\vec{\xi}^{\mu} = \{\xi_1^{\mu}, \ldots, \xi_N^{\mu}\} \qquad \text{with} \qquad \xi_i^{\mu} = \{+1, -1\} \tag{1.56}$$

is essentially a spin configuration. We choose

$$\begin{aligned} J_{ii} &= 0 \quad \text{(Hebb rule)} \\ J_{ij} &= \frac{1}{N} \sum_{\mu=1}^{P} \xi_i^{\mu} \xi_j^{\mu} \end{aligned} \tag{1.57}$$

The second requirement defines the Hopfield model and introduces *memory* in the neural network: each pattern turns out to be a fixed point of the activation dynamics,

$$S_i(t) = \xi_i^{\mu} \quad \rightarrow \quad S_i(t+1) = \xi_i^{\mu} \tag{1.58}$$

This means that patterns are solutions of the equation

$$\xi_i^{\mu} = \mathrm{sgn}\left[\sum_{j=1}^{N} J_{ij} \xi_j^{\mu}\right] \tag{1.59}$$

This can be seen by first noting that the *scalar product* between two patterns,

$$\frac{1}{N} \sum_{j=1}^{N} \xi_j^{\mu} \xi_j^{\nu} \simeq \delta_{\mu\nu} + O(N^{-1/2}) \tag{1.60}$$

yields essentially zero for $\mu \neq \nu$ if the number of pattern is small compared to the system size, $P/N \to 0$ for $N \to \infty$ even if $P \gg 1$. This limitation for P is assumed hereafter. The

second term in the previous equation takes into account the random overlap of patterns. Hence,

$$\begin{aligned}
\operatorname{sgn}\left[\sum_{j=1}^{N} J_{ij}\xi_j^\mu\right] &= \qquad \text{(with the definition of } J_{ij}\text{, (1.57))}\\
\operatorname{sgn}\left[\sum_{j=1}^{N}\left(\frac{1}{N}\sum_{\nu=1}^{P}\xi_i^\nu\xi_j^\nu\right)\xi_j^\mu\right] &= \qquad \text{(rearranging)}\\
\operatorname{sgn}\left[\sum_{\nu=1}^{P}\xi_i^\nu\left(\frac{1}{N}\sum_{j=1}^{N}\xi_j^\nu\xi_j^\mu\right)\right] &= \qquad \text{(using (1.60))}\\
\operatorname{sgn}\left[\sum_{\nu=1}^{P}\xi_i^\nu\delta_{\mu\nu}\right] &=\\
\operatorname{sgn}\left[\xi_i^\mu\right] &= \xi_i^\mu
\end{aligned} \tag{1.61}$$

which proves (1.59).

What happens if we start from a configuration slightly different from a pattern? The dynamics is mapped to that of a disordered Ising model with energy

$$\begin{aligned}
E[\vec{S}] &= -\frac{1}{2}\sum_{i,j=1}^{N} J_{ij}S_iS_j\\
&= -\frac{1}{2}\sum_{i}^{N} S_ih_i
\end{aligned} \tag{1.62}$$

that is minimized by a configuration $\vec{S}$ aligned with its local field $\vec{h}$. At finite temperature $T = 1/\beta$ the probability of a configuration $\vec{S} = \{S_1, \ldots, S_N\}$ is

$$p(\vec{S}) = \frac{1}{Z}\exp\{-\beta E[\vec{S}]\} \tag{1.63}$$

where as usual Z stands for the partition function

$$Z = \sum_{\vec{S}}\exp\{-\beta E[\vec{S}]\} \tag{1.64}$$

A zero temperature dynamics, when updating $\vec{S}(t)$, leads always to a new $\vec{S}(t+1)$ with energy $E[\vec{S}(t+1)] \leq E[\vec{S}(t)]$. When starting from an initial $\vec{S}(0)$ not too different from a given pattern $\vec{\xi^\mu}$, the energy minimization brings to $S(t) = \vec{\xi_\mu}$ at some finite time t. The configurations $\vec{S}(0)$ falling back to such pattern are in its basin of attraction, or free energy local minimum.

As an application in computer science, one may think of the neural network as a storage of P images, and of $\vec{S}(0)$ as an image copied from a $\vec{\xi^\mu}$ but with corrupted pixels. The original image may be recognized among the others by energy minimization.

Next we show that the number of energy minima of an Hopfield model is essentially the number P of patterns. The free energy per spin is

$$f = -\frac{1}{N\beta}\log Z \tag{1.65}$$

The P patterns are stationary points of the landscape of the energy function E.

By using the approximation

$$J_{ij} = (1-\delta_{ij})\frac{1}{N}\sum_{\mu=1}^{P}\xi_i^\mu\xi_j^\mu \tag{1.66}$$

$$\simeq \frac{1}{N}\sum_{\mu=1}^{P}\xi_i^\mu\xi_j^\mu \tag{1.67}$$

which is a modification forgetting about Hebb's rule encoded in (1.66) and with little relevance $\sim N$ in a sum $\sim N^2$, we rewrite J_{ij} via (1.57) in the partition function,

$$Z = \sum_{\vec{S}} \exp\left\{\frac{\beta}{2N}\sum_{ij} S_iS_j \sum_{\mu=1}^{P}\xi_i^\mu\xi_j^\mu\right\} \tag{1.68}$$

$$= \sum_{\vec{S}} \exp\left\{\frac{\beta}{2N}\sum_{\mu=1}^{P}\left(\sum_i S_i\xi_i^\mu\right)^2\right\} \tag{1.69}$$

$$= \sum_{\vec{S}}\int\prod_\mu dq_\mu \exp\left\{-\frac{1}{2}N\beta\sum_{\mu=1}^{P}q_\mu^2 + \beta\sum_{\mu=1}^{P}q_\mu\left(\sum_i S_i\xi_i^\mu\right)\right\} \tag{1.70}$$

where in the last step we removed the square by a HS transformation, for which we are forgetting the prefactors (they would be irrelevant for our purposes). Now each of the terms with $1 \le i \le N$ is independent on the others. The last term in the previous exponential becomes

$$\sum_{S_i=\pm1}\exp\left\{\beta\left(\sum_{\mu=1}^{P}q_\mu\xi_i^\mu\right)S_i\right\} = \qquad (\text{defining } \vec{q}\cdot\vec{\xi}_i = \sum_{\mu=1}^{P}q_\mu\xi_i^\mu) \tag{1.71}$$

$$2\cosh\left(\beta\vec{q}\cdot\vec{\xi}_i\right) = \tag{1.72}$$

$$\exp\{\log[2\cosh\left(\beta\vec{q}\cdot\vec{\xi}_i\right)]\} \tag{1.73}$$

and the partition function turns into

$$Z = \int\prod_\mu dq_\mu \exp\{-N\beta u(\vec{q})\} \tag{1.74}$$

with

$$u(\vec{q}) = \frac{1}{2}\sum_{\mu=1}^{P}q_\mu^2 - \frac{1}{\beta N}\sum_{i=1}^{N}\log[2\cosh\left(\beta\vec{q}\cdot\vec{\xi}_i\right)] \tag{1.75}$$

to be expanded around its stationary point $\vec{q}^*$ (minimizing $u(q)$) for performing a saddle point approximation. In fact, the free energy coincides with the u function at the saddle point,

$$f = -\frac{1}{N\beta}\log Z = u(\vec{q}^*) \tag{1.76}$$

The stationary point is found by requiring

$$\left\{\frac{\partial u}{\partial q_1} = 0, \ldots, \frac{\partial u}{\partial q_P} = 0\right\}\Big|_{\vec{q}=\vec{q}^*} \tag{1.77}$$

For any component this means

$$\frac{\partial u}{\partial q_\mu} = q_\mu - \frac{1}{N\beta}\sum_{i=1}^{N}\frac{1}{\cosh\left(\beta\vec{q}\cdot\vec{\xi}_i\right)}\sinh\left(\beta\vec{q}\cdot\vec{\xi}_i\right)\beta\xi_i^\mu = 0 \tag{1.78}$$

leading to

$$q_\mu^* = q_\mu = \frac{1}{N}\sum_{i=1}^{N}\tanh\left(\beta\vec{q}\cdot\vec{\xi}_i\right)\xi_i^\mu \qquad \forall\mu = 1,\dots,P \tag{1.79}$$

(we are dropping the star from the notation). The sum over all neurons $i = 1,\dots,N$ for $N \to \infty$ becomes an average over the possible values of ξ,

$$\frac{1}{N}\sum_i f(\xi_i) \equiv \langle f\rangle = \int d\xi P(\xi) f(\xi) \tag{1.80}$$

$P(\xi)$, the probability density of ξ, is found by noting that $\xi_i = \pm 1$ with equal chance:

$$\int d\xi P(\xi)\dots = \frac{1}{2}\delta(\xi-1)\dots + \frac{1}{2}\delta(\xi+1)\dots \equiv \mathbb{E}(\dots) \tag{1.81}$$

The solution (1.79) is thus rewritten without the index i in the notation,

$$q_\mu = \mathbb{E}\left[\tanh\left(\beta\vec{q}\cdot\vec{\xi}\right)\xi^\mu\right] \qquad \forall\mu \tag{1.82}$$

To find a solution, we assume that $\vec{q} = (q, 0, \dots, 0)$ with $P-1$ null values and only one nontrivial q value, and later check if it works.

The physical meaning of q is better understood by stepping back to the sum over discrete variables (1.70) but with function u replaced by $\tilde{u}$

$$\tilde{u}(\vec{q}, S_1, \dots, S_N) = \frac{1}{2}\sum_{\mu=1}^{P} q_\mu^2 - \frac{1}{N}\sum_{\mu=1}^{P} q_\mu\left(\sum_{i=1}^{N}\xi_i^\mu S_i\right) \tag{1.83}$$

In this case, the saddle point solution $\vec{q}$ where $\partial u/\partial q_\mu = 0$ for all μ's yields

$$q_\mu = \frac{1}{N}\sum_{i=1}^{N}\xi_i^\mu S_i \tag{1.84}$$

which reveals that q_μ is the *overlap* of the spins with the μ-th pattern of the neural network. This explains that $\vec{q} = (q_1, 0, \dots, 0)$ for $\mu = 1$ makes sense, and similarly for all other patterns $\nu \neq 1$ we should set $\vec{q} = (0, \dots, 0, q_\nu, 0, \dots, 0)$.

Focusing on the first pattern, from the above structure for q we get P equations,

$$q_1 = \mathbb{E}\left[\tanh\left(\beta q_1\xi^1\right)\xi^1\right] \qquad (1 \text{ equation}) \tag{1.85}$$

$$q_\nu = \mathbb{E}\left[\underbrace{\tanh\left(\beta q_1\xi^1\right)}_{\text{odd}}\underbrace{\xi^\nu}_{\text{odd}}\right] \qquad (P-1 \text{ equations for } \nu > 1) \tag{1.86}$$

Each of the $P-1$ equations (1.86) is an expectation of independent odd terms with an even probability distribution for ξ's, hence for parity it gives $q_\nu = 0$, consistently with the hypothesis that $\vec{q}$ contains only one nonzero element. This is given by (1.85), which explicitly is

$$\begin{aligned} q_1 &= \frac{1}{2}\tanh(\beta q_1) - \frac{1}{2}\tanh(\beta q_1)\times(-1) \\ &= \tanh(\beta q_1) \end{aligned} \tag{1.87}$$

Since this is true for all patterns, in summary the only relevant saddle point equation for the Hopfield model is

$$q = \tanh(\beta q) \tag{1.88}$$

which resembles that of the mean field Ising model, with the overlap q replacing the magnetization m. In a sense, the overlap is a generalization of the concept of magnetization, to systems where a global magnetization does not appear, but where the overlap still quantifies some local order emerging within the state space.

Similarly to what seen for the Ising model, from the saddle point equation (1.88) we learn that there is a (second order) phase transition, here at $\beta_c = 1$ where $\tanh(q)$ is tangent to the $q = q$ line. Of the two nonzero solutions $\pm\hat{q}$ appearing at $\beta > \beta_c$, the one at $-\hat{q}$ does not tells us anything more than that at $+\hat{q}$ because fully anticorrelated and fully correlated have the same information content. By going to $T \to 0$ we are left with P ground states, one for every pattern.

1.4. Sherrington-Kirkpatrick (SK) model

Here we discuss the Sherrington-Kirkpatrick (SK) model, a central model in the field of disordered systems (perhaps the *"Ising model of disordered systems"*), allowing us to introduce some ideas and techniques central in this field. In particular we will explain the reason why it is possible to compute *exactly* the free-energy of this model and how to cast this calculation into a variational problem. By doing this we will introduce the characteristic order parameter of disordered systems: the *overlap matrix* $q_{\alpha\beta} = \frac{1}{N}\sum_{i=1}^{N} S_i^\alpha S_i^\beta$. While a complete account of the (quite technical) intermediate steps of the free-energy calculation can be found in [*"Spin Glass Theory and Beyond"*, Marc Mézard, Giorgio Parisi and Miguel A. Virasoro, Wiley] and [*"Statistical Physics of Spin Glasses and Information Processing"*, Hidetoshi Nishimori, Oxford University Press], the present lesson will be concluded by an informal physical explanation of how the *hierarchical* ansatz for the matrix $q_{\alpha\beta}$ captures the hierarchical nature of the free-energy landscape (minima nested into minima) typical of the spin-glass low-temperature phase. The presence of a multi-valley free-energy landscape is indeed the characteristic feature of systems with *quenched disorder* as the SK model. While the hierarchical structure of the landscape of the SK model is represented by the highly non trivial *full* replica-symmetry-breaking ansatz (full-RSB), the more simple *one-step* replica-symmetry-breaking ansatz will be discussed in the next lecture on the *p-spin* model.

1.4.1. Derivation of saddle-point equations. We consider a magnetic system where each spin can take the values $S_i = \{+1, -1\}$, as for the Ising and Hopfield model. Again, the Hamiltonian

$$H[\vec{S}] = -\sum_{i<j} J_{ij} S_i S_j \tag{1.89}$$

contains all possible interactions between spins and is thus of fully long-range kind.

Now the couplings are quenched but not to store patterns, they are chosen at random from a Gaussian distribution

$$P(J_{ij}) = \sqrt{\frac{N}{2\pi\delta^2}} \exp\left\{-\frac{N}{2\delta^2}(J_{ij})^2\right\} \tag{1.90}$$

with zero mean (thus, no ferromagnetism is present) and variance $\langle J^2\rangle = \delta^2/N$ with constant δ. This choice leads to a free energy

$$F_J = -\frac{1}{\beta}\log Z_J \sim N \tag{1.91}$$

that is extensive for every realization of the disorder J. We can fix $\delta = 1$ and check this with a high-temperature expansion (small β expansion) of Z_J:

$$\begin{aligned} Z_J &= \sum_{\vec{S}} \exp\left\{\beta \sum_{i<j} J_{ij} S_i S_j\right\} \\ &\simeq \sum_{\vec{S}} \left[1 + \beta \sum_{i<j} J_{ij} S_i S_j\right] + \sum_{\vec{S}\times\vec{S}} \frac{\beta^2}{2} \sum_{i<j} \sum_{k<l} J_{ij} J_{kl} S_i S_j S_k S_l \end{aligned} \tag{1.92}$$

where the sum $\sum_{\vec{S}} 1 = 2^N$ because it runs over all possible 2^N spin states. The sums with spins are zero whenever a spin, say S_i, appears an odd number of times in the summand because every contribution with $S_i = +1$ is canceled by another one with $S_i = -1$. This is the case for the sum $\sum_{i<j} J_{ij} S_i S_j$. For the same reason, in the β^2 sum there survives a nonzero term when $i = k$ and $j = l$, so that

$$Z_J \simeq 2^N \left[1 + \frac{\beta^2}{2} \sum_{i<j} J_{ij}^2\right] \tag{1.93}$$

whose logarithm is

$$\begin{aligned} \log Z_J &\simeq N \log 2 + \log\left[1 + \frac{\beta^2}{2} \sum_{i<j} J_{ij}^2\right] \\ &\simeq N \log 2 + \underbrace{\frac{\beta^2}{2} \sum_{i<j} J_{ij}^2}_{\text{must be of order } N} \end{aligned} \tag{1.94}$$

In the last term there appear the variance of the disorder, $\sim N^2$ times. Imposing extensivity, $N^2\langle J^2\rangle \sim N$, we see that it must be $\langle J^2\rangle \sim N^{-1}$. This is common to many disordered systems with long-range couplings.

The goal of our calculation is to compute the free energy in the thermodynamic limit

$$F_J = \lim_{N\to\infty} -\frac{1}{\beta}\log Z_J. \tag{1.95}$$

This task might look at a first glance as a hopeless one since the partition function Z_J depends on quenched random variables. How is it possible to compute F_J while keeping the random couplings J_{ij} fixed in the Hamiltonian? And even if it was possible, how general will be the result? I.e., will it depend on the specific disorder instance, namely the choice of the coefficients $\{J_{ij}\}_{i<j}$, or will be *universal* with respect to any choice of the coefficients? The solution to all this problems/questions comes in the form of the *self-averaging* property of the free energy with respect to the disorder instance. The question on how relevant is the dependence of F_J on $\{J_{ij}\}_{i<j}$ is indeed a well posed one: considered as a function of random couplings, F_J is itself a fluctuating random variable. Two different choices of the random coupling sets, which in practice correspond to *experiments* on two

different samples, let us label them as $J^{(0)} = \{J_{ij}^{(0)}\}_{i<j}$ and $J^{(1)} = \{J_{ij}^{(1)}\}_{i<j}$, yields in general two different values of the free-energy, i.e.,

$$F_{J^{(0)}} \neq F_{J^{(1)}}. \tag{1.96}$$

The self-averaging property of the free energy tells us precisely that the difference between these two values is irrelevant in the thermodynamic limit, i.e., the standard deviation of F_J is negligible with respect to its average value. By denoting the average of F_J with respect to the probability distribution of random couplings as

$$\overline{F_J} = \int_{-\infty}^{\infty} \prod_{i<j} dJ_{ij}\ P(J_{ij})\ F_J, \tag{1.97}$$

and the standard deviation as

$$\sigma(F_J) = \sqrt{\overline{(F_J^2)} - (\overline{F_J})^2} \tag{1.98}$$

we have that the self-averaging property reads in formula as

$$\lim_{N\to\infty} \frac{\sigma(F_J)}{\overline{F_J}} \sim \frac{1}{\sqrt{N}} \tag{1.99}$$

The self-averaging property of the free energy is thus crucial: in practice it tells us that in the thermodynamic limit, *up to negligibly small fluctuations*, any choice of the disordered couplings yields the same result. From the point of view of our calculation it means that in the thermodynamic limit we can *average* over the disorder instances:

$$\lim_{N\to\infty} -\frac{1}{N\beta} \log Z_J = \lim_{N\to\infty} -\frac{1}{N\beta} \overline{\log Z_J} \equiv f, \tag{1.100}$$

where $f_J = F_J/N$ is the free energy per spin.

As for the RFIM, here it is useful to use the replica trick, here in the version based on the math identity $\log x = \lim_{n\to 0}(x^n - 1)/n$. Hence, we focus on $\overline{Z_J^n}$ rather than $\overline{\log Z_J}$. For n replicas indexed by α or β (with n remaining integer till the very last step of the calculation), we have

$$\begin{aligned} \overline{Z_J^n} &= \int_{-\infty}^{\infty} \prod_{i<j} dJ_{ij} P(J_{ij}) (Z_J)^n \\ &= \int_{-\infty}^{\infty} \prod_{i<j} dJ_{ij} P(J_{ij}) \sum_{\{S_1^\alpha,\dots,S_N^\alpha | \alpha=1,\dots,n\}} \exp\left\{ \beta \sum_{\alpha=1}^{n} \sum_{i<j} J_{ij} S_i^\alpha S_j^\alpha \right\} \end{aligned} \tag{1.101}$$

This formula contains a nice linear contribution of random J_{ij}'s. In (1.101) replicas share the same disorder but are uncoupled (no α and β together) and spins are coupled.

For a single J_{ij} with an explicit version of $P(J)$, the resulting Gaussian integral yields

$$\int_{-\infty}^{\infty} dJ_{ij} \exp\left\{ -\frac{N}{2\delta^2}(J_{ij})^2 + \beta J_{ij} \sum_{\alpha=1}^{n} S_i^\alpha S_j^\alpha \right\} = \exp\left\{ \frac{\beta^2\delta^2}{2N} \sum_{\alpha,\beta=1}^{n} S_i^\alpha S_i^\beta S_j^\alpha S_j^\beta \right\} \tag{1.102}$$

(β as an index of a replica should be not confused with the inverse temperature). Note that this average over disorder is coupling replicas. Moreover, it is decoupling spins because

$\sum_{i<j} \approx \frac{1}{2}\sum_{i,j}$ and back to the full (1.101) this gives

$$\begin{aligned}\overline{Z_J^n} &= \sum_{\{S_1^\alpha,\dots,S_N^\alpha|\alpha=1,\dots,n\}} \exp\left\{\frac{\beta^2\delta^2}{2N}\sum_{\alpha,\beta=1}^{n}\sum_{i<j}^{N} S_i^\alpha S_i^\beta S_j^\alpha S_j^\beta\right\} \\ &= \exp\left\{\frac{nN\beta^2\delta^2}{4}\right\} \sum_{\{S_1^\alpha,\dots,S_N^\alpha|\alpha=1,\dots,n\}} \exp\left\{\frac{\beta^2\delta^2}{2N}\sum_{\alpha<\beta}^{n}\left(\sum_{i=1}^{N} S_i^\alpha S_i^\beta\right)^2\right\},\end{aligned} \tag{1.103}$$

which contains no J's and displays only single site spins S_i. In order to obtain the second line from the first one of Eq. (1.103) we played a little bit with the summation indices and made an approximation which is valid only in the large N limit. For the double summation over (latin) spin indices we used

$$\begin{aligned}\sum_{i<j}^{N} S_i^\alpha S_i^\beta S_j^\alpha S_j^\beta &= \frac{1}{2}\left(\sum_{i,j=1}^{N} S_i^\alpha S_i^\beta S_j^\alpha S_j^\beta - N\right) \\ &\cong \frac{1}{2}\left(\sum_{i=1}^{N} S_i^\alpha S_i^\beta\right)\left(\sum_{j=1}^{N} S_j^\alpha S_j^\beta\right) = \frac{1}{2}\left(\sum_{i=1}^{N} S_i^\alpha S_i^\beta\right)^2,\end{aligned} \tag{1.104}$$

where we have dropped the term N because it is subleading with respect to the double sum, which contains $\mathcal{O}(N^2)$ terms. We can thus write:

$$\begin{aligned}\sum_{\alpha,\beta=1}^{n}\sum_{i<j}^{N} S_i^\alpha S_i^\beta S_j^\alpha S_j^\beta &\cong \sum_{\alpha,\beta=1}^{n}\frac{1}{2}\left(\sum_{i=1}^{N} S_i^\alpha S_i^\beta\right)^2 \\ &= \sum_{\alpha=1}^{n}\frac{1}{2}\left(\sum_{i=1}^{N} S_i^\alpha S_i^\alpha\right)^2 + \frac{1}{2}\sum_{\alpha\neq\beta}^{n}\left(\sum_{i=1}^{N} S_i^\alpha S_i^\beta\right)^2 \\ &= \frac{nN}{2} + \sum_{\alpha<\beta}^{n}\left(\sum_{i=1}^{N} S_i^\alpha S_i^\beta\right)^2,\end{aligned} \tag{1.105}$$

which is what needed to go from the first to the second line of Eq. (1.103). Note that the argument used to neglect the diagonal elements of the double summation $\sum_{ij}$, i.e. that they are N terms with respect to an overall amount of $\mathcal{O}(N^2)$ terms, cannot be applied to the double summation over replica indices $\sum_{\alpha\beta}$, where the diagonal elements are n and the off-diagonal ones are $n(n-1)$, for the simple reason that we are going to take the limit $n \to 0$ and not the limit $n \to \infty$, as we did for number of spins in the systems: in the case of replica indices the diagonal elements do not represent a subleading contribution.

Let us spend few words of comment on the last two equations: it is only thanks to the fact that in the original Hamiltonian the summation $\sum_{i<j}$ goes over *all independent pairs of indices*, i.e., over $\mathcal{O}(N^2)$ terms, that we could write the double summation typical of any model with a two-body interaction as a single summation squared. Models where this kind of simplification is possible are usually known as *mean-field* models, i.e., models where, due to the large number of mutual interactions between the degrees of freedom, one is able, after a certain number of manipulations, to write the free-energy of the systems in terms of an order parameter which carries no spatial index (the Latin index i in the present case). This means that in mean-field models any information and/or dependence

on spatial dimensions is lost. Such models yield thus only an *approximate description* of real physical model—try to figure out for a moment the difference between our SK model and a disordered Ising model on a 2D or a 3D lattice—but enjoy a very nice feature: they can be solved exactly. The possibility to unfold the double summation as shown above is thus crucial because this allows us to introduce a *global order parameter* of the kind

$$q_{\alpha\beta} = \frac{1}{N}\sum_{i=1}^{N} S_i^\alpha S_i^\beta. \tag{1.106}$$

We will comment further on the meaning of $q_{\alpha\beta}$ in what follows, here let us just note that if we plug the definition $\sum_{i=1}^{N} S_i^\alpha S_i^\beta = N q_{\alpha\beta}$ into (1.103) we get for the exponential term of the partition function

$$\exp\left\{\frac{N\beta^2\delta^2}{2}\sum_{\alpha<\beta}^{n} q_{\alpha\beta}\right\}, \tag{1.107}$$

which puts in evidence that the argument of the exponential is an extensive function, i.e., it is proportional to the number of spin N in the system, and that it depends only on a global order parameter, $q_{\alpha\beta}$. Nevertheless, this naive replacement of $\sum_{i=1}^{N} S_i^\alpha S_i^\beta$ with $q_{\alpha\beta}$ is mathematically wrong, because the spins S_i^α are summation variables inside the partition function and we must do a proper change of variables, like in an integral. The correct way to introduce the order parameter $q_{\alpha\beta}$ is by means of the Hubbard-Stratonovich transformation that we have already met discussing the Random-Field Ising Model and the Hopfield model. By doing this, the quantity $q_{\alpha\beta}$ appears simply as an auxiliary variable to linearize the squared sum $(\sum_{i=1}^{N} S_i^\alpha S_i^\beta)^2$; we will demonstrate later that, in the large-N limit, the definition of $q_{\alpha\beta}$ is really the one given in (1.106).

In the following we adopt the shorthand notation

$$\sum_{\{S_1^\alpha,\dots,S_N^\alpha|\alpha=1,\dots,n\}} \to \sum_{\S}. \tag{1.108}$$

AS we have said, the square in (1.103) is thus unfolded by a HS transformation,

$$\overline{Z_J^n} \approx \sum_{\S}\int_{-\infty}^{\infty}\prod_{\alpha<\beta} dq_{\alpha\beta}\underbrace{\exp\left\{-\frac{N\beta^2\delta^2}{2}\sum_{\alpha<\beta} q_{\alpha\beta}^2\right\}}_{[1]}\underbrace{\exp\left\{\beta^2\delta^2\sum_{\alpha<\beta} q_{\alpha\beta}\sum_{i=1}^{N} S_i^\alpha S_i^\beta\right\}}_{[2]} \tag{1.109}$$

At this step, It is worth noting that in the first line of (1.103) we had a double summation over replica indices of the kind $\sum_{\alpha,\beta=1}^{n}$ which we replaced with $2\sum_{\alpha<\beta}^{n}$, to that only independent couple of indices where effectively considered. Although we still have to prove the identity in (1.106), from it we can argue quite naturally that the matrix $q_{\alpha\beta}$ is symmetric, i.e. $q_{\alpha\beta} = q_{\beta\alpha}$, so that only half of its off-diagonal elements are independent integration variables. The appearance of terms linear in the index i allows to factorize

the trace operator. Focusing on the second term [2],

$$\begin{aligned}
&\sum_{\S} \exp\left\{\beta^2\delta^2 \sum_{\alpha<\beta} q_{\alpha\beta} \underbrace{\sum_{i=1}^{N}\left(S_i^\alpha S_i^\beta\right)}_{\text{identical for all } i}\right\} \\
&= \prod_{i=1}^{N}\left[\sum_{\S} \exp\left\{\beta^2\delta^2 \sum_{\alpha<\beta} q_{\alpha\beta}\left(S_i^\alpha S_i^\beta\right)\right\}\right] \\
&= \left[\sum_{\S} \exp\left\{\beta^2\delta^2 \sum_{\alpha<\beta} q_{\alpha\beta} S^\alpha S^\beta\right\}\right]^N \\
&\equiv \left[\sum_{\S} \exp\left\{L(q_{\alpha\beta})\right\}\right]^N \\
&= \left[\operatorname{Tr} e^{L(q_{\alpha\beta})}\right]^N \\
&= \exp\left\{N \log\left[\operatorname{Tr} e^{L(q_{\alpha\beta})}\right]\right\} \qquad (1.110)
\end{aligned}$$

with operator $L(q_{\alpha\beta}) = \beta^2\delta^2 \sum_{\alpha<\beta} q_{\alpha\beta} S^\alpha S^\beta$. The form of (1.110) is ready for the not-so-unexpected saddle point approximation, as is the full formula for the partition function,

$$\begin{aligned}
\overline{Z_J^n} &= \exp\left\{nN\beta^2\delta^2/4\right\} \int_{-\infty}^{\infty} \prod_{\alpha<\beta} dq_{\alpha\beta} \\
&\qquad \exp\left\{-\frac{N\beta^2\delta^2}{2} \sum_{\alpha<\beta} q_{\alpha\beta}^2 + N \log\left[\operatorname{Tr} e^{L(q_{\alpha\beta})}\right]\right\} \\
&= \int_{-\infty}^{\infty} \prod_{\alpha<\beta} dq_{\alpha\beta} e^{-nNA[q_{\alpha\beta}]} \qquad (1.111)
\end{aligned}$$

with the function $A[q_{\alpha\beta}]$ defined as

$$A[q_{\alpha\beta}] = \underbrace{-\frac{\beta^2\sigma^2}{4} + \frac{\beta^2\delta^2}{2n} \sum_{\alpha<\beta} q_{\alpha\beta}^2}_{\text{energetic}} \; \underbrace{-\frac{1}{n} \log\left[\operatorname{Tr} e^{L(q_{\alpha\beta})}\right]}_{\text{entropic}}, \qquad (1.112)$$

where we have emphasized which are the *energetic* and the *entropic* contributions to $A[q_{\alpha\beta}]$. Note the $1/n$ factor embedded in A.

At this point we are not at all done, but we can say that the remaining part of the calculation is only technical: we must find the correct ansatz for the structure of the matrix $q_{\alpha\beta}$ and take the limit $n \to 0$ in the appropriate way. This last part of the work is usually the subject of graduate courses and can be found in PhD summer schools lecture-notes like [*"Replica Theory and Spin Glasses"*, F. Morone, F. Caltagirone, E. Harrison, G. Parisi, *arXiv:1409.2722*]. The purpose of this lecture was just to show that the calculation of the SK free-energy can be cast as a variational problem where the variational parameter, which is, not by chance, the order parameter of the systems, is a matrix. Indeed what we are left with in (1.111) is just the integration over the matrix elements $q_{\alpha\beta}$, of which, due to the overall prefactor N, we can get rid with a saddle-point approximation. That

is, we write

$$\overline{Z_J^n} = \int_{-\infty}^{\infty} \prod_{\alpha<\beta} dq_{\alpha\beta} e^{-nNA[q_{\alpha\beta}]} \approx \exp\left\{-nNA[q^*_{\alpha\beta}]\right\}, \tag{1.113}$$

where the matrix elements $q^*_{\alpha\beta}$ are the solution of the saddle-point equations

$$\frac{\partial A}{\partial q_{\alpha\beta}} = 0, \quad \forall q_{\alpha\beta} \qquad \Longrightarrow \qquad q^*_{\alpha\beta} \tag{1.114}$$

To check the correctness of the calculation when plugging an ansatz for $q^*_{\alpha\beta}$ in A, one must always take care that in all terms of A which depend on $q_{\alpha\beta}$ a prefactor n can be singled out, so that it simplifies with the $1/n$ outside brackets on the right-hand side of (1.112). The final step to compute the free energy per spin amounts the to the switching of the two limits $N \to \infty$ and $n \to 0$, which allows us to write:

$$\begin{aligned} f &= \lim_{\substack{N\to\infty \\ n\to 0}} -\frac{1}{N\beta}\frac{\overline{Z^n}-1}{n} \\ &= \lim_{\substack{n\to 0 \\ N\to\infty}} -\frac{1}{nN\beta}\left(\overline{Z^n}-1\right) \\ &= \lim_{\substack{n\to 0 \\ N\to\infty}} -\frac{1}{nN\beta}\left(1-nNA[q^*_{\alpha\beta}]-1\right) \\ f &= \frac{1}{\beta}A[q^*_{\alpha\beta}] \end{aligned} \tag{1.115}$$

This swap of limits is a risky step that everybody is normally willing to take because mathematicians then cover it rigorously. To recover the notion of overlap in $q_{\alpha\beta}$, note that at some point after the HS transformation we had (1.109), which we rewrite as

$$\begin{aligned} \overline{Z_J^n} &= \exp\left\{nN\beta^2\delta^2/4\right\} \sum_{\S} \int_{-\infty}^{\infty} \prod_{\alpha<\beta} dq_{\alpha\beta} \\ &\qquad \exp\left\{-\frac{N\beta^2\delta^2}{2}\sum_{\alpha<\beta} q^2_{\alpha\beta} + \beta^2\delta^2 \sum_{\alpha<\beta} q_{\alpha\beta} \sum_{i=1}^{N} S_i^\alpha S_i^\beta\right\} \\ &= \exp\left\{nN\beta^2\delta^2/4\right\} \sum_{\S} \int_{-\infty}^{\infty} \prod_{\alpha<\beta} dq_{\alpha\beta} \exp\left\{-Nu(q_{\alpha\beta}, S_1^\alpha, \ldots, S_N^\alpha)\right\} \end{aligned} \tag{1.116}$$

with u defined as

$$u(q_{\alpha\beta}, S_1^\alpha, \ldots, S_N^\alpha) = \beta^2\delta^2 \left[\frac{1}{2}\sum_{\alpha<\beta} q^2_{\alpha\beta} - \sum_{\alpha<\beta} q_{\alpha\beta}\frac{1}{N}\sum_{i=1}^{N} S_i^\alpha S_i^\beta\right] \tag{1.117}$$

The saddle point requires

$$\frac{\partial u}{\partial q_{\alpha\beta}} = 0, \quad \forall q_{\alpha\beta} \qquad \Longrightarrow \qquad q_{\alpha\beta} = \frac{1}{N}\sum_{i=1}^{N} S_i^\alpha S_i^\beta \tag{1.118}$$

This $q_{\alpha\beta}$ is the overlap, i.e., again the scalar product between spins of two replicas (in the Hopfield model it was between a replica and a pattern). Therefore, the physical meaning of $q_{\alpha\beta}$ is the quantification of replicas similarity.

1.4.2. Replica-symmetric ansatz for the overlap matrix. Let us present first the explicit calculation of the mean-field free energy in the case of a replica symmetric ansatz for the matrix $q_{\alpha\beta}$, namely we consider that the order parameter is a matrix with all non-diagonal elements which are identical

$$q_{\alpha\beta} = \begin{pmatrix} 0 & q & \cdots & q \\ q & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & q \\ q & \cdots & q & 0 \end{pmatrix}. \tag{1.119}$$

It is then convenient to analize separately the energetic and the entropic part of the free energy. From the energetic contribution one can readily obtain:

$$\lim_{n\to 0} \frac{\beta^2\sigma^2}{2n} \sum_{\alpha<\beta} q_{\alpha\beta}^2 = \lim_{n\to 0} \frac{\beta^2\sigma^2}{2n} \frac{n(n-1)}{2} q^2 = -\frac{\beta^2\sigma^2}{4}\, q^2. \tag{1.120}$$

Let us notice that the only remarkable consequence of the limit $n \to 0$ is the change of sign of the energetic contribution to the free energy. This fact, following an heuristic argument that is presented below in the section dedicated to the p-spin model, allows to understand, at least on intuitive grounds, why one seeks to *maximize* rather than *minimize* the free energy $A[q_{\alpha\beta}]$ with respect to the elements of the matrix $q_{\alpha\beta}$. The explicit calculation of the entropic term, which is slightly more involved, reads as follows. The first step is to unfold, i.e., rewrite in an appropriate manner, the trace $\mathrm{Tr}[e^{L(q_{\alpha\beta})}]$

$$\begin{aligned} \mathrm{Tr}[e^{L(q_{\alpha\beta})}] &= \exp\left\{ \frac{\beta^2\sigma^2}{2} \sum_{\alpha\neq\beta} q_{\alpha\beta} S_\alpha S_\beta \right\} \overset{\mathbf{RS}}{=} \exp\left\{ \frac{\beta^2\sigma^2}{2}\, q \sum_{\alpha\neq\beta} S_\alpha S_\beta \right\} \\ &= \exp\left\{ \frac{\beta^2\sigma^2 q}{2} \sum_{\alpha\neq\beta} S_\alpha S_\beta + \frac{n\beta^2\sigma^2 q}{2} - \frac{n\beta^2\sigma^2 q}{2} \right\} \\ &= \exp\left\{ \frac{\beta^2\sigma^2 q}{2} \left(\sum_{\alpha=1}^{n} S_\alpha \right)^2 - \frac{n\beta^2\sigma^2 q}{2} \right\}. \end{aligned} \tag{1.121}$$

In order to keep the calculation ordered and not loose pieces let us start to group some of them before further steps:

$$\begin{aligned} \beta\overline{f_J} &= -\frac{\beta^2\sigma^2}{4} + A[q] \\ &= -\frac{\beta^2\sigma^2}{4} + \lim_{n\to 0} \frac{\beta^2\sigma^2}{2n} \sum_{\alpha<\beta} q_{\alpha\beta}^2 - \lim_{n\to 0} \frac{1}{n} \log\left(\mathrm{Tr}[e^{L(q_{\alpha\beta})}] \right) \\ &= -\frac{\beta^2\sigma^2}{4} - \frac{\beta^2\sigma^2}{4}\, q^2 - \lim_{n\to 0} \frac{1}{n} \log\left(\mathrm{Tr}[e^{L(q_{\alpha\beta})}] \right) \\ &= -\frac{\beta^2\sigma^2}{4} - \frac{\beta^2\sigma^2}{4}\, q^2 + \frac{\beta^2\sigma^2}{2}\, q - \lim_{n\to 0} \frac{1}{n} \log\left(\mathrm{Tr}\left[e^{\frac{q\beta^2\sigma^2}{2} \left(\sum_{\alpha=1}^{n} S_\alpha\right)^2} \right] \right) \\ &= -\frac{\beta^2\sigma^2}{4} (1-q)^2 - \lim_{n\to 0} \frac{1}{n} \log\left(\mathrm{Tr}\left[e^{\frac{q\beta^2\sigma^2}{2} \left(\sum_{\alpha=1}^{n} S_\alpha\right)^2} \right] \right). \end{aligned} \tag{1.122}$$

Now that put some order let us more to the last step of the calculation, where again we proceed by linearize dependence of spin variables in the argument of the exponential by

means of a Hubbard-Stratonovich transformation:

$$\lim_{n\to 0}\ \frac{1}{n}\log\ \mathrm{Tr}\left[e^{\frac{q\beta^2\sigma^2}{2}\left(\sum_{\alpha=1}^{n} S_\alpha\right)^2}\right] = \lim_{n\to 0}\ \frac{1}{n}\log\ \mathrm{Tr}\left[\int_{-\infty}^{\infty}\frac{dz}{\sqrt{2\pi}}\ e^{-\frac{z^2}{2}+z\beta\sigma\sqrt{q}\sum_{\alpha=1}^{n} S_\alpha}\right]$$

$$= \lim_{n\to 0}\ \frac{1}{n}\ \log\int_{-\infty}^{\infty}\frac{dz}{\sqrt{2\pi}}\ e^{-\frac{z^2}{2}}\left(\sum_{S} e^{\beta\sigma\sqrt{q}S}\right)^n = \lim_{n\to 0}\ \frac{1}{n}\ \log\ \int_{-\infty}^{\infty}\frac{dz}{\sqrt{2\pi}}\ e^{-\frac{z^2}{2}}\left[2\cosh(z\beta\sigma\sqrt{q})\right]^n$$

$$= \lim_{n\to 0}\ \frac{1}{n}\ \log\ \int_{-\infty}^{\infty}\frac{dz}{\sqrt{2\pi}}\ e^{-\frac{z^2}{2}}\exp\left\{n\ \log\left[2\cosh(z\beta\sigma\sqrt{q})\right]\right\}$$

$$\lim_{n\to 0}\ \frac{1}{n}\ \log\ \int_{-\infty}^{\infty}\frac{dz}{\sqrt{2\pi}}\ e^{-\frac{z^2}{2}}\ \left[1+n\log\left[2\cosh(z\beta\sigma\sqrt{q})\right]+\mathcal{O}(n^2)\right]$$

$$= \lim_{n\to 0}\ \frac{1}{n}\ \log\left(1+n\int_{-\infty}^{\infty}\frac{dz}{\sqrt{2\pi}}\ e^{-\frac{z^2}{2}}\log\left[2\cosh(z\beta\sigma\sqrt{q})\right]\right)$$

(1.123)

$$= \int_{-\infty}^{\infty}\frac{dz}{\sqrt{2\pi}}\ e^{-\frac{z^2}{2}}\log\left[2\cosh(z\beta\sigma\sqrt{q})\right] = \int \mathcal{D}z\ \log\cosh(z\beta\sigma\sqrt{q}).$$

By putting together all pieces of our calculation we thus have found the following expression of the free energy in the replica-symmetric ansatz, where the parameter $n \to 0$ has disappeared everywhere

$$\beta\overline{f_J} = -\frac{\beta^2\sigma^2}{4}(1-q)^2 - \int \mathcal{D}z\ \log\cosh(z\beta\sigma\sqrt{q}). \tag{1.124}$$

We have therefore finally reached an expression of the free-energy in terms of a single variational parameter, q. As in the case of the Random-Field Ising Model and the Hopfield Model one seeks the presence of a phase transition by looking for the solution of the saddle-point equation:

$$\frac{\partial\overline{f_J}}{\partial q} = 0 \quad\Longrightarrow\quad q = 1 - \frac{1}{\beta\sigma}\int \mathcal{D}z\ \frac{z}{\sqrt{q}}\ \tanh(z\beta\sigma\sqrt{q}), \tag{1.125}$$

which can be further simplified by noting that the factor z comes out from the derivative of the exponential weight

$$\begin{aligned} q &= 1 - \frac{1}{\beta\sigma\sqrt{q}}\int\frac{dz}{\sqrt{2\pi}}\ \frac{d}{dz}\left(e^{-z^2/2}\right)\ \tanh(z\beta\sigma\sqrt{q}) \\ &= 1 - \frac{1}{\beta\sigma\sqrt{q}}\int \mathcal{D}z\ \frac{d}{dz}\tanh(z\beta\sigma\sqrt{q}) \\ &= 1 - \int \mathcal{D}z\ \left[1 - \frac{\sinh^2(z\beta\sigma\sqrt{q})}{\cosh^2(z\beta\sigma\sqrt{q})}\right] \end{aligned}$$

$$q = \int \mathcal{D}z\ \tanh^2(z\beta\sigma\sqrt{q}). \tag{1.126}$$

In order to single out the presence of a face transition with q as order parameter ones has to look for the solution of the saddle-point equation Eq. (1.126). First of all it can be easily checked that $q = 0$ is always a solution: this solution is the one typical of the high-temperature ergodic phase where two configurations sampled independently from the Boltzmann measure are typically uncorrelated. By lowering the temperature one then finds a phase transition signalled by a bifurcation of the solution to the saddle point equation. Since we known that above the critical temperature T_c of this transition the solution is $q = 0$, the critical temperature itself can be found by considering a small-q

expansion of the free energy in Eq. (1.124). Due to the cancellation with the linear term in q coming from the energetic part of the free-energy, one must be only careful in retaining even the second smallest term in the expansion of the logarithm:

$$\log\cosh(z\beta\sigma\sqrt{q}) \;\approx\; \log\left(1+\frac{z^2\beta^2\sigma^2 q}{2}\right) \;\approx\; \frac{z^2\beta^2\sigma^2 q}{2}-\frac{z^4\beta^4\sigma^4 q^2}{4}, \tag{1.127}$$

so that

$$\begin{aligned}\beta\overline{f_J} &= -\frac{\beta^2\sigma^2}{4}(1-q)^2-\langle z^2\rangle\frac{\beta^2\sigma^2 q}{2}+\langle z^4\rangle\frac{\beta^4\sigma^4 q^2}{4}+\mathcal{O}(q^4)\\ &= -\frac{\beta^2\sigma^2}{4}-\frac{\beta^2\sigma^2}{4}\left(1-\beta^2\sigma^2\right)q^2+\mathcal{O}(q^4),\end{aligned} \tag{1.128}$$

where we have used the fact that

$$\langle z^{2n}\rangle = \int_{-\infty}^{\infty}\frac{dz}{\sqrt{2\pi}}\,e^{-z^2/2}\,z^{2n}=1. \tag{1.129}$$

As is customary for second order phase transitions, even for the *spin-glass* transition we can retrieve the critical temperature T_c by looking for the value at which the quadratic term in small order parameter expansion vanishes, which readily yields: $T_c=\sigma$. We have therefore found a critical temperature such that for $T>T_c$ the system is in a paramagnetic disordered phase, $q=0$, and for $T<T_c$ it is in an ordered phase with $q>0$. But we have a problem: a calculation of the free-energy of this ordered phase in the limit $T\to 0$, where $q=1-\frac{1}{\sigma}\sqrt{\frac{2}{\pi}}\,T$, yields

$$\beta\overline{f_J}=\langle U\rangle - TS = -\sqrt{\frac{2}{\pi}}\,\sigma+\frac{T}{2\pi}+\mathcal{O}(T^2), \tag{1.130}$$

from which we have

$$S(T\to 0)=-\frac{1}{2\pi}. \tag{1.131}$$

But ... it is impossible for the entropy of a sistem of discrete variables to have negative entropy, for the simple reason that in this case (discrete variables) entropy is nothig but the logarithm of a integern number, the number of microscopic states. The result of negative entropy signals that some wrong step was made in the procedure. At the beginning the most suspicious steps was considered the exchange of limits

$$\lim_{n\to 0}\lim_{N\to\infty} \;=\; \lim_{N\to\infty}\lim_{n\to 0}, \tag{1.132}$$

but then Parisi discovered the only mistake done was in the assumption of the overlap matrix $q_{\alpha\beta}$ structure: this was the discovery of replica symmetry breaking in 1979.

1.4.3. One-step Replica symmetry breaking. We have seen in the previous section that the replica-symmetric ansatz of the matrix order parameter $q_{\alpha\beta}$ leads to inconsistencies at low temperatures. It allows predict the existence of a phase transition and to characterize the critical temperature T_c, but it is not appropriate to characterize the low temperature phase. The visionary intuition of Giorgio Parisi was to realize that below the spin glass critical temperature T_c the system is always frozen in a given disordered configuration, *but* also that the degree of similarity of two different configurations obtained by a statistical sampling at $T<T_c$ cannot the parametrized by a single number. In practice, all the disordered configurations sampled with Boltzmann weight in the spin-glass phase might have different degrees of similarity/overlap. The simplest situation of that kind is the one where only two choices are given: the overlap between two independent configurations (replicas) is either q_0 with probability m or q_1 with probability $(1-m)$,

where $q_1 > q_0$ and m is a positive number $m \in [0,1]$. Such an assumption on the possible values took by the overlap of independent disordered configurations sampled below T_c goes under the name of *"one-step replica symmetry breaking"* (1-RSB). As we are going to see, this is a better approximation than the assumption of symmetry between replicas, which can be translated into "all disordered configurations are equally different", but still not the right one for the Sherrington-Kirkpatrick model. Nevertheless, since the 1-RSB scenario is realized in other models, like the p-spin to be discussed later on, and is the paradigmatic one for many glassy materials, it worth to be discussed also in the case of the Sherrington-Kirkpatrick model. From the point of view of the free-energy landscape is very simple: below T_c phase space becomes fragmented in a low of free-energy basins, each corresponding to a pool of similar configurations with mutual overlap q_1. If you compare the configurations of different basins they are not alike, their overlap is $q_0 < q_1$. Within the 1-RSB scenario, when you sample randomly two configurations with weight proportional to $\exp(-\beta H)$ you then have probability $(1-m)$ to pick up them from the same basin and probability m to pick up them from different basins. Beside its relevance for other models, the calculation of the 1-RSB free energy and saddle point equations in the SK model is quite instructive. The structure of the overlap matrix corresponding to the 1-RSB ansatz for the SK model reads as:

$$\mathcal{Q}_{1rsb} = \begin{pmatrix} 0 & q_1 & q_1 & & & & & & \\ q_1 & 0 & q_1 & & & & & q_0 & \\ q_1 & q_1 & 0 & & & & & & \\ & & & 0 & q_1 & q_1 & & & \\ & & & q_1 & 0 & q_1 & & & \\ & & & q_1 & q_1 & 0 & & & \\ & & & & & & \ddots & & \\ & & & & & & & 0 & q_1 & q_1 \\ & q_0 & & & & & & q_1 & 0 & q_1 \\ & & & & & & & q_1 & q_1 & 0 \end{pmatrix}, \tag{1.133}$$

where $\mathcal{Q}_{1rsb}$ is n x n block diagonal matrices where the submatrices are m x m square matrices. The above ansatz for the replica matrix must be then plugged into the expression of the entropic term of the free energy: $\mathrm{Tr}[\exp\left(\beta^2\sigma^2 \sum_{\alpha<\beta} S_\alpha S_\beta\right)]$. To this purpose is more convenient to rewrite the matrix in the following form:

$$\mathcal{Q}_{1rsb} = \mathcal{Q}^{fill}_{1rsb} + \mathcal{Q}^{block}_{1rsb} - \mathcal{Q}^{diag}_{1rsb}, \tag{1.134}$$

where $\mathcal{Q}^{fill}_{1rsb}$ is an n x n matrix completely filled with identical elements

$$\mathcal{Q}^{fill}_{1rsb} = \begin{pmatrix} q_0 & \cdots & q_0 \\ \vdots & \ddots & \vdots \\ q_0 & \cdots & q_0 \end{pmatrix}, \tag{1.135}$$

$\mathcal{Q}^{block}_{1rsb}$ is an n x n block diagonal matrix, each block $\mathcal{D}$ being an m x m diagonal matrix and all elements outside the blocks are zero

$$\mathcal{Q}^{block}_{1rsb} = \begin{pmatrix} \mathcal{D} & 0 & \dots & 0 \\ 0 & \mathcal{D} & 0 & \vdots \\ 0 & 0 & \ddots & 0 \\ 0 & \dots & 0 & \mathcal{D} \end{pmatrix} \qquad \mathcal{D} = \begin{pmatrix} q_1 - q_0 & \cdots & \cdots & q_1 - q_0 \\ \vdots & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ q_1 - q_0 & \cdots & \cdots & q_1 - q_0 \end{pmatrix}, \tag{1.136}$$

while $\mathcal{Q}^{diag}_{1rsb}$ is simply

$$\mathcal{Q}^{diag}_{1rsb} = \begin{pmatrix} q_1 & 0 & \dots & 0 \\ 0 & q_1 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \dots & 0 & q_1 \end{pmatrix}, \tag{1.137}$$

By then denoting as $\vec{S}$ an array in replica space, $\vec{S} = (S_1, \dots, S_n)$ we can then write

$$\begin{aligned} \sum_{\alpha<\beta} q_{\alpha\beta} &= \frac{1}{2}\sum_{\alpha\neq\beta} q_{\alpha\beta} = \\ &= \frac{1}{2}\vec{S}^T \mathcal{Q}^{fill}_{1rsb}\vec{S} + \frac{1}{2}\vec{S}^T \mathcal{Q}^{block}_{1rsb}\vec{S} - \frac{1}{2}\vec{S}^T \mathcal{Q}^{diag}_{1rsb}\vec{S} = \\ &= \frac{q_0}{2}\sum_{\alpha\beta} S_\alpha S_\beta + \frac{q_1 - q_0}{2}\sum_{k=1}^{n/m}\sum_{\alpha_k,\beta_k=1}^{m} S_{\alpha_k} S_{\beta_k} - \frac{nq_1}{2} \\ &= \frac{q_0}{2}\left(\sum_{\alpha=1}^{n} S_\alpha\right)^2 + (q_1 - q_0)\sum_{k=1}^{n/m}\left(\sum_{\alpha_k=1}^{m} S_{\alpha_k}\right)^2 - \frac{nq_1}{2} \end{aligned} \tag{1.138}$$

The calculation goes along the same lines of the replica-symmetric case, but now we have to contribution with a sum squared to be linearized:

$$\begin{aligned} \exp\left\{\frac{\beta^2\sigma^2 q_0}{2}\left(\sum_{\alpha=1}^{n} S_\alpha\right)^2\right\} &= \int_{-\infty}^{\infty}\frac{dx}{\sqrt{2\pi}}\ \exp\left\{-\frac{x^2}{2} + x\beta\sigma\sqrt{q_0}\sum_{\alpha=1}^{n} S_\alpha\right\} \\ &= \int d\mu(x)\ e^{x\beta\sigma\sqrt{q_0}\sum_{\alpha=1}^{n} S_\alpha}, \end{aligned} \tag{1.139}$$

where for brevity we have hidden the Gaussian weight of the variable x into the integration measure symbol and

$$\begin{aligned} &\exp\left\{\frac{\beta^2\sigma^2}{2}(q_1 - q_0)\sum_{k=1}^{n/m}\left(\sum_{\alpha_k=1}^{m} S_{\alpha_k}\right)^2\right\} = \\ &\int_{-\infty}^{\infty}\left[\prod_{k=1}^{n/m} dy_k\ \frac{e^{-y_k^2/2}}{\sqrt{2\pi}}\right]\ \exp\left\{\beta\sigma\sqrt{q_1 - q_0}\sum_{k=1}^{n/m} y_k \sum_{\alpha_k=1}^{m} S_{\alpha_k}\right\} = \\ &\int\prod_{k=1}^{n/m} d\mu(y_k)\ e^{\beta\sigma\sqrt{q_1-q_0}\sum_{k=1}^{n/m} y_k \sum_{\alpha_k=1}^{m} S_{\alpha_k}} \end{aligned} \tag{1.140}$$

Once the dipendence on spin variables S_α has been linearized everywhere it is possible to compute the trace over spins in the following manner:

$$
\begin{aligned}
&\mathrm{Tr}\left[e^{L(q_{\alpha\beta})}\right] = \\
&e^{-\frac{n}{2}\beta^2\sigma^2 q_1} \sum_{S_1,\dots,S_n} \int d\mu(x) \int \prod_{k=1}^{n/m} d\mu(y_k) \exp\left\{ x\beta\sigma\sqrt{q_0} \sum_{\alpha=1}^{n} S_\alpha + \beta\sigma\sqrt{q_1-q_0} \sum_{k=1}^{n/m} y_k \sum_{\alpha_k=1}^{m} S_{\alpha_k} \right\} = \\
&e^{-\frac{n}{2}\beta^2\sigma^2 q_1} \sum_{S_1,\dots,S_n} \int d\mu(x) \int \prod_{k=1}^{n/m} d\mu(y_k) \exp\left\{ \sum_{k=1}^{n/m} \beta\sigma \left[x\sqrt{q_0} + y_k\sqrt{q_1-q_0}\right] \sum_{\alpha_k=1}^{m} S_{\alpha_k} \right\} = \\
&e^{-\frac{n}{2}\beta^2\sigma^2 q_1} \int d\mu(x) \prod_{k=1}^{n/m} \int d\mu(y_k) \left[\sum_{\{S_{\alpha_k};\alpha_k=1,\dots,m\}} e^{\beta\sigma\left(x\sqrt{q_0}+y_k\sqrt{q_1-q_0}\right)\sum_{\alpha_k=1}^{m} S_{\alpha_k}} \right] = \\
&e^{-\frac{n}{2}\beta^2\sigma^2 q_1} \int d\mu(x) \prod_{k=1}^{n/m} \int d\mu(y_k) \left(\sum_{S} e^{\beta\sigma\left(x\sqrt{q_0}+y_k\sqrt{q_1-q_0}\right)S} \right)^m = \\
&e^{-\frac{n}{2}\beta^2\sigma^2 q_1} \int d\mu(x) \left[\int d\mu(y) \left(2\cosh[\beta\sigma\left(x\sqrt{q_0}+y\sqrt{q_1-q_0}\right)]\right)^m \right]^{n/m}
\end{aligned}
\tag{1.141}
$$

Befor moving further let us now briefly comment on how the way to consider the analytic continuation $n \to 0$ within the above expression is quite different from what would seem the *reasonable* way to do it. In fact since for both n and m integer we have $m < n$, one would be tempted to send both to zero while keeping fixed the ratio n/m. But this is not the right way to get the right results. One of the most misterious intermediate steps of the replica calculation, which has been historically proven mathematically by other means that leads to the correct result, consist in sending $n \to 0$ while leaving m untouched and to be determined as a variational parameter of the problem. By defining

$$
\mathcal{Z}_{m,q_0,q_1}(x) = 2 \int d\mu(y) \left[\cosh\left(\beta\sigma x\sqrt{q_0} + \beta\sigma y\sqrt{q_1-q_0}\right)\right]^m, \tag{1.142}
$$

we can finally write the entropic part of the free energy in the framework of the 1-RSB ansatz as

$$
\begin{aligned}
\beta\overline{f}_{\text{entropic}} &= \lim_{n\to 0} -\frac{1}{n} \log\left(\mathrm{Tr}[e^{L(q_{\alpha\beta})}]\right) \\
&= \frac{\beta^2\sigma^2}{2} q_1 - \lim_{n\to 0} \frac{1}{n} \log\left(\int d\mu(x) \left[\mathcal{Z}_{m,q_0,q_1}(x)\right]^{n/m} \right) \\
&= \frac{\beta^2\sigma^2}{2} q_1 - \lim_{n\to 0} \frac{1}{n} \log\left(\int d\mu(x) \left[1 + \frac{n}{m} \log[\mathcal{Z}_{m,q_0,q_1}(x)]\right] \right) \\
&= \frac{\beta^2\sigma^2}{2} q_1 - \lim_{n\to 0} \frac{1}{n} \log\left(1 + \frac{n}{m} \int d\mu(x) \ \log[\mathcal{Z}_{m,q_0,q_1}(x)] \right) \\
&= \frac{\beta^2\sigma^2}{2} q_1 - \frac{1}{m} \int d\mu(x) \ \log[\mathcal{Z}_{m,q_0,q_1}(x)]
\end{aligned}
\tag{1.143}
$$

The calculation of the energetic part of the free energy to the leading order in n within the 1-RSB is much easier and, by recalling the way to decompose the replica matrix written

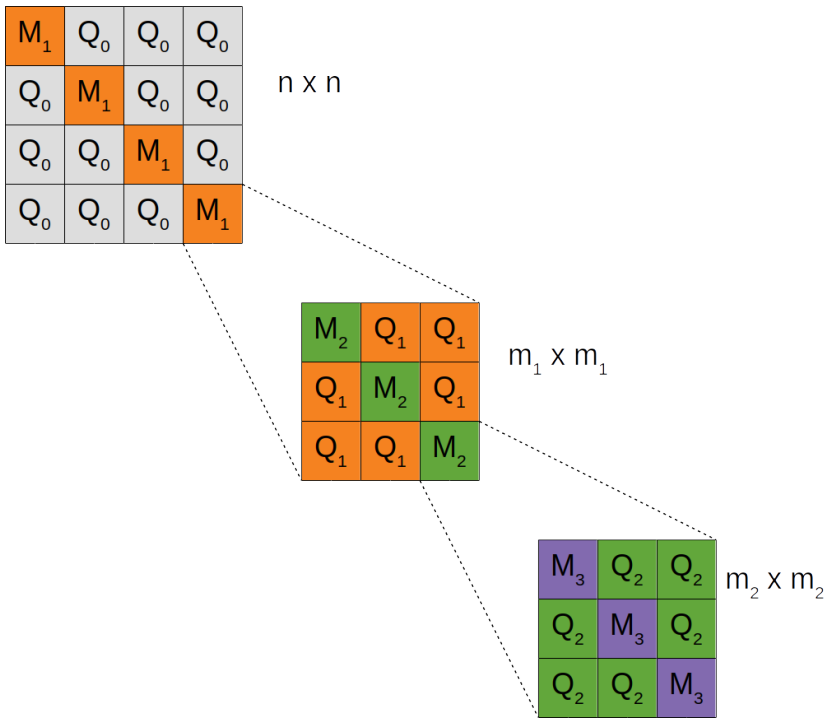

Figure 5. Sketch of the hierarchical overlap of replicas in the SK model at low temperature in the non-ergodic phase. The matrix in the top-left corner is the overlap matrix $q_{\alpha\beta}$. The diagram should be continued iterating indefinitely the magnifications.

in Eq. (1.112), easily reads out as:

$$\begin{aligned}\beta\overline{f}_{\text{energetic}} &= \lim_{n\to 0}\frac{\beta^2\sigma^2}{2n}\sum_{\alpha<\beta}q_{\alpha\beta}^2 = \lim_{n\to 0}\frac{\beta^2\sigma^2}{4n}\sum_{\alpha\neq\beta}q_{\alpha\beta}^2 = \\ &= \lim_{n\to 0}\frac{\beta^2\sigma^2}{4n}\left[n^2q_0^2 + \frac{n}{m}m^2(q_1^2-q_0^2) - nq_1^2\right] = \\ &= -\frac{\beta^2\sigma^2}{4}\left[(1-m)q_1^2 + q_0^2\right] \qquad (1.144)\end{aligned}$$

so that the complete expression of the free energy reads as:

$$\beta\overline{f} = \max_{q_0,q_1,m}\left\{-\frac{\beta^2\sigma^2}{4}\left[(1-m)q_1^2 - 2q_1 + q_0^2\right] - \frac{1}{m}\int d\mu(x)\ \log[\mathcal{Z}_{m,q_0,q_1}(x)]\right\} \qquad (1.145)$$

Historically, it has been found that the 1-RSB ansatz provides better results than the replica-symmetric one, i.e. the entropy in the limit $T \to 0$ is larger, but it is still negative, $S(T=0) < 0$, so that this is not yet the good choice for the structure overlap matrix. As will be discussed in the next chapter, Parisi discovered that a positive entropy in the $T \to 0$ limit is found only if the breaking of the replica symmetry is iterated an infinite number of times, resorting to an ansatz known as the *full-RSB* ansatz.

1.4.4. Full Replica symmetry breaking. At low temperature the SK model displays a fascinating hierarchy of replica similarities, which is summarized by a label: *full replica symmetry breaking*. It is better to think of n very large, even if the $n \to 0$ is mandatory at the end. As we have said, according to a very reasonable physical intuition, the matrix $q_{\alpha\beta}$ must be a symmetric one. Furthermore, let us notice that each row of the matrix $q_{\alpha\beta}$ represents all the possible values of the overlap between a given replica of the system, say α, with other replicas of the system, say $\beta \neq \alpha$. Since we ask that no replica is privileged with respect to the others, it is also reasonable to ask that all the elements appearing

in a row of the matrix $q_{\alpha\beta}$ also appear in all the others, up to a permutation. The only degree of freedom we are left with is thus the rule to assign the matrix elements in a row. The mesmerizing discovery of Parisi in the late '70 was that the only symmetry which is broken at the critical temperature T_c and represents the breaking of ergodicity is the *permutation symmetry* of the matrix elements within a single row of the matrix $q_{\alpha\beta}$, i.e., *different replicas might have different overlaps*. In this perspective let us fix the row index of the matrix element $q_{\alpha\beta}$ in order to parameterize the elements on a row with only one index: $q_{1i} = q_i$, where i runs from 1 to n and we have dropped the row index. The way permutation symmetry is broken is decided by the assignment rule:

$$i \longrightarrow q_i \tag{1.146}$$

Once this rule is fixed for one row of the matrix, the structure of the other rows follows, since they are just a permutation of the first. This leads to the so-called *Parisi's matrix*, which we briefly discuss in what follows.

A large n allows to introduce a sequence of similarity levels. In the first stage of symmetry breaking, the $n \times n$ matrix $q_{\alpha\beta}$ is represented as composed of $m_1 \times m_1$ submatrices M_1 and Q_0 (Figure 5). The matrix Q_0 is fully composed of elements equal to q_0, the typical lowest overlap between replicas. The matrix M_1 is found n/m_1 times on the diagonal of $q_{\alpha\beta}$.

At a second stage, by inspecting the structure of M_1, one finds that it resembles that of $q_{\alpha\beta}$ in the first stage. It is composed of submatrices M_2 and Q_1 of size $m_2 \times m_2$, where Q_1 is filled by elements $q_1 > q_0$ while M_2's are sequenced on the diagonal of M_1. The story goes on like this, forming a sequence $q_0 < q_1 < q_2 < q_3 < \ldots < q_k$ and correspondingly $n > m_1 > m_2 > m_3 > \ldots > m_k$. In order to reconnect with the discussion above on the elements of a single row, let us notice that a k-steps breaking of the permutation symmetry between replicas corresponds to the assignment of k different values to the n elements of a row for the matrix $q_{\alpha\beta}$. As long as k is finite the *number of elements* m_i in the row (with $1 < i < n$) which take the value q_i can be interpreted as being proportional to the *probability* that two replicas have the overlap q_i.

A question is then immediately in order: how to recognize which is the the correct level k of breaking of the replica permutation symmetry? And why a breaking should be considered at all?

The first thing to say is that the replica-symmetric ansatz, i.e., the assumption that all the off-diagonal elements of $q_{\alpha\beta}$ are identical ($q_{\alpha\beta} = q$ for all α and β such that $\alpha \neq \beta$), leads to a negative entropy at zero temperature: $S(0) = -\frac{1}{2\pi} \approx -0.17$. This is clearly a wrong result because we have discrete variables and the entropy has a unique and unequivocal meaning: it is the logarithm of the number $\mathcal{N}$ of microstates accessible at a given temperature and since $\mathcal{N} \geq 1$ we have $S = \log \mathcal{N} \geq 0$. If one then tries to compute the free-energy assuming one level of breaking of the permutation symmetry between replicas, i.e., to subdivide the $n \times n$ matrix in blocks of size m_1 and with possible value of the overlap q_1 and q_0, obtaining the correct value of q_0, q_1 and m_1 from the extremization of the free-energy, one finds that at $T = 0$ the entropy is still negative but *less* negative, i.e. with the *one-step-replica-symmetry-breaking* ansatz it takes the value $S(T = 0) = -0.1$. Although the result is still incorrect, this tells us that we are moving in the right direction, because the *negative* entropy has become less negative, hence less wrong. We will discuss in the next lecture a model where the assumption of only one step of replica symmetry breaking is the correct one: the p-spin model, which has two

main differences with the SK model presented here, continuous variables and non-linear interactions. By increasing further the number of levels k at which the symmetry between replicas can be broken in the SK model, one obtains a zero-temperature entropy which is closer and closer to zero. This suggested to consider the possibility of an *infinite number of breakings* of the permutation symmetry between replicas, which, at least for the SK model, turns out to be the correct assumption. Before commenting further on this, let us just quote the formal and unequivocal criterion to discriminate whether the assumption of *breaking at "k" levels* is right or wrong. The criterion to say whether a saddle-point solution is good is that it must be *stable*, namely it must be a *minimum* of the free energy. The stability of the stationary point is decided by the eigenvalues spectrum of the Hessian

$$M_{(\alpha\beta),(\gamma\delta)} = \frac{\partial^2 A}{\partial q_{\alpha\beta}\partial q_{\gamma\delta}} \tag{1.147}$$

It is only when this Hessian matrix has all the eigenvalues positive that we have a stable solution. If one computes the eigenvalues of $M_{(\alpha\beta),(\gamma\delta)}$ one finds that the smallest one, the *replicon*, is negative for any finite number k of breakings. In the SK model any saddle-point solution with a finite number of breakings is *unstable*. The name *replicon* comes from the famous paper of J.R.L de Almeida and D.J. Thouless where the (un-)stability of the replica-symmetric solution of the SK model was studied for the first time [*"Stability of the Sherrington-Kirkpatrick solution of a spin glass model"*, J.R.L. de Almeida and D.J. Thouless, *J. Phys. A* **13** (1978)] (D.J Thouless was precisely the same guy awarded with Nobel prize in 2016 for the Kosterliz-Thouless transition). For the details of all this analysis we refer the curious student to the literature quoted at the beginning of this lecture. The final outcome is, as anticipated, that the correct ansatz sends $k \to \infty$, which is called the *full*-replica-symmetry-breaking (full-RSB) ansatz. The infinite number of elements in the sequence of overlaps $q_1 < \ldots < q_k$ suggested as well the title of the paper where the full-RSB was proposed for the first time [*"Infinite number of order parameters for spin-glasses*, Giorgio Parisi, *Phys. Rev. Lett.* **43**, 1754 (1979)].

Such infinite number of order parameters reflects the hierarchical multi-valley structure of the SK energy landscape at low T, as sketched in Figure 6, with an infinite sequence of minima nested into others. Clearly, by zooming out the details of this *fractal* structure of nested minima, we find that the typical overlap between configurations trapped within the same minimum is larger the smaller is the scale at which we have zoomed the landscape, as shown in Fig. 6.

Our last comment is on the $n \to 0$ limit, on which we did not spend (on purpose) a lot of words so far. We warn the curious students that this is a highly a technical part, which might be challenging for a naive intuition of what is going on. In any case, all details are given in the references at the beginning of the lecture.

When $k \to \infty$ and at the same time $n \to 0$ one has to plug an infinite number of different choices $q_1 < \ldots < q_k$ into a zero dimensional matrix. It turns out that the most convenient thing to do (... and it works!) is to replace the numerable sequence $q_1 < \ldots < q_k$ with a continuous function

$$q(x) : [0,1] \to [0,1], \tag{1.148}$$

where the role played by the subscript index i in q_i has been taken up by the real variable x, of which the overlap q becomes a function. A function $q(x)$ in the unit interval $[0,1]$ is an element of an infinite-dimensional space, i.e., there is an infinite number of choices to define it: this is OK, because it is in fair agreement with the infinite number of order parameters we where after. Then, it can be shown (see the references) that the role played by the *breaking parameter* m_k, i.e., that of representing the probability to find the overlap

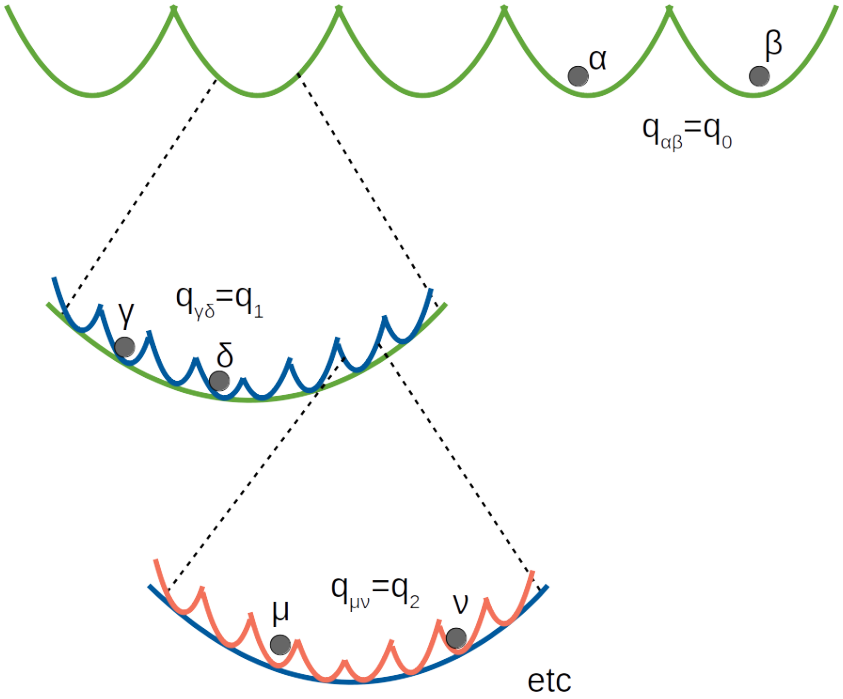

Figure 6. Sketch of the hierarchical energy landscape of the SK model at low temperature in the non-ergodic phase: $q_0 < q_1 < q_2 < \ldots$.

q_k between two replicas, is played in this sort of continuum limit by the variable x in the domain of $q(x)$. In particular it can be shown that the function $q(x) : x \in [0,1] \to q \in [0,1]$ has a well defined inverse

$$x[q] : q \in [0,1] \to x \in [0,1], \tag{1.149}$$

which can be identified as the *cumulative distribution of the overlap*, i.e., the probability density distribution $P(q)$ of the overlap turns out to be

$$P(q) = \frac{dx}{dq} \tag{1.150}$$

To conclude, let us just quote the final result, i.e., the form of the free-energy per spin when the two limits of an infinite number of breakings, $k \to \infty$, and of zero size of the overlap matrix, $n \to 0$, are taken:

$$\beta f = -\frac{\beta^2\sigma^2}{4}\left[1 + \int_0^1 dx\; q(x)^2 - 2q(1)\right] - \int_{-\infty}^{\infty} \frac{du}{\sqrt{2\pi}}\, e^{-u^2/2}\, f_0(0, u\sqrt{q(0)}), \tag{1.151}$$

where the function f_0 is the solution of the the so-called *Parisi equation*,

$$\frac{\partial f_0(x,h)}{\partial x} = -\frac{\sigma^2}{2}\frac{dq}{dx}\left[\frac{\partial^2 f_0}{\partial h^2} + x\left(\frac{\partial f_0}{\partial h}\right)^2\right], \tag{1.152}$$

where h is a magnetic field which has to be in general taken into account in the Hamiltonian, i.e., $H = -\sum_{i<j} J_{ij} S_i S_j - h\sum_{i=1}^N S_i$. In all the calculations presented in this lectures we set $h = 0$ for simplicity, but the generalization to the case of $h > 0$ is straightforward.

To conclude, let us just try to recognize the origin of the terms of the full-RSB free energy in (1.152). It is quite easy to recognize the contribution of the *energetic* term

of (1.112):

$$-\frac{\beta^2\sigma^2}{4}+\frac{\beta^2\delta^2}{2n}\sum_{\alpha<\beta}q_{\alpha\beta}^2 \quad\Longrightarrow\quad -\frac{\beta^2\sigma^2}{4}\left[1+\int_0^1 dx\; q(x)^2-2q(1)\right]. \tag{1.153}$$

Just recall that $\sum_{\alpha<\beta}=n(n-1)/2$ and that all the n rows of the matrix $q_{\alpha\beta}$ are identical up to a permutation, so that it is not difficult to believe that in the limit $n\to 0$ one has

$$\sum_{\alpha<\beta}q_{\alpha\beta}^2 \;\approx\; n(n-1)\int dq\; q(x)^2 \;\to\; -n\int dq\; q(x)^2 \tag{1.154}$$

Much less recognizable—indeed it requires several intermediate steps and tricks to be derived—is the connection of the entropic contribution with its original form at finite n:

$$\frac{1}{n}\log\left[\mathrm{Tr}\, e^{L(q_{\alpha\beta})}\right] \;\approx\; \int_{-\infty}^{\infty}\frac{du}{\sqrt{2\pi}}\; e^{-u^2/2}\; f_0(0,u\sqrt{q(0)}) \tag{1.155}$$

1.5. Spherical p-spin model

In the previous lecture we have studied the Sherrington-Kirkpatrick (SK) model, which is the mean-field version of the disordered Ising model. The most puzzling result is that the correct way to represent the breaking of ergodicity in the SK model is by an infinite sequence of breakings of the permutation symmetry between replicas, which, on physical grounds, corresponds to a low temperature phase characterized by fractal free-energy landscape with an infinite hierarchy of minima nested into each other.

The situation is in a sense much simpler in the so called spherical p-spin model, another disordered model with quenched random couplings characterized by continuous *spins* and non-linear couplings, i.e., in the Hamiltonian it is assumed $p>2$ for the order of the nonlinear p-body interaction:

$$H(\sigma)=-\sum_{i_1>\ldots>i_p=1} J_{i_1,\ldots,i_p}\sigma_{i_1}\ldots\sigma_{i_p} \tag{1.156}$$

where the N continuous spin variables are subject to the global *spherical* constraint

$$\sum_{i=1}^{N}\sigma_i^2=N, \tag{1.157}$$

Each of the quenched couplings $J_{i_1,\ldots,i_p}$ follows a Gaussian distribution

$$p(J)=\frac{N^{\frac{p-1}{2}}}{\sqrt{p!\pi}}\exp\left\{-\frac{J^2N^{p-1}}{p!}\right\}, \tag{1.158}$$

where, as we are going to show, the scaling of the variance with N is chosen precisely as to have extensive free-energy.

The peculiarity of this model, attracting a lot of interest since the end of the the 80s, is its low-temperature ergodicity-breaking transition characterized by a one-step replica-symmetry-breaking ansatz at the critical temperature T_K, which is known for this model as the Kauzmann temperature. At T_K phase space splits into many disjoint ergodic components, each corresponding to a minimum of the free energy. In the large N limit there

are only two possibilities: two configurations of the system sampled with the Boltzmann probability

$$P_J(\sigma) = e^{-\beta H_J[\sigma]}\ \delta\left(N - \sum_{i=1}^{N} \sigma_i^2\right) \tag{1.159}$$

can be either belonging to the same ergodic component, i.e., the basin of the same free-energy minimum, hence their overlap is $q_1 > 0$, or belonging to different minima, so that their overlap is $q_0 = 0$. As we are going to see the probability to have overlap q_0 or q_1 depends on the breaking parameter m, which is of course a real number in the interval $m \in [0, 1]$. Within the large-N saddle-point approximation there is no other option, not a possible continuum degree of similarity $q(x)$ like in the SK model, no structure of nested minima. Also, the different minima of the free-energy are separated by extensive, i.e., $\sim N$, free-energy barriers, at variance with the SK model.

1.5.1. Free-energy calculation. Much in the same way as in the SK model, also in the p-spin model the free-energy is *self-averaging* as defined in (1.99), so that it can be computed as the result of a saddle-point calculation of the replicated partition function:

$$f = \lim_{N\to\infty} -\frac{1}{N\beta}\overline{\log Z} = \lim_{\substack{n\to 0\\ N\to\infty}} -\frac{1}{nN\beta}(\overline{Z^n} - 1) \tag{1.160}$$

Let us now consider the case $p = 3$ till some point of the calculation because the formulae are shorter, but please keep in mind that the steps are identical for any p and so do the results. We thus assume the interaction energy

$$H = -\sum_{i<j<k} J_{ijk}\sigma_i\sigma_j\sigma_k, \tag{1.161}$$

The partition function reads as

$$\begin{aligned}\overline{Z^n} &= \int_{-\infty}^{\infty} \prod_{i<j<k} dJ_{ijk} P(J_{ijk}) \left[\int_{-\infty}^{\infty} \prod_{i=1}^{N} d\sigma_i\ e^{\beta\sum_{i<j<k} J_{ijk}\sigma_i\sigma_j\sigma_k}\ \delta\left(N - \sum_{i=1}^{N}\sigma_i^2\right)\right]^n \\ &= \int_{-\infty}^{\infty} \prod_{i<j<k} dJ_{ijk} P(J_{ijk}) \int_{-\infty}^{\infty} \mathcal{D}\sigma\ e^{\beta\sum_{i<j<k} J_{ijk}\sum_{\alpha=1}^{n}\sigma_i^\alpha\sigma_j^\alpha\sigma_k^\alpha},\end{aligned} \tag{1.162}$$

where we have used the symbol

$$\mathcal{D}\sigma = \prod_{i=1}^{N}\prod_{\alpha=1}^{n} d\sigma_i^\alpha \prod_{\alpha=1}^{n} \delta\left(N - \sum_{i=1}^{N}(\sigma_i^\alpha)^2\right), \tag{1.163}$$

to denote, with a compact notation, an integration over spin variables satisfying the spherical constraint for each replica α. It is more convenient to leave the spherical constraint hidden in the integration measure symbol.

In the Hamiltonian there is a number $\sum_{i<j<k} = N(N-2)(N-3)/3!$ of independent triplets of interacting spins, and we have a corresponding number of integrations over the disorder coefficients J_{ijk} to do. Each of them is a simple Gaussian integral, which,

dropping prefactors, is

$$\int_{-\infty}^{\infty} dJ_{ijk} \exp\left\{-J_{ijk}^2 \frac{N^{p-1}}{p!} + J_{ijk}\beta \sum_{\alpha=1}^{n} \sigma_i^\alpha \sigma_j^\alpha \sigma_k^\alpha\right\} =$$

$$= \exp\left\{\frac{\beta^2 p!}{4N^{p-1}} \left(\sum_{\alpha=1}^{n} \sigma_i^\alpha \sigma_j^\alpha \sigma_k^\alpha\right)^2\right\}$$

$$= \exp\left\{\frac{\beta^2 p!}{4N^{p-1}} \sum_{\alpha,\beta=1}^{n} \sigma_i^\alpha \sigma_i^\beta \; \sigma_j^\alpha \sigma_j^\beta \; \sigma_k^\alpha \sigma_k^\beta\right\}, \tag{1.164}$$

from which, by collecting all the $\mathcal{O}(N^3)$ integrations and by recalling that

$$p! \sum_{i<j<k}^{N} \approx \sum_{ijk}^{N} \tag{1.165}$$

we get

$$\overline{Z^n} = \int_{-\infty}^{\infty} \mathcal{D}\sigma \exp\left\{\frac{\beta^2 p!}{4N^{p-1}} \sum_{i<j<k} \sum_{\alpha,\beta=1}^{n} \sigma_i^\alpha \sigma_i^\beta \; \sigma_j^\alpha \sigma_j^\beta \; \sigma_k^\alpha \sigma_k^\beta\right\}$$

$$= \int_{-\infty}^{\infty} \mathcal{D}\sigma \exp\left\{\frac{\beta^2}{4N^{p-1}} \sum_{\alpha,\beta=1}^{n} \left(\sum_{i=1}^{N} \sigma_i^\alpha \sigma_i^\beta\right) \left(\sum_{j=1}^{N} \sigma_j^\alpha \sigma_j^\beta\right) \left(\sum_{k=1}^{N} \sigma_k^\alpha \sigma_k^\beta\right)\right\}$$

$$= \int_{-\infty}^{\infty} \mathcal{D}\sigma \exp\left\{\frac{\beta^2 N}{2} \sum_{\alpha<\beta}^{n} \left(\frac{1}{N} \sum_{i=1}^{N} \sigma_i^\alpha \sigma_i^\beta\right)^p\right\} \tag{1.166}$$

In the expression of the replicated partition function in the last line of Eq. (1.166) it is easy to recognize the expression of the matrix order parameter that we have already introduced in the discussion of the SK model, the *overlap*:

$$Q_{\alpha\beta} = \frac{1}{N} \sum_{i=1}^{N} \sigma_i^\alpha \sigma_i^\beta, \tag{1.167}$$

As in the case of the SK, this expression shows that the overlap is a quantity (a) free from spin indices, i.e., a *global* order parameter; and (b) of order $\mathcal{O}(1)$, so that we correctly have that the argument of the exponential is an extensive function, i.e., $\beta^2 N/2 \sum_{\alpha<\beta}^{n} Q_{\alpha\beta}^p$. The possibility to write things in terms of a global order parameter comes once again due to the fact that the sum in the Hamiltonian of Eq. (1.161) is over all the independent p-uplets of spins, whose number is of order $\mathcal{O}(N^p)$. Hence each spin participates to a number of interactions of order $\mathcal{O}(N^{p-1})$, i.e., infinite in the thermodynamic limit $N \to \infty$. Due to this property, we known that once again we are dealing with a *mean-field* model, which has no space structure but can be solved exactly.

Since $p > 2$, in the p-spin model we cannot introduce anymore the variables $Q_{\alpha\beta}$ by means of a Hubbard-Stratonovich transformation, as we did for the RFIM, Hopfield model and the SK model. The change of variables from spins to overlaps is pursued by exploiting the formal identity

$$1 = \int dQ_{\alpha\beta} \delta\left(NQ_{\alpha\beta} - \sum_{i=1}^{N} \sigma_i^\alpha \sigma_i^\beta\right), \tag{1.168}$$

where we have assumed the following normalization for the Dirac-delta integral

$$\int_{-\infty}^{\infty} dx\ \delta(x - x_0) = N, \tag{1.169}$$

so that

$$\int_{-\infty}^{\infty} dx\ \delta(Nx - x_0) = \frac{1}{N}\int_{-\infty}^{\infty} dx\ \delta\left(x - \frac{x_0}{N}\right) = 1 \tag{1.170}$$

To carry on easily the integration over the spins, it is worth recalling the integral representation of the Dirac delta,

$$\delta(x) = \int_{-\infty}^{\infty} \frac{dk}{2\pi}\ e^{ikx}. \tag{1.171}$$

For the purpose of the following calculation it is convenient to regard the integral on the right-hand side of Eq. (1.171) as an integral over a contour in the complex k plane, but rotated of $\pi/2$ radians (Wick rotation), which amounts to change variables from k to $s = e^{i\pi/2}k$, and then shift the (now parallel to the $Im(k)$ axis) contour to the point s_0, which is assumed to be on the *right* of any singularity in the complex k plane (any deformation of the integration contour which does not pass through a singular point is allowed by complex analysis). This means we can rewrite

$$\delta(x) = \int_{s_0-i\infty}^{s_0+i\infty} \frac{ds}{2\pi i}\ e^{sx}. \tag{1.172}$$

The need to account for a shift of the contour to a (still) unknown value of s_0 is due to the following reason. In general the Dirac delta is used within an expression of the kind,

$$\begin{aligned}\mathcal{I}(\beta) &= \int dx f_\beta(x)\delta(x)\\ &= \int_{s_0-i\infty}^{s_0+i\infty} \frac{ds}{2\pi i}\int dx f_\beta(x)e^{sx}\\ &= \int_{s_0-i\infty}^{s_0+i\infty} \frac{ds}{2\pi i} g(s),\end{aligned} \tag{1.173}$$

where $g(s)$ is, in full generality, a function in the complex s plane which might have singularities. In order for the Wick rotation to be done correctly the value s_0 must lie on the right of any singularity of $g(s)$. Since the analytic structure of $g(s)$ depends on $f_\beta(x)$, which is not yet intervening when the integral definition of $\delta(x)$ is introduced first, one must leave the value s_0 generic. By repeating the above steps with the Dirac delta appearing in Eq. (1.168) and taking into account that we need to introduce the integration over $n(n-1)/2$ elements of matrix $Q_{\alpha\beta}$, we get:

$$\begin{aligned}1 &= \int_{-\infty}^{\infty}\prod_{\alpha<\beta} dQ_{\alpha\beta}\prod_{\alpha<\beta}\delta\left(NQ_{\alpha\beta} - \sum_{i=1}^{N}\sigma_i^\alpha\sigma_i^\beta\right)\\ &= \int_{-\infty}^{\infty}\prod_{\alpha<\beta} dQ_{\alpha\beta}\int_{\lambda^0_{\alpha\beta}-i\infty}^{\lambda^0_{\alpha\beta}+i\infty}\prod_{\alpha<\beta} d\lambda_{\alpha\beta}\ \exp\left\{N\sum_{\alpha<\beta}\lambda_{\alpha\beta}Q_{\alpha\beta} - \sum_{\alpha<\beta}\lambda_{\alpha\beta}\sum_{i=1}^{N}\sigma_i^\alpha\sigma_i^\beta\right\},\end{aligned} \tag{1.174}$$

where the meaning of the extremes of integration for $\lambda_{\alpha\beta}$ variables should be clear from the previous discussion. Summarizing, we get:

$$\overline{Z^n} = \int \mathcal{D}Q\mathcal{D}\lambda\mathcal{D}\sigma \exp\left\{\frac{\beta^2 N}{2}\sum_{\alpha<\beta} Q_{\alpha\beta}^p + N\sum_{\alpha<\beta} Q_{\alpha\beta}\lambda_{\alpha\beta} - \sum_{\alpha<\beta}\lambda_{\alpha\beta}\sum_{i=1}^{N}\sigma_i^\alpha\sigma_i^\beta\right\}, \tag{1.175}$$

where we have used the abbreviation

$$\int_{-\infty}^{\infty}\prod_{\alpha<\beta} dQ_{\alpha\beta}\int_{\lambda^0_{\alpha\beta}-i\infty}^{\lambda^0_{\alpha\beta}+i\infty}\prod_{\alpha<\beta} d\lambda_{\alpha\beta} = \int \mathcal{D}Q\mathcal{D}\Lambda \tag{1.176}$$

From the expression in Eq. (1.175) it is clear that the integration over the spin variables can be easily carried out and amounts to N identical integrals. By recalling that $\sum_{\alpha<\beta} = \frac{1}{2}\sum_{\alpha,\beta=1}^{n}$ we can write

$$\begin{aligned}
&\int_{-\infty}^{\infty}\prod_{i=1}^{N}\prod_{\alpha=1}^{n} d\sigma_i^\alpha \ \exp\left\{-\frac{1}{2}\sum_{\alpha,\beta=1}^{n}\lambda_{\alpha\beta}\sum_{i=1}^{N}\sigma_i^\alpha\sigma_i^\beta\right\}\\
&= \prod_{i=1}^{N}\left[\int_{-\infty}^{\infty}\prod_{\alpha=1}^{n} d\sigma_i^\alpha \ \exp\left\{-\frac{1}{2}\sum_{\alpha,\beta=1}^{n}\lambda_{\alpha\beta}\sigma_i^\alpha\sigma_i^\beta\right\}\right]\\
&= \left[\int_{-\infty}^{\infty}\prod_{\alpha=1}^{n} d\sigma^\alpha \exp\left\{-\frac{1}{2}\boldsymbol{\sigma}^T\Lambda\ \boldsymbol{\sigma}\right\}\right]^N\\
&= \left(\frac{1}{\sqrt{\det\Lambda}}\right)^N, \qquad (1.177)
\end{aligned}$$

where $\boldsymbol{\sigma} = (\sigma^1, \ldots, \sigma^n)$ indicates a vector in replica space and Λ is the $n \times n$ matrix with elements $\lambda_{\alpha\beta}$. By finally exponentiating the determinant in Eq. (1.177) our replicated partition function has become an integral where, due to the N factor appearing in front of every term in the argument of the exponential, the possibility to complete the calculation with a saddle-point approximation seems really at hand. Indeed we have the expression

$$\overline{Z^n} = \int \mathcal{D}Q\mathcal{D}\Lambda \exp\left[-NS(Q,\Lambda)\right]\prod_{\alpha=1}^{n}\delta\left(1 - Q_{\alpha\alpha}\right), \tag{1.178}$$

with

$$S(Q,\Lambda) = -\frac{\beta^2}{4}\sum_{\alpha\beta=1}^{n} Q_{\alpha\beta}^p - \frac{1}{2}\operatorname{Tr}\left(Q\Lambda\right) + \frac{1}{2}\log\det\Lambda, \tag{1.179}$$

at which we arrive via the following identities (we have also used the symmetry property of the matrix Λ, or equivalently of Q):

$$\sum_{\alpha<\beta} Q_{\alpha\beta}\lambda_{\alpha\beta} = \frac{1}{2}\sum_{\alpha\beta=1}^{n} Q_{\alpha\beta}\lambda_{\alpha\beta} = \frac{1}{2}\sum_{\alpha\beta=1}^{n} Q_{\alpha\beta}\lambda_{\beta\alpha} = \frac{1}{2}\operatorname{Tr}(Q\Lambda) \tag{1.180}$$

Since we are interested in the large-N limit, we can evaluate the integration over the elements of Λ by means of a saddle point approximation,

$$\int \mathcal{D}\Lambda \exp\left[-NS(Q,\Lambda)\right] \approx \exp\left[-NS(Q,\Lambda^*[Q])\right], \tag{1.181}$$

where the symbol $\Lambda^*[Q]$ indicates the dependence of Λ on Q. The $\Lambda^*[Q]$ is found by the saddle-point equations

$$\frac{\partial S}{\partial \lambda_{\alpha\beta}} = 0 \quad \Longrightarrow \quad \Lambda^*[Q]. \tag{1.182}$$

By then recalling the identity

$$\log\det\Lambda = \mathrm{Tr}\log\Lambda, \tag{1.183}$$

and letting the derivation operator $\partial/\partial\lambda_{\alpha\beta}$ pass through the "Tr" operator for both the two terms $\mathrm{Tr}\,(Q\Lambda)$ and $\mathrm{Tr}\log\Lambda$ one can finally obtain the identity

$$Q_{\alpha\beta} = \frac{\partial}{\partial\lambda_{\alpha\beta}}\log\Lambda = (\Lambda^{-1})_{\alpha\beta}. \tag{1.184}$$

The reader has to recall at this point that the elements of the matrix Q are real numbers. This means (as soon as we set $\lambda^*_{\alpha\beta} = (Q^{-1})_{\alpha\beta}$) that the saddle-point of the function $\exp\{-NS(Q,\Lambda)\}$ with respect to each integration $\int_{\lambda^0_{\alpha\beta}-i\infty}^{\lambda^0_{\alpha\beta}+i\infty} d\lambda_{\alpha\beta}$ lies precisely at the intersection between the (vertical) integration contour and the real axis, i.e. a point which is shifted with respect to the origin. Finally, according to a somehow *a posteriori* argument, we have an explanation of why to choose $\lambda^0_{\alpha\beta} \neq 0$ in the in the integration symbol of Eq. (1.174) and that this quantity turns out to be precisely

$$\lambda^0_{\alpha\beta} = \lambda^*_{\alpha\beta} = (Q^{-1})_{\alpha\beta}. \tag{1.185}$$

As a conclusion we can simply replace Λ with Q^{-1} into the expression of $S(Q,\Lambda)$ of Eq. (1.179), getting (apart from terms constant with respect to Q and β)

$$\overline{Z^n} \approx \int \mathcal{D}Q \exp\left[-NS(Q)\right] \prod_{\alpha=1}^{n} \delta\left(1 - Q_{\alpha\alpha}\right) \approx \exp\left[-nNA(Q^*)\right], \tag{1.186}$$

with

$$A(Q) = \underbrace{-\frac{\beta^2}{4n}\sum_{\alpha\beta=1}^{n} Q^p_{\alpha\beta}}_{\text{energetic}} \underbrace{-\frac{1}{2n}\log\det Q}_{\text{entropic}}, \tag{1.187}$$

and where Q^* in the last term of Eq. (1.186) is the solution of the saddle-point equations

$$\frac{\partial A}{\partial Q_{\alpha\beta}} = 0 \quad \Longrightarrow \quad Q^*_{\alpha\beta}. \tag{1.188}$$

As we did in the SK model, also here in Eq. (1.187) we have emphasized which are the *energetic* and the *entropic* contributions to the free energy. Let us also stress that the matrix Q^* that satisfies the saddle-point equations must also satisfy the spherical constraint conditions implemented by the Dirac deltas in Eq. (1.186): whatever is the ansatz for the matrix Q^*, it must be one with all elements equal to 1 on the diagonal.

The free-energy of the system is retrieved by taking the limit $n \to 0$ of the replicated partition function:

$$\beta f = \lim_{\substack{n\to 0\\ N\to\infty}} -\frac{1}{nN}\left(\overline{Z^n} - 1\right) = \lim_{\substack{n\to 0\\ N\to\infty}} -\frac{1}{nN}\left(e^{-nNA(Q^*)} - 1\right) = \tag{1.189}$$

$$= \lim_{\substack{n\to 0\\ N\to\infty}} -\frac{1}{nN}\left(1 - nNA(Q^*) - 1\right) = A(Q^*) \tag{1.190}$$

1.5.2. Replica Symmetric Solution. In order to write the free energy in Eq. (1.187) one needs an assumption on the matrix $Q_{\alpha\beta}$. The simplest is a replica-symmetric assumption, i.e., all elements outside the diagonal are identical to each other and equal to q_0, whereas on the diagonal they are equal to 1:

$$Q_{\alpha\beta} = \mathrm{Diag}_n(1-q_0,\dots,1-q_0) + q_0\ \mathbf{1}_n \otimes \mathbf{1}_n, \tag{1.191}$$

where $\mathrm{Diag}_n(1-q_0,\dots,1-q_0)$ denotes a $n\times n$ diagonal matrix with all elements equal to $1-q_0$ on the diagonal, while $\mathbf{1}_n \otimes \mathbf{1}_n$ denotes an $n\times n$ matrix with *all* elements identical to 1.

In order to compute the free-energy one takes advantage of the known formula for the determinant of a matrix with elements identical outside the diagonal, i.e., a matrix A of the kind

$$A = \mathrm{Diag}_n(a_1-b,\dots,a_n-b) + b\ \mathbf{1}_n \otimes \mathbf{1}_n. \tag{1.192}$$

There is a general formula for the determinant of the matrix A, which reads:

$$\det(A) = \prod_{i=1}^{n}(a_i-b) + b\sum_{i=1}^{n}\prod_{\substack{j=1\\ j\neq i}}(a_j-b) \tag{1.193}$$

Adapting the result for the determinant of A to the replica symmetric matrix Q of Eq. (1.191) one gets

$$\det(Q) = (1-q_0)^n + nq_0(1-q_0)^{n-1} = (1-q_0)^n\left(1+n\frac{q_0}{1-q_0}\right), \tag{1.194}$$

so that

$$\begin{aligned}\lim_{n\to 0}\frac{1}{n}\log\det(Q) &= \lim_{n\to 0}\frac{1}{n}\left[n\log(1-q_0)+\log\left(1+n\frac{q_0}{1-q_0}\right)\right]\\ &= \log(1-q_0)+\frac{q_0}{1-q_0}\end{aligned} \tag{1.195}$$

Much more easily we obtain the *energetic* contribution in the limit $n\to 0$:

$$\sum_{\alpha\beta=1}^{n} Q_{\alpha\beta}^p = n + n(n-1)q_0^2. \tag{1.196}$$

Putting together the pieces we thus have:

$$\begin{aligned}f &= \lim_{n\to 0}\ -\frac{1}{\beta}\left\{\frac{\beta^2}{4n}\sum_{\alpha\beta=1}^{n}Q_{\alpha\beta}^p + \frac{1}{2n}\log\det Q\right\}\\ &= -\frac{1}{2\beta}\left\{\frac{\beta^2}{2}\left[1-q_0^p\right]+\log(1-q_0)+\frac{q_0}{1-q_0}\right\}.\end{aligned} \tag{1.197}$$

The value of the free energy in the replica symmetric phase can be finally obtained by replacing q_0 in Eq. (1.197) with the value q_0^* which satisfies the saddle point equation $\partial f/\partial q_0 = 0$. A peculiarity of the number of replicas $n\to 0$ is that the correct q_0^* in this limit *maximizes* rather than *minimizes* the expression in Eq. (1.197). This can be intuitively understood in light of the *change of sign* of the energetic contribution implied by the limit $n\to 0$. Very heuristically, one has that the analytic continuation to values $n<1$ implies

$$-\frac{1}{4\beta}Q_{\alpha\beta}^p \quad\Longrightarrow\quad \frac{1}{4\beta}q_0^p. \tag{1.198}$$

Then, we must recall the general thermodynamic relation

$$F = U - TS \tag{1.199}$$

where F is free energy, U the internal energy and TS the temperature times entropy. From Eq. (1.199) it is clear that the *convexity* of free energy function must have the same sign as that of the energy, thus a change of sign of the latter implies the same for the former.

1.5.3. Replica Symmetry Breaking. It is well known that the p-spin model has a critical temperature T_K at which ergodicity is broken: phase space splits in a multiplicity of disjoint ergodic components corresponding to different free-energy minima. In particular, what happens at T_K is that the free-energy obtained with a one-step replica symmetry breaking (1-RSB) ansatz becomes lower than the replica-symmetric energy. Not only, but it can be checked that the 1-RSB saddle point is a *stable one*, so that no further levels of breakings of the permutation symmetry between replicas are needed to characterize the glassy phase at $T < T_K$.

The typical structure of the matrix order parameter $Q_{\alpha\beta}$ under the 1-RSB hypothesis consists of a square block circulant matrix made of n/m blocks which are themselves $m \times m$ matrices of two kinds: $A_{ij} = \delta_{ij} + (1 - \delta_{ij})\, q_1$ and $B_{ij} = q_0$, where the block structure reads as

$$Q = \begin{pmatrix} A & B & \cdots & B \\ B & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & B \\ B & \cdots & B & A \end{pmatrix} \tag{1.200}$$

More in detail one has something of the kind

$$Q = \begin{pmatrix} 1 & q_1 & q_1 & & & & & & \\ q_1 & 1 & q_1 & & & & & q_0 & \\ q_1 & q_1 & 1 & & & & & & \\ & & & 1 & q_1 & q_1 & & & \\ & & & q_1 & 1 & q_1 & & & \\ & & & q_1 & q_1 & 1 & & & \\ & & & & & & \ddots & & \\ & & & & & & & 1 & q_1 & q_1 \\ & q_0 & & & & & & q_1 & 1 & q_1 \\ & & & & & & & q_1 & q_1 & 1 \end{pmatrix}, \tag{1.201}$$

where we have

$$\begin{cases} Q_{\alpha\alpha} = 1 & \\ Q_{\alpha\beta} = q_0 & \text{(different free-energy minima)} \\ Q_{\alpha\beta} = q_1 & \text{(same free-energy minimum)} \end{cases} \tag{1.202}$$

so that in each row one has $(n - m)$ elements equal to q_0 and $(m - 1)$ elements equal to q_1. In the matrix in Eq.(1.201) the choice $m = 3$ is just for representation purposes.

The calculation of the energetic contribution to the free-energy from the 1-RSB ansatz is quite easy. From the above considerations on the number of elements equal to q_0 and

the number of elements equal to q_1 in each row and taking into account that all the n rows provide the same contribution we have

$$\lim_{n\to 0} \; -\frac{\beta^2}{4n}\sum_{\alpha\beta} Q^p_{\alpha\beta} = \lim_{n\to 0} \; -\frac{\beta^2}{4n}\cdot n\left[1+(m-1)q_1^p+(n-m)q_0^p\right]$$

$$= -\frac{\beta^2}{4}\left[1-mq_0^p-(1-m)q_1^p\right] \tag{1.203}$$

Then, in order to evaluate the *entropic* contribution $\log\det(Q)$, one needs to compute the eigenvalues of $Q_{\alpha\beta}$ within the 1-RSB ansatz and their corresponding multiplicities. From the calculation shown in the appendix here below it turns out that $Q_{\alpha\beta}$ has three different eigenvalues, with different multiplicities, which read:

$$\begin{aligned}
&\lambda_1 = 1-q_1 && \text{multiplicity} = n-\frac{n}{m}\\
&\lambda_2 = m(q_1-q_0)+(1-q_1) && \text{multiplicity} = \frac{n}{m}-1\\
&\lambda_3 = nq_0+m(q_1-q_0)+(1-q_1) && \text{multiplicity} = 1
\end{aligned}$$

so that

$$\log\det(Q) = n\left(1-\frac{1}{m}\right)\log(1-q_1)+\left(\frac{n}{m}-1\right)\log\left[1-mq_0-(1-m)q_1\right]+$$
$$+\log[nq_0+1-mq_0-(1-m)q_1]. \tag{1.204}$$

Then in the limit $n\to 0$ we have, using the properties of the logarithm,

$$\begin{aligned}
&\lim_{n\to 0} \; \log[1-mq_0-(1-m)q_1+n\;q_0] =\\
&\lim_{n\to 0} \; \log[1-mq_0-(1-m)q_1]+\log\left(1+n\;\frac{q_0}{1-mq_0-(1-m)q_1}\right) =\\
&= \log[1-mq_0-(1-m)q_1]+n\;\frac{q_0}{1-mq_0-(1-m)q_1},
\end{aligned} \tag{1.205}$$

The contribution $\log[1-mq_0-(1-m)q_1]$ not proportional to n coming from the last term added on the right hand side of Eq. (1.204) cancels with the same term with opposite sign coming from the second addendum on the right of Eq. (1.204). In conclusion, as expected within a 1-RSB scheme, we have that the entropic term is proportional to n,

$$\log\det(Q) = n\Big[\frac{m-1}{m}\log(1-q_1)+\frac{1}{m}\log\left(1+n\;\frac{q_0}{1-mq_0-(1-m)q_1}\right)+$$
$$+\frac{q_0}{1-mq_0-(1-m)q_1}\Big] \tag{1.206}$$

Putting together the pieces, in the limit $n\to 0$ the free energy from the 1-RSB ansatz formally reads as:

$$-2\beta F = \frac{\beta^2}{2}\left[1-mq_0^p-(1-m)q_1^p\right]+\frac{m-1}{m}\log(1-q_1)+$$
$$+\frac{1}{m}\log(1-mq_0-(1-m)q_1)+\frac{q_0}{1-mq_0-(1-m)q_1}. \tag{1.207}$$

The expression in Eq. (1.207) is of course a formal one for the reason that we have not yet decided which physical meaning to attach to the number m. Indeed, for integer n, the parameter m is such that $1<m\leq n$ ... but what about the limit $n\to 0$?

1.5.4. The breaking parameter m. In order to provide a correct interpretation for m in the limit $n \to 0$ let us stick on its meaning when n is integer and large. Already in the case of the SK model we have mentioned that *"the number m_k of times that the value q_k appears in a row of the matrix $Q_{\alpha\beta}$ is proportional to its probability"*. Being $q^*_{\alpha\beta}$ a solution of the saddle-point equations we have thus that a meaningful definition of the overlap probability reads as

$$\overline{P_J(q)} = \frac{2}{n(n-1)} \sum_{\alpha<\beta} \delta\left(q - q^*_{\alpha\beta}\right), \tag{1.208}$$

where the average over the disorder for $P_J(q)$ is mandatory for two reasons: 1) first, the probability of the overlap is not a self-averaging quantity (not proved here), which means that even in the $N \to \infty$ limit it depends on the instance of random couplings, hence the subscript J is needed; 2) second, since the 1-RSB (or whatever k-RSB) ansatz for $Q_{\alpha\beta}$ is made inside formulas coming *when the free energy has been already averaged over the disorder*, for consistency the overlap matrix $q^*_{\alpha\beta}$ chosen for the saddle point must be related to the the *disorder average* of the overlap probability distribution $\overline{P_J(q)}$. This said, let us show how $\overline{P_J(q)}$ reads explicitly according to its definition in Eq. (1.208) in the case of a 1-RSB ansatz:

$$\begin{aligned} \overline{P_J(q)} &= \frac{1}{n(n-1)} \sum_{\alpha\neq\beta} \delta\left(q - q^*_{\alpha\beta}\right) \\ &= \frac{m-1}{n-1}\delta(q-q_1) + \frac{n-m}{n-1}\delta(q-q_0). \end{aligned} \tag{1.209}$$

From the expression in Eq. (1.209) we have that the *naïve* way to take the limit $n \to 0$, i.e., sending n to zero while leaving m untouched, which seems in contradiction with the inequality $m < n$, leads to:

$$\lim_{n\to 0} \overline{P_J(q)} = m\ \delta(q-q_0) + (1-m)\ \delta(q-q_1) \tag{1.210}$$

We are thus at the end. From the expression of $\overline{P_J(q)}$ in Eq. (1.210) it is clear that in order to have it well defined as a normalized probability distribution there is only one possible interpretation for the parameter m: it must be a number in the interval $[0, 1]$.

We have thus learned that the limit $n \to 0$ implies a kind of *switch* in the definition of m, i.e., from a natural number in the interval $1 < m \leq n$ to a real number in the interval $m \in [0, 1]$. At this stage, in order to completely solve the thermodynamics of the p-spin we just have to fix q_0, q_1 and m at the different temperatures.

1.5.5. The critical temperatures T_K and T_d. Now that the nature of the variational parameter m has been established we have to look for solutions of the 1-RSB equations:

$$\frac{\partial f}{\partial q_0} = 0, \qquad \frac{\partial f}{\partial q_1} = 0, \qquad \frac{\partial f}{\partial m} = 0. \tag{1.211}$$

The usual strategy is to solve the first two equations at fixed m, and then plug m, $q_0(m)$ and $q_1(m)$ into the free-energy, inspecting numerically for which value $m = m^*$ it has a *maximum*. Clearly m, q_0 and q_1 are variational parameters that depend on the temperature. A peculiarity of the p-spin model is that, at variance with the SK, the study of the mean-field equations for m, q_0 and q_1 tells us that there are two critical temperatures, T_d and T_K with $T_d > T_K$, which represent respectively the critical temperature for *dynamical ergodicity breaking* (T_d) and the critical temperature for *thermodynamic* ergodicity

breaking (T_K), also known as the *ideal glass transition* temperature.

Before studying the saddle-point equations there is an aspect worth noticing, namely that by plugging $m = 1$ into the 1-RSB free energy of Eq. (1.207) we get back exactly the RS free energy of Eq. (1.197). It was therefore more economic to study directly the 1-RSB free energy, keeping in mind that the RS free energy is its special case.

Summarizing, from the study of the 1-RSB saddle-point equations in (1.211) there are three interesting regimes emerging:

- $T > T_d$: The saddle-point equations admit only one trivial solution: $q_1 = q_0 = 0$ and $m = 1$. This is the ergodic replica symmetric phase. There are no free energy minima in the landscape.
- $T_K < T < T_d$: There is still the trivial solution that we have for $T > T_d$ but also a new non-trivial one appears: $q_1 > q_0$ and $m = 1$. This solution, which has the same free-energy of that at $T > T_d$, is the signature that *many* free-energy minima separated by extensive *energetic* barriers arise as soon as $T < T_d$. For a detailed study of this regime, a more advanced course is needed. Here we can only say (without proving it) that the number of such minima at a given free energy, $\mathcal{N}(f)$, is found to be *exponentially large* in the size of the system, i.e.

$$\mathcal{N} \approx \exp\left[N\Sigma(f)\right], \tag{1.212}$$

  where $\Sigma(f)$ is a quantity known in the jargon of glassy systems as *configurational entropy*. One can attach to these minima the notion of *states*, much like the positive or negative magnetization *states* of a ferromagnet at low temperature. The reason for this is that, due to the infinite energy barrier separating them, any dynamics starting with an initial condition in one of these states remains trapped there forever (no tunneling is allowed, at least in a men-field model). They thus represent *disjoint ergodic components* of phase space, a sort of prelude of the ergodicity breaking transition taking place at the lower temperature $T_K < T_d$. Nevertheless, from a thermodynamic point of view, states are irrelevant above T_K since the probability that the system ends up into one of them is exponentially small, precisely in force of of their exponential abundance, i.e.,

$$p_{state} \sim \exp\left[-N\Sigma(f)\right] \tag{1.213}$$

  This thermodynamic *irrelevance* of the states above T_K is reflected precisely by the fact that the solution with $q_1 > q_0$ and $m = 1$ has the same free-energy of the one with $q_0 = q_1 = 0$ and that if we ask for the probability to find two configurations at overlap $q_1 > 0$ this is zero even according to the definition of $\overline{P}_J(q)$ given in Eq. (1.210).
- $T < T_k$: In this regime we find both the solution with $q_0 = q_1 = 0$ and $m = 1$ and the replica-symmetry-broken one with $q_1 > q_0$ and $m < 1$, corresponding to finding with finite probability two configurations of the system at finite overlap, see Eq. (1.210). In particular, if one compares the free energy of the RS solution and of the 1-RSB one, when $T < T_K$ the 1-RSB solution has a smaller free energy,

$$T \leq T_K \quad \Longrightarrow \quad f(q_1 > q_0, m < 1) \; < \; f(q_1 = q_0 = 0, m = 1). \tag{1.214}$$

  Hence, for temperatures below T_K, thermodynamic equilibrium is determined by the glass phase, where the systems remains trapped forever in one of the states. The ergodicity-breaking transition taking place in the p-spin model at T_K is usually known as the *ideal glass transition* or, more precisely the *Random First-Order Transition* (RFOT). The RFOT name comes from the fact that the ergodicity-breaking

transition at T_K has a mixed character between a first and second order transition. This happens because the distribution $\overline{P_J(q)}$ is bimodal already at the transition temperature $T = T_K$, with a secondary peak at a finite distance from the origin, corresponding to $q_1 > q_0$ already at $T = T_K$ [see Eq.(1.210)]—a feature typical of a *first-order* transition—while at the same time for the transition at T_K there is no latent heat, much like a *second-order* or *continuous* transition. "Random First-Order" was therefore introduced to refer to something which is a kind of "smoothed first-order transition".

1.5.6. Appendix. Consider an $m \times m$ matrix:

$$\begin{pmatrix} \alpha & \beta & \cdots & \beta \\ \beta & \alpha & \cdots & \beta \\ \vdots & \vdots & \ddots & \vdots \\ \beta & \beta & \cdots & \alpha \end{pmatrix}$$

Its generic element can be written as:

$$c_{ij} = \alpha\delta_{ij} + \beta\,(1 - \delta_{ij}) \tag{1.215}$$

The eigenvalues equation reads:

$$\sum_{j=1}^{n} c_{ij}v_j = \sum_{j=1}^{n} \alpha\delta_{ij}v_j + \beta\,(1 - \delta_{ij})\,v_j = \lambda v_i \tag{1.216}$$

$$(\alpha - \beta)\,v_i + \beta \sum_{j=1}^{m} v_j = \lambda v_i \tag{1.217}$$

If $\sum_{j=1}^{n} v_j = 0$ (can be done in $m - 1$ ways by choosing different values for v_j):

$$\lambda = \alpha - \beta$$

else sum over i (and $\sum_{j=1}^{m} v_j \neq 0$):

$$(\alpha - \beta) \sum_{i=1}^{m} v_i + \beta \sum_{i=1}^{m} \sum_{j=1}^{m} v_j = \lambda \sum_{i=1}^{m} v_i \tag{1.218}$$

$$(\alpha - \beta) \sum_{i=1}^{m} v_i + m\beta \sum_{i=1}^{m} v_i = \lambda \sum_{i=1}^{m} v_i \tag{1.219}$$

divide by $\sum_{j=1}^{m} v_j$:

$$\lambda = \alpha + (m - 1)\beta$$

So we have $\lambda_1 = \alpha - \beta$ with $\deg\lambda_1 = m - 1$ and $\lambda_2 = \alpha + (m - 1)\beta$ with $\deg\lambda_2 = 1$. The replica symmetric matrix is a block circulant matrix where the first row block is:

$$\left(A \underbrace{B \ldots B}_{f - 1 \text{ times}} \right)$$

$$A_{ij} = \delta_{ij} + q_1\,(1 - \delta_{ij})$$

$$B_{ij} = q_0$$

$$f = n/m$$

where n is the number of replicas. The eigenvalues of A are:

$$\lambda_{1A} = 1 - q_1 \quad \deg_1 = m - 1$$

$$\lambda_{2A} = 1 + (m - 1)q_1 \quad \deg_2 = 1$$

The eigenvalue of B are:

$$\lambda_{1B} = 0 \quad \deg_1 = m - 1$$
$$\lambda_{2B} = mq_0 \quad \deg_2 = 1$$

Now A and B commute and they are the same kind of circulant matrix of first row $(\alpha, \beta, \ldots, \beta)$ solved above. The block matrix is the block analogue of this kind of circular matrix and because A and B commute the eigenvalues of this matrix have the same form when we make the substitution $\alpha \to \lambda_A$ and $\beta \to \lambda_B$; there is only one requirement: we can mix only eigenvalues having the same set of eigenvectors. In the end we get

$$\lambda_A = 1 - q_1 \quad \lambda_B = 0 \tag{1.220}$$

$$\lambda_1 = \lambda_A - \lambda_B = 1 - q_1 \tag{1.221}$$

$$\lambda_1 = \lambda_A + (f-1)\lambda_B = 1 - q_1 \tag{1.222}$$

and

$$\lambda_A = 1 + (m-1)q_1 \quad \lambda_B = mq_0 \tag{1.223}$$

$$\lambda_2 = \lambda_A - \lambda_B = 1 + (m-1)q_1 + mq_0 = 1 - q_1 + m(q_1 - q_0) \tag{1.224}$$

$$\lambda_3 = \lambda_A + (f-1)\lambda_B = 1 - q_1 + m(q_1 - q_0) + nq_0 \tag{1.225}$$

The eigenvalue λ_2 has degeneracy $f - 1 = \frac{n}{m} - 1$ and λ_3 has 1; since it is assured that the matrix is diagonalizable the degeneracy of λ_1 is $n - \frac{n}{m}$.

Problems

Following the notation of the lecture notes:

Exercise 1.1. Consider the Random Field Ising Model (RFIM), in which the disorder has variance δ^2. Proceed from (1.41) to arrive at the formula where the number n of replicas appears explicitly in the magnetization m,

$$m = \frac{1}{Z_1(m)} \int \frac{d\nu}{\sqrt{2\pi}} \exp\left[\frac{1}{2}\nu^2 + n \ln 2 \cosh(2\beta J m + \beta\delta\nu)\right] \tanh(2\beta J m + \beta\delta\nu)$$

Exercise 1.2. With the self-consistent solution $m_{SC}(m) = m$ of the RFIM, by using the condition $\partial m_{SC}/\partial m = 1$ for the critical point, show that the phase transition between paramagnetic phase and ferromagnetic phase takes place where this condition is satisfied:

$$2\beta J \int dh\, p(h) \frac{1}{[\cosh(\beta h)]^2} = 1 \tag{1.226}$$

Exercise 1.3. Show that at zero temperature in the RFIM there is a disorder-driven para-ferromagnetic transition where the random field standard deviation δ and the coupling J satisfy $2J/\delta = \sqrt{\pi/2}$. For simplicity one may take $\delta = 1$.

Exercise 1.4. For the matrix

$$A = \begin{bmatrix} 5 & \sqrt{3} \\ \sqrt{3} & 3 \end{bmatrix}$$

derive that

$$\log(A) = \begin{bmatrix} \log\left(2 \times 3^{3/4}\right) & \frac{\sqrt{3}}{4}\log(3) \\ \frac{\sqrt{3}}{4}\log(3) & \frac{1}{4}\log(48) \end{bmatrix}$$

Moreover, verify that $\log(\det(A)) = \mathrm{Tr}(\log(A)) = \log(12)$. Note that the log of a matrix is not the matrix of the logarithms of its elements. Hint: look for how to compute the log of *symmetric* matrices.