0.1 Disordered Systems - Ergodicity Breaking without Symmetry Breaking

References:

1. "Spin glass theory and beyond" (1987), Marl Mézard, Giorgio Parisi, M. A. Virasoro, *very technical, difficult to follow*

2. "Information, Theory, Computation" (2009), Marl Mézard, Andrea Montanari, *more clear*

3. "Statistical Physics of Spin Glasses and Information Processing" (2001), Hidetoshi Nishimori, *very clear, today's lecture comes from one of its chapter*

4. "Random fields and spin glasses" (2006), Irene Giardina, Cirand De Dominicis, *sometime not rigorous, requires attention*

Reviews Papers:
(*Applications of disordered system results to Condensed Matter*)

1. "Theoretical Perspective on the glass transition and amorphous materials" Ludovic Berthier, Giulio Biroli, Rev. Mod. Phys 83 (2011)

2. "Supercooled Liquids for Pedestrians", Andrea Cavagna, Phys. Rep 476 (2009)

(Lesson ? of 05/12/19) Compiled: December 5, 2019

Disordered Systems are important in physics because they are examples of systems that exhibit *ergodicity breaking without symmetry breaking.*
There will be 4 lectures for this part, following this outline:

1. Neural Network

2. Sherrington Kirkpatric model, p-spin

3. Franz-Parisi Potential, calculations $\sim$ p-spin

0.2 Introduction

In physics, usually we deal with potentials like the following:

- **Harmonic potential** with a single **global** minimum:

$$\mathcal{E} = \frac{1}{2}m\dot{x}^2 + V(x) \qquad V(x) = \frac{1}{2}kx^2$$

- **Bistable potential** with two equivalent minima:

$$\mathcal{E} = K + V(x) \qquad V(x) = \frac{a}{2}x^2 + \frac{b}{4}x^4$$

However, when examining a real macroscopic system in statistical mechanics we deal with more complicated potentials. For example, consider $N \sim 10^{23}$ particles, with positions $\{\boldsymbol{x}_1, \dots, \boldsymbol{x}_N\}$ ($\boldsymbol{x}_i \in \mathbb{R}^3$), one way to model *close-range* interactions is thanks to the following:

$$V(\boldsymbol{x}_i, \boldsymbol{x}_j) = \left(\frac{\sigma_{ij}}{|\boldsymbol{x}_i - \boldsymbol{x}_j|}\right)^{12} - \left(\frac{\sigma_{ij}}{|\boldsymbol{x}_i - \boldsymbol{x}_j|}\right)^{6}$$

where σ_{ij} are random numbers, with a distribution $p(\sigma_{ij}) \sim \sigma_{ij}^{-3}$.
In these cases, we have *many* ($\sim e^{N^2}$) local minima, both for the energy $\mathcal{E}$ and the *free energy* $F = \mathcal{E} - TS$. While theoretically we could *label* each minima, macroscopically we cannot distinguish them, because they "all look the same": this is the main problem of studying *disordered systems*.
Consider now a set of interacting particles with spins $\boldsymbol{S} = \{S_1, \dots, S_N\}$. The average $\langle S_i \rangle$ measures the overall *magnetization* of the system: if $\langle S_i \rangle = 0$ we are in the *disordered state*, while if $\langle S_i \rangle \neq 0$ there is some *preferred* alignment of the spins.
Suppose the potential has many local minima, labelled with greek letters: α, β, γ, etc. We can try to distinguish them by the value of $\langle S_i \rangle$. For example, if we focus on α and δ, supposing that:

$$\langle S_i \rangle_\alpha \neq \langle S_i \rangle_\delta$$

by measuring $\langle S_i \rangle$ we can know if the system is in α or δ. The average *over a local minimum* is defined as the following:

$$\langle \cdots \rangle = \frac{1}{Z_\alpha} \sum_{\{\boldsymbol{S} \in \alpha\}} \left(e^{-\beta H[\boldsymbol{S}]} S_i\right) \qquad Z_\alpha = \sum_{\boldsymbol{S} \in \alpha} e^{-\beta H[\boldsymbol{S}]}$$

where $\boldsymbol{S} \in \alpha$ means a sum over all possible collections of spins that result in the same minimum α for the potential.
When doing this calculation in practice, however, we find that:

$$\langle S_i \rangle_\alpha = \langle S_i \rangle_\beta = \cdots = 0$$

meaning that this approach yields no result.
We can use this qualitative result to *quantify* whether a system is in a *ergodic* or *non-ergodic* phase, depending on how much the local minima *overlap* with each other. So, we introduce an **order parameter** for disordered systems called **overlap**, and defined as following:

$$q^{\alpha\beta} = \frac{1}{N} \sum_{i=1}^{N} S_i^\alpha S_i^\beta \qquad \begin{aligned} \boldsymbol{S}^\alpha &= \{S_1^\alpha, \dots, S_N^\alpha\} \\ \boldsymbol{S}^\beta &= \{S_1^\beta, \dots, S_N^\beta\} \end{aligned}$$

Then:

- If the system is in a **ergodic phase** (high temperature) we have $\langle q^{\alpha\beta} \rangle = 0$
- If the system is in a **non-ergodic phase** (low temperature) then $\langle q^{\alpha\beta} \rangle \neq 0$

0.3 Neural Network

Consider a network of units, **neurons**, connected by **synapses**. We denote the state of each neuron with a *spin*-like number $S_i = \{-1, +1\}$ with the following meaning:

- $S_i = +1$: the i-th neuron is *excited*
- $S_i = -1$: the i-th neuron is *at rest*

Connections between neurons have a certain *weight* J_{ij} , called **synaptic efficacy**. Each neuron receives an *input impulse* equal to the *activity of all other neurons* weighted by the *strength of their connection* to that neuron:

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

Let's limit J_{ij} to only two values:

$$J_{ij} = \begin{cases} +1 & \text{Excitatory synapse} \\ -1 & \text{Inhibitory synapse} \end{cases}$$

meaning that neurons can contribute to *activate* or *turn off* other neurons.
Then, we introduce a *time evolution* in the system, following the **dynamic rule**:

$$S_i(t+1) = \operatorname{sgn}(h_i(t) - \theta_i^*) \tag{1}$$

This means that the i-th neuron's state at time $t + 1$ will be *excited* if the input impulse is greater than a threshold θ_i^*, or *at rest* otherwise.
If we make the simplifying assumption that:

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

the dynamic rule becomes:

$$(1) = \operatorname{sgn}\Big(\underbrace{\sum_{j=1}^{N} J_{ij}(1 + S_j(t))}_{h_i(t)} - \underbrace{\sum_{j=1}^{N} J_{ij}}_{\theta_i^*} \Big) \Rightarrow S_i(t+1) = \operatorname{sgn}\left(\sum_{j=1}^{N} J_{ij} S_j(t) \right)$$

We then assume that *all neurons* are connected to each other, meaning that:

$$J_{ij} \neq 0 \qquad \forall (i, j) \text{ such that } i \neq j$$

and we prohibit *self connections*:

$$J_{\iota} = 0 \text{ (Hebb rule)} \tag{2}$$

We then need a rule to choose J_{ij} with $i \neq j$. The idea is that a neural network's purpose is to *store patterns*. We define a **pattern** as a vector of spins:

$$\boldsymbol{\xi}^\mu = \{\xi_1^\mu, \dots, \xi_N^\mu\} \qquad \xi_i^\mu = \{+1, -1\}$$

Suppose we have p patterns ($\mu = 1, \dots, p$), and we want to store them in the neural network. We then define:

$$J_{ij} = \frac{1}{N} \sum_{\mu=1}^{p} \xi_i^\mu \xi_j^\mu \tag{3}$$

This is so that ξ^μ are all *fixed point* of the activation dynamics, meaning that neurons storing these patterns *keep them* during the time evolution:

$$S_i(t) = \xi_i^\mu \Rightarrow S_i(t+1) = \xi_i^\mu$$

That is, ξ_i^μ are *solutions* of the dynamic rule equation:

$$\xi_i^\mu = \operatorname{sgn}\left(\sum_{j=1}^{N} J_{ij} \xi_j^\mu\right)$$

In fact:

$$\operatorname{sgn}\left(\sum_{j=1}^{N} J_{ij}\xi_j^\mu\right) \underset{(3)}{=} \operatorname{sgn}\left(\sum_{j=1}^{N} \frac{1}{N} \sum_{\nu=1}^{p} \xi_i^\nu \xi_j^\nu \xi_j^\mu\right) = \operatorname{sgn}\left(\sum_{\nu=1}^{p} \underbrace{\left(\frac{1}{N}\sum_{j=1}^{N} \xi_j^\mu \xi_j^\nu\right)}_{\delta_{\mu\nu} + O(1/\sqrt{N})}\right) \underset{(a)}{=}$$

$$= \operatorname{sgn}\left(\sum_{\nu=1}^{p} \xi_i^\nu \delta_{\mu\nu}\right) = \operatorname{sgn}(\xi_i^\mu) \underset{(a)}{=} \xi_i^\mu$$

in (a) the idea is that we are taking the average of $\xi_j^\mu \xi_j^\nu$. As $\xi_j^{\mu,\nu} = \{\pm 1\}$, if they are two *different (uncorrelated) vectors*, that mean will tend to 0, and it's exactly 1 if $\xi_j^\mu = \xi_j^\nu$ for all j. Then in (b) we used the fact that $\xi_i^\mu = \{\pm 1\}$. Also, here we are assuming that $p/N \xrightarrow[N\to\infty]{} 0$.

The constraint (3) together with Hebb's rule (2) form the **Hopfield Model** for a neural network.
If we now define the **energy** of the network as:

$$\mathcal{E} \equiv -\frac{1}{2} \sum_{ij} S_i S_j J_{ij} = -\frac{1}{2} \sum_i S_i h_i$$

We can choose the initial spin configuration with a Boltzmann distribution:

$$p(S_1, \dots, S_N) = \frac{1}{Z} \exp\left(-\beta \mathcal{E}(S_1, \dots, S_N)\right) \qquad \beta = \frac{1}{T}$$

The energy has *lots* of minima, each one corresponding to a different *pattern* stored in the network. At low energy, the network dynamics will make the network *converge* to the closest minimum.

We now show search the minima of the *free energy*. First, the partition function Z is defined as:

$$Z \equiv \sum_{\{S_1,\ldots,S_N\}} \exp(-\beta\mathcal{E}(S_1,\ldots,S_N)) \tag{4}$$

and the **free energy** f:

$$f = -\frac{1}{N\beta}\log(Z)$$

We want to show that the "p" patterns embedded in J_{ij} are p stationary points of the free energy.
Inserting (3) in (4) leads to:

$$Z = \sum_{\{\boldsymbol{S}\}} \exp\left(\frac{\beta}{2N}\sum_{i,j=1}^{N} S_i S_j \sum_{\mu=1}^{p} \xi_i^\mu \xi_j^\mu\right)$$

Note that, *to be precise*, we should use:

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

However, this would lead to harder computations. So we *omit* the δ_{ij}, letting the diagonal elements be non-zero. This can be shown to be a *good approximation* of the general case (but we will not prove it).
Then, using:

$$\sum_{i,j=1}^{N} x_i x_j = \left(\sum_{j=1}^{N} x_i\right)^2$$

leads to:

$$Z = \sum_{\{\boldsymbol{S}\}} \exp\left[\frac{\beta}{2N}\sum_{\mu=1}^{p}\left(\sum_{i=1}^{N}\xi_i^\mu S_i\right)^2\right]$$

To deal with the square we use the Hubbard-Stratonovich transformation (which is just the *inverse of the completation of the square*):

$$\exp\left(\frac{\beta}{2N}x^2\right) = \int_{-\infty}^{+\infty} \mathrm{d}q \exp\left(-\frac{1}{2}N\beta q^2 + \beta q x\right)$$

so that:

$$Z = \sum_{\{\boldsymbol{S}\}} \int_{-\infty}^{+\infty} \prod_{\mu=1}^{p} \mathrm{d}q_\mu \exp\left[-\frac{1}{2}N\beta\sum_{\mu=1}^{p} q_\mu^2 + \beta\sum_{\mu=1}^{p} q_\mu \sum_{i=1}^{N}\xi_i^\mu S_i\right]$$

Note that now S_i appears in a *linear term*. We can finally compute the sum, term by term:

$$\sum_{S_i=\{+1,-1\}} \exp\left[\beta\left(\sum_{\mu=1}^{p} q_\mu \xi_i^\mu\right) S_i\right] = \exp\left(\beta\sum_{\mu=1}^{p} q_\mu \xi_i^\mu\right) + \exp\left(-\beta\sum_{\mu=1}^{p} q_\mu \xi_i^\mu\right) =$$

$$= 2\cosh(\beta \boldsymbol{q}\cdot\boldsymbol{\xi}_i) \qquad \Big| \, \boldsymbol{q}\cdot\boldsymbol{\xi}_i = \sum_{\mu=1}^{p} q_\mu \xi_i^\mu$$

where the *greek* coordinates denote the *patterns*, and the *latin* ones the *spins* inside each pattern.

$$= \exp\left(log(2\cosh(\beta \boldsymbol{q}\cdot\boldsymbol{\xi}_i))\right)$$

Substituting back:

$$Z = \int_{-\infty}^{+\infty} \prod_{\mu=1}^{p} \mathrm{d}p_\mu \exp\left[-\beta N u(q_1,\ldots,q_p)\right]$$

$$u(\boldsymbol{q}) = \frac{1}{2}\sum_{\mu=1}^{p} q_\mu^2 - \frac{1}{\beta N}\sum_{i=1}^{N} \ln(2\cosh(\beta \boldsymbol{q}\cdot\boldsymbol{\xi}_i))$$

All that's left is to compute the *stationary points* of $u(\boldsymbol{q})$, that is the points satisfying:

$$\left(\frac{\partial u}{\partial q_1}, \frac{\partial u}{\partial q_2}, \ldots, \frac{\partial u}{\partial q_2}\right) = \mathbf{0}$$