Lecture-09: The Curie-Weiss Model, and a Word on Combinatorial Optimization

1 The Curie-Weiss (or Mean-Field) Model

Last lecture, we looked at the Ising model, in which the N-fold configuration space, $\mathcal{X}_N$, was a set of points on a d-dimensional lattice, i.e.,

$$\mathcal{X}_N = \{-1,+1\}^{\mathbb{L}}, \text{ for } \mathbb{L} = \{1,2,\ldots,L\}^d.$$

Further, the energy function for a configuration $\boldsymbol{\sigma} \in \mathcal{X}_N$ was given by:

$$E(\boldsymbol{\sigma}) = -\sum_{i\sim j}\sigma_i\sigma_j - \sum_{i=1}^{N} B\sigma_i,$$

where $i \sim j$ indicates that the co-ordinate i is adjacent to the co-ordinate j, for $i, j \in \mathbb{L}$. Note that the summation in the first term counts the product, $\sigma_i\sigma_j$, for $i \sim j$ exactly once. We shall now look at yet another setting—the **Curie-Weiss** model. As before, the configuration space, $\mathcal{X}$, is the set $\{-1,+1\}$, and $\mathcal{X}_N$ denotes the N-fold Cartesian product of $\mathcal{X}$ with itself. Let $\boldsymbol{\sigma} = (\sigma_1,\ldots,\sigma_N) \in \mathcal{X}_N$, with $\sigma_i \in \{-1,+1\}$, for $i \in [N]$. Then, the energy function, $E(\boldsymbol{\sigma})$, for the Curie-Weiss model is:

$$E(\boldsymbol{\sigma}) = \frac{-1}{2N}\left(\sum_i\sum_{j\neq i}\sigma_i\sigma_j\right) - B\sum_{i=1}^{N}\sigma_i, \tag{1}$$

where B represents the external magnetic field, as before.

Remark 1.1. The Curie-Weiss model is an example of the more general Mean-Field model.

Analogous to the notion of average magnetization (defined earlier), we introduce the following definition:

Definition 1.2. (Empirical Magnetization) For a configuration $\boldsymbol{\sigma} = (\sigma_1,\ldots,\sigma_N) \in \mathcal{X}_N$, the empirical (or instantaneous) magnetization is given by:

$$m(\boldsymbol{\sigma}) = \frac{1}{N}\sum_{i=1}^{N}\sigma_i.$$

An immediate observation is that $\langle m(\boldsymbol{\sigma})\rangle = \frac{1}{N}\sum_{i=1}^{N}\langle\sigma_i\rangle = M_N(\beta, B)$. Modulo this definition, we can recast the expression for the energy function, in the following form:

$$E(\boldsymbol{\sigma}) = \left(\frac{N}{2} - \frac{N}{2}(m(\boldsymbol{\sigma}))^2\right) - BNm(\boldsymbol{\sigma}).$$

The expression follows from the facts that:

$$\begin{aligned}
\frac{-1}{2N}\left(\sum_i\sum_{j\neq i}\sigma_i\sigma_j\right) &= \frac{-1}{2N}\left(\sum_{i=1}^{N}\sigma_i\sum_{j=1}^{N}\sigma_j - \sum_{i=1}^{N}\sigma_i^2\right)\\
&= \frac{N}{2} - \frac{N}{2}(m(\boldsymbol{\sigma}))^2\\
\text{and } -B\sum_{i=1}^{N}\sigma_i &= -BNm(\boldsymbol{\sigma}).
\end{aligned}$$

Hereinafter, we shorten $m(\boldsymbol{\sigma})$ to m, as the argument is implicit. With this in place, we can write the partition function as

$$\begin{aligned} Z_N(\beta,B) &= e^{-N\beta/2}\sum_{\sigma} e^{\beta N m^2/2+\beta B N m} \\ &\stackrel{(a)}{=} e^{-N\beta/2}\sum_{m}\binom{N}{N(m+1)/2} e^{\beta N m^2/2+\beta B N m} \\ &\doteq_N e^{-N\beta/2}\sum_{m} e^{N\mathcal{H}((m+1)/2)} e^{\beta N m^2/2+\beta B N m} \end{aligned}$$

where equation (a) follows from the observation that the number of positive spins in $\boldsymbol{\sigma}$ is $N(m+1)/2$. Moreover, we use $\mathcal{H}(.)$ to represent the binary entropy function, expressed in nats.

Further approximation in the limit of large N leads us to:

$$Z_N(\beta,B) \doteq_N \int_{-1}^{1} e^{N\phi_{mf}(m;\beta,B)} dm, \tag{2}$$

where

$$\phi_{mf}(m;\beta,B) := \frac{-\beta}{2}(1-m^2) + \beta B m + \mathcal{H}(\frac{m+1}{2}). \tag{3}$$

The largest contribution to the integral in (**??**) comes from the largest exponent, and hence, the free energy density,

$$\phi(\beta,B) = \sup_{m\in[-1,1]} \phi_{mf}(m;\beta,B).$$

1.1 Analysis of $\phi_{mf}(m;\beta,B)$

In what follows, we shall take a look at how ϕ_{mf} behaves on varying β and m. To obtain the point of maximum, m^*, of $\phi_{mf}(.)$, we note from (**??**) that:

$$0 = \frac{\partial\phi_{mf}(m;\beta,B)}{\partial m}\Big|_{m=m^*} = \beta m^* + \beta B + \frac{1}{2}log\left(\frac{2}{m^*+1}-1\right),$$

which yields:

$$m^* = \tanh(\beta m^* + \beta B) \tag{4}$$

Now that we have calculated the point of maximum of $\phi_{mf}(m;\beta,B)$, we seek to characterize the variation of $\phi(\beta,B) = \phi_{mf}(m^*;\beta,B)$, with β. From (**??**), it is easy to arrive at the fact that

$$\frac{\partial m^*}{\partial \beta} = \frac{(1-(m^*)^2)(B+m^*)}{1-\beta(1-(m^*)^2)} \tag{5}$$

As a simplifying assumption, we take $B=0$. Let us denote by $g(\beta)$, the partial derivative of $\phi(\beta,0)$ with respect to β, i.e.,

$$g(\beta) = \frac{\partial\phi(\beta,0)}{\partial\beta} = \frac{\partial\phi_{mf}(m^*;\beta,0)}{\partial\beta}.$$

With the aid of (**??**), we note that $g(\beta)$ changes sign at $\beta=\beta_c=1$, with $g(1)=0$. This point, β_c, at which a *phase transition* occurs, is the inverse of the **Curie Temperature**, T_c, when scaled to appropriate units. The veracity of the derivations presented above, can be ascertained with the help of the following plots:

2 The Ising Spin-Glass (or Edwards-Anderson) Model

In this section, we briefly go over yet another model for the interactions between particles. As before, $\boldsymbol{\sigma} \in \mathcal{X}_N = \{-1,+1\}^{\mathbb{L}}$ represents a configuration of the N-particle system, with $\mathbb{L} = \{1,\ldots,L\}^d$ representing a d-dimensional lattice. In the Ising spin-glass model,

$$E(\boldsymbol{\sigma}) = -\sum_{(ij)} J_{i,j}\sigma_i\sigma_j - B\sum_{i\in\mathbb{L}}\sigma_i.$$

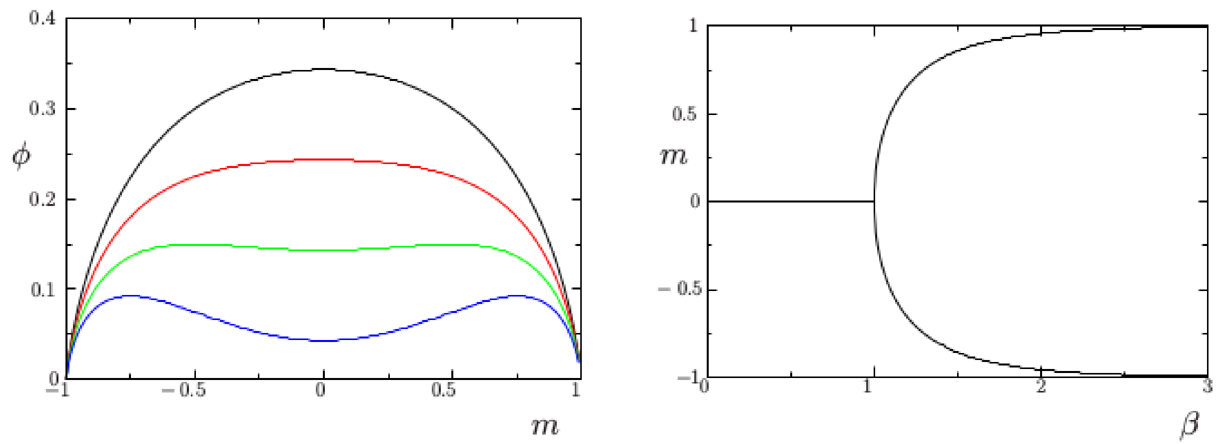

Figure 1: The plot on the left shows the variation of $\phi_{mf}(m;\beta,0)$ with m, for different values of β. For $\beta < 1$, there is a unique maximum, and for $\beta > 1$, there are two maxima—indicating a phase transition at $\beta = \beta_c = 1$. On the right, the plot shows the variation of the values of m that maximize $\phi_{mf}(m;\beta,0)$, with β. The phase transition at $\beta = 1$ is indicated by a bifurcation.

Here, the first summation runs over each edge of the lattice, and the multiplying factor, $J_{i,j} \in \mathbb{R}$, for $i, j \in \mathbb{L}$. Note the difference in the energy function from the Ising model, in that now, each 2-particle interaction is multiplied by a (possibly) different factor. We state here that it is not straightforward to arrive at a low energy configuration in this model (by satisfying each local constraint), as elucidated in the example below:

Example 2.1. Consider the Ising spin-glass model, for an $L = 2$, $d = 2$ system, with $B = 0$. The lattice is hence a 2-dimensional square, with vertices $V = \{(1,1),(1,2),(2,1),(2,2)\}$. Let $J_{(1,1)\sim(1,2)} = 1$, $J_{(1,1)\sim(2,1)} = 1$, $J_{(2,1)\sim(2,2)} = 1$ and $J_{(2,2)\sim(1,2)} = -1$, where the notation $(a,b) \sim (c,d)$ is used to represent the edge between (a,b) and (c,d), for $(a,b),(c,d) \in \mathbb{L}$. We observe that the two configurations,

$$\boldsymbol{\sigma}^1 = \left(\sigma_{(1,1)} = 1, \sigma_{(1,2)} = 1, \sigma_{(2,1)} = 1, \sigma_{(2,2)} = 1\right)$$
$$\text{and } \boldsymbol{\sigma}^2 = \left(\sigma_{(1,1)} = 1, \sigma_{(1,2)} = -1, \sigma_{(2,1)} = 1, \sigma_{(2,2)} = 1\right)$$

are degenerate, with $E(\boldsymbol{\sigma}^1) = E(\boldsymbol{\sigma}^2) = 2$. This is, however, a **frustrated** system, since it is impossible to satisfy each local constraint induced by the individual $J_{i,j}$s, $i, j \in \mathbb{L}$.

3 Optimization and Statistical Physics

Combinatorial optimization problems present inherent difficulties owing to the "discreteness" (or lack of smoothness) of the space. In general, in such problems, given a configuration space $\mathcal{X}$, we wish to find $C \in \mathcal{X}$ with the smallest cost. It is possible to introduce a Boltzmann probability distribution on the space of configurations as:

$$\mu_\beta(C) = \frac{1}{Z(\beta)} e^{-\beta E(C)}, \quad Z(\beta) = \sum_{C \in \mathcal{X}} e^{-\beta E(C)},$$

for $C \in \mathcal{X}$. In the limit as $\beta \to \infty$, the probability distribution concentrates on the ground states—which is the case when all optimization constraints are satisfied. We conclude the lecture with two different examples, to drive home the obstacles posed by such optimization problems:

Example 3.1. (Min-Cuts on Graphs) We consider again, the Ising spin-glass model, with $B = 0$, and $E(\boldsymbol{\sigma}) = \sum_{(ij)} J_{i,j}\sigma_i\sigma_j$. Each configuration partitions the set $\{\sigma_1, \ldots, \sigma_N\}$ into 2 subsets:

$$V_+ = \{i : \sigma_i = +1\}, \quad V_- = \{i : \sigma_i = -1\}.$$

Thus,

$$E(\boldsymbol{\sigma}) = -C + 2 \sum_{(ij)\in\gamma(V_+)} J_{i,j},$$

where $C = \sum_{(ij)} J_{i,j}$, and $\gamma(V_+) = \{(i,j) : i \in V_+, j \in V_-\}$. Solving for the lowest energy configuration, is hence, exactly equivalent to finding the min-cut in the graph
$\mathcal{G} = (V_+ \cup V_-, \mathcal{E})$, with $\mathcal{E}$ being the set of edges induced by the particle interactions.

Example 3.2. (Error-Correcting Codes) In this example, we illustrate the potential hardness of the decoding problem, for binary codes.

Recall the setting of a communication system, which consists of an encoder $e : m \in \{0, \ldots, 2^M - 1\} \mapsto \mathbf{x} \in \{0,1\}^N$, that maps the output of an i.i.d. uniformly distributed source, to a codeword $\mathbf{x}$. Let $\mathbf{y} \in \{0,1\}^N$ be the output of the channel described by the conditional distribution, $Q(\mathbf{y}|\mathbf{x})$, for $\mathbf{y}, \mathbf{x} \in \{0,1\}^N$.

The decoder, $d : \mathbf{y} \in \{0,1\}^N \mapsto \hat{\mathbf{x}} \in \{0,1\}^N$, outputs an estimate, $\hat{\mathbf{x}}$, of the codeword $\mathbf{x}$. The average probability of error,

$$P_B^{avg} = \frac{1}{2^M} \sum_m \sum_{\mathbf{y}:d(\mathbf{y})\neq\mathbf{e}(\mathbf{m})} Q(\mathbf{y}|\mathbf{e}(\mathbf{m})) = 1 - \frac{1}{2^M} \sum_m \sum_{\mathbf{y}:d(\mathbf{y})=\mathbf{e}(\mathbf{m})} Q(\mathbf{y}|\mathbf{e}(\mathbf{m})).$$

It is, therefore, obvious (after interchanging summations) that the optimal decoder (that minimizes P_B^{avg}) must map the received word $\mathbf{y}$ to that codeword, $\hat{\mathbf{x}}$, that maximizes $Q(\mathbf{y}|\hat{\mathbf{x}})$.
However, one can note that this procedure of finding the "most likely" codeword, involves searching over all possible 2^M codewords, leading to exponential time complexity. In fact, the general problem of decoding codes that admit a concise specification (polynomial in the block-length) is NP-hard.