

LUCAS SILVA SIMÕES

EMERGENT COLLEC-  
TIVE PROPERTIES IN  
SOCIETIES OF NEURAL  
NETWORKS

## 0.1 Summary

We model a society of agents who interact in pairs by exchanging for/against opinions about issues using a perceptron algorithm. Employing a maximum entropy analysis one can describe the interacting pair as a dynamics along the gradient of the logarithm of the evidence. This permits introducing an energy like quantity and an approximate global hamiltonian. Knowledge of the expected value of the hamiltonian is relevant information for the state of the society, inducing a canonical distribution by maximum entropy. We study the phase diagram of the society using a mean field approximation which indicate a phase transitions separating ordered and disordered phases depending on the society parameters.

## 0.2 Bayesian Machine Update Learning

Consider a machine prepared to work with a fixed (*hardcoded*) model  $\mathcal{M}$  which depends on some internal variables  $\mathbf{x}$ . Such machine is presented with examples  $\mathcal{D}_t = \{(\xi_\mu, \sigma_\mu)_{\mu=1}^t\}$  and seeks to update its knowledge about  $\mathbf{x}$  in such a way to describe correctly (with  $\mathcal{M}$ ) the received examples while still improving its ability to predict future inputs (also known as capability of *generalization*).

That machine needs to somehow codify which values of  $\mathbf{x}$  are more appropriate to the classification task being learned. As this is an incomplete information situation (that is, we do not have full access to the optimal parameters  $\mathbf{x}^*$ ), we use probability theory tools to describe the plausibility (or beliefs) the machine attributes - through its mechanical-logical-electronic processes - to each value of  $\mathbf{x}$ . This we call the distribution  $Q(\mathbf{x})$ .

By means of practicality we are going to consider that the distribution  $Q(\mathbf{x})$  belongs to some *parametric family* of probability distributions. That is, there is a set of conditions (from now on called **generators**) satisfied by and, more than that, *defining* the utilized distribution:

$$\exists \{F_a(\mathbf{x})\}_{a=1}^G \text{ such that } \mathbb{E}_Q[F_a] = \eta_a$$

The technological (and evolutionary) justification supporting the choice of a parametric family comes from noticing that the machine's memory (or the animal's brain, in a biology analogy) is limited and

thus can only store a finite amount of information usable by its processing. Therefore it is viable that we imagine a numberless of different machines with different architectures (which differ on the parametric family used) and then compare their performances in different settings.

In possession of the generators we can do Maximum Entropy and obtain:

$$Q(\mathbf{x}) = P(\mathbf{x}|\theta) = \frac{1}{\zeta} \exp \left( - \sum_{a=1}^G \theta_a F_a(\mathbf{x}) \right)$$

where  $\zeta = \int d\mathbf{x} \exp \left( - \sum_{a=1}^G \theta_a F_a(\mathbf{x}) \right)$  is the normalization of the probability distribution  $Q$ , also called the **partition function**.

We calculate some identities that are going to be useful later on:

$$\begin{aligned} \frac{\partial \log \zeta}{\partial \theta_b} &= \frac{1}{\zeta} \frac{\partial \zeta}{\partial \theta_b} = \frac{1}{\zeta} \int d\mathbf{x} (-F_b(\mathbf{x})) e^{-\sum_a \theta_a F_a(\mathbf{x})} \\ \therefore \frac{\partial \log \zeta}{\partial \theta_b} &= -\eta_b \end{aligned} \quad (1)$$

$$\begin{aligned} \frac{\partial Q}{\partial \theta_b} &= \frac{1}{\zeta} e^{-\sum_a \theta_a F_a(\mathbf{x})} (-F_b(\mathbf{x})) + e^{-\sum_a \theta_a F_a(\mathbf{x})} \left( \frac{-1}{\zeta^2} \right) \frac{\partial \zeta}{\partial \theta_b} \\ \therefore \frac{\partial Q}{\partial \theta_b} &= [\eta_b - F_b(\mathbf{x})] Q \end{aligned} \quad (2)$$

Knowing the probability distribution the machine attributes to the parameters of its interior model, we might calculate how it is that the "knowledge" of the machine will change when a new example is presented to it. Using Bayes' Theorem:

$$P_{n+1}(\mathbf{x}) = P(\mathbf{x}|\mathcal{D}_{n+1}) = \frac{L_{n+1} Q_n}{Z_{n+1}} = \frac{P(\mathcal{D}_{n+1}|\mathbf{x}) P(\mathbf{x}|\theta_n)}{\int d\mathbf{x} P(\mathcal{D}_{n+1}|\mathbf{x}) P(\mathbf{x}|\theta_n)}$$

Although this update is exact, it does not suit our needs since the new distribution  $P_{n+1}$  usually will not belong to the initial parametric family. We need an alternative step that takes into account the relevant data from Bayes' rule and remains with the same functional form as the prior.

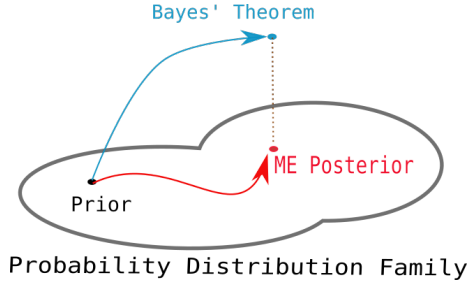
To do that one can minimize the distance (maximize the entropy) between the distributions  $D_{KL}[Q_{n+1}||Q_n]$  while enforcing the expectation values of  $P_n$  with  $\mathbb{E}_{P_n}[F_a(\mathbf{x})] = \eta_a$ . We want to minimize the following Lagrangian<sup>1</sup>:

$$\Lambda[P, Q, \{\Delta_a\}] = \int d\mathbf{x} Q_{n+1} \log \frac{Q_{n+1}}{Q_n} - \Delta^a \left[ \int d\mathbf{x} F_a Q_{n+1} - \eta_a \right] \quad (3)$$

$$- \Delta_0 \left[ \int d\mathbf{x} Q_{n+1} - 1 \right] \quad (4)$$

Since both  $Q_n$  and  $P_n$  are already fixed, one can only minimize the Lagrangian varying the posterior distribution  $Q_{n+1}$ . Taking a functional derivative, one finds:

$$\frac{\delta \Lambda}{\delta Q_{n+1}} = \int d\mathbf{x} \delta Q_{n+1} [\log Q_{n+1} + 1 - \log Q_n - \Delta^a F_a - \Delta_0]$$



Equating this to zero, one finds the expression for the Maximum Entropy (ME) posterior:

$$Q_{n+1}(\mathbf{x}) = Q_n(\mathbf{x}) e^{-1+\Delta_0+\Delta^a F_a} = \frac{1}{\zeta_{n+1}} \exp \left( - \sum_{a=1}^G \theta_a^{n+1} F_a(\mathbf{x}) \right)$$

where  $\theta_a^{n+1} = \theta_a^n + \Delta_a$  and  $\zeta_{n+1}$  is the new normalization factor. If one takes a derivative with respect to  $\Delta_b$  it becomes evident that the constraint adopted was:

$$\mathbb{E}_{Q_{n+1}}[F_b(\mathbf{x})] \equiv \eta_{n+1}^b = \mathbb{E}_{P_{n+1}}[F_b(\mathbf{x})]$$

Subtracting  $\eta_n^b$  from both sides and working out the equation with results eq. 1 and eq. 2 we find an update rule to the parameters of the distribution.

<sup>1</sup> Here we start using Einstein summation convention

$$x_a y^a \equiv \sum_a x_a y_a$$

Figure 1: Schematic representation of the update procedure done to revise the distribution  $Q_n$ . It goes as follows: one uses Bayes' Theorem (blue path) to get the new constraints and then updates the distribution through maximum entropy (red path), therefore minimizing the distance relative to the prior while enforcing the new expected values of the posterior.

$$\eta_{n+1}^b - \eta_n^b = \mathbb{E}_{P_{n+1}}[F_b(\mathbf{x})] - \eta_n^b = \int d\mathbf{x} P_{n+1} F_b(\mathbf{x}) - \eta_n^b \int d\mathbf{x} P_{n+1} \quad (5)$$

$$= \int d\mathbf{x} P_{n+1} [F_b(\mathbf{x}) - \eta_n^b] = \int d\mathbf{x} \frac{L_{n+1}}{Z_{n+1}} Q_n [F_b(\mathbf{x}) - \eta_n^b] \quad (6)$$

$$= \int d\mathbf{x} \frac{L_{n+1}}{Z_{n+1}} \left( -\frac{\partial Q_n}{\partial \theta_n^b} \right) = -\frac{1}{Z_{n+1}} \frac{\partial}{\partial \theta_n^b} \left( \int d\mathbf{x} L_{n+1} Q_n \right) \quad (7)$$

Remarkably, the schematics represented by fig. 1 is simply a **gradient descent** equation:

$$\eta_{n+1}^b = \eta_{(n)}^b - \frac{\partial \log Z_{n+1}}{\partial \theta_{(n)}^b} = \eta_{(n)}^b + \frac{\partial \mathcal{E}_n}{\partial \theta_{(n)}^b} \quad (8)$$

where we defined  $\mathcal{E}_n = -\log Z_n$  the cost function of our problem. This cost function, or "*free energy*", depends on the *intrinsic* mechanisms of the machine, not taking into account the generator family chosen. We will deal more with this topic in section sec. 0.4.

### 0.3 Gaussian Parametric Family

Now we analyze separately a special case of parametric family. The assumed relevant generators to make inference over a vector  $\mathbf{B} \in \mathbb{R}^K$  will be

$$\begin{cases} \mathbb{E}_n[B^i] &= J_n^i \\ \mathbb{E}_n[B^i B^j] &= C_n^{ij} + J_n^i J_n^j \end{cases}$$

and the resulting ME distribution is a **Multivariate Gaussian**:

$$Q_n(\mathbf{B}) \equiv \mathcal{N}(\mathbf{B}|\mathbf{J}_n, \mathbf{C}_n) = |2\pi\mathbf{C}_n|^{-1/2} \exp \left[ -\frac{1}{2}(\mathbf{B} - \mathbf{J}_n) \cdot \mathbf{C}_n^{-1}(\mathbf{B} - \mathbf{J}_n) \right] \quad (9)$$

$$= \frac{1}{Z_{\mathcal{N}_n}} \exp \left[ -\vec{\theta}_n \cdot \mathbf{B} - \mathbf{B} \cdot \underline{\theta}_n \mathbf{B} \right] \quad (10)$$

where  $\theta_n^i = -\sum_j (C_n^{-1})_{ij} J_n^j$  and  $\underline{\theta}_n^{ij} = \frac{1}{2} (C_n^{-1})_{ij}$  are the Lagrange multipliers relevant to define the distribution.

In this section we analyze how the parameters of the gaussian distribution update under the framework developed in section sec. 0.2. In particular, we want to find two simplified equations for the evolution of  $J_n^i$  and  $C_n^{ij}$ .

First of all, let us analyze the Lagrange multipliers associated with the former:

$$\theta_n^i = - \sum_j (C_n^{-1})_{ij} J_n^j \Rightarrow \vec{\theta}_n = -\mathbf{C}_n^{-1} \mathbf{J}_n \Rightarrow \mathbf{J}_n = -\mathbf{C}_n \vec{\theta}_n$$

From this equation we can obtain:

$$\frac{\partial}{\partial \theta_n^i} = \sum_{l=1}^K \frac{\partial J_n^l}{\partial \theta_n^i} \frac{\partial}{\partial J_n^l} = - \sum_l C_n^{li} \frac{\partial}{\partial J_n^l}$$

And finally

$$J_{n+1}^i = J_n^i + \frac{\partial \mathcal{E}_{n+1}}{\partial \theta_n^i} = J_n^i - \sum_l C_n^{li} \frac{\partial \mathcal{E}_{n+1}}{\partial J_n^l} \quad (11)$$

In vectorial form (noticing that  $\mathbf{C}$  is symmetric):

$$\mathbf{J}_{n+1} = \mathbf{J}_n - \mathbf{C}_n \cdot \nabla_{\mathbf{J}_n} \mathcal{E}_{n+1} \quad (12)$$

We could follow the same procedure to study the evolution of  $C_n^{ij}$  but in that case we would need to study the derivative of  $\mathcal{E}_{n+1}$  with respect to  $(C_n^{-1})_{ij}$ , which can be quite complicated. An easier approach is to match the generators' expected values. Either way, this evolution will not be important to the work we develop in this report so we only state the result:

$$C_{n+1}^{ij} = C_n^{ji} - \sum_k \sum_l C_n^{jk} C_n^{il} \frac{\partial}{\partial J_n^k} \frac{\partial}{\partial J_n^l} \mathcal{E}_{n+1} \quad (13)$$

In vectorial form (noticing that  $\mathbf{C}$  is symmetric since it is a covariance matrix):

$$\mathbf{C}_{n+1} = \mathbf{C}_n - \mathbf{C}_n \cdot (\mathbf{H}_{\mathbf{J}_n} \mathcal{E}_{n+1}) \cdot \mathbf{C}_n \quad (14)$$

where  $\mathbf{H}_{\mathbf{J}_n} \mathcal{E}_{n+1}$  is the Hessian matrix of second derivatives of  $\mathcal{E}_{n+1}$  with respect to the elements of  $\mathbf{J}_n$ .

#### 0.4 Bayesian & Gaussian Perceptron

Proceeding a bit further, we consider the following machine  $\mathcal{M}$  hard-coded to infer the value of  $\mathbf{B} \in \mathbb{R}^K$  (what we have been calling  $\mathbf{x}$ ) and make predictions  $\sigma$  about  $\xi \in \mathbb{R}^K$  using the following mechanisms:

$$\mathcal{M} : \begin{cases} \text{input/data:} & y = (\xi \in \mathbb{R}^K; \sigma \in \{-1, +1\}) \\ \text{architecture:} & \tau = \text{sign}(\xi \cdot \mathbf{B}) \\ & \sigma = \begin{cases} -\tau & \text{with probability } \varepsilon \\ \tau & \text{with probability } 1 - \varepsilon \end{cases} \\ \text{inference:} & \mathbb{E}[B^i] = J^i, \mathbb{E}[B^i B^j] = C_{ij} + J^i J^j \end{cases}$$

Since the set of generators being used is the same as last section's, we know that the Maximum Entropy distribution will be a multivariate gaussian as given by eq. 9 and the update equations will be given by eq. 12 and eq. 14.

The learning situation described by  $\mathcal{M}$  can be interpreted as follows: consider a pair of vectors (*perceptrons*)  $\mathbf{J}, \mathbf{B} \in \mathbb{R}^K$  where  $\mathbf{B}$  is called **professor** and  $\mathbf{J}$  is called **student**. The student  $\mathbf{J}$  will learn to imitate the professor  $\mathbf{B}$  through examples  $(\xi_\mu, \sigma_\mu)$ , where  $\xi_\mu \in \mathbb{R}^K$  is typically called an **issue** and  $\tau_\mu = \text{sign}(\mathbf{B} \cdot \xi_\mu)$  is the professor's **opinion** over the respective question. This opinion can be corrupted by a multiplicative noise  $\varepsilon$  (a **distrust**) when it is passed to the student. The set of  $n$  pairs issue-opinion  $\mathcal{D}_n = \{(\xi_\mu, \sigma_\mu)_{\mu=1}^n\}$  is called **learning set** at time  $n$ .

As we have already laid out our inference problem in section sec. 0.2 and solved it for the gaussian parametric family in section sec. 0.3, all that is left is to calculate the free energy  $\mathcal{E}_n = -\log Z_n$ .

$$Z_{n+1} = \int d\mathbf{x} P(\mathcal{D}_{n+1}|\mathbf{x}) P(\mathbf{x}|\theta_n) = \int d\mathbf{B} P(\xi) P(\sigma|\xi, \mathbf{B}) Q_n(\mathbf{B}) \quad (15)$$

$$= k_\xi \langle P(\sigma|\xi, \mathbf{B}) \rangle_{Q_n(\mathbf{B})} \quad (16)$$

where  $k_\xi = P(\xi)$  is a constant because we are considering the issues  $\xi$  are sampled uniformly over some manifold (for example, the  $K$ -sphere) and are independent of  $\mathbf{B}$ .

Under this learning paradigm we also postulate the likelihood distribution to be:

$$P(\sigma|\xi, \mathbf{B}, \varepsilon) = \varepsilon \Theta(-\sigma \xi \cdot \mathbf{B}) + (1 - \varepsilon) \Theta(\sigma \xi \cdot \mathbf{B}) \quad (17)$$

$$= \varepsilon + (1 - 2\varepsilon) \Theta(\sigma \xi \cdot \mathbf{B}) \quad (18)$$

The only part depending on  $\mathbf{B}$  will be the Heaviside function  $\Theta$ , therefore we proceed to calculate its expected value:

$$\langle \Theta(\sigma \xi \cdot \mathbf{B}) \rangle_{Q_n(\mathbf{B})} = \int d\mathbf{B} \Theta(\sigma \xi \cdot \mathbf{B}) \frac{1}{|2\pi \mathbf{C}_n|^{\frac{1}{2}}} \exp \left[ -\frac{1}{2} (\mathbf{B} - \mathbf{J}_n) \cdot \mathbf{C}_n^{-1} (\mathbf{B} - \mathbf{J}_n) \right] \quad (19)$$

$$= \int db \Theta(\sigma b) \int \frac{d\mathbf{B}}{|2\pi \mathbf{C}_n|^{\frac{1}{2}}} \exp \left[ -\frac{1}{2} (\mathbf{B} - \mathbf{J}_n) \cdot \mathbf{C}_n^{-1} (\mathbf{B} - \mathbf{J}_n) \right] \delta \left( b - \frac{1}{\sqrt{K}} \xi \cdot \mathbf{B} \right) \quad (20)$$

$$= \int db \Theta(\sigma b) \int \frac{d\hat{b}}{2\pi} e^{i\hat{b}b} \int \frac{d\mathbf{B}}{|2\pi \mathbf{C}_n|^{\frac{1}{2}}} \exp \left[ -\frac{1}{2} (\mathbf{B} - \mathbf{J}_n) \cdot \mathbf{C}_n^{-1} (\mathbf{B} - \mathbf{J}_n) - \frac{i\hat{b}}{\sqrt{K}} \xi \cdot \mathbf{B} \right] \quad (21)$$

$$= \int db \Theta(\sigma b) \int \frac{d\hat{b}}{2\pi} e^{i\hat{b}b} \exp \left[ -\frac{1}{2} \left( \frac{\hat{b}^2}{K} \xi \cdot \mathbf{C}_n \xi + \frac{2i\hat{b}}{\sqrt{K}} \xi \cdot \mathbf{J}_n \right) \right] \quad (22)$$

$$= \int db \Theta(\sigma b) \int \frac{d\hat{b}}{2\pi} \exp \left[ -\frac{1}{2} (\hat{b}^2 \Gamma_n^2 + 2i\hat{b}(h_n - b)) \right] \quad (23)$$

$$\langle \Theta(\sigma \xi \cdot \mathbf{B}) \rangle_{Q_n(\mathbf{B})} = \int db \Theta(\sigma b) \frac{1}{\sqrt{2\pi \Gamma_n^2}} \exp \left[ -\frac{1}{2} \left( \frac{h_n - b}{\Gamma_n} \right)^2 \right]$$

where we used the definition of  $h_n$  (a projection of the issue  $\xi$  over the mean value  $\mathbf{J}_n$ ) and defined a new variable  $\Gamma_n^2 = \frac{1}{K} \xi \cdot \mathbf{C}_n \xi$ , as the norm of the issue presented to the student *weighted by it's correlation matrix*  $\mathbf{C}_n$ .

Analyzing each case  $\sigma = \pm 1$  separately we end up with:

$$Z_{n+1} = k_\xi \left[ \varepsilon + (1 - 2\varepsilon) \Phi \left( \frac{\sigma h_n}{\Gamma_n} \right) \right]$$

where  $\Phi$  is the cumulative distribution function of the gaussian distribution.

Finally, our "free energy" cost function (discarding the constant part  $k_\xi$ ) will be:

$$\mathcal{E}_{n+1} = -\log \left[ \varepsilon + (1 - 2\varepsilon) \Phi \left( \frac{\sigma h_n}{\Gamma_n} \right) \right]$$



### 0.5 Mean Field approximation

A possible strategy to study a society of the agents developed in section 0.4 is to define conserved quantities - such as the Hamiltonian associated with the interaction potential of those agents - and proceed with the usual Statistical Mechanics tools. At some point the calculation might become intractable and that is when one must transition to approximate results and/or computational methods. In this section we develop a Mean Field approach to an specific canonical ensemble of social agents in a noisy society.

Let us consider that a society of agents  $\{\mathbf{J}_i\}$  which we suppose can be described mostly by one specific Hamiltonian  $\mathcal{H}$ :

$$\mathcal{H} = -\Gamma^2 \sum_{\langle i,j \rangle} \log \left[ \varepsilon + (1 - 2\varepsilon) \Phi \left( \frac{1}{\Gamma} (\sigma_i h_j + \sigma_j h_i) \right) \right] = \sum_{\langle i,j \rangle} V_{ij} \quad (24)$$

We also suppose that the mean value of this quantity  $\langle \mathcal{H} \rangle$  is conserved throughout the configuration evolution of the society, that is,  $\mathcal{H}$  remains close to some fixed value  $E$  of energy but has the possibility of oscillating to higher/lower energy values depending on some "temperature" parameter (which can also be seen as a "social pressure"). In a Maximum Entropy framework, we can say that the probability distribution describing this society with this information paradigm is given by the canonical (Boltzmann) distribution:

$$P_B(\{\mathbf{J}_i\}) = \frac{1}{Z_B} \exp(-\beta \mathcal{H}(\{\mathbf{J}_i\}))$$

This description, however, is rather complex depending on the kind of applications we want to follow. A common procedure to simplify our model is to consider a *mean field approximation*, which projects our solution (the probability distribution  $P_B$ , which depends on the whole set  $\{\mathbf{J}_i\}$  all at once) into a parametric family of separable probability distributions  $P_0 = \prod_i P_i(\mathbf{J}_i)$  much simpler to work with.

In that case, we do not wish to choose a separable distribution indiscriminately; we want to pick as close as possible the specific  $P_0$  which best approximates  $P_B$  given the constraints we have assigned to it. That is a calculation we can do maximizing the entropy  $S$  (or minimizing the Kullback-Leibler divergence), as follows:

$$S[P_0||P_B] = - \iint \left( \prod_{i=1}^N d\mathbf{J}_i \right) P_0 \log \left( \frac{P_0}{P_B} \right) \quad (25)$$

$$= - \iint \left( \prod_i d\mathbf{J}_i \right) P_0 [\log(Z_B) + \beta \mathcal{H} + \log(\prod_i P_i)] \quad (26)$$

$$= \left\langle \log Z_B - \beta \mathcal{H} - \sum_i \log P_i \right\rangle_{P_0} \quad (27)$$

We want to find  $P_0$  so that variations of  $S$  equals to zero,  $\delta S = 0$ . Usual functional calculus arguments lead to:

$$\delta S = \int dJ_k \delta P_k \left[ \log Z_B + \log P_k + 1 + \beta \sum_{i \in \partial k} \int d\mathbf{J}_i P_i V_{ik} \right]$$

This result can only be possible for any variation  $\delta P_k$  if the term in brackets  $[\dots]$  is identically zero itself, so we get the following result:

$$P_k(\mathbf{J}_k) = \frac{1}{Z_k} \exp \left( -\beta \sum_{i \in \partial k} \int d\mathbf{J}_i P_i V_{ik} \right) \quad (28)$$

Unfortunately, due to the rather complex form of the potential  $V_{ij}$ , the equation above is intractable. In that case we are going to *choose* a mean field probability distribution family (instead of *selecting* the best one).

We proceed comparing the exact mean field projection done above and the hamiltonian in eq. 24 to an approximate distribution that takes into account only an effective number of neighbors  $\nu_{(i)}$  with effective interactions  $m_{(i)}$  and  $r_{(i)}$ :

$$P_j(\mathbf{J}_j) = \frac{1}{Z_j} \left[ \varepsilon + (1 - 2\varepsilon) \Phi \left( \frac{1}{\Gamma} \left( r_{(i)} h_j + \sigma_j m_{(i)} \right) \right) \right]^{\beta \nu_{(j)} \Gamma^2}$$

In fact, one can think of  $m$  and  $r$  as parameters that describe the overall behavior of the society, such that one agent  $i$  receives signals from its neighbors irrespective of label  $i$ . This does not mean that all the agents are identical, but that they draw their moral vector from the same probability distribution.

We can represent this self-consistently with the following set of equations to be obeyed:

$$m = \int d\mathbf{J} h(\mathbf{J}) P_{\text{MF}}(\mathbf{J}) \quad (29)$$

$$r = \int d\mathbf{J} \sigma(\mathbf{J}) P_{\text{MF}}(\mathbf{J}) \quad (30)$$

Then the mean field probability distribution becomes:

$$P_{\text{MF}}(\mathbf{J}) = \frac{1}{Z} \left[ \varepsilon + (1 - 2\varepsilon) \Phi \left( \frac{r}{\Gamma} h(\mathbf{J}) + \frac{m}{\Gamma} \sigma(\mathbf{J}) \right) \right]^{\beta\nu\Gamma^2} \quad (31)$$

We can now come back to the definition of our order parameters and evaluate explicitly their values by integration:

$$m = \int d\mathbf{J}_i h_i P_i = \frac{1}{Z_k} \int d\mathbf{J}_i h_i \left[ \varepsilon + (1 - 2\varepsilon) \Phi \left( \frac{1}{\Gamma} (r h_i + \sigma_i m) \right) \right]^{\beta\nu\Gamma^2} \quad (32)$$

$$= \frac{1}{Z_i} \frac{1}{\sqrt{K}} \int d\mathbf{J}_i \xi \cdot \mathbf{J}_i \left[ \varepsilon + (1 - 2\varepsilon) \Phi \left( \frac{1}{\Gamma} \left( r \frac{\xi \cdot \mathbf{J}_i}{\sqrt{K}} + \text{sign}(\xi \cdot \mathbf{J}_i) m \right) \right) \right]^{\beta\nu\Gamma^2} \quad (33)$$

$$(34)$$

$$r = \int d\mathbf{J}_i \sigma_i P_i = \frac{1}{Z_k} \int d\mathbf{J}_i \sigma_i \left[ \varepsilon + (1 - 2\varepsilon) \Phi \left( \frac{1}{\Gamma} (r h_i + \sigma_i m) \right) \right]^{\beta\nu\Gamma^2} \quad (35)$$

$$= \frac{1}{Z_i} \int d\mathbf{J}_i \text{sign}(\xi \cdot \mathbf{J}_i) \left[ \varepsilon + (1 - 2\varepsilon) \Phi \left( \frac{1}{\Gamma} \left( r \frac{\xi \cdot \mathbf{J}_i}{\sqrt{K}} + \text{sign}(\xi \cdot \mathbf{J}_i) m \right) \right) \right]^{\beta\nu\Gamma^2} \quad (36)$$

Since we can always rotate the coordinate system to a given orientation, we choose one in which  $\xi = |\xi| \hat{e}_5$ , so that  $\xi \cdot \mathbf{J}_i = \sqrt{K} \cos \theta$  and all the other angular integrals are trivial besides the one in  $\theta$ :

$$m = \frac{1}{Z} \int_0^\pi d\theta \sin^3 \theta \cos \theta B(\theta|\varepsilon, \Gamma, m, r, \beta\nu) \quad (37)$$

$$r = \frac{1}{Z} \int_0^\pi d\theta \sin^3 \theta \text{sign}(\cos \theta) B(\theta|\varepsilon, \Gamma, m, r, \beta\nu) \quad (38)$$

$$Z = \int_0^\pi d\theta \sin^3 \theta B(\theta|\varepsilon, \Gamma, m, r, \beta\nu) \quad (39)$$

where we implicitly defined the function  $B(\theta) \equiv B(\theta|\varepsilon, \Gamma, m, r, \beta\nu)$  :

$$B(\theta) = \left[ \varepsilon + (1 - 2\varepsilon) \Phi \left( \frac{1}{\Gamma} (r \cos \theta + \text{sign}(\cos \theta) m) \right) \right]^{\beta\nu\Gamma^2}$$

Although there is no analytical solution to those equations they can be solved numerically.