The TAP approximation

Caterina De Bacco

1 Sherrington-Kirpatrick Model

We saw in the previous lecture that the Mean Field approximation is valid for weak interactions, the Curie-Weiss model was an example of that. Now we turn our attention to the case where we cannot neglect these fluctuations anymore. We consider the case of a fully connected network (yes, also the Curie-Weiss model had a *complete* network) but now the interaction couplings J_{ij} are not constant anymore. Rather, let's consider them iid random variables themselves:

$$J_{ij} \sim \mathcal{N}\left(0, \frac{J_0}{N}\right) \quad . \tag{1}$$

The energy of the system is then:

$$H(\mathbf{s}) = -\sum_{i<j} J_{ij} s_i s_j - \sum_i s_i h_i \quad , \tag{2}$$

and the associated probability distribution is the usual Boltzmann with energy $H(\mathbf{s})$. This is called the Sherrington-Kirpatrick *spin glass* model.

A network representation is given in Fig. 1.

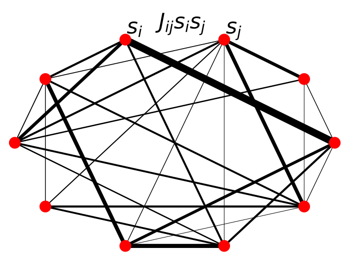

Figure 1: The complete network representation of a SK model. Each edge is an interaction of magnitude $J_{ij}s_is_j$, where the J_{ij} are iid $J_{ij} \sim \mathcal{N}\left(0, \frac{J_0}{N}\right)$. Here, interactions with the external field are not drawn.

Also in this case we ask about accessing quantities like marginals or expected values of statistical quantities. Recall that for the MF case we used the intuition of substituting:

$$\begin{aligned} h_i(\mathbf{s}_{\setminus \mathbf{i}}) := \sum_{j \neq i} J_{ij} s_j \approx \mathbb{E}_P\left[h_i(\mathbf{s}_{\setminus \mathbf{i}})\right] &= \sum_{j\neq i} J_{ij} \mathbb{E}_P\left[s_j\right] & (3)\\ &= \sum_{j \neq i} J_{ij} m_j =: h^{MF} \quad , & (4) \end{aligned}$$

c.f. for the Curie-Weiss model, i.e. $J_{ij} = \frac{J}{N}$ and external field h, we had a simpler expression: $h^{MF} := \frac{J}{2}\langle m \rangle + h$.

Obs1: the variances of is $h_i(\mathbf{s}_{\setminus \mathbf{i}})$ are not decaying fast with N as in MF.
Repeating similar calculations as we the exercise left for MF, where we neglect dependencies among s_i's and between J_{ij}'s and s_j's, we obtain:

$$
\begin{aligned}
Var\left(J_{ij}s_j\right) &= \left(\mathbb{E}_{J,s}\left[(J_{ij}s_j)^2\right] - \mathbb{E}_{J,s}\left[(J_{ij}s_j)\right]^2\right) && (5)\\
&= \mathbb{E}_J\left[J_{ij}^2\right]\mathbb{E}_{s_j}\left[s_j^2\right] - \mathbb{E}_J\left[J_{ij}\right]^2\mathbb{E}_{s_j}\left[s_j\right]^2 && (6)\\
&= \mathbb{E}_J\left[J_{ij}^2\right] && (7)\\
&= \frac{J_0}{N} \quad , && (8)
\end{aligned}
$$

which implies that summing over N terms we get $Var(h_i(\mathbf{s}_{\setminus \mathbf{i}})) = Var\left(\sum_j J_{ij}s_j\right) \sim O(1)$.
Hence, MF fails in the SK model.

Remember that we obtained similar results by considering the Variational approach:

$$
KL(Q||P) = \log Z + \beta F[Q] \tag{9}
$$

and considered a $\beta F[Q_{MF}]$, which is an upper bound for $-\log Z$, a quantity also called Helmoltz free energy which, when known in a tractable form, gives access to all information needed about the full distribution P.

We can nevertheless use the same procedure as before:

1. Write the energy function $H(\mathbf{s})$.
2. Build an approximate Gibbs free energy $F[Q] = E[Q] - \frac{1}{\beta}S[Q]$ (c.f. as in Eq. (30) of Lecture 7).
3. Minimize it w.r.t. the parameters of Q.
4. The optimal Q is the best approximation for $P(\mathbf{s}) = \frac{e^{-\beta E(\mathbf{s})}}{Z}$.

1.1 Correcting the Mean Field approximation

This time though, we also have J_{ij} random variables and with relatively big variance. Because of the additional noise introduced by non-weak random J_{ij} the Mean-Field approach is not helping anymore and the above approximation is too rough.

For simplicity, let's focus on the case $h = 0$. It turns out that if we define a new Gibbs free energy:

$$
\begin{aligned}
\beta F[Q_{TAP}] &= -\beta \sum_{ij} J_{ij}m_i m_j + \sum_i \left[\frac{1+m_i}{2}\log\left(\frac{1+m_i}{2}\right) + \frac{1-m_i}{2}\log\left(\frac{1-m_i}{2}\right)\right] \\
&\quad -\frac{\beta^2}{2}\sum_{ij} J_{ij}^2\left(1-m_i^2\right)\left(1-m_j^2\right) && (10)
\end{aligned}
$$

and you follow the above procedure (find the stationary point of it, etc...) the final Q_{TAP} approximates P way better than Q_{MF}.

At this point, you might have noticed that:

$$
\beta F[Q_{TAP}] = \beta F[Q_{MF}] - \frac{\beta^2}{2}\underbrace{\sum_{ij} J_{ij}^2\left(1-m_i^2\right)\left(1-m_j^2\right)}_{\text{Onsanger term}} \quad , \tag{11}
$$

where the last term is called *Onsanger* term, and it is a correction of the MF variational energy that gives a better approximation for the joint probability in all the situations where MF does not work. If we now minimize $F[Q_{TAP}]$ w.r.t. the magnetizations m_i, we would get the so called *TAP equations*:

$$m_i = \tanh\left(\beta \sum_j J_{ij} m_j - \beta^2 m_i \sum_{j\neq i} J_{ij}^2 (1-m_j^2)\right) \quad , \tag{12}$$

where $m_i \sum_{j\neq i} J_{ij}^2 (1-m_j^2)$ is called the *Onsanger reaction term*.

Question: where did the *Onsanger* term come from?!?! Intuitively, it is hard to have guessed it... However, it comes out mathematically if one expands the Helmholtz free energy $F_{true} = -\frac{\log Z}{\beta}$ around small β as we will see below.

1.2 TAP: physics intuition approach

We now want to give a physical intuition as we did for the Curie-Weiss model.

Recall: we assumed that s_i felt an *average* contribution from the others, plus the external field (which is fixed as a parameter). This means to approximate the local field with a mean field:

$$h_i^{MF} := \mathbb{E}_P\left[h_i(\mathbf{s}_{\setminus\mathbf{i}})\right] = \sum_j J_{ij} \mathbb{E}_P\left[s_j\right] = \sum_j J_{ij} m_j^{MF} \quad , \tag{13}$$

which led to self-consistency equations (for a more general Hamiltonian than the Curie-Weiss one):

$$m_i^{MF} = \tanh\left(\beta \sum_{j\neq i} J_{ij} m_j^{MF} + \beta\, h_i\right) \quad . \tag{14}$$

Assume an energy:

$$H(\mathbf{s}) = -\sum_{i\neq j} J_{ij} s_i s_j \quad , \tag{15}$$

where for simplicity we set $h_i = 0$, so that:

$$h_i(\mathbf{s}_{\setminus\mathbf{i}}) = \sum_{j\neq i} J_{ij} s_j \quad . \tag{16}$$

Now we refine the MF approximation by accounting for fluctuations of the surrounding spins around their expected value of the magnetization:

$$h_i^{\lambda}(\mathbf{s}_{\setminus\mathbf{i}}) := \underbrace{\mathbb{E}_P\left[h_i(\mathbf{s}_{\setminus\mathbf{i}})\right]}_{h_i^{MF}} + \lambda_i \delta h_i(\mathbf{s}_{\setminus\mathbf{i}}) \quad , \tag{17}$$

where we define the fluctuations felt by s_i as:

$$\delta h_i(\mathbf{s}_{\setminus\mathbf{i}}) := \sum_{j\neq i} J_{ij} \underbrace{\left(s_j - m_j\right)}_{\delta s_j} \quad , \tag{18}$$

and the λ_i is a perturbation parameter that scales the contribution of the fluctuation.
We can then repeat the same argument for the total Hamiltonian, again separating into a Mean Field term $H_{MF}(\mathbf{s})$ (where variables are independent) and a perturbation term accounting for the previous fluctuations:

$$\begin{aligned} H_\lambda(\mathbf{s}) &= H_{MF}(\mathbf{s}) - \lambda \sum_{i,j} J_{ij} (s_i - m_i)\left(s_j - m_j\right) && (19) \\ &= H_{MF}(\mathbf{s}) - \lambda \sum_i (s_i - m_i)\, \delta h_i(\mathbf{s}_{\setminus\mathbf{i}}) \quad , && (20) \end{aligned}$$

where we defined

$$H_{MF}(\mathbf{s}) = -\sum_{ij} J_{ij}\left[s_i m_j - m_i m_j\right] \quad , \tag{21}$$

where the last term inside the summation is subtracted so to obtain a zero-mean variable $\mathbb{E}\left[H_{MF}(\mathbf{s})\right] = 0$, which makes calculations easier and does not change the results.
With this energy $H_\lambda(\mathbf{s})$, we obtain the equation for the expected magnetization as:

$$m_i^\lambda := \mathbb{E}_{H_\lambda}\left[s_i\right] = \mathbb{E}_{H_\lambda}\left[\tanh\left(\beta h_i^\lambda(\mathbf{s}_{\setminus \mathbf{i}})\right)\right] \tag{22}$$

Obs1: if $\lambda = 0$ we get the Mean Field case.
Obs2: if $\lambda = 1$ we get the original model.
Obs3: using perturbation theory (not shown in this class), the solution of Eq. (22) corresponds to minimizing $F[Q_{TAP}]$ as in (10), i.e. Eq. (12).

1.3 Deriving the TAP equation with variational methods

Suppose that we now adopt again the variational approach, i.e. look for a distribution Q which well approximates P in terms of KL divergence. Contrarily to the Mean Field case, here we do not make any assumption about how Q factorizes. Instead, we only assume that the expected individual magnetizations $\mathbb{E}_Q\left[s_i\right]$ should match a given value m_i, which for the moment we fix arbitrarily.

We can enforce this constraint by introducing Lagrange multipliers $\lambda_i(\beta)$ and get a modified Gibbs free energy as:

$$F[Q_{TAP}] := E[Q] - S[Q] - \sum_i \lambda_i(\langle S_i\rangle_Q - m_i) \quad , \tag{23}$$

where the λ_i are Lagrange multipliers enforcing the constraint that the magnetization have their desired values. This is equivalent to a system with Hamiltonian equal to $H_{TAP}(\mathbf{s}) := H(\mathbf{s}) - \frac{1}{\beta}\sum_i \lambda_i s_i$.

This is equivalent to introducing Lagrange multipliers $\lambda_i(\beta)$ inside the Helmholtz free energy $F_{true} = -\log Z/\beta$:

$$-\beta F_{TAP} = \log Z_{TAP} = \log \sum_{s_1\dots,s_N} \exp\left(\beta \sum_{ij} J_{ij} s_i s_j + \sum_i \lambda_i(\beta)(s_i - m_i)\right) \quad , \tag{24}$$

where the $\lambda_i(\beta)$ are the Lagrange multipliers. To see this, consider $\frac{\partial}{\partial \lambda_i(\beta)} \log Z \equiv 0$. We also made explicit the dependence of the multipliers on β. We can imagine them as “auxiliary” external fields acting on each of the variables independently (i.e. similar to terms like $h_i s_i$ as we saw before).

Let’s now expand the expression (24) around $\beta = 0$ and for simplicity write F_{TAP} instead of $F[Q_{TAP}]$:

$$-\beta F_{TAP} = -(\beta F_{TAP})_{\beta=0} - \beta\left(\frac{\partial(\beta F_{TAP})}{\partial \beta}\right)_{\beta=0} - \frac{\beta^2}{2}\left(\frac{\partial^2(\beta F_{TAP})}{\partial \beta^2}\right)_{\beta=0} + \dots \quad . \tag{25}$$

Each of the terms in this expansions can be derived (see below), the results are:

$$\begin{aligned}
-(\beta F_{TAP})_{\beta=0} &= -\sum_i \left[\frac{1+m_i}{2}\log\left(\frac{1+m_i}{2}\right) + \frac{1-m_i}{2}\log\left(\frac{1-m_i}{2}\right)\right] && (26)\\
\beta\left(\frac{\partial(\beta F_{TAP})}{\partial \beta}\right)_{\beta=0} &= \beta \sum_{ij} J_{ij} m_i m_j && (27)\\
-\frac{\beta^2}{2}\left(\frac{\partial^2(\beta F_{TAP})}{\partial \beta^2}\right)_{\beta=0} &= \frac{\beta^2}{2}\sum_{ij} J_{ij}^2\left(1-m_i^2\right)\left(1-m_j^2\right) \quad . && (28)
\end{aligned}$$

Obs1: notice that $-(\beta F_{TAP})_{\beta=0} \equiv S[Q_{MF}]$, the entropy of the Mean Field distribution.
Obs2: notice that $\sum_{ij} J_{ij} m_i m_j \equiv -E[Q_{MF}]$, the internal energy of the Mean Field distribution.
Obs3: considering only terms up to the first order, we obtain: $F[Q_{TAP}] = F[Q_{MF}]$.
Obs4: the last term is the Onsanger correction term we saw before. We then see that this term comes from expanding the free energy up to the second term! If we had stopped at the first-order, we would have gotten instead the Mean Field result

Proof of Section 1.3

When $\beta = 0$, the system only depends on the "auxiliary" fields, i.e. all the variables are independent. Recall now the constraint that $\mathbb{E}_Q[s_i] \equiv m_i$ should be valid for any β, in particular for $\beta = 0$. In this case, we can directly compute:

$$m_i \equiv \mathbb{E}_Q[s_i]_{\beta=0} = \frac{\sum_{s_i=\pm 1} s_i \, e^{\lambda_i(0)s_i}}{\sum_{s_i=\pm 1} e^{\lambda_i(0)s_i}} = \tanh(\lambda_i(0)) \tag{29}$$

This equations maps the parameters m_i to the Lagrange multipliers $\lambda_i(0)$.
From definition (24) we obtain[1]:

$$-(\beta F_{TAP})_{\beta=0} = \sum_i \log[\cosh(\lambda_i(0))] - \lambda_i(0)\, m_i \tag{30}$$

We can now work either in terms of m_i or $\lambda_i(0)$. It is easier to consider m_i, so that we get[2]:

$$-(\beta F_{TAP})_{\beta=0} = -\sum_i \left[\frac{1+m_i}{2} \log\left(\frac{1+m_i}{2}\right) + \frac{1-m_i}{2} \log\left(\frac{1-m_i}{2}\right) \right] \quad . \tag{31}$$

Next, we consider the first order derivative:

$$\beta \left(\frac{\partial(\beta F_{TAP})}{\partial \beta} \right)_{\beta=0} = \beta\, \mathbb{E}_{\beta=0}\left[\sum_{ij} J_{ij} s_i s_j \right] + \beta\, \mathbb{E}_{\beta=0}[s_i - m_i] \left(\frac{\partial \lambda_i}{\partial \beta} \right)_{\beta=0} \tag{32}$$

$$= \beta \sum_{ij} J_{ij} m_i m_j \quad , \tag{33}$$

where the second row comes from taking the expectations when $\beta = 0$ (factorized distribution).

Let's finally take also the second order term. After some algebra (not shown here) we get:

$$-\frac{\beta^2}{2} \left(\frac{\partial^2(\beta F_{TAP})}{\partial \beta^2} \right)_{\beta=0} = \frac{\beta^2}{2} \sum_{ij} J_{ij}^2 \left(1 - m_i^2\right)\left(1 - m_j^2\right) \quad , \tag{34}$$

this is the Onsanger correction term we saw before.

1.4 Inverse Ising model

So far we have been interested in solving the *forward* problem: given a fixed set of parameters $\mathbf{J}$ and $\mathbf{h}$, estimate statistical observables like the moments of a distribution, i.e. expected magnetization and correlations. We saw two approximations to perform these estimates, MF and TAP.

Here we briefly touch on the *inverse* problem: given a set of M data samples $\mathbf{s}_1, \dots, \mathbf{s}_M$, estimate the values of $\mathbf{J}$ and $\mathbf{h}$ that better explains the observed data.

[1] Recall that $\tanh^{-1}(x) = \frac{1}{2}\log\frac{1+x}{1-x}$ for $|x| < 1$.
[2] Recall that: $\log\cosh\tanh^{-1}(x) = -\frac{1}{2}[\log(1+x) + \log(1-x)]$.

This means that now we have access to quantities like:

$$m_i^D = \frac{1}{M}\sum_m s_{i,m} = \langle s_i \rangle_D \tag{35}$$

$$\chi_{ij}^D = \frac{1}{M}\sum_m s_{i,m}s_{j,m} = \langle s_i s_j \rangle_D \quad , \tag{36}$$

where the D stands for "data", so the averages are over the M empirical samples and $s_{i,m}$ is the i-th variable of the m-th sample.

Objective: estimate the posterior $P(\mathbf{J}, \mathbf{h}|\mathbf{s}_1, \ldots, \mathbf{s}_M)$.

Adopting a maximum likelihood approach, we could use $P(\mathbf{J}, \mathbf{h}|\mathbf{s}_1, \ldots, \mathbf{s}_M) \propto P(\mathbf{s}_1, \ldots, \mathbf{s}_M|\mathbf{J}, \mathbf{h})$ and assume that the samples are independent, given a realization of $\mathbf{J}$ and $\mathbf{h}$, yielding:

$$P(\mathbf{J}, \mathbf{h}|\mathbf{s}_1, \ldots, \mathbf{s}_M) \propto \prod_{m=1}^{M} P(\mathbf{s}_m|\mathbf{J}, \mathbf{h}) \quad . \tag{37}$$

We have already learned how to approximate $P(\mathbf{s}_m|\mathbf{J}, \mathbf{h})$, so we are going to start from the MF and TAP results for that. Recall that we obtained self-consistent equations for the expected magnetizations: Eq. (14) (MF) and Eq. (12) (TAP).

We can then start by imposing that these expected magnetizations match the empirical ones. In other words:

$$m_i^{model} \equiv \langle s_i \rangle_D \tag{38}$$

where the superscript "model" stands for either MF or TAP (or others you can think of...). For instance, for the MF case we obtain:

$$\tanh\left(\beta \sum_j J_{ij}\langle s_j \rangle_D + \beta\, h_i\right) \equiv \langle s_i \rangle_D \quad , \tag{39}$$

which can be inverted to find the first set of parameters, for $\beta\ = 1$, the external fields:

$$h_i^{MF} = \operatorname{arctanh}\langle s_i \rangle_D - \sum_j J_{ij}\langle s_j \rangle_D \quad . \tag{40}$$

We are left with estimating the second set of parameters, the J_{ij}.
For this we need an extra equation, which should involve the other piece of information, the $\langle s_i s_j \rangle_D$. Fortunately, there is the linear response (or fluctuation-dissipation) relation (not proven here) that states:

$$C_{ij} := \langle s_i s_j \rangle_D - \langle s_i \rangle_D \langle s_j \rangle_D = \frac{\partial \langle s_i \rangle_{model}}{\partial h_j} \quad . \tag{41}$$

Exercise: prove it. Hint: use the fact that $\langle s_i \rangle_{model} = \frac{1}{Z}\sum_{\mathbf{s}} s_i e^{-H(\mathbf{s})}$.

In words, this states that if you apply a perturbation to a system, in this case the external field h_j on s_i, the way the system responds, i.e. measured here by $\frac{\partial \langle s_i \rangle_{model}}{\partial h_j}$, is related to the correlation function C_{ij}.

If we now derive the RHS of Eq. (41) using Eq. (14)[3], we get:

$$C_{ij} = \frac{\partial \langle s_i \rangle_{model}}{\partial h_j} = (1 - m_i^2)\left[\delta_{ij} + \sum_k J_{ik} C_{kj}\right] \quad , \tag{42}$$

which can be rewritten as:

$$\mathbf{J}^{MF} = \mathbf{P}^{-1} - \mathbf{C}^{-1} \tag{43}$$

where $P_{ij} := (1 - m_i^2)\delta_{ij}$.
One can derive similar results for the TAP estimates $\mathbf{h}^{TAP}$ and $\mathbf{J}^{TAP}$.

[3] Recall that $\frac{d \tanh x}{dx} = 1 - \tanh^2(x)$.

1.5 TAP: summary

- TAP approach consists in correcting the Mean Field approximation by adding explicit correction term to the free energy, i.e. the Onsanger term.
- From a Variational point of view, it considers a variational distribution Q such that the expected values of the variables are fixed as a parameter; no factorization assumption is made (as in MF).
- As a physical intuition, this can be obtained by expanding the local field surrounding a variable s_i around its expected value plus fluctuations. It is exactly by accounting for the fluctuations that TAP improves from Mean Field.
- In principle, more terms can be kept during the expansion, however, high-order terms are useful only for some models. In the Curie-Weiss model, the Onsanger term is negligible; for the SK model, terms beyond second order are negligible for large N, etc...
- For the SK model, TAP works because the couplings have small enough variance, which leads to weak correlations between variables. If this was not the case, then $F[Q_{TAP}]$ does not have a unique solution anymore, but instead is multimodal and the landscape is complex.

One main reference for this Lecture is Opper and Saad (2001).

References

M. Opper and D. Saad, *Advanced mean field methods: Theory and practice* (MIT press, 2001).