3 Bonus questions: Fisher Information Matrix and multi-dimensional fluctuation-dissipation theorem

We would like to check analytically that there is a unique vector $\boldsymbol{g}^\star$ solution of $\partial\mathcal{L}(\boldsymbol{g})/\partial g_\mu = 0$ for all the parameters g_μ in the vector $\boldsymbol{g}$, and that this solution indeed corresponds to a maximum of the log-likelihood function $\mathcal{L}$ (and not a minimum). To do so, we must check the sign of the second order derivatives of $\mathcal{L}$. We introduce the Hessian matrix, which is the matrix H with elements:

$$H_{\mu\nu}(\boldsymbol{g}) = \frac{\partial^2 \mathcal{L}(\boldsymbol{g})}{\partial g_\mu\, \partial g_\nu}\,. \tag{9}$$

In this equation, g_μ and g_ν are two parameters of the vector $\boldsymbol{g}$, they indifferently denote the parameters h_i and J_{ij} (i.e., they could be either h_i's or J_{ij}'s). The matrix H is thus a $K \times K$ matrix, where K is the number of parameters in $\boldsymbol{g}$, i.e. $K = n(n+1)/2$ (n fields h_i, plus $n(n-1)/2$ pairwise couplings J_{ij}).

Q1. Can you show that:

$$\frac{\partial^2 \mathcal{L}(\boldsymbol{g})}{\partial g_\mu\, \partial g_\nu} = -\frac{\partial^2 \log Z(\boldsymbol{g})}{\partial g_\mu\, \partial g_\nu} \ \ ? \tag{10}$$

The matrix $\mathcal{I}(\boldsymbol{g}) = \left(\frac{\partial^2 \log Z(\boldsymbol{g})}{\partial g_\mu\, \partial g_\nu}\right)_{\mu,\nu}$, whose elements appear on the right-hand-side of this equation, is some type of susceptibility matrix. We recall that in the Ising model, the susceptibility is related to the log-partition function by $\chi = k_B T\, \frac{\partial^2 \log Z}{\partial H^2}$, where H is an external field. $\mathcal{I}(\boldsymbol{g})$ tells you how "susceptible" is the model under fluctuations of the parameters around a chosen value of $\boldsymbol{g}$, i.e. how much the probability distribution associated to the Ising model with parameters $\boldsymbol{g}$ will change if we modify a little bit the values of the parameters. Properties of this matrix will characterize the "flexibility" of the model to data, i.e. how easily it will be able to fit various types of data.

Q2. For some of you who are interested in Information theory: $\mathcal{I}(\boldsymbol{g})$ corresponds to the Fisher information matrix. By definition the element of the Fisher information matrix are given by:

$$\mathcal{I}_{\mu\nu}(\boldsymbol{g}) = -\left\langle \frac{\partial^2 \log\, p_{\boldsymbol{g}}(\boldsymbol{s})}{\partial g_\mu\, \partial g_\nu} \right\rangle\,, \tag{11}$$

where $\langle\cdot\rangle$ denotes the usual ensemble average (with the model probability distribution $p_{\boldsymbol{g}}(\boldsymbol{s})$). Can you show that, when $p_{\boldsymbol{g}}(\boldsymbol{s})$ is the Ising model distribution, we indeed have that

$$\mathcal{I}_{\mu,\nu}(\boldsymbol{g}) = \frac{\partial^2 \log Z(\boldsymbol{g})}{\partial g_\mu\, \partial g_\nu}\ ? \tag{12}$$

To be more general, we will also use a common notation for the single spin interactions s_i, and the pairwise interactions $s_i s_j$. We introduce the set $\boldsymbol{\phi}$ of the K spin interactions:

$$\boldsymbol{\phi} = \{s_1, s_2, \cdots, s_n, s_1 s_2, \cdots, s_{n-1}s_n\}\,. \tag{13}$$

The elements $\phi_\mu(\boldsymbol{s})$ of $\boldsymbol{\phi}$ are called *spin operators*. We recall that $\boldsymbol{g} = (h_1, \cdots, h_n, J_{12}, \cdots, J_{n-1,n})$. The element in $\boldsymbol{g}$ are organized in the same order than the elements in $\boldsymbol{\phi}$, such that for any choice of index $\mu \le K$, the parameters g_μ is the one that parametrizes the interaction $\phi_\mu(\boldsymbol{s})$. For example, $g_1 = h_1$ and $\phi_1(\boldsymbol{s}) = s_1$. Another example: $g_{n+1} = J_{12}$ and $\phi_{n+1}(\boldsymbol{s}) = s_1 s_2$.

Q3. Can you show that the energy function in the exponential of Eq. (1) can be re-written as:

$$E(\boldsymbol{s}) = -\sum_{i=1}^{n} h_i\, s_i \,-\, \sum_{pair(i,j)} J_{ij}\, s_i s_j \tag{14}$$

$$E(\boldsymbol{s}) = -\sum_{\mu=1}^{K} g_\mu\, \phi_\mu(\boldsymbol{s})\ ? \tag{15}$$

Observe that this expression is quite general. One could easily imagine using spin interactions $\phi_\mu(\boldsymbol{s})$ of order higher than pairwise. For instance, why not using a triplet interaction between the spins s_1, s_2 and s_3? This would simply corresponds to an operator $\phi_\mu(\boldsymbol{s}) = s_1 s_2 s_3$ associated with a parameter $g_\mu = J_{123}$!

Q4. Using the expression in Eq. (15) for the energy of the Ising model, can you show that:

$$\frac{\partial^2 \log Z(\boldsymbol{g})}{\partial g_\mu\, \partial g_\nu} = \langle \phi_\mu(\boldsymbol{s})\, \phi_\nu(\boldsymbol{s}) \rangle - \langle \phi_\mu(\boldsymbol{s}) \rangle \, \langle \phi_\nu(\boldsymbol{s}) \rangle \ ? \tag{16}$$

We observe that the matrix with elements $\langle \phi_\mu(\boldsymbol{s})\, \phi_\nu(\boldsymbol{s}) \rangle - \langle \phi_\mu(\boldsymbol{s}) \rangle \, \langle \phi_\nu(\boldsymbol{s}) \rangle$ is the covariance matrix of $\boldsymbol{\phi}$:

$$\mathrm{cov}(\boldsymbol{\phi}) = \langle [\boldsymbol{\phi} - \langle \boldsymbol{\phi} \rangle][\boldsymbol{\phi} - \langle \boldsymbol{\phi} \rangle]^\top \rangle \tag{17}$$

As a property of covariance matrices, the matrix $\mathrm{cov}(\boldsymbol{\phi})$ is semi-definite positive, i.e. that its determinant is always positive or null. And the determinant will be null only in cases for which the values taken by one of the operators is constant in the model (i.e. its variance is zero), which is not possible with the Ising model (that would be the case only if one takes a parameter to $\pm\infty$: for instance $h_1 = +\infty$ corresponds to a spin s_1 taking only the value $+1$). As a result the determinant of Hessian of $\mathcal{L}(\boldsymbol{g})$ is strictly negative (see Eq. (10)), and the function $\mathcal{L}(\boldsymbol{g})$ is therefore a concave function, with a single maximum at $\boldsymbol{g}^\star$. Note that, if in the data, the variance of one of the operator $\phi_\mu(\boldsymbol{s})$ is null, then the maximum will be in $\pm\infty$ for the associated parameter g_μ.