Mini-project: Residual entropy of a frustrated magnet

1 Microscopic model of a magnet

In introductory courses to statistical physics, classical magnets – systems whose microscopic degrees of freedom are classical spins – are often considered as paradigmatic models for *thermal phase transitions*. In a minimal scenario, such a phase transition originates from the competition between energy and entropy in a many-body system: to minimize the energy of the system the interaction between the microscopic magnets favor ordered states while maximizing the entropy at finite temperature favors disordered configurations. Notably, this competition typically results is a sharp transition point, at a critical temperature T^*, below which the magnet exhibits order, e.g., in the form of a spontaneous magnetization, while it is a paramagnet at higher temperatures. This leads to the question whether any system of interacting microscopic spins orders at low temperatures or whether there are mechanisms that prevent a phase transition.

In this project you will investigate the effect of *geometric frustration* on the low-temperature behavior of a magnet. The model system will be the antiferromagnetic Ising model on a *triangular* lattice as shown in Fig. 1a), which is described by the Hamilton function

$$H_J(\mathbf{s}) = J \sum_{\langle i,j \rangle} s_i s_j \ . \tag{1}$$

Here, $s_i = \pm 1$ represents a microscopic Ising magnet with two possible orientations ("up" or "down"). We will in the following refer to these degrees of freedom as *spins*. The vector $\mathbf{s} = (s_1, \ldots, s_N)$ constitutes the microscopic state of the system composed of N spins and J is the interaction energy. In the following we fix $J = 1$ as the unit of energy. The first sum is over pairs of neighboring lattice sites $\langle i, j \rangle$ in the triangular lattice and we generally consider periodic boundary conditions.

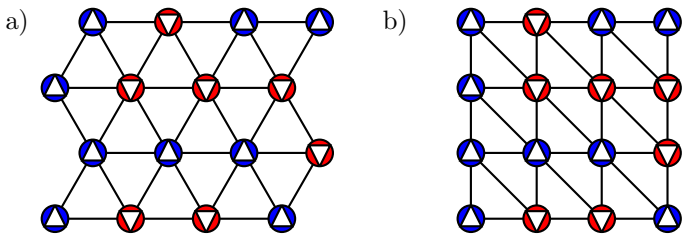

Figure 1 – a) Microscopic Ising magnets on a triangular lattice. b) The triangular lattice can also be viewed as a square lattice with additional connections along one of its diagonals.

When implementing simulations the alternative view of the triangular lattice in Fig. 1b) can be useful for easier indexing of the lattice sites. For our purpose, we restrict ourselves to square dimensions of the lattice, i.e., $N = L^2$ with edge length L.

2 Variational free energy

At a given inverse temperature $\beta = \frac{1}{T}$, the probability of a spin configuration s is given by the Boltzmann distribution

$$p_\beta(\mathbf{s}) = \frac{1}{Z} e^{-\beta H_J(\mathbf{s})} \,, \tag{2}$$

where $Z = \sum_{\mathbf{s}} e^{-\beta H_J(\mathbf{s})}$ is the *partition sum*. A central quantity of interest in statistical mechanics is the free energy

$$F = -\frac{1}{\beta} \ln Z = E - \frac{1}{\beta} S \tag{3}$$

with the energy $E = \langle H(\mathbf{s}) \rangle_{\mathbf{s}\sim p_\beta}$ and the entropy $S = \langle \ln p_\beta(\mathbf{s}) \rangle_{\mathbf{s}\sim p_\beta}$.

Consider an ansatz $q_\theta(\mathbf{s})$ for a probability distribution, which is parametrized by a set of parameters θ. The Kullback-Leibler (KL) divergence

$$D_{\mathrm{KL}}\big(q_\theta || p\big) = \sum_{\mathbf{s}} q_\theta(\mathbf{s}) \ln \left(\frac{q_\theta(\mathbf{s})}{p(\mathbf{s})} \right) \tag{4}$$

quantifies the similarity between q_θ and another distribution p.[1] By plugging in the Boltzmann distribution, we find

$$D_{\mathrm{KL}}\big(q_\theta || p_\beta\big) = \beta \big(F_\theta - F\big) \,, \tag{5}$$

where

$$F_\theta = \frac{1}{\beta} \sum_{\mathbf{s}} q_\theta(\mathbf{s}) \big(\beta H_J(\mathbf{s}) + \ln q_\theta(\mathbf{s})\big) \tag{6}$$

is the *variational free energy* and F is the exact free energy from Eq. (3). Since $D_{\mathrm{KL}}\big(q_\theta || p_\beta\big) = \beta \big(F_\theta - F\big) \geq 0$, the variational free energy is a strict upper bound of the free energy. This means we can approximate F by minimizing F_θ. Moreover, since the optimization will yield $q_\theta(\mathbf{s}) \approx p_\beta(\mathbf{s})$,

$$E \approx \langle H_J(\mathbf{s}) \rangle_{\mathbf{s}\sim q_\theta} \,, \quad S \approx \langle \ln q_\theta(\mathbf{s}) \rangle_{\mathbf{s}\sim q_\theta} \,. \tag{7}$$

This means that we can extract various quantities of interest, if we manage to minimize the variational free energy for a given energy function $H(\mathbf{s})$.

3 Autoregressive neural networks

One remaining challenge for the variational approach outlined above is to choose an ansatz $q_\theta(\mathbf{s})$, which is normalized, i.e., $\sum_{\mathbf{s}} q_\theta(\mathbf{s}) = 1$. One efficient way of enforcing the normalization is to endow the ansatz with an *autoregressive* structure. This means to express the joint distribution as a product of conditionals,

$$q_\theta(\mathbf{s}) = q_\theta^{(1)}(s_1) q_\theta^{(2)}(s_2|s_1) q_\theta^{(3)}(s_3|s_1,s_2) \dots q_\theta^{(N)}(s_N|s_1,\dots,s_{N-1}) \,. \tag{8}$$

[1] D_{KL} is not a distance, because it is not symmetric. But it is positive and $D_{\mathrm{KL}}(q||p) = 0 \Leftrightarrow p = q$.

Since the individual variables have only two possible outcomes, $s_j = \pm 1$, we can explicitly normalize every conditional distribution by computing the probability for each outcome. And if all conditionals are normalized, the same holds for the product that gives the joint distribution. In neural networks, the normalization is typically implemented using the softmax (or log-softmax) activation function.

There are various neural network architectures that exhibit the autoregressive structure, for example, RNNs, PixelCNNs and Transformers.

4 Tasks

1. Implement an autoregressive neural network ansatz for $q_\theta(\mathbf{s})$ suited for the two-dimensional Ising magnet (1). Convince yourself for a small example, that the ansatz is indeed normalized.

2. Implement a function that generates samples from the distribution $q_\theta(\mathbf{s})$, i.e., a *batch* of configurations $\mathcal{S}_q = \{\mathbf{s}^{(1)}, \ldots, \mathbf{s}^{(M)}\}_{\mathbf{s}^{(i)} \sim q_\theta}$. For this purpose, you can use the Metropolis algorithm.

3. Implement the gradient-based optimization of the variational free energy (6). Notice that F_θ is an expectation value with respect to $q_\theta(\mathbf{s})$. Since the sum of all possible configurations $\mathbf{s}$ has exponentially many terms, you have to approximate it with a Monte Carlo estimate using the batches from step 2, i.e.,

$$\beta F_\theta \approx \frac{1}{M} \sum_{\mathbf{s} \in \mathcal{S}_q} \big(\beta H_J(\mathbf{s}) + \ln q_\theta(\mathbf{s})\big) \,. \tag{9}$$

   In conventional machine learning applications one would call $\mathcal{S}_q$ a batch of *training data*. By contrast, the data is in this case generated by sampling from the neural network in every optimization step.

4. Use the variational approach to compute the entropy S as a function of temperature. What happens in the limit of small temperatures, $T \ll 1$?

5. Prepare your results in a suitable form for a short discussion (e.g., notes, notebook, or slides).