Ising Model: Markov chain-based vs Boltzmann distribution analysis

Tomasz Ślusarczyk
*Massachusetts Institute of Technology**
(Dated: May 16, 2025)

In this paper we will study the celebrated Curie-Weiss model of a spin lattice, which can be viewed as a mean-field version of the Ising model. It is known that this system exhibits a phase transition from an ordered, low-temperature phase to a disordered, high-temperature phase.

We will investigate two theoretical approches to define and study the phase transition. The first "static" method assumes the system is stationary in Boltzmann distribution, and investigates non-analyticity of free energy and magnetization in the limit of a large system size. The second "dynamic" approach treats the system as evolving over time, and its rate of convergence to Boltzmann equilibrium is studied.

I. INTRODUCTION

The Ising model was introduced in the early 20th century [1] as a model of ferromagnetism and has been widely studied by physicists, mathematicians, and computer scientists. One of its key features is the prediction of a phase transition at a critical temperature T_c (notably absent in the original work of Ising, but analyzed in great detail later), such that for $T > T_c$ the system is in andisordered (paramagnetic) state with zero net magnetization, and for $T < T_c$ the system is in an ordered (ferromagnetic) state with spontaneous nonzero magnetization. Understanding of the critical temperature and its dependence on interaction strength will be the main focus of this paper.

Formally, the Ising model is defined as follows. The system consists of $N \gg 1$ spins $s_1, \dots, s_N$, each of which can take the value of -1 or $+1$. The spins interact with each other according to the (symmetric) matrix J, and with a uniform external field h, so that the enthalpy can be written as

$$\mathcal{H} = -\sum_{i,j} J_{ij} s_i s_j - h \sum_i s_i$$

where here and throughout all indices in sums are assumed to run over $[N] \coloneqq \{1, 2, \dots, N\}$ unless specified otherwise. Hence, we obtain the partition function as

$$\begin{aligned} Z = Z(s) &= \sum_{\{s_i\}} e^{-\beta \mathcal{H}} \\ &= \sum_{\{s_i\}} \exp\left(\beta \sum_{i,j} J_{ij} s_i s_j + \beta h \sum_i s_i\right) \end{aligned} \tag{1}$$

where $\sum_s$ indicates the summation over the space Ω of all 2^N possible states $s = (s_1, \dots, s_N)$ where each spin takes value ± 1. This general model allows for various types of interactions, both ferromagnetic ($J_{ij} > 0$) and antiferromagnetic ($J_{ij} < 0$), but we will focus on the ferromagnetic case where all nonzero interactions have the same strength $J > 0$. That is, $J_{ij} = JM_{ij}$ where M_{ij} is a symmetric matrix with entries in $\{0, 1\}$, and only zeroes on the diagonal. Such matrix can be viewed as an adjacency matrix of a simple graph G on the vertex set $V = [N]$, so accordingly, for each spin s_i, we will define the set of its neighbors

$$N(i_= \{j : M_{ij} = 1\}.$$

For simplicity, we will also assume $h = 0$, though most results can be extended to nonzero h as well.

II. CRITICAL TEMPERATURE

We are now ready to describe phase transitions in a system described by eq. (1). Our main goal is to compare two ways to define a critical temperature (phase transition).

A. Non-analyticity of the partition function

The most prevalent approach used by physicists to study critical phenomena is the divergence of the partition function. In this paper we will call it the *static* approach, as it studies the system in thermal equilibrium (defined by Boltzmann distribution). More formally, let $f = \frac{\ln Z}{N}$ be the free energy per particle. Whenever for some fixed $T = T_c$ we observe that $\lim_{N\to\infty} f$ is non-analytic as a function of T_c, we call T_c a *static* critical temperature.

B. Mixing time of Glauber dynamics

One can also drop the assumption of initial "ergodicity" and instead assume initially the system is in some arbitrary state $\{s_i\}_0$. It is then allowed to evolve over time, which, for simplicity, is assumed to be discrete.[2] At each step of evolution, one spin s_i is selected uniformly at random, and then resampled according to the Boltzmann

* tomaszsl@mit.edu

distribution with remaining spins $\{s_j\}_{j\neq i}$ fixed. That is, if at stage t spin i was selected, we have

$$\begin{aligned} s_j(t+1) &= s_j(t) \qquad \text{for } j \neq i \\ \mathbb{P}[s_i(t+1)=1] &\propto \exp\big(-\beta\mathcal{H}(s, s_i \leftarrow 1)\big) \\ \mathbb{P}[s_i(t+1)=-1] &\propto \exp\big(-\beta\mathcal{H}(s, s_i \leftarrow -1)\big)\,. \end{aligned} \tag{2}$$

Here we used the notation $\mathcal{H}(s, s_i \leftarrow x)$ to denote enthalpy evaluated at the spin state $s(t) = (s_1(t), \dots, s_N(t))$ at time t, with $s_i(t)$ replaced by x (independently of its previuos value value). Notice that probabilities in eq. (2) need to be normalized so that they sum to 1. The dynamics described by eq. (2) is called the *Glauber dynamics*, and was notably introduced [3] to study specifically the Ising model, but has been applied to numerous random systems since. One can view the Glauber dynamics as one way to justify the ergodic hypothesis, as one can easily show that independently of the starting state $s(0)$, the distribution of $s(t)$ converges to the Boltzmann distribution when $t \to \infty$. Indeed, by direct verification one checks that Boltzmann distribution is the steady state of Glauber dynamics, and the Markov chain defined on the set of all states by Glauber dynamics is reversible. Then, by Perron-Frobenius theorem we conclude that the transition matrix $\mathcal{T} \in \mathrm{Sym}(\mathbb{R}^{2^N \times 2^N})$ has a largest unique eigenvalue equal to 1. This is proof is a very standard part of Markov chain theory so we will omit the details here.

Even though Glauber dynamics always convergence to the Boltzmann distribution, the rate of this convergence heavily depends on the parameters of the problem. To formally define the convergence rate, let $\mathcal{B}$ be the Boltzmann distribution, and let $\mathcal{D}(t, s)$ be the distribution of $s(t)$ resulting from t steps of Glauber dynamics, starting from the given state $s(0) = s$. Recall that the total variation (TV) distance between distributions $\mathcal{P}$ and $\mathcal{Q}$ over a finite set Ω is defined as

$$\|\mathcal{P} - \mathcal{Q}\|_{\mathrm{TV}} \coloneqq \max_{A\subseteq\Omega} |\mathcal{P}(A) - \mathcal{Q}(A)| = \frac{1}{2}\sum_{x\in\Omega} |\mathcal{P}(x) - \mathcal{Q}(x)|\,.$$

Finally, we define the mixing time t_{mix} as follows.

$$\begin{aligned} t_{\mathrm{mix}}(\epsilon, s) &= \min\{t \geq 0 : \|\mathcal{B} - \mathcal{D}(t, s)\| \leq \epsilon\} \\ t_{\mathrm{mix}}(\epsilon) &= \max_s t_{\mathrm{mix}}(\epsilon, s)\,. \end{aligned} \tag{3}$$

Finally, if we don't specify ϵ, we take $t_{\mathrm{mix}} = t_{\mathrm{mix}}(1/4)$. In fact, any constant strictly smaller than $1/2$ woudl work in place of $1/4$, as $t_{\mathrm{mix}}(\epsilon) \lesssim t_{\mathrm{mix}}(1/4)\log(1/\epsilon)$ (which is a folklore fact in the theory of Markov chains and we will omit its proof). Recall also another folklore lemma known as the data processing inequality[4]:

$$\|\mathcal{B} - \mathcal{D}(t, s)\| \leq \|\mathcal{B} - \mathcal{D}(t', s)\| \qquad \text{for } t \geq t'\,. \tag{4}$$

Hence, for any $t \geq t_{\mathrm{mix}}$ we have $\|\mathcal{D}(t) - \mathcal{B}\|_{\mathrm{TV}} \leq 1/4$, further motivating the definition of t_{mix}.

Now we are finally ready to define the critical temperature. Unfortunately, there is now clear notion of what the critical temperature should be. However, it is observed in practice that for many relevant systems the mixing time increases when T decreases. Usually t_{mix} is polynomial in N for large T and exponential in N for small T. Remarkably, usually there exists a unique temperature T_c, such that for $T_c > T$, the mixing time is polynomial, while for $T_c < T$, the mixing time is exponential. We will call such T_c the *dynamic* critical temperature. Notice that this definition corresponds to the behavior expected from the systems undergoing a phase transition from a low-temperature ordered phase to a high-temperature disordered phase: in a disordered state we would expect the system to rapidly explore most of the state space, while in an ordered state the system will be likely confined to the region of the state space corresponding to some equilibrium values of the order parameters. This notion will be made precise in the following sections.

III. CURIE–WEISS MODEL

To exemplify the two definitions of the critical temperature, let us consider the Curie–Weiss model, that is

$$M_{ij} = 1 - \delta_{i,j}\,.$$

While this model is difficult to realize experimentally due to the short range of magnetic interactions, we can view the Curie–Weiss model as a formal approach to the mean field heuristic, with each spin experiencing the same averaged field from other spins. With this perspective in mind, we will normalize J to J/N, so that the total interaction strength between s_i and its neighbors is independent of N (since s_i has $N - 1 \approx N$ neighbors).

A. Static critical temperature

Let us define the average magnetization $m = \frac{1}{N}\sum_i s_i$. The numbers of $+1$ and -1 spins (denoted by N_+ and N_-) are given by $N_+ = \frac{1+m}{2}N$ and $N_- = \frac{1-m}{2}N$. Notice that

$$\begin{aligned} &-\beta\mathcal{H}(s) \\ &\quad= 2\beta\frac{J}{N}\sum_{i<j} s_i s_j \\ &\quad= 2\beta\frac{J}{N}\left(\binom{N_+}{2} + \binom{N_-}{2} - N_+N_-\right) \\ &\quad= \beta\frac{J}{N}(N_+ - N_-)^2 + O(1) \\ &\quad= \beta J N m^2 + O(1) \end{aligned}$$

where the $O(1)$ term depends on β, J, and m, and is non-increasing in N. Notice also that there are exactly $\binom{N}{N_+}$ states for a given $m = 2N_+/N - 1$. Using the standard entropy approximation (or several Stirling approximations)

we have $\binom{N}{N_+} = e^{NH(N_+/N)+o(N)} = e^{NH((1+m)/2)+o(N)}$ where

$$H(x) = -x\ln x - (1-x)\ln(1-x)\,.$$

Finally, m ranges from -1 to 1 at steps of width $2/N$ (which we will hide under $\sum_m$). We can write

$$Z = \sum_m \binom{N}{N_+} \exp(-\beta\mathcal{H}(m)) = \sum_m e^{N\left(\beta J m^2 + H((1+m)/2)\right)+o(N)}$$

so that (approximating sum by integral and dropping non-leading terms)

$$f \approx \frac{1}{N}\ln\int_{-1}^{1}\frac{N}{2}e^{N\left(\beta J m^2+H((1+m)/2)\right)}dm\,. \tag{5}$$

Elementary calculus shows that the leading (in N) term in eq. (5) depends only on the peak of the integrand. Since its width is roughly $1/N$, after taking the logarithm the width term becomes lower-order and can be ignored. Hence

$$f \approx \max_{m\in[-1,1]}\left(\beta J m + H\left(\frac{1+m}{2}\right)\right)\,. \tag{6}$$

Routine calculation shows that for $\beta J < 1/2$ the maximum in eq. (6) is achieved at $m = 0$, so that $f(T) = \ln 2$ for $T > 2J$ (in the unit system where $k_{\rm B} = 1$). However, for $\beta J < 1/2$, the maximum is achieved at $\pm m^*$ where m^* is the positive solution of $4\beta J m + 2 + \ln(1-m^2) = 0$. Hence $T_c = 2J$ is the static critical temperature.

B. Dynamic critical temperature

Let us now show a sketch of the proof that $T_c = 2J$ is also the dynamic critical temperature. We need to provide a polynomial upper bound for $t_{\rm mix}$ when $T > 2J$, and an exponential lower bound for $t_{\rm mix}$ when $T < 2J$. Both parts rely on somewhat sophisticated theory, so we will only discusses sketches of proofs.

1. *High temperature phase: $T > 2J$*

We will rely here on the *coupling method* of Bubley and Dyer [5]. Let $\|s - s'\|_0$ denote the number of individual spins differing between states s and s'. Suppose that there is a way to sample two (non-independent) Glauber dynamics $s(t)$ and $s(t')$, so that whenever $\|s(t) - s'(t)\| = 1$, then

$$\mathbb{E}[\|s(t+1) - s'(t+1)\|_0] \le 1 - \frac{1}{CN}\,. \tag{7}$$

Then the Bubley-Dyer path coupling theorem would imply $t_{\rm mix} \le CN\log(4N)$.

Hence we need to find the correct way to sample s and s'. The first natural assumption is to select the same spin in both s and s' to resample at each stage. Suppose that in such resampling of s we get $\mathbb{P}[s_i(t+1) = 1] = p$ and $\mathbb{P}[s'_i(t+1) = 1] = p' > p$. Then instead of resampling s_i and s'_i independently, we could pick $s_i = s_{i+1} = 1$ with probability 1, $s_i = s_{i+1} = -1$ with probability $1 - p'$, and otherwise $s_i = -1$, $s'_i = 1$. This guarantees that both s and s' remain marginally Glauber dynamics, but at the same time distance between s and s' is minimized: $\mathbb{P}[s_i(t+1) \ne s'_i(t+1)] = |p - p'|$. Now if $s(t)$ and $s'(t)$ differ only at spin i, resampling i will always lead to $s(t+1) = s'(t+1)$ by construction. Resampling $j \ne i$ leads to $s(t+1) \ne s'(t+1)$ with probability

$$|p_j - p'_j| = \left|\frac{e^{-\beta\mathcal{H}(s,s_j\leftarrow 1)}}{e^{-\beta\mathcal{H}(s,s_j\leftarrow 1)} + e^{-\beta\mathcal{H}(s,s_j\leftarrow -1)}} - \frac{e^{-\beta\mathcal{H}(s',s'_j\leftarrow 1)}}{e^{-\beta\mathcal{H}(s',s'_j\leftarrow 1)} + e^{-\beta\mathcal{H}(s',s'_j\leftarrow -1)}}\right| \le \tanh(2\beta J/N) \tag{8}$$

where the last inequality follows from elementary but quite tedious optimization over s, s' (as one might expect, the difference in eq. (8) is maximized when $\sum_{k\ne i,j} s_k = 0$). Hence for $\|s(t) - s'(t)\|_0 = 1$

$$\begin{aligned}
&\mathbb{E}[\|s(t+1) - s'(t+1)\|_0] \\
&\quad = 1 + \mathbb{P}[\|s(t+1) - s'(t+1)\|_0 = 2] \\
&\qquad - \mathbb{P}[\|s(t+1) - s'(t+1)\|_0 = 0] \\
&\quad = 1 + \frac{1}{N}\sum_{j\ne i}|p_j - p'_j| - \frac{1}{N} \\
&\quad \le 1 - \frac{1 - N\tanh(2\beta J/N)}{N} \le 1 - \frac{1-2\beta J}{N}\,.
\end{aligned}$$

So indeed when $T > T_c = 2J$ we have $1 - 2\beta J = c > 0$, and hence eq. (7) is satisfied, and we obtain $t_{\rm mix} \lesssim$

$N \log N$.

2. Low temperature phase: $T < 2J$

This case is more closely related to the static computation. We will use the *conductance method*, which formalizes the intuition that Glauber dynamics should mix slowly, since the system becomes "trapped" in $m = m^*$ macrostate, and will take a long time to flip to $m = -m^*$. Let Ω_+ be the subspace of the state space with $m > 0$, $\Omega_- = \Omega \setminus \Omega_+$, and suppose that $s(0) \in \Omega_+$. Let t^*_{mix} be the first time when $\mathbb{P}[s(t^*_{\text{mix}}) \in \Omega_+] \le 3/4$, notice that $t^*_{\text{mix}} < t_{\text{mix}}$. We have

$$\begin{aligned}\frac{1}{4} &\le \mathbb{P}[s(t^*_{\text{mix}}) \notin \Omega_+] \\ &\le \sum_{t=0}^{t^*_{\text{mix}}-1} \mathbb{P}[s(t+1) \notin \Omega_+ \mid s(t) \in \Omega_+] \cdot \mathbb{P}[s(t) \in \Omega_+] \\ &\lesssim t^*_{\text{mix}} \mathbb{P}_{s'(0)\sim\mathcal{B}}[s'(1) \notin \Omega_+ \mid s'(0) \in \Omega_+]\end{aligned}$$

where the last inequality is only a heuristic: probability of moving from Ω_+ to Ω_- in s shouldn't be much larger than the probability that a stationary system (under Boltzmann distribution $\mathcal{B}$) moves to Ω_- from Ω_+ in one step. This would suggest that

$$t_{\text{mix}} \ge \frac{c}{\mathbb{P}_{s(0)\sim\mathcal{B}}[s(1) \notin \Omega_+ \mid s(0) \in \Omega_+]}. \tag{9}$$

Equation (9) turns out to be true (up to constants) in general. Let $1 = \gamma_1 > \gamma_2 \ge \cdots \ge \gamma_N > -1$ be the eigenvalues of the transition matrix $\mathcal{T}$ of the Markov chain. Let $\gamma = 1 - \max(\gamma_2, -\gamma_N)$ be the spectral gap. Then

$$t_{\text{mix}} \ge \left(\frac{1}{\gamma} - 1\right) \log \frac{1}{8} \tag{10}$$

(see [6]). To bound γ, we can use (a version of) the Cheeger's inequality (see [6] for a folklore proof):

$$\gamma \le 2 \frac{\mathbb{P}_{s(0)\sim\mathcal{B}}[s(1) \notin \Omega_+ \mid s(0) \in \Omega_+]}{\mathbb{P}_{s(0)\sim\mathcal{B}}[s(0) \in \Omega_+]} \tag{11}$$

(if we replace Ω_+ with a more optimal set, we would obtain the *conductance* of the state space, hence the name of the method). The combination of eq. (10) and eq. (11) essentially recovers eq. (9). Now we can finish the computation.

Notice that by symmetry

$$\mathbb{P}_{s(0)\sim\mathcal{B}}[\mid s(0) \in \Omega_+] = \frac{1}{2} - \mathbb{P}_{s(0)\sim\mathcal{B}}[m(s(0)) = 0]$$

where $m = \sum_i s_i$ as before. At the same time for a state to leave Ω_+ it needs to have $m = 0$, so

$$\mathbb{P}_{s(0)\sim\mathcal{B}}[s(1) \notin \Omega_+ \mid s(0) \in \Omega_+] \le \mathbb{P}_{s(0)\sim\mathcal{B}}[m(s(0)) = 0].$$

With $\mathbb{P}_{s(0)\sim\mathcal{B}}[m(s(0)) = 0] = p$, we have then

$$t_{\text{mix}} \gtrsim \frac{1}{\gamma} \gtrsim \frac{1/2 - p}{p}.$$

Finally, notice that for $T < T_c$ we have

$$\begin{aligned}p &= \frac{\binom{N}{N/2} e^{-\beta \mathcal{H}(0)}}{Z} \\ &= e^{N \ln 2 - N\left(\beta J m^* + H((1+m^*)/2)\right) + o(N)}.\end{aligned}$$

As discussed before, for $T < T_c$ we have $m^* > 0$, and $\beta J m^* + H((1+m^*)/2) = f > \ln 2$, so that $p = e^{-cN + o(N)}$ for $c = f - \ln 2 > 0$. Hence $t_{\text{mix}} > e^{cN + o(N)}$ as desired.

All in all we conclude that the dynamic critical temperature is the same as the static critical temperature $T_c = 2J$.

IV. CONCLUSION

We demonstrated that the static and dynamic critical temperatures of the Curie–Weiss model are identical. The static computation is largely considered folklore in statistical mechanics, but the dynamic part relies on more sophisticated theory related to Markov chains. However, in the low-temperature regime, both approaches are closely related and result in the same final optimization. This finding suggests there is a deeper connection between the two methods, which indeed has been studied but is beyond the scope of this paper. Since the static approach is preferred by physicists, and the dynamic approach is preferred by computer scientists, many results consider different regimes or limits (e.g. physicists usually study geometrically-constrained constructions, while in computer science, arbitrary or random interaction graphs are of greater interest, and so are algorithmic considerations). However, in certain cases all approaches converge, such as in the beautiful theory of the exact solution of the 2-dimensional Ising model by Onsager. The summation over paths executed by Onsager can be related to a version of the correlation decay on the computation tree of a regular graph. This, however, is a much more recent approach (see e.g [7] for application of the method to antiferromagnetic systems), and is much more complex than the results in this paper.

[1] E. Ising, Contribution to the Theory of Ferromagnetism, Z. Phys. **31**, 253 (1925).

[2] The continuous time setting is largely equivalent to discrete time. Indeed, if evolution steps come as Poisson point events with rate γ, after large time t the system can be approximated as having experienced exactly $\lfloor \gamma t \rfloor$ evolution steps. Then various probabilistic tools can be used to formally justify why deviations from the mean are negligible – which usually is not very difficult, but annoying technically and hence beyond the scope of this paper.

[3] R. J. Glauber, Time-dependent statistics of the ising model, Journal of Mathematical Physics **4**, 294 (1963).

[4] Note that this inequality is somewhat different from another inequality on mutual information and entropy also known as data processing inequality.

[5] R. Bubley and M. Dyer, Path coupling: A technique for proving rapid mixing in markov chains, in *Proceedings 38th Annual Symposium on Foundations of Computer Science* (1997) pp. 223–231.

[6] D. A. Levin, Y. Peres, and E. L. Wilmer, *Markov chains and mixing times* (American Mathematical Society, 2006).

[7] L. Li, P. Lu, and Y. Yin, Correlation decay up to uniqueness in spin systems, in *Proceedings of the 2013 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pp. 67–84, https://epubs.siam.org/doi/pdf/10.1137/1.9781611973105.5.