Exploring the Temperature-Dependent Phase Transition in Modern Hopfield Networks

Felix Koulischer   **Cédric Goemaere**   **Tom Van Der Meersch**
Johannes Deleu   **Thomas Demeester**
IDLab, Ghent University – imec
Ghent, Belgium
`first.last@ugent.be`

Abstract

The recent discovery of a connection between Transformers and Modern Hopfield Networks (MHNs) has reignited the study of neural networks from a physical energy-based perspective. This paper focuses on the pivotal effect of the inverse temperature hyperparameter β on the distribution of energy minima of the MHN. To achieve this, the distribution of energy minima is tracked in a simplified MHN in which equidistant normalised patterns are stored. This network demonstrates a phase transition at a critical temperature β_{c}, from a single global attractor towards highly pattern specific minima as β is increased. Importantly, the dynamics are not solely governed by the hyperparameter β but are instead determined by an effective inverse temperature β_{eff} which also depends on the distribution and size of the stored patterns. Recognizing the role of hyperparameters in the MHN could, in the future, aid researchers in the domain of Transformers to optimise their initial choices, potentially reducing the necessity for time and energy expensive hyperparameter fine-tuning.

1   Introduction

Energy based models (EBMs) have been around for over twenty years, but were until recently regarded by many as historical artefacts, due to overwhelming successes in deep neural networks trained with backpropagation [1]. However, recent developments have once again brought EBMs into the spotlight [2], aligning with the surging interest in more energy-efficient [3] or biologically plausible neural networks [4]. The main constraint of the prototypical EBM, the Hopfield Network (HN) [5, 6], namely its low storage capacity, was remedied by increasing the interaction order inside the hamiltonian [7, 8]. Such networks, known as Dense Associative Memory models (DAMs), possess a polynomial storage capacity C, i.e. with d the dimension of states within the network and n the interaction order, one has that $C \propto d^{n-1}$. Going a step further and using an exponential interaction, effectively considering an infinitely large interaction order, an exponential storage capacity was obtained, i.e. $C \propto \alpha^d$ where α is a constant [9, 10, 11]. Interestingly, the EBM consisting of such an exponential interaction vertex, referred to as the Modern Hopfield Network (MHN), displays a strong connection to a widespread neural network architecture, the Transformer [10, 12]. The connection between the two arises from the specific mathematical form of the energy minimising update rule, consisting of a Softmax averaging operation reminiscent of the attention mechanism present in Transformers. While the Transformer has become the go-to architecture in many fields [13], the understanding of its inner workings remains limited [14]. This is in stark contrast with the MHN, which is very well described from a theoretical point of view, while currently remaining rarely used in practice [10, 14]. It is hoped that the theoretical understanding of MHNs could help guide research towards novel Transformer architectures [14].
The energy function of the MHN, whose expression strongly resembles the Helmholtz free energy in

the canonical ensemble from statistical physics, is very sensitive to the inverse temperature hyper-parameter β (sometimes referred to as the interaction strength [11]). Amongst other, β determines the peakedness of the arising Softmax function by rescaling the dot-product inside the exponential interaction. It is also known that β has an impact on the distribution of the minima of the energy function [10, 11]. At large β, the minima reside close to single stored patterns, while at lower values the minima often reside in the linear combination of multiple similar patterns [10]. This paper, akin to the approach taken by C. Lucibello and M. Mézard [11], investigates the transformation of the distribution of the energy minima of the MHN as β varies from zero to infinity. At very high temperatures ($\beta \approx 0$) the energy landscape is maximally blurred, and a single global minimum is present at the origin [10, 11]. Conversely, at very low temperatures ($\beta \approx \infty$) the energy landscape consists of many sharp minima residing in each of the stored patterns. Between these two limiting scenarios, a certain phase transition occurs. In contrast to C. Lucibello and M.Mézard's work, we will assume the patterns stored in the MHN to be equidistant and normalized, rather than Gaussian or uniformly distributed spherical vectors. This particular simplifying assumption allows for an analytical treatment, which yields useful insights into MHNs' inner workings. Specifically, it allows computing the critical temperature β_c at which the phase transition occurs. Crucially, it appears that the network dynamics are governed by a physical quantity we will refer to as the *effective inverse temperature*, denoted as β_{eff}. We will show how β_{eff} depends on the hyperparameter β, as well as on the distribution and size of the stored patterns.

Understanding how β affects the network dynamics could guide researchers working on MHNs, and potentially even on Transformers, towards better hyperparameter selection, which is known to be an energy and time consuming aspect to practical neural networks training [15].

2 Theoretical considerations on phase transitions in MHNs

Modern Hopfield Networks
Primarily designed to work as an associative memory model, the MHN stores N patterns $\mathbf{x_i}$ of dimension d. These patterns are grouped together as the columns of a single matrix $\mathbf{X}$. As all EBMs, its dynamics are governed by the minimisation of its energy function in the state space represented by the state vector $\boldsymbol{\xi} \in \mathbb{R}^d$. The considered energy function is given by:

$$E\left(\boldsymbol{\xi}, \mathbf{X}, \beta\right) = -\frac{1}{\beta}\ln\sum_{i=1}^{N}\exp\left(\beta \mathbf{x_i}^T\boldsymbol{\xi}\right) + \frac{1}{2}\boldsymbol{\xi}^T\boldsymbol{\xi} \tag{1}$$

Application of the concave-convex procedure (CCCP) [16] leads to the following energy minimising state update rule that expresses the subsequent state $\boldsymbol{\xi}'$ as a function of the current state $\boldsymbol{\xi}$ and the given patterns $\boldsymbol{X}$:

$$\boldsymbol{\xi}' = \mathbf{X}\boldsymbol{p}, \quad \text{with} \ \ \boldsymbol{p} = \frac{\exp\left(\beta\boldsymbol{X}^T\boldsymbol{\xi}\right)}{\sum_{k=1}^{N}\exp\left(\beta\boldsymbol{x}_k^T\boldsymbol{\xi}\right)} \tag{2}$$

Here, the stochastic probability $\boldsymbol{p}$-vector has been introduced, containing the recognizable Softmax operator. It is clear that as the state $\boldsymbol{\xi}$ converges towards a stable state $\boldsymbol{\xi}^*$, the $\mathbf{p}$-vector also converges towards a stable state $\mathbf{p}^{*\,1}$. It is therefore sufficient to look at how the $\mathbf{p}$-vector changes under repeated application of the Softmax update rule to understand how β affects the equilibration dynamics. Moreover, the $\mathbf{p}$-vector serves as a direct representation of the proximity of a given state to the stored patterns, making it a valuable tool to understand how the distribution of energy minima depends on the positions of these patterns.

Effective inverse temperature
In order to facilitate the analytical analysis of the role of the inverse temperature β, the following simplifying assumptions on the nature of the stored patterns $\boldsymbol{x}_i$ are made. First, the stored patterns are assumed to have identical norms, i.e. $\|\mathbf{x_i}\| = \|\mathbf{x}\|, \forall i = 1, .., N$. Second, it is assumed that all patterns have the same separation angle between them, i.e. $\boldsymbol{x}_i^T\boldsymbol{x}_j = \|\boldsymbol{x}\|^2\cos(\theta), \forall i \neq j$, with θ the angle between both patterns[2]. Geometrically, the stored patterns can be thought of as equidistant points on the d-sphere. The above assumptions can be expressed as

$$\left(\boldsymbol{X}^T\boldsymbol{X}\right)_{ij} = \|\boldsymbol{x}\|^2\big(\cos\theta + (1 - \cos\theta)\delta_{ij}\big) \tag{3}$$

---

[1]Technically, this is only correct up to the kernel of the matrix $\mathbf{X}$ (i.e. $\boldsymbol{\xi}' = \boldsymbol{\xi} \iff \mathbf{p}' = \mathbf{p} + \ker(\mathbf{X})$). It was however verified that as ξ converged towards its equilibrium value, the same held for $\mathbf{p}$.

[2]This imposes a restriction on the number of stored patterns N, namely $N \leq d + 1$.

with δ_{ij} the Kronecker delta. For the constrained patterns satisfying (3) we find, by substitution of the state update equation (2),

$$\boldsymbol{X}^T \boldsymbol{\xi}' = \boldsymbol{X}^T \boldsymbol{X} \boldsymbol{p} = \|\boldsymbol{x}\|^2 \big(\cos\theta \, \boldsymbol{1} + (1 - \cos\theta)\, \boldsymbol{p} \big) \qquad (4)$$

such that the update equation for $\boldsymbol{p}$ can be written as

$$\boldsymbol{p}' = \frac{\exp\big(\beta \boldsymbol{X}^T \boldsymbol{\xi}'\big)}{\sum_{k=1}^{N} \exp\big(\beta \boldsymbol{x}_k^T \boldsymbol{\xi}'\big)} = \frac{\exp\big(\beta \|\boldsymbol{x}\|^2 (1 - \cos\theta)\, \boldsymbol{p}\big)}{\sum_{k=1}^{N} \exp\big(\beta \|\boldsymbol{x}\|^2 (1 - \cos\theta)\, p_k\big)} = \frac{\exp\big(\beta_{\text{eff}}\, \boldsymbol{p}\big)}{\sum_{k=1}^{N} \exp\big(\beta_{\text{eff}}\, p_k\big)} \qquad (5)$$

with the introduction of the *effective* inverse temperature $\beta_{\text{eff}} = \beta \|\mathbf{x}\|^2 (1 - \cos(\theta))$, which can be seen as the relevant parameter that directly governs the equilibrium dynamics. Due to the simplicity of the pattern correlations, the MHN dynamics mimic those of the iterated softmax [17, 18].

The obtained expression for β_{eff} implies that when the stored patterns are highly similar ($\cos(\theta) \to 1$), the hyperparameter β has to be chosen significantly higher to exhibit the same dynamics (i.e., a similar β_{eff}) as a network consisting of markedly distinct patterns. As explained by Ramsdauer et al., when multiple patterns are in close proximity they tend to form an energy minimum lying close to their mean, referred to as metastable state [10]. To further enhance a pattern specific memory, it makes sense that β should be increased further. The dependence of β_{eff} on the patterns norms stems from the quadratic scaling of the dot-product on the former. Crucially, this implies that distinct β values are required to obtain similar dynamics depending on whether normalisation is used in the network.

Phase transition and critical inverse temperature
Another important note is that the uniform **p**-vector, defined by $p_i = \frac{1}{N}$, $\forall i = 1, .., N$, is always a fixed point of the update rule defined by (5). This fixed point further corresponds to the high temperature limit, as $\beta_{\text{eff}} \to 0$ one has that $p_i \to \frac{1}{N}$. To understand the low temperature limit, it is useful to consider that the **p**-vector has one component, say p_I, that is larger than all others, i.e. $p_I > p_i$, $\forall i \neq I$. In the limit of $\beta_{\text{eff}} \to \infty$ the sum can be approximated by a single term, namely $\sum_{i=1}^{N} \exp(\beta p_i) \approx \exp(\beta p_I)$. One can then write:

$$p_i' \approx \frac{\exp(\beta_{\text{eff}} p_i)}{\exp(\beta_{\text{eff}} p_I)} = \exp(\beta_{\text{eff}}(p_i - p_I)) \approx \delta_{iI} \text{ as } \beta_{\text{eff}} \to \infty \qquad (6)$$

This observation indicates that, at low temperatures, the energy minima of the Modern Hopfield Network are situated close to the stored patterns. The value of β_c for which the uniform **p**-vector becomes unstable in this model can be computed. Due to the high degree of symmetry present in the patterns, arising due to our assumptions, N different minima will appear simultaneously when the uniform fixed point loses its stability.

To find at which value of β this happens, a cob-web like analysis can be followed [19, 20], such that the following set of equations needs to be solved:

$$\boldsymbol{p}' = \boldsymbol{p} = \boldsymbol{p}^*$$
$$\nabla p_1' = (1, 0, ..., 0) \qquad (7)$$

These equations can be transformed into

$$\begin{cases} 1 = p_c \left(1 + (N-1) \exp \dfrac{1 - N p_c}{N p_c (1 - p_c)} \right) \\ \beta_c = \dfrac{N-1}{N} \dfrac{1}{p_c (1 - p_c)} \end{cases} \qquad (8)$$

with the introduction of the critical values p_c, β_c.

Figure 1: N-dependence of critical inverse temperature β_c, according to (8).

Details on how the above equations were derived, are provided in Appendix A. Numerically solving these equations, taking into account that $p_c \in\ ]0, 1]$, yields the critical values p_c and β_c for any given N, as visualized in Fig. 1.
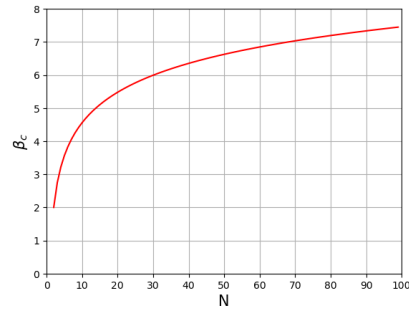
3 Empirical validation

Phase transition in MHN with equally spaced and equally normed patterns
To understand the transition between both limiting cases (uniform vs. one-hot $\boldsymbol{p}$ vector), the Kullback-
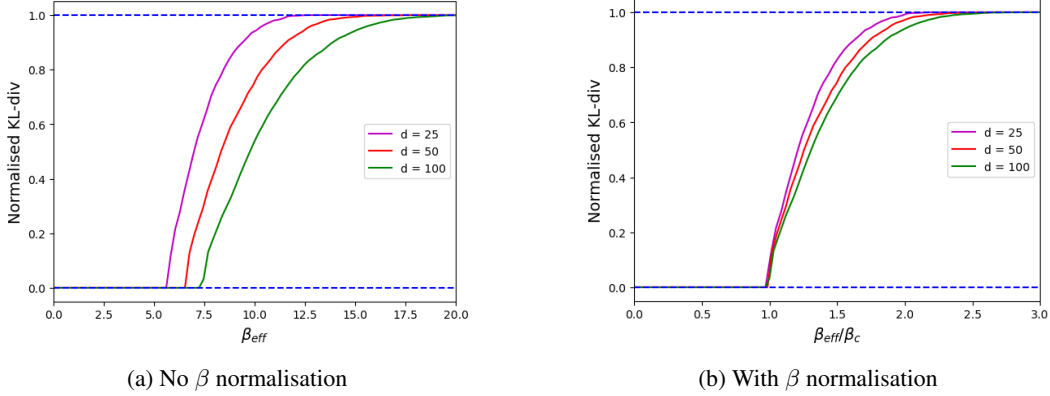
(a) No β normalisation   (b) With β normalisation

Figure 2: β dependence of the normalised KL-divergence for different values of d. In (b) β is normalised with respect to a critical value β_c which is obtained analytically by solving (8).

Leibler divergence (D_{KL}) of $\boldsymbol{p}^*$ w.r.t a uniform distribution is analysed. To make the results independent of the dimension, D_{KL} is normalised such that the divergence at large β is equal to one [3]. How D_{KL} varies as a function of β_{eff} is shown in Fig. 2. At the critical point of β_c the uniform **p**-vector becomes unstable. As β is further increased, the **p**-vector becomes increasingly peaked. From these results, it is clear that for an MHN to work as an associative memory model β_{eff} should be chosen larger than β_c, which is easily achieved by increasing β.

Phase transition in MHN with MNIST patterns

In a final empirical result, we show that the presence of a phase transition as described above is not merely a consequence of the strict conditions imposed on the stored patterns [11]. To illustrate this, an MHN containing $N = 25$, 50 and 100 randomly chosen MNIST digits was designed. The number of minima reached starting from noisy versions of the stored patterns was then counted, and displayed in Fig. 3. From this figure, it becomes clear that the MHN consisting of freely chosen patterns displays the same kind of phase transition as the simplified network treated in this paper. How the critical parameter β_c varies as a function of N also seems to correspond to the observations above. Precisely how an effective inverse temperature could be introduced in such a *free* setting is left for future work.
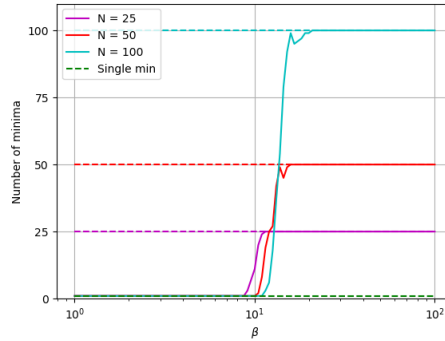

Figure 3: Number of minima in an MHN containing MNIST images as a function of the hyperparameter β.

4 Conclusion

An analysis of the Modern Hopfield Network's behavior when tasked with storing equidistant patterns of identical norms provided valuable insights into its core properties. In particular, we described the presence of a phase transition limiting the regime of high temperatures, at which the energy function displays a single global attractor, to that of low temperatures, at which the energy minima lie infinitesimally close to the stored patterns. The critical value at which this transition appeared in the simplified model was analytically obtained. In addition, we discussed the dependence of the dynamics on the *effective inverse temperature*, which also depends on the distribution and norm of the stored patterns. As a practical take-away, we have provided insights on how the optimal choice of the hyperparameter β for MHNs may vary, depending on the properties of the stored patterns.

---

[3]Mathematically the D_{KL} is divided by the divergence between a delta distribution and a uniformly distributed vector, i.e. by $\log(d)$

Aknowledgement

References

[1] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. `http://www.deeplearningbook.org`. MIT Press, 2016.

[2] Dmitry Krotov. "A new frontier for Hopfield networks". In: *Nature Reviews Physics* 5.7 (2023). DOI: `10.1038/s42254-023-00595-y`.

[3] Gianluca Zoppo, Francesco Marrone, and Fernando Corinto. "Equilibrium Propagation for Memristor-Based Recurrent Neural Networks". In: *Frontiers in Neuroscience* 14 (2020). DOI: `10.3389/fnins.2020.00240`.

[4] Yoshua Bengio et al. "Towards Biologically Plausible Deep Learning". In: *CoRR* (2015). arXiv: `1502.04156`.

[5] J J Hopfield. "Neural networks and physical systems with emergent collective computational abilities." In: *Proceedings of the National Academy of Sciences* 79.8 (1982). DOI: `10.1073/pnas.79.8.2554`.

[6] J J Hopfield. "Neurons with graded response have collective computational properties like those of two-state neurons." In: *Proceedings of the National Academy of Sciences* 81.10 (1984). DOI: `10.1073/pnas.81.10.3088`.

[7] Dmitry Krotov and John J. Hopfield. "Dense Associative Memory for Pattern Recognition". In: *Advances in Neural Information Processing Systems*. Vol. 29. 2016. URL: `https://proceedings.neurips.cc/paper_files/paper/2016/file/eaae339c4d89fc102edd9dbdb6a28915-Paper.pdf`.

[8] Dmitry Krotov and John Hopfield. "Dense Associative Memory is Robust to Adversarial Inputs". In: *Neural Computation* 30.12 (2018). DOI: `10.1162/neco_a_01143`.

[9] Mete Demircigil et al. "On a Model of Associative Memory with Huge Storage Capacity". In: *Journal of Statistical Physics* 168.2 (2017). DOI: `10.1007/s10955-017-1806-y`.

[10] Hubert Ramsauer et al. "Hopfield Networks is All You Need". In: *International Conference on Learning Representations*. 2021. URL: `https://openreview.net/forum?id=tL89RnzIiCd`.

[11] Carlo Lucibello and Marc Mézard. "The Exponential Capacity of Dense Associative Memories". In: *ArXiv* (2023). arXiv: `2304.14964`.

[12] Ashish Vaswani et al. "Attention is All you Need". In: *Advances in Neural Information Processing Systems*. Vol. 30. 2017. URL: `https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf`.

[13] Tianyang Lin et al. "A survey of transformers". In: *AI Open* 3 (2022). DOI: `10.1016/j.aiopen.2022.10.001`.

[14] Benjamin Hoover et al. "Energy Transformer". In: *ArXiv* (2023). arXiv: `2302.07253`.

[15] Emma Strubell, Ananya Ganesh, and Andrew McCallum. "Energy and Policy Considerations for Modern Deep Learning Research". In: *Proceedings of the AAAI Conference on Artificial Intelligence* 34.9 (2020). DOI: `10.1609/aaai.v34i09.7123`.

[16] A. L. Yuille and Anand Rangarajan. "The Concave-Convex Procedure". In: *Neural Computation* 15.4 (2003). DOI: `10.1162/08997660360581958`.

[17] Peter Tino. "Equilibria of iterative softmax and critical temperatures for intermittent search in self-organizing neural networks". In: *Neural Computation* 19.4 (2007). DOI: `https://doi.org/10.1162/neco.2007.19.4.1056`.

[18] Peter Tiňo. "Bifurcation structure of equilibria of iterated softmax". In: *Chaos, Solitons Fractals* 41.4 (2009). DOI: `https://doi.org/10.1016/j.chaos.2008.07.026`.

[19] Steven H Strogatz. *Nonlinear dynamics and chaos*. 2nd ed. London, England: Routledge Cavendish, 2018. Chap. 2.

[20] J. J. Binney et al. *The Theory of Critical Phenomena: An Introduction to the Renormalization Group*. USA: Oxford University Press, Inc., 1992.

A Computing the Critical Inverse Temperature β_c

In this appendix, the mathematics leading to the expression of the critical parameters are described in more detail. The considered network is again an MHN containing normalised equidistant patterns. To have a better understanding of how the critical values in such a system can be found, it is helpful to first look at the situation with a unique degree of freedom. Due to the normalising condition, this situation is present when $N = 2$, as the system is then entirely characterised by the first component of the $\boldsymbol{p}$-vector, namely $\boldsymbol{p} = (p, 1 - p)$.

In this simple case, the update equation reduces to:

$$p' = f(p) = \frac{\exp(\beta p)}{\exp(\beta p) + \exp(\beta(1 - p))} = \frac{\exp(\beta p)}{Z} \qquad (9)$$

To analyse the fixed point of such a system, it is easiest to do a cobweb analysis [19]. The idea is to simultaneously plot the function $y = f(p)$ and the first bisector $y = p$ in the given range of interest, in this case $p \in [0, 1]$. This is shown for different values of β on Fig. A1a. It should be clear that the intersection of both curves represent the fixed points of the system, described by the equality $p^* = f(p^*)$. The stability of a fixed point is then characterised by the slope of the function f at the given fixed point [19]. If the slope is smaller than one, the considered fixed point is stable, while the opposite holds when the slope is larger than one.

Two distinct phases of the system are recognisable. If $\beta < \beta_c$ a unique stable fixed point is found. This fixed point lies precisely in the uniform $\boldsymbol{p}$-vector, i.e. in $p = \frac{1}{2}$. For low values of β this uniform fixed point is stable, i.e. $\frac{\mathrm{d}f}{\mathrm{d}p}\big|_{p*} < 1$. At a certain value of β, which is defined as the critical value β_c, a pitchfork bifurcation occurs [19]. Two stable fixed points appear, while the uniform fixed point loses its stability. As β is increased further, the two stable fixed points go closer and closer to the values of 0 and 1. This highlights that as the temperature decreases, the fixed point of the $\boldsymbol{p}$-vector becomes increasingly peaked.

This phase transition can also be understood by analysing how the energy function changes for different values of β. It should be noted that within the made assumptions the energy function introduced by Ramsauer et al. [10], can be rewritten as a function of the $\boldsymbol{p}$-vector, i.e. $E \equiv E(\boldsymbol{p}, \beta)$. Neglecting the physically irrelevant constants, one finds:

$$
\begin{aligned}
E &= -\frac{1}{\beta} \ln \left(\sum_{i=1}^{N} \exp(\beta \boldsymbol{x_i^T} \boldsymbol{\xi}) \right) + \frac{1}{2} \boldsymbol{\xi^T} \boldsymbol{\xi} \\
&= -\frac{1}{\beta} \ln \left(\sum_{i=1}^{N} \exp(\beta p_i) \prod_{j \neq i} \exp(\beta \cos(\theta) p_j) \right) + \frac{1}{2} \sum_{i=1}^{N} \left(p_i^2 + \cos(\theta) \sum_{j \neq i}^{N} p_i p_j \right)
\end{aligned}
\qquad (10)
$$
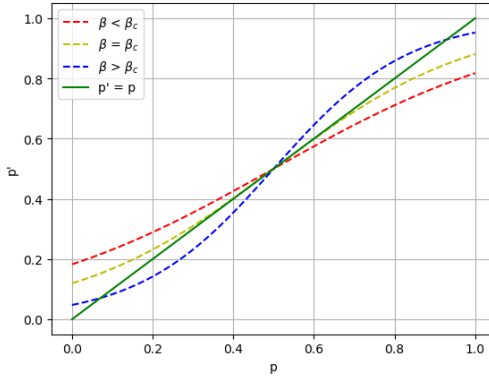
(a) Update rule for cobweb analysis
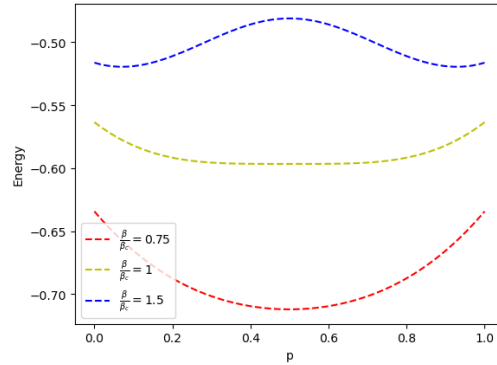
(b) Energy function for different values of β

Figure A1: In (a), graphical representation of the update rule in one dimension for different values of β. In (b), the energy function is illustrated for diferent values of β

This energy function is plotted in 1D for different values of β in Fig. A1b. On these plots, the physical cause of the phase transition becomes clear. At high temperatures ($\beta < \beta_\mathrm{c}$) a single minimum is present residing in the uniform $\boldsymbol{p}$-vector. As the temperature is decreased, this minimum becomes flatter and flatter until at some critical value it separates into two minima.

The critical value β_c can be found by solving the following system:

$$
\begin{cases}
p' = f(p) = p = p* \\
\dfrac{\mathrm{d}f}{\mathrm{d}p}\Big|_{p*} = 1
\end{cases}
\tag{11}
$$

The first equation represents the fixed point condition. The second equation encompasses the requirement that the uniform fixed point changes stability. This set of equation will not explicitly be solved in this simple illustrative case. Instead, it will first be outlined how this method can be generalised to N dimensions.

It should be noted that due to the normalisation condition of the $\boldsymbol{p}$-vector, in N-dimensions there are $N - 1$ degrees of freedom, i.e. $\boldsymbol{p} = (p_1, ..., p_{N-1}, 1 - p_1 - ... - p_{N-1})$. The update equation for each component is then given by:

$$
p_i' = \frac{\exp(\beta p_i)}{Z}
\tag{12}
$$

Where Z is analogue to the partition function from statistical physics given by $Z = \sum_{k=1}^{N} \exp(\beta p_k)$. Notice that, thanks to the symmetry present in the system, it can be understood that when the uniform fixed point loses its stability, N different minima will appear simultaneously, in some sense a pitchfork bifurcation is still present. This originates from the fact that all degrees of freedom can be interchanged freely in the energy function defined by (10). This implies that analysing the evolution of one of the degrees of freedom, say p_1, should tell us everything there is to know about the system. Imposing the same conditions as previously done in the cob-web analysis gives:

$$
\begin{cases}
\boldsymbol{p}' = f(\boldsymbol{p}) = \boldsymbol{p} = \boldsymbol{p}* \\
\nabla f(\boldsymbol{p})|_{p*} = (1, 0, ..., 0)
\end{cases}
\tag{13}
$$

It should be noted that this simple scheme is not sufficient in systems where such a high degree of symmetry is not present. In practice, one should apply other techniques based on the *Renormalisation Group* [20]. The application of such more intricate methods is however beyond the scope of this article and is left for future work.

To solve this system, it proves useful to first consider $j \neq 1$:

$$
\begin{aligned}
\frac{\partial p_1'}{\partial p_j} &= \frac{\partial}{\partial p_j} \frac{\exp(\beta_\mathrm{c} p_1)}{Z} \\
&= -\frac{\exp(\beta_\mathrm{c} p_1)}{Z^2} \frac{\partial Z}{\partial p_j} \\
&= -\beta_\mathrm{c} \frac{\exp(\beta_\mathrm{c} p_1)}{Z^2} \left[\exp \beta_\mathrm{c} p_j - \exp \beta_\mathrm{c} p_N\right]
\end{aligned}
\tag{14}
$$

These simplifications are easily understood when realising that the only two terms in Z depending on p_j are $\exp(\beta_\mathrm{c} p_j)$ and $\exp(\beta_\mathrm{c} p_N)$. Asking that the last line vanishes, as required by (13), one finds that $\exp \beta_\mathrm{c} p_j = \exp \beta_\mathrm{c} p_N$. Or thus that $p_j = p_N, \forall j \neq 1$. Hence, as expected, due to the symmetry of the considered model the original N degrees of freedom reduce to a unique degree of freedom which will further simply be denoted as $p = p_1$. Using the fact that $\sum_{i=1}^{N} p_i = 1$ it is then easily found that $p_N = \frac{1-p}{N-1}$.

The case of $i = 1$ can then be analysed:

$$
\begin{aligned}
\frac{\partial p\prime}{\partial p} &= \frac{\partial}{\partial p} \frac{\exp(\beta_\mathrm{c} p)}{Z} \\
&= \frac{\beta_\mathrm{c}}{Z^2} \left[Z \exp(\beta_\mathrm{c} p) - \exp(\beta_\mathrm{c} p) \left(\exp(\beta_\mathrm{c} p) - \exp\left(\beta_\mathrm{c} \frac{1-p}{N-1}\right)\right)\right] \\
&= -\beta_\mathrm{c} \left[p - p^2 + p\frac{1-p}{N-1}\right] = 1
\end{aligned}
\tag{15}
$$
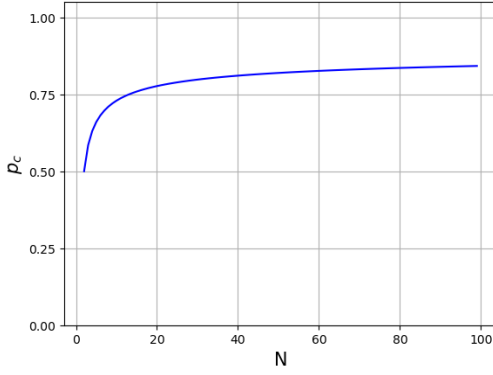
(a) N-dependence of p_c
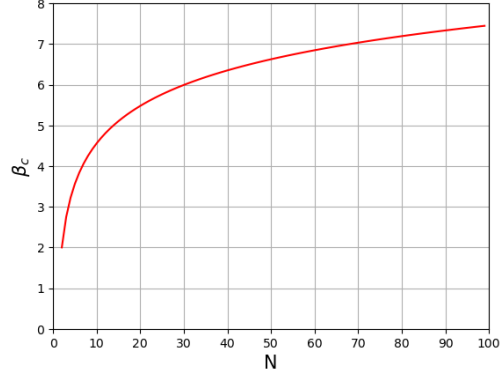
(b) N-dependence of β_c

Figure A2: N-dependence of both p_c and β_c computed by numerical solving of (17) and then using (16)

In the above calculations the fixed point condition was used, i.e. $p' = \frac{\exp(\beta p)}{Z} = p$.
Solving the last line for β_c:

$$\beta_c = \frac{N-1}{N} \frac{1}{p_c(1-p_c)} \tag{16}$$

Finally, by plugging this into the fixed point condition and applying some careful manipulations, one obtains:

$$p_c \left(1 + (N-1)\exp \frac{1 - Np_c}{Np_c(1-p_c)} \right) = 1 \tag{17}$$

This equation is easily solved numerically as it is known that $p_c \in \,]0, 1]$, such that the range over which solutions are searched is very limited. Once p_c is known, it is sufficient to plug this value back into the equation for β_c to obtain the critical temperature.
Solving this equation repetitively for different values of N the N-dependence of β_c is obtainable. The N-dependence of p_c and β_c are shown in Fig. A2. On this graph, a log-like increasing trend for β_c can be recognised.