Controlling the bifurcations of attractors in modern Hopfield networks

Maria Yampolskaya
Department of Physics
Boston University
Boston, MA 02135
mariay@bu.edu

Pankaj Mehta [*]
Department of Physics
Boston University
Boston, MA 02135
pankajm@bu.edu

Abstract

Hopfield networks model complex systems with attractor states. However, there are many systems where attractors are not static. Attractors may undergo bifurcations under certain conditions; for example, cell fates have been described as attractor states that can be stabilized or destabilized by signalling. In the case of neural networks, retrieving a sequence of memories involves changing attractor states. We provide an extension to the modern Hopfield network that connects network dynamics to the landscape of any potential. With our model, it is possible to control the bifurcations of attractors and simulate the resulting neuron dynamics. By introducing controlled bifurcations, our formulation expands the application of Hopfield models to real-world contexts where attractors do not remain static.

1 Introduction

Across disciplines, there are many systems where the interactions between many parts lead to distinct patterns. Hopfield networks are useful for modelling the dynamics of such systems, from memory retrieval to cell fate specification. Attractors do not change in the typical Hopfield network. However, many biological contexts involve attractors that are stabilized or destabilized through bifurcations due to external forces. For example, the process of cells differentiating into mature cell fates is often described as a a ball rolling down valleys in a landscape, with attractor basins corresponding to the cell fates (Waddington [1957]). In Rand et al. [2021], changes in cell fate mediated by signalling are described by a landscape with bifurcating attractors. There has been significant work on formalizing the geometry of cell fate landscapes (Raju and Siggia [2023]). However, these landscapes exist in an abstract space whereas the actual biological process of differentiation occurs in a cell's gene regulatory network. We present an attractor network model that bridges the gap between changing landscapes and network dynamics.

We propose an extension of modern Hopfield networks that involves attractor bifurcations controlled by changing the parameters of a pseudo-potential. This work contributes the following:

- We generalize the Kanter and Sompolinsky [1987] construction for attractor networks with correlated patterns to the case of modern Hopfield networks
- We show how the transformer-like version of the modern Hopfield model as described by Ramsauer et al. [2020] can be interpreted as gradient descent along a pseudo-potential, which is distinct from the Lyapunov/energy function commonly described
- We show how the pseudo-potential can be manipulated by changing an external parameter, leading to attractors being created or destroyed by bifurcations

[*] Center for Regenerative Medicine of Boston University and Boston Medical Center; Faculty of Computing and Data Science, Boston University; Biological Design Center

- We demonstrate network dynamics in the case of pseudo-potentials constructed using classes of bifurcations, such as the heteroclinic flip

2 Kanter and Sompolinsky construction

There are three relevant spaces in our model:

1. The space of neurons, described by continuous values $\{x_i\}$. Let N be the total number of neurons.
2. The space of attractors. This is described by m_μ, which measures the alignment of the network with pattern μ.
3. The space of parameters that control the bifurcations and pseudo-potential. These parameters represent external factors, such as signals received by a cell.

In our formulation, the update rule determines the dynamics of neurons in the network while minimizing the pseudo-potential in the space of attractors. Meanwhile, the control parameters determine the attractors of the pseudo-potential. Varying the control parameters causes bifurcations in attractor space, and the change in attractors causes the neurons to adopt a different pattern.

We will use the form of m_μ described by Kanter and Sompolinsky [1987]. This is a general form for the order parameters which reproduces the correct pattern retrieval dynamics whether or not the patterns are correlated. Let ξ_i^{mu} indicate the value of the i^{th} neuron in the μ^{th} attractor state. Then we define the correlation matrix $A_{\mu\nu} = \sum_i \xi_i^\mu \xi_i^\nu$ and its inverse $g_{\mu\nu} = (A^{-1})_{\mu\nu}$. m_μ takes the following form:

$$\begin{aligned} m_\mu &= \frac{1}{N} \sum_{i\nu} g_{\mu\nu} \xi_i^\nu x_i \\ &= \sum_\nu g_{\mu\nu} m^\nu \end{aligned}$$

With a raised index, $m^\nu = \frac{1}{N} \sum_i \xi_i^\nu x_i$. In the case of uncorrelated patterns, $g_{\mu\nu} = \delta_{\mu\nu}$ and $m_\mu = m^\nu$. In this notation, we have a covariant form m_μ, a contravariant form m^μ, and a metric tensor $g_{\mu\nu}$ that is dependent on the correlations between patterns.

3 Modern Hopfield networks with bifurcations

We explain our model by starting from a well-known formulation of the modern Hopfield network. From Krotov and Hopfield [2020], the update rule for the modern Hopfield network is given by:

$$\begin{aligned} \tau_f \frac{dx_i}{dt} &= \sum_\mu \xi_i^\mu \sigma(\beta \frac{1}{N} \sum_i \xi_i^\mu x_i) - x_i \\ &= \sum_\mu \xi_i^\mu \sigma(\beta m_\mu) - x_i \end{aligned}$$

where $\sigma(\beta m_\mu) = \frac{\exp{(\beta m_\mu)}}{\sum_\nu \exp{(\beta m_\nu)}}$ is the softmax function. τ_f is the time it takes to update a neuron. This update rule states that the values of neurons shift towards a given stored pattern ξ^μ according to the m_μ, which measures the similarity between the network state and the stored pattern. β is inverse temperature; at high temperatures, all attractor states are equally likely, and we no longer have a pattern-retrieval model. Throughout this paper, we assume the temperature is on the order of N^{-1}, where N is the number of neurons. At this low temperature, the only fixed points of softmax are values of m_μ that correspond to one-hot vectors (Tiňo [2009]). In other words, only pure individual patterns act as attractors.

We can rewrite this update rule as follows:

$$\tau_f \frac{dx_i}{dt} = \sum_\mu \xi_i^\mu \sigma(-\beta \frac{\partial V}{\partial m_\mu}) - x_i \tag{1}$$

$$V(\vec{m}) = -\sum_\mu \frac{1}{2} m^\mu m_\mu \tag{2}$$

In this format, we can consider a new interpretation: this equation determines the steepness of the pseudo-potential V in the directions of the stored states, then moves the system towards the stored state with the steepest corresponding direction.

In the usual modern Hopfield network, softmax pushes the system towards the stored state μ with the highest-value m_μ. This creates an attraction towards the highest-overlapping state. The pseudo-potential has only quadratic terms, which correspond to a simple well potential. Consider a more complicated pseudo-potential, $V_{\{a\}}$ with corresponding parameters $\{a\}$. For example, for a cusp bifurcation with three attractors, we may have $V_{a,b}(\vec{m}) = (m_2 - m_1)^4 + a(m_2 - m_1)^2 + b(m_2 - m_1) + \frac{1}{2} m_0^2$. In this case, the parameters controlling the potential (and therefore the bifurcations) are a and b.

To minimize the parameterized pseudo-potential $V_{\{a\}}$, we add it to the quadratic potential terms: $V = -\sum_\mu \frac{1}{2} m^\mu m_\mu + V_{\{a\}}$. The resulting update rule is:

$$\tau_f \frac{dx_i}{dt} = \sum_\mu \xi_i^\mu \sigma(\beta(m_\mu - \frac{\partial V_{\{a\}}}{\partial m_\mu})) - x_i$$

By adding terms corresponding to the gradient of the pseudo-potential, we are now able to control bifurcations of the attractors of the Hopfield network.

This maintains the original attractors of the usual Hopfield network while also including the dynamics of a changing landscape. At the minima of $V_{\{a\}}$, $\frac{\partial V_{\{a\}}}{\partial m_\mu} = 0$ and the update rule takes the form of the usual modern Hopfield model. In other words, near the minima of the parameterized pseudo-potential, the effective pseudo-potential is once again a simple well.

4 Heteroclinic flip example

In this section, we demonstrate the model with a particular choice of pseudo-potential. In cell fate differentiation, two cell types with a shared progenitor have previously been modelled with two types of bifurcations: the heteroclinic flip and the double cusp (Raju and Siggia [2023]). We take the heteroclinic flip as defined by Sáez et al. [2022]:

$$V_{a,b}(x,y) = x^4 + y^4 + y^3 - 4x^2 y - ax + by - y^2$$

This potential is shown in 1(a) for varying values of a, b. The three attractors of this system are located at (x_0, y_0) and $(x_{1\pm}, y_1)$ (let $x_1 = |x_{1\pm}|$). We need a change of coordinates from x, y to $\vec{m}$. Since there are three attractors, $\vec{m} = (m_0, m_1, m_2)$. Let $y = y_0 m_0 - y_1(m_0 - 1) = (y_0 - y_1)m_0 + y_1$ and $x = x_1(m_2 - m_1)$. With this change of coordinates, (x_0, y_0) corresponds to $\vec{m} = (1, 0, 0)$, (x_{1+}, y_1) corresponds to $\vec{m} = (0, 1, 0)$, and (x_{1-}, y_1) corresponds to $\vec{m} = (0, 0, 1)$. Using Einstein notation for implicit summation, we can rewrite the potential as follows:

$$\begin{aligned}
h_0 &= (y_0 - y_1, 0, 0) \\
h_{12} &= (0, -x_1, x_1) \\
V_{a,b}(\vec{m}) &= (h^{12,\mu} m_\mu)^4 + (h^{0,\mu} m_\mu + y_1)^4 + (h^{0,\mu} m_\mu + y_1)^3 - 4(h^{12,\mu} m_\mu)^2 (h^{0,\mu} m_\mu + y_1) \\
&\quad - a h^{12,\mu} m_\mu + b(h^{0,\mu} m_\mu + y_1) + c(h^{0,\mu} m_\mu + y_1)^2
\end{aligned}$$

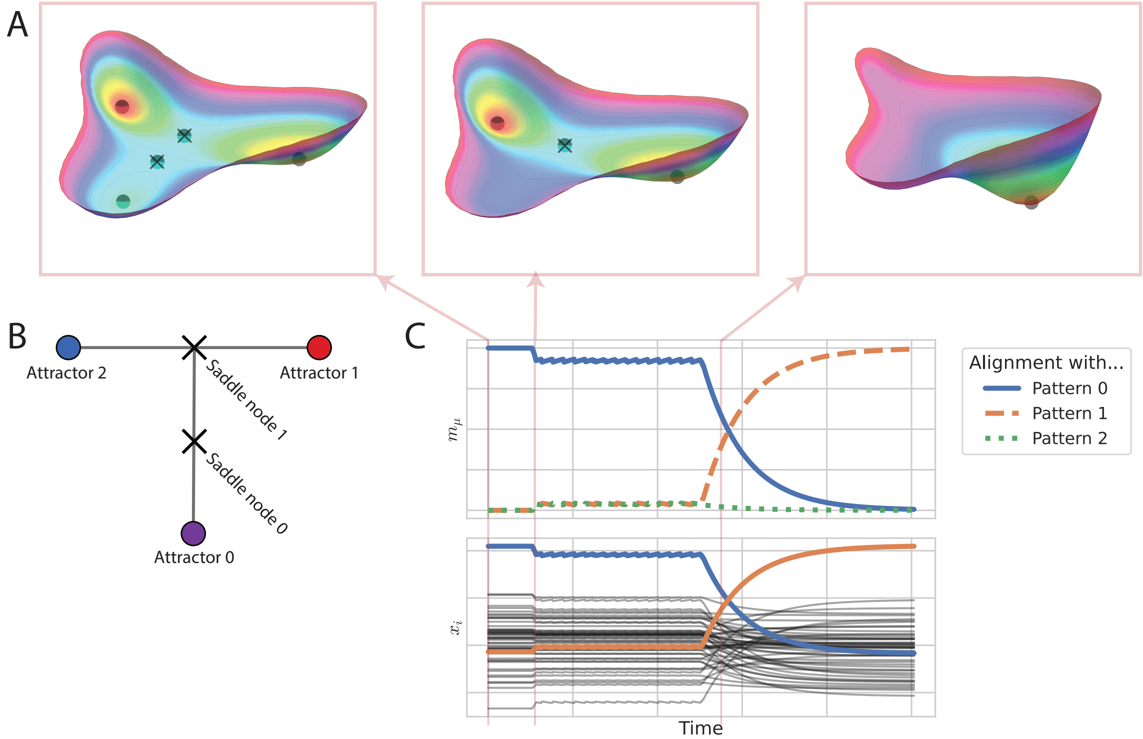

Figure 1: Modern Hopfield network with bifurcations controlled by a varying pseudo-potential. (a) The potential describing a heteroclinic flip. The spheres indicate points where the gradient is zero (saddle points are marked with X's). In the first box, there are three attractors and two saddle points. Parameter b is varied first, causing a bifurcation where the attractor corresponding to pattern 0 and the adjacent saddle point are destroyed. Then, parameter a is varied so that the attractor corresponding to pattern 2 collides with the other saddle point and they are both destroyed, leaving just the attractor corresponding to pattern 1. (b) A diagram showing the connections between saddle nodes (marked by X's) and attractors (marked by circles). In (a), attractor 0 is destabilized by colliding with saddle node 0, and attractor 2 is destabilized by colliding with saddle node 1. (c) The dynamics in attractor space (top) and neuron space (bottom) corresponding to the changes in potential. Pink lines indicate the time points where each of the potential are in the states shown in (a). m_μ measures alignment with pattern μ. The x_i values plotted in blue (orange) correspond to the neuron with the highest value in pattern 0 (1).

Now we have a pseudo-potential described in the space of attractors. The dynamics are determined by $\tau_f \frac{dx_i}{dt} = \sum_\mu \xi_i^\mu \sigma(\beta(m_\mu - \frac{\partial V_{a,b}}{\partial m_\mu})) - x_i$. For the sake of simplicity, we assumed uncorrelated patterns for Figure 1. We can vary the parameters a, b and see the corresponding changes in attractor space and neuron space (Figure 1(c)).

We show an example using correlated patterns and real gene expression data in section S1. The correlation between patterns changes the geometry and the metric tensor $g_{\mu\nu}$.

5 Conclusion

Associative memory networks are a rich set of models that provide a way to understand dynamic systems containing attractor states. There are many directions to expand these models. In this paper, we presented a construction that connects the world of attractor networks to bifurcations and potential landscapes. By connecting the spaces of neurons, attractors, and bifurcation parameters, we provide a framework for modelling complex systems where processes occur in all three spaces at once. In the supplementary information, we show how this construction can be used to simulate cell fate dynamics using real gene expression data for the stored patterns. We also discuss the model's relation to the bipartite graph formulation of modern Hopfield networks introduced by Krotov and Hopfield [2020].

References

Qiaolin Deng, Daniel Ramsköld, Björn Reinius, and Rickard Sandberg. Single-cell rna-seq reveals dynamic, random monoallelic gene expression in mammalian cells. *Science*, 343(6167):193–196, 2014.

I Kanter and Haim Sompolinsky. Associative recall of memory without errors. *Physical Review A*, 35(1):380, 1987.

Dmitry Krotov and John Hopfield. Large associative memory problem in neurobiology and machine learning. *arXiv preprint arXiv:2008.06996*, 2020.

Dandan Liu, Yidong Chen, Yixin Ren, Peng Yuan, Nan Wang, Qiang Liu, Cen Yang, Zhiqiang Yan, Ming Yang, Jing Wang, et al. Primary specification of blastocyst trophectoderm by scrna-seq: New insights into embryo implantation. *Science Advances*, 8(31):eabj3725, 2022.

Archishman Raju and Eric D Siggia. A geometrical perspective on development. *Development, Growth & Differentiation*, 2023.

Hubert Ramsauer, Bernhard Schäfl, Johannes Lehner, Philipp Seidl, Michael Widrich, Thomas Adler, Lukas Gruber, Markus Holzleitner, Milena Pavlović, Geir Kjetil Sandve, et al. Hopfield networks is all you need. *arXiv preprint arXiv:2008.02217*, 2020.

David A Rand, Archishman Raju, Meritxell Sáez, Francis Corson, and Eric D Siggia. Geometry of gene regulatory dynamics. *Proceedings of the National Academy of Sciences*, 118(38): e2109729118, 2021.

Meritxell Sáez, Robert Blassberg, Elena Camacho-Aguilar, Eric D Siggia, David A Rand, and James Briscoe. Statistically derived geometrical landscapes capture principles of decision-making dynamics during cell fate transitions. *Cell Systems*, 13(1):12–28, 2022.

Peter Tiňo. Bifurcation structure of equilibria of iterated softmax. *Chaos, Solitons & Fractals*, 41(4): 1804–1816, 2009.

Conrad Hal Waddington. *The strategy of the genes*. Routledge, 1957.

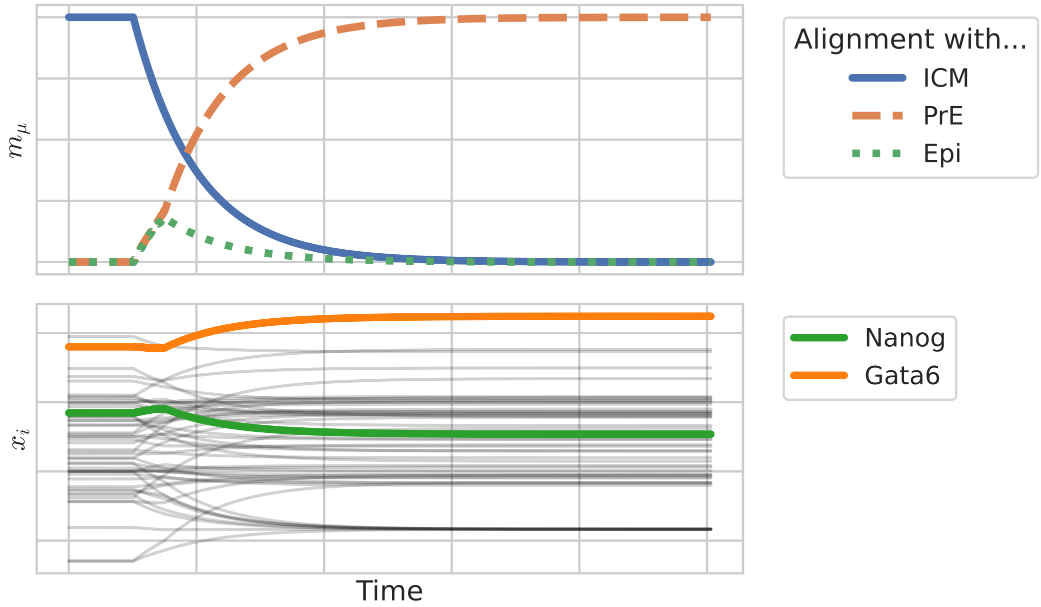

Figure S1: Modern Hopfield network with a heteroclinic flip, using gene expression data as the stored patterns. The same bifurcations occur as in Figure 1, causing inner cell mass (ICM) cells to differentiate into primitive endoderm (PrE) cells instead of epiblast (Epi). Gene expression dynamics for two genes in particular are shown. Gata6 is associated with the PrE fate while Nanog is associated with Epi.

Supplementary Information

S1 Heteroclinic flip with gene expression data

We can apply the concept of attractor networks controlled by parametrized potentials to cell fate specification. In this case, neurons correspond to genes and patterns correspond to the gene expression profiles of cell types. Related cell types express many genes in common, so $g_{\mu\nu} \neq \delta_{\mu\nu}$. Raju and Siggia [2023] suggest that one of the earliest cell fate decisions – the specification of inner cell mass into epiblast or primitive endoderm – can be modelled by either a heteroclinic flip or double cusp. We used gene expression data in the form of single-cell RNA-sequencing (scRNA-seq) corresponding to these three cell types from Deng et al. [2014] and Liu et al. [2022]. We used these gene expression profiles as the values of ξ_i^μ. Then, starting in the inner cell mass state, we varied the potential the same way as in Figure 1.

Figure S1 shows the dynamics in cell fate space and gene expression space. The correlation between cell types causes a bias towards the primitive endoderm state, which causes the network to leave the separatrix leading towards the remaining saddle node before the second bifurcation occurs. The gene expression dynamics of two genes are shown. Mature primitive endoderm cells are known to highly-express Gata6 while mature epiblast cells highly-express Nanog.

S2 Relation to bipartite formulation

The Krotov and Hopfield [2020] formulation of modern Hopfield networks describes a bipartite graph of hidden neurons h_μ and visible neurons v_i with continuous values governed by the equations:

$$\tau_f \frac{dv_i}{dt} = \sum_{\mu=1}^{N_h} \xi_i^\mu f(\{h^\mu\}) - v_i$$

$$\tau_h \frac{dh^\mu}{dt} = \sum_{i=1}^{N_f} \xi_i^\mu g(\{v_i\}) - h^\mu$$

$$f(\{h^\mu\}) = \frac{\partial L_h}{\partial h^\mu}$$

$$g(\{v_i\}) = \frac{\partial L_v}{\partial v_i}$$

where ξ_i^μ denotes the connection between hidden neuron μ and visible neuron i. When considering the non-bipartite version of the modern Hopfield network, where any neuron may be connected to any other neuron, ξ_i^μ is the value of the μ^{th} pattern at the i^{th} neuron. In order to retrieve the typical Hopfield dynamics, τ_f is approximated as zero and so $h_\mu \approx \sum_i \xi_i^\mu g_i$

In this notation, $m_\mu = \sum_j \xi_j^\mu v_j$. In order to introduce the $\frac{\partial V}{\partial m_\mu}$ as written in the main text, we note that:

$$\frac{\partial}{\partial m_\mu} V(\{m_\mu\}) = \frac{\partial v_i}{\partial m_\mu} \frac{\partial}{\partial v_i} V(\{\sum_j \xi_j^\mu v_j\})$$

$$\frac{\partial v_i}{\partial m_\mu} = \sum_j B_{ij} \xi_j^\mu$$

$$B = (\xi^T \xi)^{-1}$$

Thus, our equations can be written in the same formulation as Krotov and Hopfield [2020] if we make the following changes:

$$\tau_f \frac{dv_i}{dt} = \sum_\mu \xi_i^\mu f_\mu - v_i$$

$$\tau_h \frac{dh_\mu}{dt} = \sum_{i,j} g_i B_{ij} \xi_j^\mu - h_\mu$$

$$E(t) = [\sum_i v_i g_i - L_v] + [\sum_\mu h_\mu f_\mu - L_h] - \sum_{\mu,i,j} g_i B_{ij} \xi_j^\mu f_\mu$$

We make the following choices for the Lagrangian functions L_v, L_h:

$$L_v = -V(\{\sum_j \xi_j^\mu v_j\})$$

$$L_h = log(\sum_\mu e^{h_\mu})$$

$$f_\mu = \frac{\partial L_h}{\partial h_\mu} = \sigma(h_\mu)$$

$$g_i = \frac{\partial L_v}{\partial v_i} = -\frac{\partial V}{\partial v_i}$$

For $\tau_h \approx 0$, $h_\mu \approx \sum_{i,j} g_i B_{ij} \xi_j^\mu$, we retrieve the same update rule as before.

In the bipartite formulation, our model introduces an asymmetry between the visible and hidden neurons. In Krotov and Hopfield [2020], $E(t)$ is shown to monotonically decrease as time passes on the condition that the Hessians of the Lagrangians are positive definite. Since L_v is no longer necessarily positive definite, we cannot make the same argument.