SCALING NETWORKS THROUGH CATEGORICAL AND GEOMETRIC LENS

MILTON LIN

Personal Experiences

My research[1] in the geometric Langlands program has centered on bridging discrete structures (number theory) with continuous spaces (topology) through *algebraic formalism*. Building on this, I aim to explore how **algebraic and categorical models** can reveal the **qualitative dynamics** of modern language models and neural circuits. This involves examining the extent to which **biological realism** can be mirrored in computational frameworks.

$$\begin{tikzcd}
& \text{Categorical} \arrow[dl, bend right] \arrow[dr, bend left] & \\
\text{Cognitive} \arrow[rr, leftrightarrow, bend right] & & \text{Computational}
\end{tikzcd}$$

The overarching goal is to unify insights from *cognitive science*, *category theory*, and *computation* to understand the *limiting behaviors* of complex networks, both artificial and biological. Specifically, my research will offer new perspectives on *scaling laws* in large language models [Kap+20], [SMK23] and propose biologically grounded designs for neural networks, guided by experimentally observed properties of neural circuits [Ber+23]. In Section 1, I introduce *associative memory networks* like *Hopfield Networks* and *Simplicial Hopfield Netowrks*. Our research studies the weight space decomposition of these models and evaluates their scaling behavior. In Section 2, I highlight how my background in algebra and geometry positions me to contribute to this interdisciplinary research.[2]

1. Associative Memory Networks

Associative Memory Networks, or Hopfield Networks, are well-known models where memories are stored and recalled based on Lyapunov energy functions that govern the network's dynamics [Hop84]. The local minima in the energy landscape correspond to stored memories. Recent models, such as Dense Associative Memories, improve memory capacity exponentially to $2^{d/2}$ by modifying the energy dynamics [KH16; Dem+17]. These have been connected to attention mechanisms in transformers [Ram+21], though their full relations remain largely unexplored.

Research Goal: Scaling Properties of Associative Memory and Modern Models. The two key research areas are, joint with Chris Hillar (Redwood Research), Tenzin Chan (Algebraic) and Muhan Gao (Johns Hopkins University)

1. *Polytopal Decomposition of Weight Spaces in Toy Models*: We will extend the polytopal weight space decomposition, as present in literature on threshold linear networks, [CLM20], [CGM23], to

[1] My mathematical research statement is available at `https://cwlin4916.github.io/Trees/Application/Postdoc/Research.pdf`.

[2] I am indebted to my discussions with Prof. Daniel Khashabi and Prof. Leyla Isik.

higher order memory networks, such as *simplicial Hopfield networks* inspired by setwise connection [BF23] or *dense associative memories*. This connects to recent approaches using spline theory to understand neural networks, [Bb18], [Bla+22]. We will study how the decomposition changes as the *network size increases*.

2. *Energy Transformer Experiments*: We will evaluate the scaling properties of the *Energy Transformer* [Hoo+23], a model based on dense associative memory principles, hypothesizing across three dimensions: - i) Parameters: increasing parameters increases number of attractors (local minima). - ii) Data: scaling data size will refine attractors, resulting in more meaningful minima. - iii) Compute: increasing compute (depth or iterations) will lead to faster convergence to low-energy retrievals.

We will also investigate how *data augmentation* during training affects memory storage and retrieval efficiency, as in prior work on transformers [AL24]. Finally, the end goal is to provide both empirical and theoretical comparison with modern networks [Hoo+24]; some works which studied scaling properties of memory networks include [ND21], [Niu+24], and [CDB24].

Expected impact: This research will highlight the limitations of synthetic memory networks, especially in their use as proxies for explaining biological networks, see also [KH21]. Finally, it is an interesting problem to create hybrid models that respect biological constraints while maintaining the computational power of synthetic networks.

2. Categorical Models and Homotopy Theory

Categorical approaches have gained momentum as a systematic framework for studying network structures [Gav+24]. This has been particularly successful in the field of *geometric deep learning* [Bro+21], where abstract mathematical structures help describe complex neural networks. We propose to explore *Hopfield networks* using a recent formalism by Manin et al. [MM24], which uses *summing functors* and *Gamma spaces* to model the allocation of resources in neural networks. These concepts will allow us to understand how the complexity of memory networks scales as network size increases. The formalism allows us to study a *homotopy type* - a mathematical construct at a deeper level than *homology*[3]. Homotopy captures invariants of network up to continuous deformations. Previous studies have shown that stimulus space can be reconstructed up to homotopy [Man15].

Research Goal: Homotopical Complexity Under Scaling. This research will investigate how the **homotopical complexity** of memory networks evolves as their size increases. Specifically, we will examine how *memory capacity* correlates with *homotopical invariants* like *Betti numbers* (which measure the number of independent cycles in a space) and *simplicial complexes* (which provide a higher-dimensional generalization of networks). Burns and Fukai have already done early work in this direction [BF23], but much remains to be explored.

Expected Impact. Understanding the scaling behavior of homotopic information could provide a lens through which to study the behavior of neural networks algebraically.

References

[AL24] Allen-Zhu, Zeyuan and Li, Yuanzhi. *Physics of Language Models: Part 3.1, Knowledge Storage and Extraction*. 2024. arXiv: 2309.14316 [cs.CL]. URL: https://arxiv.org/abs/2309.14316 (cit. on p. 2).

[3] which is commonly used in topological data analysis (TDA). For a short survey of topology and neural code, see [Cur16].

[Bb18] Balestriero, Randall and baraniuk, richard. “A Spline Theory of Deep Learning”. In: *Proceedings of the 35th International Conference on Machine Learning*. Ed. by Jennifer Dy and Andreas Krause. Vol. 80. Proceedings of Machine Learning Research. PMLR, Oct. 2018, pp. 374–383. URL: https://proceedings.mlr.press/v80/balestriero18b.html (cit. on p. 2).

[Ber+23] Bernáez Timón, Laura et al. “How to incorporate biological insights into network models and why it matters”. In: *The Journal of Physiology* 601.15 (2023), pp. 3037–3053 (cit. on p. 1).

[BF23] Burns, Thomas F and Fukai, Tomoki. “Simplicial Hopfield networks”. In: *The Eleventh International Conference on Learning Representations*. 2023. URL: https://openreview.net/forum?id=_QLsH8gatwx (cit. on p. 2).

[Bla+22] Black, Sid et al. *Interpreting Neural Networks through the Polytope Lens*. 2022. arXiv: 2211.12312 [cs.LG]. URL: https://arxiv.org/abs/2211.12312 (cit. on p. 2).

[Bro+21] Bronstein, Michael M. et al. *Geometric Deep Learning: Grids, Groups, Graphs, Geodesics, and Gauges*. 2021. arXiv: 2104.13478 [cs.LG]. URL: https://arxiv.org/abs/2104.13478 (cit. on p. 2).

[CDB24] Cabannes, Vivien, Dohmatob, Elvis, and Bietti, Alberto. *Scaling Laws for Associative Memories*. 2024. arXiv: 2310.02984 [stat.ML]. URL: https://arxiv.org/abs/2310.02984 (cit. on p. 2).

[CGM23] Curto, Carina, Geneson, Jesse, and Morrison, Katherine. *Stable fixed points of combinatorial threshold-linear networks*. 2023. arXiv: 1909.02947 [q-bio.NC]. URL: https://arxiv.org/abs/1909.02947 (cit. on p. 1).

[CLM20] Curto, Carina, Langdon, Christopher, and Morrison, Katherine. *Combinatorial Geometry of Threshold-Linear Networks*. 2020. arXiv: 2008.01032 [math.CO]. URL: https://arxiv.org/abs/2008.01032 (cit. on p. 1).

[Cur16] Curto, Carina. *What can topology tell us about the neural code?* 2016. arXiv: 1605.01905 [q-bio.NC]. URL: https://arxiv.org/abs/1605.01905 (cit. on p. 2).

[Dem+17] Demircigil, Mete et al. “On a Model of Associative Memory with Huge Storage Capacity”. In: *Journal of Statistical Physics* 168.2 (May 2017), pp. 288–299. ISSN: 1572-9613. URL: http://dx.doi.org/10.1007/s10955-017-1806-y (cit. on p. 1).

[Gav+24] Gavranović, Bruno et al. *Position: Categorical Deep Learning is an Algebraic Theory of All Architectures*. 2024. arXiv: 2402.15332 [cs.LG]. URL: https://arxiv.org/abs/2402.15332 (cit. on p. 2).

[Hoo+23] Hoover, Benjamin et al. *Energy Transformer*. 2023. arXiv: 2302.07253 [cs.LG]. URL: https://arxiv.org/abs/2302.07253 (cit. on p. 2).

[Hoo+24] Hoover, Benjamin et al. *Memory in Plain Sight: Surveying the Uncanny Resemblances of Associative Memories and Diffusion Models*. 2024. arXiv: 2309.16750 [cs.LG]. URL: https://arxiv.org/abs/2309.16750 (cit. on p. 2).

[Hop84] Hopfield, John J. “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), pp. 3088–3092 (cit. on p. 1).

[Kap+20] Kaplan, Jared et al. *Scaling Laws for Neural Language Models*. 2020. arXiv: 2001.08361 [cs.LG]. URL: https://arxiv.org/abs/2001.08361 (cit. on p. 1).

[KH16] Krotov, Dmitry and Hopfield, John J. *Dense Associative Memory for Pattern Recognition*. 2016. arXiv: 1606.01164 [cs.NE]. URL: https://arxiv.org/abs/1606.01164 (cit. on p. 1).

[KH21] Krotov, Dmitry and Hopfield, John. *Large Associative Memory Problem in Neurobiology and Machine Learning*. 2021. arXiv: 2008.06996 [q-bio.NC]. URL: https://arxiv.org/abs/2008.06996 (cit. on p. 2).

[Man15] Manin, Yuri I. *Neural codes and homotopy types: mathematical models of place field recognition*. 2015. arXiv: 1501.00897 [math.HO]. URL: https://arxiv.org/abs/1501.00897 (cit. on p. 2).

[MM24] Manin, Yuri and Marcolli, Matilde. "Homotopy Theoretic and Categorical Models of Neural Information Networks". In: *Compositionality* Volume 6 (2024) (Sept. 2024). ISSN: 2631-4444. URL: http://dx.doi.org/10.46298/compositionality-6-4 (cit. on p. 2).

[ND21] Nichol, Alex and Dhariwal, Prafulla. *Improved Denoising Diffusion Probabilistic Models*. 2021. arXiv: 2102.09672 [cs.LG]. URL: https://arxiv.org/abs/2102.09672 (cit. on p. 2).

[Niu+24] Niu, Xueyan et al. *Beyond Scaling Laws: Understanding Transformer Performance with Associative Memory*. 2024. arXiv: 2405.08707 [cs.LG]. URL: https://arxiv.org/abs/2405.08707 (cit. on p. 2).

[Ram+21] Ramsauer, Hubert et al. "Hopfield Networks is All You Need". In: *International Conference on Learning Representations*. 2021. URL: https://openreview.net/forum?id=tL89RnzIiCd (cit. on p. 1).

[SMK23] Schaeffer, Rylan, Miranda, Brando, and Koyejo, Sanmi. *Are Emergent Abilities of Large Language Models a Mirage?* 2023. arXiv: 2304.15004 [cs.AI]. URL: https://arxiv.org/abs/2304.15004 (cit. on p. 1).