Cooperative Phenomenon: Statistical Physics Origins of Connectionist Learning

Mehmet Süzen

Member ACM and IoP

E-mail: `suzen@acm.org;mehmet.suzen@physics.org`

May 2020

Abstract. A short account of origins of mathematical formalism of neural networks is presented for physicists and computer scientist in basic discrete mathematical setting informally. The discourse of the development of mathematical formalism on the dynamics of lattice models in statistical physics and learning internal representations of neural networks as discrete architectures as quantitative tools evolve in two almost distinct fields more than half a century with limited overlap. We aim at bridging the gap by claiming that the analogy between two approaches are not artificial but naturally occuring due to how modelling *cooperative phenomenon* is constructed. We define the *Lenz-Ising architectures* (ILAs) for this purpose.

Keywords: Neural Networks, Statistical Physics, Neurophysiology, Spin-glasses, Lattice Models, Dynamics, Learning, Statistics of Cooperative Phenomenon, Lenz-Ising Architectures

1. Introduction

Understanding natural or artificial phenomenon in the language of discrete mathematics is probably one of the most powerful toolbox scientist use [1]. Large portion of computer science and statistical physics deals with such finite structures. One of the most prominent successful usage of such approach was Lenz and Ising's work on modelling ferromagnetic materials [2–5] and neural networks as a model to biological neuronal structures [6–8].

The analogy between two areas of distinct research have been pointed out by many researchers [9–13]. However, the discourse and evolution of these approaches were kept as two distinct research fields and many innovative approaches rediscovered under different names.

2. Cooperative phenomenon

Statistical definition of cooperative phenomenon pioneered by Wannier and Kremer [14–16]. Even though their technical work focused on extension of Ising model to 2D with cyclic boundary condition and introduction of exact solutions with matrix algebra, they were the first to document the potential of how Lenz-Ising model actually represent a more generic system than marely model to ferromagnets, namely anything falls under *cooperative phenomenon* can be addressed with Lenz-Ising type model, summarized in Definition 1.

Definition 1. *Cooperative phenomenon of Wannier type [14]*
Set of N discrete units, $\mathscr{U}$, identified with a function s_i, i=1,..,N forms a collection or assembly. The function that identifies the units is a mapping $s_i : \mathbb{R} \rightarrow \mathbb{R}$. A statistic $\mathscr{S}$ applied on $\mathscr{U}$ is called cooperative phenomenon of Wannier type $\mathscr{W}$.

A statistic $\mathscr{S}$ can be any mapping or set of operations on the assembly of units $\mathscr{U}$. For example inducing ordering on the assembly of units and summation over s_i values, would corresponds to non-interacting magnetic system with unit external field or non-connected set of neurons capacity of inhibition or exhibition. However, amazingly, Definition 1 is so generic that Rosenblatt's perceptron [17], current deep learning systems [18] and complex networks [19] falls into this category as well.

The originality of Cooperative phenomenon of Wannier type comes on a secondary concept, so called event propagation as given in Definition 2.

Definition 2. *Event propagation [14]*
An event is defined as a snapshot of cooperative phenomenon of Wannier type $\mathscr{W}$. If an event takes place of one unit of assembly $\mathscr{U}$, the same event will be favored by other units, this is expressed as event propagation between two disjoint set of units $\mathscr{E}(u_1, u_2)$, and $u_1 \cap u_2 = \varnothing$ and $u_1, u_2 \in \mathscr{U}$ and with an additional statistic $\mathscr{S}$ is defined.

The parallels between Wannier's event propagations are remarkably the same as of neural network formalism defined by McCulloch-Pitts-Kleene [6, 7], not only conceptually but matematical treatment is identical and originates from Lenz-Ising model's treatment of discrete units. As we mentioned, this goes beyond doubt not a simple analogy but forms a generic framework as envisioned by Wannier. The similarity between ferromagnetic systems and neural networks is probably first documented directly by Little [8]: Spin states of magnetic spins corresponds to firing state of a neuron. Unfortunately, Little only see it as simple analogy, and missed the opportunity provided by Wannier as a generic natural phenomenon of cooperation.

The conceptual similarity and inference on Wannier's event propagation appears to be quite close to Hebb's learning [20] and gives natural justification for backpropagation for multilayered networks. History of backpropagation is exhaustively studied elsewhere [18].

3. Lenz-Ising architectures: Ferromagnets to Nerve Nets

As we establised two basic definitions of cooperative phenomenon, we can now define a generic setting of Lenz-Ising model that captures both physics literature that extensively used this in so called spin-glasses research and for neural networks. A guiding principle will be based on Wannier's definition of cooperative phenomenon.

Definition 3. *Lenz-Ising Architectures*
Given Wannier type cooperative phenomenon $\mathscr{W}$, imposing constrains on the discrete units, $\mathscr{U}^c$ that they should be spatially correlated on the edges E of an arbirary graph $\mathscr{G}(E, V)$ with ordering and with vertices V of the arbitrary graph carring coupling weight between connected two units with biases. Set of event propagations $\mathscr{E}^c$ defined

on the cooperative phenomeon can induce dynamics on defining vertice weights, or vice versa. ILAs are defined as statistic $\mathscr{S}$ applied to $\mathscr{U}^c$ with propagations $\mathscr{E}^c$.

Lenz-Ising Architectures (ILAs) should not be confused with graph neural networks as it does not model data structures. It could be seen as subset of graph dynamical systems in some sense but formal connections should be established elsewhere. However, primary characteristic of ILAs are that it is conceptual and mathematical representation of spin-glass systems (including Lenz-Ising, Anderson, Sherrington-Kirkpatrick, Potts systems) and neural networks (including recurrent and convolutional networks) under the same umbrella.

4. Learning representations inherent in Metropolis-Glauber dynamics

The primary originality in any neural network research papers lies in so called *learning representation* from data and generalisation. However, it isn't obvious to the that community that actually spin-glasses are capable of learning representations inherently by induced dynamics such as Metropolis or Glauber dynamics by construction, as an inverse problem.

In physics literature this appears as finding a solution to the problem of how to express free energy and minimisation of this with respect to weights or coupling coefficients, This is noting but a learning represenations. Usually a simulation approach is taken as a route, for example Monte Carlo techniques [5, 21, 22] via Metropolis or Glauber dynamics. The intimate connection between concepts of ergodicity and learning in deep learning is recently shown [13, 23, 24] in this context.

As we argued earlier the generic definition provided by Wannier on *cooperative phenomenon* and ILAs; there is an intimate connection with learning and so called solving spin-glasses that usually boils down to computing free energies as mentioned. And a link between two distinct fields, computing backpropagation and free energies are natural candidates to establish equivalance relations.

5. Conclusions and Outlook

Apart from honouring physicists Lenz and Ising, based on understanding of cooperative phenomenon's origins, naming the research outpus from of spin-glasses and neural networks under an umbrella term *Lenz-Ising architectures* (ILAs) is historically accurate and technically a resonable naming scheme under the overwhelming evidence given in the literature. This is akin to naming current computers with *von Neumann architectures.* This forms the origins of connectionist learning from statistical physics, where this approach currently enjoying vast engineering success today.

The rich connection between two areas in computer science and statistical physics should be celebrated. For more fruitful collaborations, both literatures, embracing large statistics literature as well, should converge much more closely. This would help communities to avoid awkward situations of reinventing the wheel again and hindering recognition of the work done by physicists decades earlies, i.e., Ising and Lenz.

Acknowledgements and Notes

Author is grateful to Jürgen Schmidhuber for correspondence, pointing out the work of S.C.Kleene on the dynamics of neural networks and encouraging the inceptions of this manuscript. Author has no competing or other kind of conflict of interests. This work is produced solely with the aim of scholarly work and does not have any personal nature at all. Manuscript is dedicated in memory of Ernst Ising for his contribution to physics of ferromagnetic materials, now seems to have far more implications.

[1] Kenneth H Rosen. *Handbook of Discrete and Combinatorial Mathematics*. CRC Press, 1999.

[2] W.Lenz. Beitrag zum Verstländnis der Magnetischen Erscheinungen in Festen Körpern. *Phys.Z*, 21:613, 1920.

[3] Ernst Ising. Beitrag zur Theorie des Ferromagnetismus. *Zeitschrift für Physik*, 31(1):253–258, 1925.

[4] Thomas Ising, Reinhard Folk, Ralph Kenna, Bertrand Berche, and Yurij Holovatch. The fate of Ernst Ising and the fate of his model. *arXiv preprint arXiv:1706.01764*, 2017.

[5] David P Landau and Kurt Binder. *A guide to Monte Carlo Simulations in Statistical Physics*. Cambridge University Press, 2014.

[6] W.S. McCulloch and W.H. Pitts. A Logical Calculus of the Ideas Imminent in Nervous Activity. *Bull. Math. Biophys.,(5)*, pages 115–133.

[7] Stephen Cole Kleene. Representation of Events in Nerve Nets and Finite Automata. Technical report, RAND Project, Santa Monica, 1951.

[8] W. A. Little. The Existence of Persistent States in the Brain. *Mathematical Biosciences*, 19(1-2):101–120, 1974.

[9] P Peretto. Collective Properties of Neural Networks: a Statistical Physics Approach. *Biological Cybernetics*, 50(1):51–62, 1984.

[10] Jan L van Hemmen. Spin-glass Models of a Neural Network. *Physical Review A*, 34(4):3435, 1986.

[11] Haim Sompolinsky. Statistical Mechanics of Neural Networks. *Physics Today*, 41(21):70–80, 1988.

[12] David Sherrington. Neural Networks: the Spin Glass Approach. In *North-Holland Mathematical Library*, volume 51, pages 261–291. Elsevier, 1993.

[13] Yasaman Bahri, Jonathan Kadmon, Jeffrey Pennington, Sam S Schoenholz, Jascha Sohl-Dickstein, and Surya Ganguli. Statistical Mechanics of Deep Learning. *Annual Review of Condensed Matter Physics*, 2020.

[14] Gregory H Wannier. The Statistical Problem in Cooperative Phenomena. *Reviews of Modern Physics*, 17(1):50, 1945.

[15] Hendrik A Kramers and Gregory H Wannier. Statistics of the two-dimensional ferromagnet. Part I. *Physical Review*, 60(3):252, 1941.

[16] Hendrik A Kramers and Gregory H Wannier. Statistics of the two-dimensional ferromagnet. Part II. *Physical Review*, 60(3):263, 1941.

[17] C van der Malsburg. Frank Rosenblatt: principles of neurodynamics: perceptrons and the theory of brain mechanisms. In *Brain theory*, pages 245–248. Springer, 1986.

[18] Jürgen Schmidhuber. Deep learning in Neural Networks: An overview. *Neural networks*, 61:85–117, 2015.

[19] Duncan J Watts and Steven H Strogatz. Collective dynamics of 'small-world'networks. *Nature*, 393(6684):440, 1998.

[20] Donald Olding Hebb. *The Organization of Behavior: a Neuropsychological Theory*. J. Wiley; Chapman & Hall, 1949.

[21] Mehmet Süzen. Effective ergodicity in single-spin-flip dynamics. *Physical Review E*, 90(3):032141, 2014.

[22] Mehmet Süzen. Anomalous diffusion in convergence to effective ergodicity. *arXiv preprint arXiv:1606.08693*, 2016.

[23] Mehmet Süzen, Cornelius Weber, and Joan J Cerdà. Spectral ergodicity in deep learning architectures via surrogate random matrices. *arXiv preprint arXiv:1704.08303*, 2017.

[24] Mehmet Süzen, JJ Cerdà, and Cornelius Weber. Periodic Spectral Ergodicity: A Complexity Measure for Deep Neural Networks and Neural Architecture Search. *arXiv preprint arXiv:1911.07831*, 2019.