# Higher-order neural networks, Polyà polynomials, and Fermi cluster diagrams

K. E. Kürten
*Institut für Experimentalphysik, Universität Wien, Austria*

J. W. Clark
*McDonnell Center for the Space Sciences and Department of Physics, Washington University, St. Louis, Missouri 63130, USA*

(Received 27 June 2003; published 18 September 2003)

The problem of controlling higher-order interactions in neural networks is addressed with techniques commonly applied in the cluster analysis of quantum many-particle systems. For multineuron synaptic weights chosen according to a straightforward extension of the standard Hebbian learning rule, we show that higher-order contributions to the stimulus felt by a given neuron can be readily evaluated via Polyà's combinatoric group-theoretical approach or equivalently by exploiting a precise formal analogy with fermion diagrammatics.

DOI: 10.1103/PhysRevE.68.037701 PACS number(s): 02.70.Rr, 07.05.Mh, 02.20.−a, 07.05.Pj

In attempting to unravel the mechanisms of information processing and attendant adaptive behavior in neurobiological systems, considerable attention is currently being directed to nonlinear processing in dendritic trees and to the computational power that can be gained from multiplicative or higher-order interactions between neurons [1,2]. This focus is supported by a large body of theoretical work demonstrating enhanced performance in artificial neural networks involving such higher-order or multineuron interactions, as applied to a variety of information-processing tasks, most notably memory storage, and recall Refs. [3–13]. Introduction of higher-order couplings is accompanied, however, by the threat of a combinatoric explosion that may strongly inhibit analysis, evaluation, and optimization. In this note we expose some simple techniques based on group-theoretic symmetry arguments that serve, in some cases, to reduce the severity of these problems and give access to the advantages of higher-order networks for problem domains involving complex correlations. Our study is guided by interesting parallels with the diagrammatic analysis of fermion clusters in many-body physics.

We consider the following simple but standard model of a higher-order neural network. The network consists of $N$ binary-output hard-threshold units (model neurons) $i$ whose state variables $\sigma_i$ take the value $+1$ if the unit is active (''firing'') and $-1$ if the unit is inactive (''not firing''). Model neuron $i$ receives inputs from exactly $K_i$ other units of the network, with self-interactions excluded so that $1\leqslant K_i \leqslant N-1$. A given neuron updates its state on a discrete time grid according to the deterministic threshold rule

$$\sigma_i(t+1)=\mathrm{sgn}[h_i(t)], \quad i=1,\dots,N. \tag{1}$$

Here $h_i(t)$ is the net stimulus felt by the neuron at time $t$, coming from internal and external inputs but reduced by a threshold parameter. For our purposes it is immaterial whether sequential or parallel updating is imposed. The general higher-order synaptic structure of the network model is expressed in the assumed form

$$\begin{aligned} h_i(t)&=c_{i0}(t)+\sum_{j_1} c_{ij_1}(t)\sigma_{j_1}++\sum_{j_1j_2} c_{ij_1j_2}(t)\sigma_{j_1}(t)\sigma_{j_2}(t)\\ &\quad+\cdots+\sum_{j_1<j_2<\cdots<j_{K_i}} c_{ij_1j_2\dots j_{K_i}}(t)\sigma_{j_1}(t)\\ &\quad\times\sigma_{j_2}(t)\cdots\sigma_{j_{K_i}}(t)\\ &=C_0(t)+C_1(t)+C_2(t)+\cdots+C_{K_i}(t), \end{aligned} \tag{2}$$

where the sums include only those $K_i$ neurons from which neuron $i$ receives inputs. The first term represents any external input to neuron $i$ (reduced by its threshold), while the second term is the usual one representing binary interactions, a simple linear sum of states of input neurons weighted by synaptic strengths $c_{ij_1}$. The higher-order terms in the expansion, for $n\geqslant 2$, represent ''multiplicative'' interactions in that they are linear combinations of the *products* of two or more input-neuron states. One also speaks of a ''sum-of-products'' form for such interactions.

We observe that the general $n$th-order contribution

$$C_n=\sum_{j_1<j_2\cdots<j_n} c_{ij_1j_2\cdots j_n}\sigma_{j_1}\sigma_{j_2}\cdots\sigma_{j_n}, \tag{3}$$

representing the irreducible interaction of $n$ neurons with neuron $i$, introduces $\binom{K_i}{n}=K_i!/n!(K_i-n)!$ weight parameters. Accordingly, specification of the net stimulus (2) requires $2^{K_i}$ parameters. The exponential explosion of parameters with increasing connectivity $K_i$ has deterred widespread application of higher-order networks, in spite of their theoretical advantages.

Indeed, complete optimization of a network having all possible combinations of higher-order terms is patently impractical for sizable values of $K_i$ typically needed in real-world applications. However, a restricted optimization problem has been attacked by retaining only a strongly reduced pattern-specific connectivity [14,15], while otherwise implementing the extended Hebbian learning rule to be introduced below. A similar strategy based on a connection-pruning scheme adapted to the pattern domain has been employed to tame the combinatoric explosion of parameters in higher-order probabilistic perceptrons [16].

Of course, if the entire array of coefficients $c_{ij_1j_2\cdots j_n}$ is specified at the outset, the explosive combinatoric optimization problem becomes moot. In this note we shall focus on the fully connected network in an important special case of "one-shot" learning in which it is feasible and straightforward to evaluate the general term $C_n$ of the series (2). In fact, by exploiting standard group-theoretic results, we are actually able to sum this series in the limit of asymptotically large connectivity ($K_i \to \infty$, implying an infinitely large network).

We consider the familiar task of storage and recall of $p$ random patterns $S^\mu = \{S_1^\mu, S_2^\mu, \ldots, S_2^N\}$ in the firing activities of the neuronal units, where again $S_j \in \{-1,1\}$. As is well known [4,7,8], such patterns can be faithfully stored as fixed points of the dynamics (1) of the network model to a capacity $p = O(N^K)$ (with $K = \min_i K_i$), if the weight parameters of the stimulus expression (2) are chosen according to an extension of the classical Hebbian learning rule to the presence of interactions of all orders up to $K_i$:

$$c_{ij_1j_2\cdots j_n} = \sum_{\mu=1}^{p} S_i^\mu S_{j_1}^\mu S_{j_2}^\mu \cdots S_{j_n}^\mu, \quad n = 1, \ldots, K_i. \tag{4}$$

The efficacy of memory storage is commonly analyzed in terms of the overlaps

$$m^\mu(t) = \sum_j S_j^\mu \sigma_j(t) \tag{5}$$

of the current network configuration

$$\{\sigma_1(t), \sigma_2(t), \ldots, \sigma_N(t)\}$$

with a given pattern $S^\mu$. When a relative-entropy cost function is adopted [17], the weight specification (4) can be shown to be optimal among the class of simple local learning rules (where "local" implies that changes of synaptic strength depend only on the states of the neurons interacting at the given synapse).

To evaluate the generic term (3) in the stimulus expansion (2) under the extended Hebbian ansatz (4), it is convenient to define "generalized" overlaps

$$m_\alpha^\mu(t) = \sum_j [S_j^\mu \sigma_j(t)]^\alpha \tag{6}$$

of the current network configuration with one of the prescribed patterns, where $\alpha$ is a positive integer. Since $S_j^2 = \sigma_j^2 = 1$, the quantity $m_\alpha^\mu(T)$ reduces to $K_i$ for $\alpha$ even and to the ordinary overlap (5) for $\alpha$ odd. Proceeding with direct evaluation of the right-hand side of Eq. (3) for $n = 1$–4, we establish the pattern of behavior for the higher orders:

$$C_1 = \sum_{\mu=1}^{p} S_i^\mu [m_1^\mu], \tag{7}$$

$$C_2 = \sum_{\mu=1}^{p} S_i^\mu \frac{1}{2!}[(m_1^\mu)^2 - m_2^\mu], \tag{8}$$

$$C_3 = \sum_{\mu=1}^{p} S_i^\mu \frac{1}{3!}[(m_1^\mu)^3 - 3m_1^\mu m_2^\mu + 2m_3^\mu], \tag{9}$$

and

$$C_4 = \sum_{\mu=1}^{p} S_i^\mu \frac{1}{4!}[(m_1^\mu)^4 - 6(m_1^\mu)^2 m_2^\mu + 8m_1^\mu m_3^\mu + 3(m_2^\mu)^2 - 6m_4^\mu]. \tag{10}$$

It is seen that the generic term $C_n$ is built as a sum over all patterns of individual terms of the form

$$S_i^\mu \frac{1}{n!} \gamma(\alpha_1, \ldots, \alpha_n) \prod_{l=1}^{n} (m_l^\mu)^{\alpha_l}, \tag{11}$$

where $\gamma(\alpha_1, \ldots, \alpha_n)$ is a statistical weight factor and the generalized overlaps $m_l^\mu$ enter with positive integral powers satisfying the partitioning condition

$$\sum_{l=1}^{n} l\alpha_l = n. \tag{12}$$

The statistical factor is found to obey the sum rules

$$\sum_{(\underline{\alpha})} \gamma(\alpha_1, \ldots, \alpha_n) = 0 \quad \text{and} \quad \sum_{(\underline{\alpha})} |\gamma(\alpha_1, \ldots, \alpha_n)| = n!, \tag{13}$$

and can be constructed as

$$\gamma(\alpha_1, \ldots, \alpha_n) = n! \Bigg/ \left[\prod_{l=1}^{n} (-1)^{\alpha_l+1} (l^{\alpha_l}) \alpha_l!\right]. \tag{14}$$

Thus, for arbitrary $n$, the contribution $C_n$ can be written explicitly as

$$C_n = \sum_{\mu=1}^{p} S_i^\mu \overline{\mathcal{P}}_n(m_1^\mu, \ldots, m_n^\mu), \tag{15}$$

where

$$\overline{\mathcal{P}}_n(m_1, \ldots, m_n) = \frac{1}{n!} \sum_{(\underline{\alpha})} \prod_{l=1}^{n} \gamma(\alpha_1, \ldots, \alpha_n) m_l^{\alpha_l}. \tag{16}$$

The sum over $\underline{\alpha}$ in definition (16) extends only over those $n$-dimensional vectors $\underline{\alpha} = (\alpha_1, \ldots, \alpha_n)$ whose components satisfy constraint (12). The quantity $\overline{\mathcal{P}}_n(m_1, \ldots, m_n)$ is identified as a generalized Polyà polynomial [18] of the symmetric group $\mathcal{S}_n$, with the signs $(-1)^{\alpha_l+1}$ of the corresponding cyclic permutations incorporated.

For given $n$, the total number of solutions $P(n)$ of condition (12) can be determined by induction from the recurrence relation [19]

$$P(n) = \frac{1}{n} \sum_{q=1}^{n} \rho(q) P(n-q), \tag{17}$$

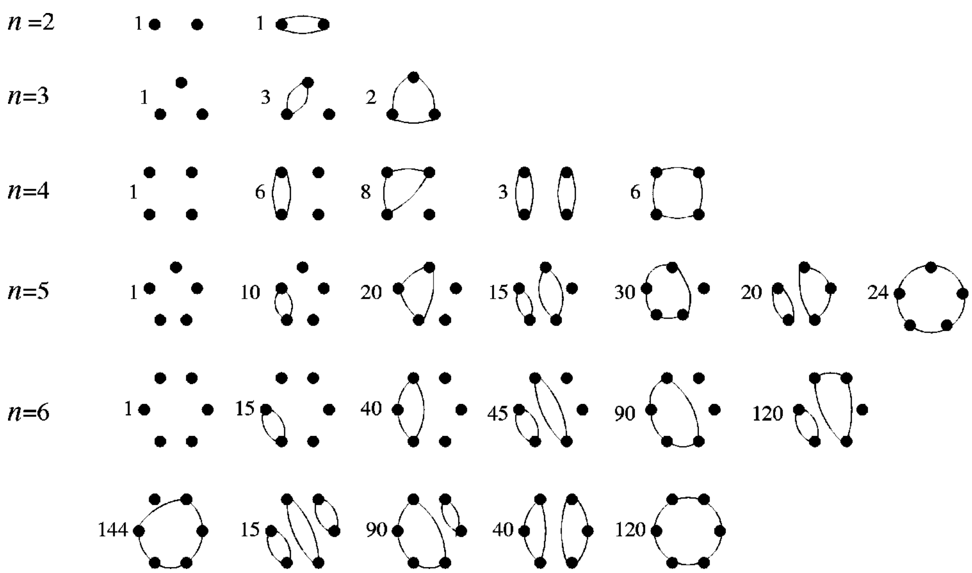


FIG. 1. All possible fermion cluster diagrams for $n=2,3,\ldots,6$, in the absence of dynamical correlations.

in which the divisor function $\rho(l)$ is the sum of the first powers of the divisors of $q$. For large $n$, $P(n)$ behaves asymptotically as

$$P(n)=\frac{1}{4n\sqrt{3}}e^{\pi\sqrt{2n/3}}. \tag{18}$$

Importantly, the generating function of the Polyà polynomials may be employed to calculate the sum over $n$ of all individual contributions $C_n$ in the limit of large connectivity $K_i$, which is equivalent to the thermodynamic limit. We obtain thereby a closed formula for the net internal stimulus defined in Eq. (2),

$$h_i=\sum_{n=0}^{\infty} C_n=\sum_{\mu=1}^{p} S_i^{\mu}\exp\left[\sum_{l=1}^{\infty}\frac{(-1)^{l+1}}{l}m_l^{\mu}\right] \quad (K_i\to\infty). \tag{19}$$

In contrast to this formal result, practical neural-network applications often work with a single fixed order or with a few low orders adapted to the complexity of the problem (see, for example, Ref. [20]).

Combinatoric group-theoretical considerations reveal an interesting one-to-one correspondence between the $n$th-order contribution $C_n$ to the stimulus sum (2) and the sum of planar $n$-particle cluster diagrams for noninteracting particles obeying Fermi statistics. Each fermion cluster diagram is uniquely defined by an $n$-dimensional vector $(\alpha_1,\ldots,\alpha_n)$ satisfying relation (12) and specifying a partitioning of the $n$-particle cluster into subclusters correlated by exchange, namely, into $\alpha_1$ 1-cycles, $\alpha_2$ 2-cycles, $\ldots$, $\alpha_n$ $n$-cycles. The statistical weight factor $\gamma(\alpha_1,\ldots,\alpha_n)$ is the number of ways in which $n$ particles can be assigned to $\alpha_l$ exchange clusters of size $l$, with $l$ running from 1 to $n$. Figure 1 shows all possible cluster diagrams up to order $n=6$. Each contribution diagram consists of $n$ filled dots and the associated exchange lines. Reflecting the Fermi (or Bose) symmetry of the wave function, the exchange lines only occur in closed loops: the particles belonging to a given exchange cluster appear as nodes in a continuous circuit of lines that represents a transposition or cyclic permutation. Cluster diagrams of this type (though with additional lines representing dynamical correlations) are used in the description of noninteracting fermions or bosons in the correlated wave-function and correlated density-matrix formalisms [21,22].

A large number of computer experiments [23] have established the following behavior of higher-order networks when applied to problems in pattern recognition. When the patterns to be recognized are structured rather than random, the network dynamics usually converges to the pattern with *structural* similarity closest to the initial pattern, rather than to (or to a state very near) the pattern having largest overlap with the initial state. This behavior contrasts with that of first-order networks having only binary synapses [24]; relative to these conventional systems, higher-order networks demonstrate a greatly enhanced capability for structural discrimination of arbitrarily complex patterns. Moreover, when functioning in the regime of dilute pattern storage (i.e., far from saturation, thus $p\sim N<<N^K$, $K\geqslant 2$), the basins of attraction of the memorized patterns are dramatically enlarged. Finally, it is to be emphasized that in the model we have considered, the combinatoric explosion of weight coefficients is obviated, since the network only needs to know the overlaps of the present state with all the patterns to be embedded.

In closing, it may be remarked that support-vector machines [25] share with the above construction the salient feature of avoiding explicit evaluation of higher-order terms. In the feedforward architecture characterizing support-vector machines, a hidden layer of suitably chosen inner-product kernels is introduced to establish the optimal hyperplane in feature space without having to address the feature space explicitly [26]. In particular, polynomial kernels will automatically incorporate the effects of multiplicative interactions to the specified order. Beyond this commonality of result, deeper relations between the two constructions are not transparent. We note, in particular, that the Hebbian choice of weights is quite special, and in general implies all-to-all connections between the neuronal units.

This paper is a contribution to the ZiF Research Year on the Sciences of Complexity: From Mathematics to Complexity to a Sustainable World. The research was supported in part by the U.S. National Science Foundation under Grant Nos. PHY-9900713 and PHY-0140316.


[1] *Dendrites*, edited by G. Stuart, N. Spruston, and M. Hausser (Oxford University Press, Oxford, 1999).

[2] B. Mel, Neural Comput. **6**, 1031 (1994).

[3] Y.C. Lee, G. Doolen, H.H. Chen, G.Z. Sun, T. Maxwell, H.Y. Lee, and C.L. Giles, Physica D **22**, 276 (1986).

[4] P. Peretto and J.J. Niez, Biol. Cybern. **54**, 53 (1986).

[5] H. Sompolinsky, Phys. Rev. A **34**, 2571 (1986).

[6] P. Baldi and S.S. Venkatesh, Phys. Rev. Lett. **58**, 913 (1987).

[7] L. Personnaz, I. Guyon, and G. Dreyfus, Europhys. Lett. **4**, 863 (1987).

[8] M.V. Feigel'man and M. Tsodyks, Int. J. Mod. Phys. B **1**, 51 (1987).

[9] E. Gardner, J. Phys. A **20**, 3453 (1987).

[10] L.F. Abbott and Y. Arian, Phys. Rev. A **36**, 5091 (1987).

[11] D. Horn and M. Usher, J. Phys. (Paris) **49**, 389 (1988).

[12] D. Psaltis, C.H. Park, and J. Hong, Neural Networks **1**, 149 (1988).

[13] J.J. Arenzon and R.M.C. de Almeida, Phys. Rev. E **48**, 4060 (1993).

[14] K.E. Kürten, Lect. Notes Phys. **368**, 461 (1990).

[15] K.E. Kürten, J. Phys. (France) **51**, 1585 (1990).

[16] J.W. Clark, K.A. Gernoth, S. Dittmar, and M.L. Ristig, Phys. Rev. E **59**, 6161 (1999).

[17] M.P. Qian, G. Gong, and J.W. Clark, Phys. Rev. A **43**, 1061 (1991).

[18] R. Lund, Math. Comput. **36**, 267 (1981).

[19] *Handbook of Mathematical Functions*, edited by M. Abramowitz and I.A. Stegun (Dover, New York, 1970).

[20] M. Minsky and S. Papert, *Perceptrons* (MIT Press, Cambridge, MA, 1969).

[21] J.W. Clark, Prog. Part. Nucl. Phys. **2**, 89 (1979).

[22] M.L. Ristig, G. Senger, M. Serhan, and J.W. Clark, Ann. Phys. (N.Y.) **243**, 247 (1995).

[23] K.E. Kürten, in *Neural Network Dynamics*, edited by J.G. Taylor, E.R. Caianiello, R.M.J. Cotterill, and J.W. Clark (Springer-Verlag, Berlin, 1992).

[24] D.J. Amit, *Modeling Brain Function: The World of Attractor Neural Networks* (Cambridge University Press, Cambridge, 1989).

[25] V. Vapnik, *The Nature of Statistical Learning Theory* (Springer-Verlag, Heidelberg, 1995).

[26] S. Haykin, *Neural Networks: A Comprehensive Foundation* (McMillan, New York, 1998).