Backpropagation Training in Adaptive Quantum Networks

Christopher Altman · Romàn R. Zapatrin

Abstract We introduce a robust, error-tolerant adaptive training algorithm for generalized learning paradigms in high-dimensional superposed quantum networks, or *adaptive quantum networks*. The formalized procedure applies standard backpropagation training across a coherent ensemble of discrete topological configurations of individual neural networks, each of which is formally merged into appropriate linear superposition within a predefined, decoherence-free subspace. Quantum parallelism facilitates simultaneous training and revision of the system within this coherent state space, resulting in accelerated convergence to a stable network attractor under consequent iteration of the implemented backpropagation algorithm. Parallel evolution of linear superposed networks incorporating backpropagation training provides quantitative, numerical indications for optimization of both single-neuron activation functions and optimal reconfiguration of whole-network quantum structure.

Keywords Neural networks · Quantum topology · Adaptive learning

1 Introduction

Artificial neural networks are routinely applied to resolve unstructured or multivariate machine learning problems such as high-speed pattern recognition, image processing and associative pattern matching tasks. Due to their novelty, however, quantum neural networks remain relatively uncharted in the artificial intelligence and quantum algorithms communities. Several groups [1–4] have outlined preliminary quantum network architectures; each novel approach contributes significant insights towards application methodology, alternative implementations and underlying modal interpretation—however, widespread and effective

C. Altman
Applied Physics, Kavli Institute of Nanoscience, Delft University of Technology, Postbus 5,
2600 AA Delft, The Netherlands
e-mail: altman@casimirinstitute.net

R.R. Zapatrin (✉)
Dept. of Informatics, The State Russian Museum, Inženernaya, 4, 191186, St. Petersburg, Russia
e-mail: Roman.Zapatrin@gmail.com

universal implementation of quantum neural networks remains an open research question, as both theoretical and experimental toolsets are still in the incipient stages of development and maturation.

A common underlying thread shared by quantum network proposals to date is that each implements superposition of neural transition functions upon a single, fixed topological foundation. In this letter, we expand upon a novel framework initially introduced in [5], which diverges from prior network models by fine-tuning not only optimization of neural transition functions—but by fully reconfiguring the connective physical *topology* of the quantum network itself. The mathematical formalism employed for this approach descends from the Rota algebraic spatialization procedure of evolving reticular quantum structures, which was initially developed in [6] to address superposed topological manifolds of space-time foam as described in quantum gravity. Subsequently, not only neuron weighting and *transition functions*—but also linear superposition of the *network topology* itself—are subject to training and revision within this coherent state space.

We formally incorporate the standard backpropagation training algorithm initially introduced by Werbos in [7]. Repeated iteration of training series results in convergence of sample output to a stable network attractor corresponding to the lowest energy configuration space between *given input* and *desired output* layer. Following convergence to this minimum, the superposed linear network is converted upon measurement to a conventional, classical neural network by consequent application of the Rota algebraic projection formalism.

2 'Linear' Neural Networks—Dual contexts

Traditional computers execute algorithms—that is, they follow a specific set of instructions to arrive at a solution to a given problem. Artificial neural networks, by contrast, learn by trial and error: training through example. In a specific class of neural networks—multilayered, *feedforward* neural networks—signals are allowed to propagate only forward: there is no feedback process in the training phase. These connection patterns are formally classified as mathematical structures known as POSETS—partially ordered sets, or DIRECTED ACYCLIC GRAPHS—DAGS.

In this letter, we convert the primary component of artificial neural networks—directed acyclic graphs—into a set of linear matrices. The training optimization protocol is then reclassified in terms of unitary matrices and matrix operations. The result of training optimization is then a matrix, rather than a directed acyclic graph. In order to recover the initial graph structure upon derivation of the appropriate solution, the Rota algebraic spatialization procedure [8] is applied to recover the appropriate graph.

Herein we use the term *adaptive neural networks*. However, the primary objects subject to revision and training are operators in linear spaces—which can be realized, for instance, as quantum observables, rather than as explicitly defined neural networks. In our paper, we shall use the term 'linear' under two separate contexts. The first definition implies restriction to linear artificial neural networks, as we defer application of nonlinear network optimization problems to a subsequent manuscript. In this letter, we focus solely upon optimized approach and applications of *linear* neural networks.

The second usage of *linearization* is the standard linear formalism central to quantum mechanics. This is the basis of our mathematical approach, and is outlined in further detail. The topology of a feedforward artificial neural network, $\mathcal{N}$, is described by the TEMPLATE

MATRIX $\mathbf{A}$ of the appropriate directed acyclic graph, which is formed as follows:

$$\mathbf{A}_{jk} = \begin{cases} *, & \text{if } j \to k \text{ in } \mathbf{M}, \\ 0, & \text{otherwise}, \end{cases} \tag{1}$$

where $*$ stands for a wildcard—any number, and the set of such numerical matrices form an algebra [10] as it is closed under multiplication. The main property of $\mathbf{A}$ is that the synaptic weights 'follow' it—namely, if $\mathbf{A}_{jk} = 0$, then $w_{jk} = 0$.

Returning to picture (2) of signal propagation in $\mathcal{N}$, taking various matrices $\mathbf{A}$—and allowing them only to comply with the template matrix (1)—we form various products (2). The resulting set of matrices is called the ROTA ALGEBRA of $\mathbf{M}$. It can be verified that this set is closed under sums and matrix products, and thus qualifies as a closed algebra. This description is explicated under greater detail in [5].

Given a feedforward neural network $\mathcal{N}$ with the set of nodes $\mathbf{M}$, consider the linear space $\mathcal{H}$, whose basis is labeled by the elements of $\mathbf{M}$. A vector $\mathbf{x} \in \mathcal{H}$ is associated with a state of the network, namely the j-th component of $\mathbf{x}$ is the activity in node j. Note that the state of $\mathcal{N}$ captures the activities of *all* nodes, so a convention is required to define the default activity value of the nodes—which are thus set to zero.

Initial activation corresponds to changing values of only a part of nodes—specifically those in the input layer. The signal then propagates through $\mathcal{N}$, and we describe this activation by a linear[1] operator W in $\mathcal{H}$. The time in which the signal requires to advance is described as the following evolution of the state of the simulator: We start with an initial input vector $\mathbf{x}_0$, thus:

$$\mathbf{x}_0 \mapsto \mathbf{x}_1 = \mathbf{x}_0 W \mapsto \cdots \mapsto \mathbf{x}_{\text{out}} = \mathbf{x}_n = \mathbf{x}_0 W^n \tag{2}$$

following n steps. Recall that postfix notation is employed in this instance.

The crucial feature of the suggested approach is that signal propagation is now described by iteration of *the same* operator W. From this step forward, the operator W, rather than the collection of weights as in the traditional backpropagation approach, will be subject to training.

3 Performing Standard Operations in Linearized Form

In this section, we reformulate classical neuron activity, propagation, and error backpropagation methods in linear terms. A first detail to consider is the process of signal propagation in terms of Rota algebras. We initially have a register—a sequence of unique numbers defining the artificial neural network state—and view it as a vector $\mathbf{x}$ in the appropriate linear space. Following formula (2), we consecutively act upon the register, following elements of the Rota algebra. Initially, setup is defined as $x = 0$. Recall that the entries of $\mathbf{x}$ correspond to activities of a single neuron. When the initial signal is input, this results in the fragment of $\mathbf{x}$ corresponding to the input layer acquiring the appropriate values, while the remainder is still set to zero. The application of W activates the neurons of the second layer. Following this action, the first fragment bears the input values, and the second fragment corresponds to the net inputs of the second layer. This activation function is then applied forward throughout the network.

---

[1] Recall that in this paper we focus solely on linear neural networks.

The final vector, $\mathbf{x}_n$, possesses the following structure, consisting of blocks corresponding to layers—the input block contains corresponding input data, and the intermediate blocks contain net inputs from the previous layers. This captures the property that as the signal propagates through the network, the intermediate neurons retain their values. On the other hand, this requirement is not essential—within the same model, we may also consider artificial neural networks with simultaneous signal inputs at the propagation point of a single input pattern. When presented with a given training set, the overall error function may be minimized by solving the equation,

$$\frac{\partial}{\partial W} A W^n - T = 0. \tag{3}$$

One disadvantage at this point is that as the dimensionality of the state space grows, the exact solution of this equation may become intractable, at which point approximation methods such as gradient descent may need to be employed. To implement this, an error backpropagation method is required. The initial error vector δ is set to be zero. Just as in (2), the backpropagation process can then be described. The input, δ_0, is defined as the difference between the fragment of the output $\mathbf{x}_n$ of the forthpropagation process and the target vector $\mathbf{t}$. The expression for updated errors is the product of these two factors: the first factor can be expressed as $W\mathbf{x}_{n-t}$. The second factor can be expressed as $\delta_t W^T$, where W^T is the transposed matrix. In accordance with agreement that the states of $\mathcal{N}$ are diagonal matrices, rather than vectors, the product of the two factors is well-defined. As a result,

$$\delta_{t+1} = W\mathbf{x}_{n-t}\delta_t W^T. \tag{4}$$

The standard weighting adjustment formula can then be written in compact form, using the adjacency matrix $\mathbf{A}$ of the directed acyclic graph associated with our network:

$$W^{\text{upd}} = W^{\text{ini}} + \delta_n \cdot \mathbf{A}\mathbf{x}_n. \tag{5}$$

This procedure is then applied recurrently until required error values converge to a minimum within predesignated boundary tolerances. It should be noted that there is no novel formalism at this early stage of the procedure—this step is but a reformulation of a standard algorithm in supervised learning.

4 Mathematics of Spatialization

The key point of the spatialization procedure (first introduced in a different context in [6]) is the following: consider the full matrix algebra $\mathcal{A}$ as a linear space. From this perspective, the Rota algebras Ω described by (1) are just linear subspaces of $\mathcal{A}$. Having a procedure which—starting from a subset of $\mathcal{A}$—creates a topological space, providing the capability to discuss superposed configurations of differing neural networks. In this section we present a procedure which—starting from a given subspace of $\mathcal{A}$—produces a set, and endows it with a topology that can be associated with certain acyclic directed graphs.

When the set of nodes is fixed, Rota algebras provide a suitable machinery to describe topology changes, which are expressed in terms of creation and annihilation of edges. In terms of template matrices (1), these operations are adding or removing asterisks from the appropriate templates. In a fixed basis, a Rota algebra is a subspace of a full matrix algebra $\mathcal{A}$—therefore adding or removing asterisks is equivalent to adding or removing basis vectors. Thus, any particular Rota algebra is a linear subspace of Ω. The basic idea of our approach is to associate *any* subspace of Ω with an appropriate directed acyclic graph.

Spectra—the emergence of nodes.   The notion of a spectrum is tightly related with that of an ideal. Suppose again that we have an algebra of functions $\mathcal{A}$ on a set **M**. Fix a point $m \in \mathbf{M}$, and consider the subset $\mathcal{I} \subseteq \mathcal{A}$ of functions f such that $f(m) = 0$. First, $\mathcal{I}$ itself is an algebra. Furthermore, for any $f \in \mathcal{A}$ and any $h \in \mathcal{I}$ the product fh is always an element of $\mathcal{I}$. Such subsets $\mathcal{I}$ are called ideals of the algebra $\mathcal{A}$. Points of **M** are in one-to-one correspondence with the MAXIMAL IDEALS of $\mathcal{A}$[2] and **M** is called SPECTRUM[3] of $\mathcal{A}$.

In the case wherein $\mathcal{A}$ has infinite dimension, such as in a linear space, there are different, non-equivalent and nontrivial topologies which can be defined on $\mathcal{A}$. There are standard recipes for this step: for example, the Zariski topology on the prime spectrum of Ω. Unfortunately, on finite-dimensional algebras this topology is always discrete—which leaves us no chance to fit the requirement of being non-Hausdorff. In terms of graphs, that means that the standard recipes can be utilized to create directed arrows. Thus, we are compelled to find another topology. For these purposes, the Rota topology was suggested in [8].

Suppose we are given a finite-dimensional associative—and non-commutative, in general—algebra Ω. According to standard conceptions and methods of modern algebraic geometry, as well as the general algebraic approach to physics, we introduce the points of Ω as its irreducible representations (IRs). The first step of the spatialization procedure is then creating or finding points of Ω, which will become nodes of the future graph:

$$\{\text{points}\} = \{\text{IRs}\}. \tag{6}$$

Rota topology.   Denote by **M** the set of points of Ω, each of which we shall associate with a prime ideal in Ω. Consider two points (representations of Ω) $i, j \in \mathbf{M}$ and denote by $\ker i, \ker j$ their kernels. Both of them, being kernels of representations, are two-sided ideals in Ω, in particular, subsets of Ω, hence both of the following expressions make sense:

$$\ker i \cap \ker j \subset \Omega \quad \text{and} \quad \ker i \cdot \ker j \subset \Omega.$$

The latter denotes the product of subsets of Ω: $\ker i \cdot \ker j = \{a \in \Omega \mid \exists u \in \ker i, v \in \ker j : uv = a\}$. Since $\ker i, \ker j$ are ideals, we always have the inclusion $\ker i \cdot \ker j \subseteq \ker i \cap \ker j$, which may be strict or not. Define the relation $\propto$ on **M** as follows:

$$i \propto j \quad \text{if and only if} \quad \ker i \cdot \ker j \neq \ker i \cap \ker j. \tag{7}$$

The Rota topology is then the weakest one in which $i \propto j$ implies convergence $i \to j$ of the point i to the point j. Explicitly, the necessary and sufficient conditions for i to converge to j in the Rota topology reads:

$$i \to j \quad \text{if and only if} \quad \exists k_0, \ldots, k_t, \ldots, k_n \mid k_0 = i, \; k_n = j; \; k_{t-1} \propto k_t. \tag{8}$$

This operation is called the transitive closure of the relation $\propto$. Note that in general, the Rota topology can be defined upon any set of ideals, and it is not necessary for Ω to be an algebra. Any linear subspace is suited for this purpose.

As outlined in the mathematical formalism, it should be noted that *adaptive quantum networks*, as coherent quantum metastructures, are not standard neural networks. Rather, they

---

[2]A characterization of all two-sided ideals required for spatialization was recently provided in [9].

[3]Note that there may be different spectra for the same algebra. A simple and degenerate example is given by the set **C** of complex numbers. Its spectrum with respect to complex numbers consists of one point, $\mathbf{M_C} = \{1\}$, while its spectrum with respect to real numbers consists of two points, $\mathbf{M_R} = \{1, i\}$.

are defined as superposed linear spaces, a metasystem whose subspaces consist of neural networks—just as a superposed wavefunction possesses an identity distinct from its constituent classical counterpart under influence of environmental interaction. In the ideal case, only following convergence to the predesignated, target network minima under backpropagation training is projection via the spatialization procedure applied to convert the metasystem into a classical artificial neural network. As graphs are automatically produced as a consequence of this spatialization procedure, no *ab initio* association of states with graphs is requisite to implement the model into a candidate physical structure.

5 Concluding Remarks

We have integrated standard neural network backpropagation training [7] into a linear, decoherence-free subspace of superposed quantum network topologies [5]—or *adaptive quantum networks*—to simultaneously optimize both neural transition functions and topological network configuration. The framework is both robust and error-tolerant against local permutations, and as such is ideally suited to applications in rapid pattern matching, signal processing, image recognition and associative learning.

Coherent quantum information processing architectures that are both well-suited for implementation of adaptive quantum networks and demonstrate promising experimental progress to date include highly-entangled, cluster state quantum computers suggested by Raussendorf and Briegel [11] and implement recently by research group headed by Zeilinger [12], high-dimensional Josephson junction qubit arrays [13], optical lattice traps in Bose-Einstein condensates [14], thin-layer diamond nanofilms [15], as well as novel materials currently under development by research teams active worldwide.

Implementation of the training algorithm requires a relatively limited number of physical constraints to be satisfied—high-dimensional superposition, multiple degrees of internal freedom, and sufficient timescale under decoherence-free coherent state evolution. Consequently, adaptive quantum networks may be implemented both in state-of-the-art quantum technologies, as well as unexpectedly demonstrated within naturally-occurring dynamic molecular complexes such as brain microtubes [18] and proteomics in molecular biology [19, 21]. Several natural systems have been observed to exhibit interference properties and high degrees of internal freedom—including fullerenes [17], tetraphenyl-porphyrin [17], small biomolecules [17], and chlorophyll [19]. Contrary to conventional expectations, Briegel et al. recently demonstrated persistent, driven entanglement in non-equilibrium quantum systems coupled to a hot and noisy environment, in which the presence of environmental noise played a constructive role in maintaining recurring quantum entanglement [21]. As such, promising candidates in coherent quantum information processing technologies, condensed-matter systems, and biophysics are all potential areas of investigation for experimental implementation of adaptive quantum networks [16, 20].

Superposed quantum network topologies—*adaptive quantum networks*—as outlined in this paper, are unique in allowing for simultaneous training of both transition functions and optimization of whole-network topological structure. Quantitative backpropagation training provides predictive indications for optimal reconfiguration of network performance, which can be applied towards a broad class of problems currently addressed by classical neural networks. Continued progress will focus upon predictive simulation methods to determine convergence series in standardized sample classes, complex and chaotic attractor states under dynamic entanglement coupling strengths, alternative training paradigms, and nonlinear dynamics in the presence of environmental coupling through partially-induced weak measurements.

References

1. Ventura, D., Martinez, T.: An artificial neuron with quantum mechanical properties. In: Proceedings of
   the International Conference on Artificial Neural Networks and Genetic Algorithms, pp. 482–485 (1997)
2. Behrman, E.C., Steck, J.E., Skinner, S.R.: A spatial quantum neural computer. In: Proceedings of the
   International Joint Conference on Neural Networks, Washington D.C. (1999)
3. Chrisley, R.L.: Quantum Computation and Quantum Learning. Workshop on Brain, Mind and Physics,
   Prague (1993)
4. Meyer, D., Wallach, N., Freedman, M., Wang, Z.: Topology and foundations of quantum algorithms. US
   ARDA/NSA/ARO Quantum Computing Program Review. Nashville, TN (2003)
5. Altman, C., Pykacz, J., Zapatrin, R.R.: Superpositional quantum network topologies. Int. J. Theor. Phys.
   43, 2029 (2004)
6. Raptis, I., Zapatrin, R.R.: Algebraic description of spacetime foam. Class. Quantum Gravity **18**, 4187
   (2001)
7. Werbos, P.J.: The Roots of Backpropagation, from Ordered Derivatives to Neural Networks and Polit-
   ical Forecasting. Adaptive and Learning Systems for Signal Processing, Communications, and Control
   Series. Wiley, New York (1994)
8. Zapatrin, R.R.: Finitary algebraic superspace. Int. J. Theor. Phys. **37**, 799–815 (1998)
9. Sorkin, R.D.: Indecomposable ideals in incidence algebras. Mod. Phys. Lett. A **18**, 2491–2500 (2003);
   eprint math.CO/0309126
10. Rota, G.-C.: On the foundation of combinatorial theory. I. The theory of Möbius functions. Zeitschrift
    für Wahrscheinlichkeitstheorie **2**, 340 (1968)
11. Raussendorf, R., Briegel, H.J.: Phys. Rev. Lett. **86**, 5188 (2001)
12. Tame, M.S., Prevedel, R., Paternostro, M., Bohi, P., Kim, M.S., Zeilinger, A.: Experimental realization
    of Deutsch's algorithm in a one-way quantum computer. Phys. Rev. Lett. **98**, 140501 (2007)
13. Mooij, H.: The road to quantum computing. Science **307**, 5713, 1210–1211 (2005)
14. Zwierlein, M.W., Stan, C.A., Schunck, C.H., Raupach, S.M.F., Gupta, S., Hadzibabic, Z., Ketterle, W.:
    Observation of Bose-Einstein condensation of molecules. Phys. Rev. Lett. **91**, 250401 (2003)
15. Hanson, R., Awschalom, D.: Coherent manipulation of single spins in semiconductors. Nature **453**,
    1043–1049 (2008)
16. Ferreira, A., Guerreiro, A., Vedral, V.: Macroscopic thermal entanglement due to radiation pressure.
    Phys. Rev. Lett. **96**, 060407 (2006)
17. Arndt, M., Nairz, O., Voss-Andreae, J., Keller, C., van der Zouw, G., Zeilinger, A.: Wave-particle duality
    of C60. Nature **401**, 680–682 (1999)
18. Hagan, S., Hameroff, S.R., Tuszynski, J.A.: Quantum computation in brain microtubules: decoherence
    and biological feasibility. Phys. Rev. E **65**, 061901 (2002)
19. Engel, G., Calhoun, T., Read, E., Ahn, T., Manal, T., Cheng, Y., Blankenship, R., Fleming, G.: Evidence
    for wavelike energy transfer through quantum coherence in photosynthetic systems. Nature **446**, 782–
    786 (2007)
20. Nogales, E., Bouwmeester, D.: Structure of the tubulin dimer by electron crystallography. Nature **391**,
    1099 (1998)
21. Cai, J., Popsecu, S., Briegel, H.J.: Dynamic entanglement in oscillating molecules. LANL technical
    report (2008). arXiv:0809.4906v1 [quant-ph]