Quantization in Neural Networks

Alexander Weissblut[1]

[1] *Kherson State University, 27, Universitetska st., Kherson, 73000, Ukraine*

Abstract
Like the human brain, an artificial neural network is a complex nonlinear parallel processor; it is often called a neurocomputer. Accordingly, mathematical models of a neural network are usually continuous and stochastic, naturally associated with fuzzy logic. Classical systems of artificial intelligence are always naturally associated with classical logic and discrete mathematics. Thus, the representations and models of knowledge, undeniable at least since Aristotle, do not correspond to the cognitive models that are obtained as a result of studying the human brain. In view of Niels Bohr, quantization is a phenomenon of a discrete, sequential process, that inherent in continuous and stochastic systems. However, the traditional mathematical model of quantum mechanics did not imply generalization to dissipative systems. The corresponding generalization, called the Dynamic quantum model (DQM), was proposed by author. It is defined for any dynamic system, given by ordinary differential equation or by diffeomorphism, or for dynamic systems that using logical operations. The neural network is exactly the DQM in the space of input signals. In this paper DQM is defined and constructed universally for both Hamiltonian systems and systems with the fuzzy logic truth function on phase space. The paper goal is to demonstrate quantization on DQM, i.e. actually on neural networks, and to extend the classical Bohr-Sommerfeld condition to the general case, in particular, to systems with a fuzzy truth function.

Keywords
neural network, artificial intelligence, quantization, dynamical system, dynamic quantum model, Markov cascade, fuzzy logic.

1. Introduction

In recent years, artificial intelligence technologies are increasingly based on artificial neural networks, often they are already simply identified with neural networks [1]. Like the human brain, an artificial neural network is a complex nonlinear parallel processor; it is often called a neurocomputer [2].

A neural network has a special form of robustness. If the calculations are distributed among many neurons, then it does not matter that the state of individual neurons in the network is different from what is expected. Noisy or incomplete input signal can still be recognized; the damaged network can perform its functions at a satisfactory level; learning doesn't have to be perfect [3]. Accordingly, mathematical models of a neural network are usually continuous and stochastic, naturally associated with fuzzy logic [4].

Meanwhile, this method of information processing is fundamentally different from the methods used by a conventional digital computer (von Neumann's machine). And the cognitive models of the neural network are fundamentally different from the traditional models of artificial intelligence, corresponding to the von Neumann machine [1]. Classical systems of artificial intelligence (based on proof theory, theory of algorithms, etc.) always involve the use of a symbolic language to represent knowledge, and cognition in them is carried out as a sequential processing of symbolic information. Therefore, their mathematical models are naturally and inevitably associated with classical logic and discrete mathematics [5].

ICTERI-2021, Vol II: Workshops, September 28 – October 2, 2021, Kherson, Ukraine
EMAIL: veitsblit@gmail.com (A. 1)
ORCID: 0000-0002-1476-8884 (A. 1)

Thus, the representations and models of knowledge, undeniable at least since Aristotle, apparently clearly do not correspond to the cognitive models that are obtained as a result of studying the human brain [6].

This paradox was already realized by Niels Bohr. In his view, quantization is a phenomenon of a discrete and sequential process, that inherent in continuous, stochastic and parallel systems. It refers to systems not only physical, but also cognitive [7]. This idea remained a philosophical idea, not embodied in the concrete model. This happened mainly because the ultimately obtained mathematical foundation of quantum mechanics was based on the ideas not of Bohr, but of Schrödinger and was associated with the peculiarities of the physical microcosm [8].

As a result, the traditional mathematical model of quantum mechanics did not imply generalization to dissipative and stochastic systems [9].

The corresponding generalization, called the Dynamic quantum model (DQM), was proposed in [10]. From the assumption that quantum effects are caused by unrecoverable "white noise", this mathematical model of quantum mechanics already follows and is essentially unambiguous. Dynamics in it is described by Markov cascades (time is discrete). This model is simply connected with the traditional one: there is a simple correspondence between Markov cascades and quasisolutions of the corresponding Schrödinger equation. Thus, in a sense, DQM is a bridge between the traditional calculus of quantum mechanics and the intuitive vision of physicists. On the other hand, in this model spectral problems are reduced to the usual perturbation theory of smooth dynamical systems. Thus, the construction of such models can be considered as an asymptotic method for solving spectral problems [11].

But the definition of DQM is not formally related to Hamiltonian systems. It is defined for any dynamic system, given by ordinary differential equation or by diffeomorphism on any smooth Riemannian manifold, or for dynamic systems that using logical operations: algorithms, theorems, software applications. As a Markov cascade, the DQM is approximated by a Markov chain and on a compact set by a finite Markov chain arbitrarily exactly. This allows you to clearly understand the DQM dynamics and build effective algorithms for the study of concrete systems. On the other hand, when fluctuations tend to zero, i.e. in the semiclassical limit, the dynamics of the DQM goes into the initial smooth dynamics. The equivalence of structural stability and hyperbolicity for smooth discrete dynamical systems is established along this path [12].

If we build an approximate model of a given theorem or software application using neural network training, then we will get exactly the DQM of this object. Indeed, when fixing synaptic weights, we get some realization of a Markov cascade, i.e. some realization of this DQM by its definition. (For the rigorous definition of DQM realization and more detailed descriptions see [12].) Thus, a neural network is an example of DQM; moreover, this is its general example.

Thus in [12] the general definition of DQM are given and its use to prove the equivalence of structural stability and hyperbolicity. In [11] the relationship between DQM and the traditional calculus of quantum mechanics was demonstrated. In this paper DQM is defined and constructed universally for both Hamiltonian systems and systems with the fuzzy logic truth function on phase space. Here we generalize the quantization from the traditional quantum mechanics onto this general case.

The paper goal is 1) to demonstrate quantization on DQM (i.e. actually on neural networks) 2) to extend the classical Bohr-Sommerfeld condition to the general case, in particular, to systems with a fuzzy truth function.

The paper is organized as follows: in part 2 we present the basic concepts of the DQM and along the way some necessary lemmas (the detailed description of DQM is given in [11], [12]); in part 3 we prove Bohr-Sommerfeld condition for DQM; part 4 concludes.

We had to omit proofs of some lemmas in order to fit the paper format.

2. The Dynamic Quantum Model: Basic Definitions

2.1. DQM Definition

Let $p(x)$ be an n-dimensional smooth vector field on an n-dimensional smooth Riemannian manifold M, where $x(x_1, x_2, ..., x_n)$ are local Euclidean coordinates on M, $p_i(x) \in C^\infty(R^n)$ ($i = 1,...,n$). On each phase curve $x(t) \in M$ of the dynamical system generated by this vector field

$$\frac{dx_i}{dt} = p_i(x), \quad (i = 1,...,n) \tag{1}$$

consider the integral of the "shortened action" $s(t) = \int_{x(t)} p(x)dx = \int_0^t \|p(\tau)\|^2 d\tau$, where $\|p(\tau)\|^2 = \sum_{i=1}^n p_i^2(\tau)$. The value of $s(t)$ on each curve $x(t)$, which is different from a fixed point, is diffeomorphically expressed in t and is called "optical time". Let ρ be a metric such that $s(t) = \int_{x(t)} d\rho$: $d\rho = \|p(t)\|^2 dt$.

A dynamic quantum model first shifts each point along the phase curve of a given dynamic system over the optical time σ^2 (or ρ – length σ^2) and in a neighborhood of the caustic this shift increases abruptly by $\frac{1}{4}\mu\sigma^2 = \mu\frac{\pi h}{2}$. And then randomly shifts on a distance not exceeding the length of the trajectory from the original to the new point. The following rigorous definition summarizes this description. The definition of a dynamic quantum model is given for an arbitrary dynamic system (1) on an arbitrary compact Riemannian manifold M.

Definition 1. By a dynamic quantum model (DQM) for dynamical system (1) we mean the Markov cascade with the transition function $P(x, A)$, which associates with each point x of the trajectory of (1) and an open subset A of the configuration space probability of getting from x to A in one iteration:

$$P(x, A) = \frac{1}{\sqrt{2\pi t}\sigma} \int_A e^{-(y-Gx)^2/2\sigma^2 t} dy, \tag{2}$$

where t is the shift time from x to Gx along the path of the ρ-length $2\pi h$ or $2\pi h$ + $\mu\frac{\pi h}{2}$ in the neighborhood of the caustic, $\sigma^2 = 2\pi h$. Given the initial distribution, we obtain a Markov process P with this initial distribution and the transition function $P(y, A)$: if μ_t is the distribution at time t, Δt is the lag between the two nearest measurements, then the DQM sets new distribution $P(\mu_t) = \mu_{t+\Delta t}$ at time $t + \Delta t$.

2.2. DQM eigenvalues and Markov deviations

Our goal is to determine pure states and eigenvalues of DQM. And now, along with the discreteness of the measurement process, its limited time will be essential. Of course, the measurement process cannot continue indefinitely, but here its duration is dictated by the very definition of DQM. Namely, the duration of the measurement, in principle, cannot exceed on order $\frac{1}{h}$ since further the measurement errors with dispersion $\sigma^2 t$ (where the diffusion coefficient σ^2 is small of order h) are no longer small and the notion of trajectory loses its meaning. (And you can only talk about the average values for the ensemble, as in statistical physics). Therefore, we limit the time to a certain limiting value T of order $\frac{1}{h}$ ($T \sim \frac{1}{h}$): $T \le \frac{B}{h}$, where B > 0 is a constant. (In general, we say that the quantity u = u (h) in a DQM is of order h^k ($u \sim h^k$ or $u = O(h^k)$), if

$|u| \le h^k$. And u = u (h) is exactly of the order h^k ($u \sim= h^k$), if $ch^k \le |u| \le Ch^k$ for some constants C, c > 0).

Definition 2. The Markov deviation *Z*(*z*) is a smooth vector field on phase space such that 1)

$$\|Z(z)\| \le B\frac{h^2}{\|p\|^2}$$

($B > 0$) is a constant of dynamical system (1) (i.e., the length *Z*(*z*) does not exceed in order $\frac{h^2}{\|p\|^2}$ for all points *z* of the phase space);

2) for any initial point $z_0 = z(t_0)$ on the phase curve *z*(*t*) of the dynamical system (1) and the time instant $\bar{t}$ in optical time ($|\bar{t}| <$ T ~ $\frac{1}{h}$) we have

$$|\int_{t_0}^{\bar{t}}(\int_{t_0}^{t}(Z(z(s),e(s))ds)dt \le Bh^2|, \qquad \textbf{(3)}$$

where e(*t*) is the unit normal vector of the closed phase curve at the point *z*(*t*), $B > 0$ is the constant of dynamical system (1).

Property (3) of the Markov deviation is due to the fact that, by construction, the vector *Z*(*z*(*t*)) has a random orientation, therefore, the pluses and minuses of the accumulations of its projections on the unit vectors are compensated. Therefore, the integral of the accumulation of projections along the phase curve is experimentally indistinguishable from zero.

Corollary of condition 2). If the DQM trajectory z(t) is closed and $z(0) = z(\hat{t})$ ($\hat{t} > 0$) then

$$|\int_0^{\hat{t}}(Z(z(t),e(t))dt \;| \le ch^3, \qquad (4)$$

$$|\int_0^{\hat{t}}(\int_0^{t}(Z(z(s),e(s))ds)dt| \le ch^3 \qquad (5)$$

$$|\int_0^{\hat{t}}\int_t^{t+\hat{t}}(Z(z(s),e(s)ds)dt \;| \le ch^4 \qquad (c > 0). \qquad (6)$$

If instead of a given time limit $T \sim \frac{1}{h}$ we take $T_1 \sim \frac{1}{h}$ ($T_1 \ne T$), then we obtain another Markov deviation; similarly, when replacing the zero point in time. So the Markov deviation is a smooth vector field that depends on the parameters; further, it can be assumed to be a general view field.

2.3. DQM pure states and eigenvalues. Quantization of spectrum in DQM.

The physical meaning of the eigenvalues is that these are all values of energy that can be the result of reliable, i.e. the most accurate measurement (ideally of the order of h^2). But as a result of the most accurate experiments, as we have seen, in reality the dynamics is studied not of the diffeomorphism G, but of its perturbation $\overline{G} = G + Z$. Let *J* = *J* (*z*) is a given smooth function on phase space. We can interpret it as the Hamiltonian (energy in the phase space) or as a function of truth (0 ≤ *J* (*z*) ≤1), equal to 1 on the true trajectory and 0 outside some neighborhood of it. Given the irremovable errors of the Markov deviation, the discreteness of the measurement process and its limited time, we arrive at the maximum number of the most accurate measurements $\frac{1}{N_t}\sum_{i=0}^{N_t} J(\overline{G}^i z)$, where z is the point of phase space, Z is the general view Markov deviation, $\overline{G} = G + Z$ is a diffeomorphism, N_t is the maximum number of significantly different measurements over time $t \le T$.

Definition 3. Let $\overline{G} = G + Z$, where Z is a general view Markov deviation; N_t is the number of all iterations of the diffeomorphism in time t; α is a real number. Let $D = D_{\alpha h}$ be the set of points z of the phase space such that for all sufficiently large $t < T \sim= \frac{1}{h}$

$$\left| \frac{1}{N_t} \sum_{i=0}^{N_t} J(\overline{G}^i z) - \alpha \right| < Bh^2 , \tag{7}$$

where B is a constant. Then, if for any Z of a general view and sufficiently small h, the set $D_{\alpha h}$ contains a ball with a diameter of exactly the order of h, then α will be called the eigenvalue of the DQM for dynamical system (1), and $D_{\alpha h}$ will be called the carrier of the pure state corresponding to this eigenvalue.

Thus, all points of a DQM spectrum are formally determined only with an accuracy of the order of h^2, but this corresponds precisely to their meaning. By definition, the domain $D_{\alpha h}$ is an open $\overline{G}$ - invariant subset of the phase space.

So, to define DQM means to set: 1) the Markov process in accordance with Definition 1; 2) the Markov deviation Z of general view or, what is the same, diffeomorphism $\overline{G} = G + Z$ in accordance with Definition 2.

2.4. Quantizing the spectrum in DQM

Consider the two–dimensional dynamical system (1), the compact phase space Λ of which is filled with closed phase curves. After the smooth change of variables, in canonical coordinates $z(p; x)$, this is the dynamics of uniform rotation along concentric circles. If we interpret J as a function of truth, then its values on each circle, concentric to the true path (true circle), are constants (i.e., they do not depend on a point on this circle): $J(p, x) = J(p)$. At the semantic level, with such interpretation, we are talking about transitions to equivalent propositions.

Theorem. The DQM eigenvalues of the given dynamical system, with accuracy of the order h^2, are equal to the values of $J(z)$ on the phase circles in Λ, the ρ - length of which satisfies the Bohr - Sommerfeld condition

$$\int p(x)\,dx = \pi I = \pi h(n + \frac{1}{2})$$

and only they.

Let $z_0 \in \Lambda$, then by DQM definition $z_1 = \overline{G} z_0 = Gz_0 + Z(Gz_0)$, $x(Gz_0) = x(z_0) + p(z_0)\Delta t_0$, where Δt_0 is time of transition from z_0 to z_1, $x(z_1) = x(Gz_0) + x(Z(Gz_0))$, $p(z_1) = p(z_0) + p(Z(Gz_0))$. Likewise $z_2 = \overline{G} z_1 = \overline{G}^2 z_0 = G(z_1) + Z(Gz_1)$, whence $p(z_2) = p(z_1) + p(Z(Gz_1))$, $x(Gz_1) = x(z_1) + p(z_1)\Delta t_1$, $x(z_2) = x(Gz_1) + x(Z(Gz_1))$. In general $z_k = \overline{G}^k z_0 = \overline{G} z_{k-1} = G(z_{k-1}) + Z(Gz_{k-1})$, whence $p(z_k) = p(z_{k-1}) + p(Z(Gz_{k-1}))$, $x(Gz_{k-1}) = x(z_{k-1}) + p(z_{k-1})\Delta t_{k-1}$, $x(z_k) = x(Gz_{k-1}) + x(Z(Gz_{k-1}))$, Δt_{k-1} is time of transition from z_{k-1} to z_k $\Rightarrow$

$$p(\overline{G}^k z_0) = p(z_0) + \sum_{i=0}^{k-1} p(Z(Gz_i)),$$

$$x(\overline{G}^k z_0) = \sum_{i=0}^{k-1} p(z_i)\Delta t_i + \sum_{i=0}^{k-1} x(Z(Gz_i)) . \tag{8}$$

From property 2) of the Markov deviation, i.e. (3) it follows that for the perturbed DS with these conditions, the averaging theorem in single-frequency systems [13] is valid. From this theorem for the considered DQM follows the usual derivation of the perturbation theory: trajectories and therefore iterations of the perturbed diffeomorphism $\overline{G}$, starting on the circle of the dynamical system (DS), at any time $t < T \sim \frac{1}{h}$ remain in its neighborhood of order h: for some constant b

$$\left\| \overline{G}^n z - G^n z \right\| \le bh \qquad (2\pi hn \le T \sim \frac{1}{h}) \tag{9}$$

Lemma 1. Let $z_0 \in \Lambda$, K be the phase circle from Λ, $z_0 \in$K, $\hat{t}$ is the traversal time of K, $N_{\hat{t}}$ is the number of iterations of the map G in time $\hat{t}$. Then for some DQM constant A > 0

$$\left| p(G^{N_{\hat{t}}} z_0) - p(z_0) \right| \le Ah^2 \quad \Leftrightarrow \quad p(G^{N_{\hat{t}}} z_0) - p(z_0) \sim h^2 .$$

In accordance with Lemma 1, for the general form of the DQM (i.e., for its general form of the Markov deviation), we assume on each closed trajectory of the DS(1) such z_0, that

$$p(G^{N_{\hat{t}}} z_0) - p(z_0) \sim= h^2 \Leftrightarrow \left| p(G^{N_{\hat{t}}} z_0) - p(z_0) \right| = dh^2 , \tag{10}$$

where $d_0 > 0$ is a constant.

Lemma 2. Under the conditions of Lemma 1, for some DQM constant B > 0

$$\left| x(\overline{G}^{N_{\hat{t}}} z_0) - x(G^{N_{\hat{t}}} z_0 \right| \le Bh^2 \Leftrightarrow x(\overline{G}^{N_{\hat{t}}} z_0) - x(G^{N_{\hat{t}}} z_0) \sim h^2 .$$

Lemma 3. Under the conditions of Lemma 1, let $\Delta p_j = p(\overline{G}^j z_0) - p(z_0)$, $\gamma_j = p(\overline{G}^{N_{\hat{t}}+j} z_0) - p(\overline{G}^j z_0)$ (j = 1, …, $N_{\hat{\tau}}$). Then for some DQM constants $d_1 > 0$ and $d_2 > 0$

$$\beta = \sum_{j=0}^{N_{\hat{t}}-1} \Delta p_j \sim h \Leftrightarrow |\beta| \le d_1 h , \tag{11}$$

$$\delta = \sum_{j=0}^{N_{\hat{t}}-1} \gamma_j \sim h^2 \Leftrightarrow |\delta| \le d_2 h . \tag{12}$$

Lemma 4. Let $z_1, z_2 \in \Lambda$, $x(z_2) - x(z_1) \sim h$, $p(z_2) - p(z_1) \sim h^2$, natural number m $\sim \frac{1}{h}$. Then

$$\left\| (\overline{G}^m z_2 - \overline{G}^m z_1) - (z_2 - z_1) \right\| \sim h^2 \tag{13}$$

If $x(\overline{G}^m z_2) - x(\overline{G}^m z_1) \sim h$, $p(\overline{G}^m z_2) - p(\overline{G}^m z_1) \sim h^2$, then (13) is also true for m $\sim \frac{1}{h}$.

Consequence. If $z_1, z_2 \in \Lambda$, m $\sim \frac{1}{h}$ and $z_2 - z_1 \sim h^2$ or $\overline{G}^m z_2 - \overline{G}^m z_1 \sim h^2$, then $(p(\overline{G}^m z_2) - p(\overline{G}^m z_1)) - (p(z_2) - p(z_1)) \sim h^3 \Leftrightarrow$ for any b > 0 such that $\|z_2 - z_1\| \le bh^2$ there is such a DQM constant C = C(b), that $|p(\overline{G}^m z_2) - p(\overline{G}^m z_1)) - (p(z_2) - p(z_1)| \le Ch^3$.

Proof of the theorem. Let the ρ - length $2\pi I_0$ be a multiple of 2πh on the phase circle $K_0 \subset \Lambda$, taking into account two turning points: $2\pi I - 2 \cdot \frac{\pi h}{2} = 2n\pi h$. Then $G^n z = z$ for all $z \in K_0$. Let $\alpha_n = J(z)$ at $z \in K_0$, where J is a function of truth, $K_1 = \overline{G}^n K_0 \subset \Lambda$, $K_2 = \overline{G}^n K_1 \subset \Lambda$, …, $K_i = \overline{G}^n K_{i-1} \subset \Lambda$, … . Further, let the set of points $D = D_n$ contain all regions between K_i and K_0 for i = 1,…, m ~ = $\frac{1}{h}$, where the choice of m is indicated below. Let us show that

1) D_n contains a ball with a diameter of the order of h;

2) (7) holds on D_n .

Then D_n is the support of the pure state corresponding to the eigenvalue α_n by Definition 3.

3) There are no other eigenvalues besides α_n (n = 1, 2,…) in this DQM.

1) Let $z_0 \in K_0$ is such that $d = p(\overline{G}^n z_0) - p(z_0) = d_0 h^2$ in accordance with (19); here $d > 0$: the case d < 0 is identical and is reduced to d > 0 by reversing the time. We put $z_i = \overline{G}^{in} z_0 \in K_i$ for all i = 1, 2, … . Here $z_1 = \overline{G}^n z_0 \in K_1$, $z_2 = \overline{G}^n z_1 \in K_2$, $d = p(z_1) - p(z_0)$. By lemma 2 $x(z_1) - x(z_0) \sim h^2$ in view of $x(G^{N_{\hat{t}}} z_0) = x(G^n z_0) = x(z_0)$. From here $z_1 - z_0 \sim h^2$. Then by the corollary to Lemma 4 $(p(z_2) - p(z_1)) - (p(z_1) - p(z_0)) = (p(\overline{G}^n z_1) - p(\overline{G}^n z_0)) - (p(z_1) - p(z_0)) \sim h^3 \Rightarrow p(z_2) - p(z_1) = d + O(h^3)$. And $x(z_2) -$

$x(z_1) = x(\overline{G}^n z_1) - x(\overline{G}^n z_0) = (x(G^n z_1) + O(h^2)) - (x(G^n z_0) + O(h^2))$ by Lemma 4. Hence $x(z_2) - x(z_1) = (x(G^n z_1) - x(G^n z_0)) + 2O(h^2) = (p(z_1) - p(z_0))\sum_{i=1}^{n}\Delta t_i + 2O(h^2)$, $\Delta t_i = 2\pi h$ for all i except two turning points, where $\Delta t_i = 2\frac{1}{2}\pi h$, so that $\sum_{i=1}^{n}\Delta t_i = 2\pi I_0$. So

$$x(z_2) - x(z_1) = 2\pi I_0 d + 2O(h^2), \quad p(z_2) - p(z_1) = d + O(h^3).$$

This constitutes the basis of induction. Consider the general case.

Let B is the constant from Lemma 2, C is the constant from the corollary of Lemma 4 that for $v_1, v_2 \in \Lambda$

$$|p(v_2) - p(v_1)| \le 2d_0h^2, \quad |x(v_2) - x(v_1)| \le (4\pi I_0 d_0 + 2B)h^2 \tag{14}$$

implies $|(p(\overline{G}^n v_2) - p(\overline{G}^n v_1)) - (p(v_2) - p(v_1))| \le Ch^3$. Let m be the integer part of the $\frac{d_0}{2Ch}$: this is the value of m used to define the area $D = D_n$. Let by induction

$$|(p(z_i) - p(z_{i-1})) - d| \le C(i-1)h^3, \quad |x(z_i) - x(z_{i-1})| \le (4\pi I_0 d_0 + 2B)h^2 \tag{15}$$

Then $|p(z_i) - p(z_{i-1})| \le d_0h^2 + C(i-1)h^3 \le d_0h^2 + Cmh^3 \le 2d_0h^2 \Rightarrow$ for z_i, z_{i-1} conditions (31) are satisfied and the corollary of Lemma 4 with the constant C is applicable to them:

$$|(p(z_{i+1}) - p(z_i)) - (p(z_i) - p(z_{i-1}))| = |p\overline{G}^m z_i) - p(\overline{G}^m z_{i-1})) - (p(z_i) - p(z_{i-1})| \le Ch^3.$$ Hence

$$p(z_{i+1}) - p(z_i) \le (p(z_i) - p(z_{i-1})) + Ch^3 \le (d + C(i-1)h^3) + Ch^3 = d + Cih^3,$$
$$p(z_{i+1}) - p(z_i) \ge (p(z_i) - p(z_{i-1})) - Ch^3 \ge (d - C(i-1)h^3) - Ch^3 = d - Cih^3.$$

Hence $|(p(z_{i+1}) - p(z_i)) - d| \le Cih^3$. This proves the induction hypothesis (32) on p. Now on x: $x(z_{i+1}) - x(z_i) = x(\overline{G}^n z_i) - x(\overline{G}^n z_{i-1})$, whence by Lemma 2

$$|(x(z_{i+1}) - x(z_i)) - (x(G^n z_i) - x(G^n z_{i-1}))| \le 2Bh^2 \Rightarrow |x(z_{i+1}) - x(z_i)| \le |x(G^n z_i) - x(G^n z_{i-1})| + 2Bh^2 =$$

$$|p(z_i) - p(z_{i-1})| \cdot \sum_{i=1}^{n}\Delta t_i + 2Bh^2 = 2\pi I_0 |p(z_i) - p(z_{i-1})| + 2Bh^2.$$

But $|p(z_i) - p(z_{i-1})| \le 2|d_0|\, h^2$ by the induction hypothesis, whence $|x(z_{i+1}) - x(z_i)| \le (4\pi I_0|d_0| + 2B)h^2$. This proves the induction hypothesis, i.e. conditions (14) for all $i = 1,\dots, m$.

It follows from (14), in particular, that for all $i = 0, 1, \dots, m$

$$p(z_{i+1}) - p(z_i) \ge d - Cih^3 \ge d - Cmh^3 \ge d_0h^2 - \frac{d_0}{2}h^2 = \frac{d_0}{2}h^2 \quad \Rightarrow$$

$$p(z_m) - p(z_0) = ((p(z_m) - p(z_{m-1})) + \dots + ((p(z_1) - p(z_0)) \ge m\frac{d_0}{2}h^2 \sim= h$$ in view of m ~ = $\frac{1}{h}$. Hence in view of (14)

$$p(z_m) - p(z_0) \sim= h \tag{16}$$

On x, from (14) we obtain for all $i = 1, \dots, m$

$$|x(z_i) - x(z_0)| \le |x(z_i) - x(z_{i-1})| + \dots + |x(z_1) - x(z_0)| \le m \cdot (4\pi I_0 d_0 + 2B)h^2 \sim h. \tag{17}$$

Let V be an arc of a circle K_0 centered at a point z_0 of length exactly on the order of h, the radius of which exceeds all $|x(z_i) - x(z_0)|$ ($i = 1, \dots, m$) in accordance with (17). Let $\gamma_1 = \{z(x) \mid x \in [x(z_0); x(z_1)], p(z(x)) \in [p(z_0); p(z_1)]\}$ is a continuous curve from z_0 to z_1 in the region between K_1 and K_0 and thus in D_n. Thus $\gamma_2 = \overline{G}^n\gamma_1 = \{z(x) \mid x \in [x(z_1); x(z_2)], p(z(x)) \in [p(z_1); p(z_2)]\}$, …, $\gamma_m = \overline{G}^{mn}\gamma_1 =$

$\{z(x) \mid x \in [x(z_{m-1}); x(z_m)], p(z(x)) \in [p(z_{m-1}); p(z_m)]\}$ are continuous curves from z_1 to z_2 and from z_{m-1} to z_m accordingly. And then $\gamma = \bigcup_{i=1}^{m} \gamma_i$ is continuous curve in D_n from z_0 to z_m which is projected along x into an arc V. Consider the flow of the perturbed trajectories starting at points from γ, projected along x onto the arc V, i.e. trajectories in a time of the order of h. During this time, the trajectory with the initial dynamics diverges from the perturbed trajectory with the same initial point by a distance of the order of h^2. Hence, in view of (16), on the trajectory of this flow with origin at z_m, the minimum distance $\bar{d}$ to K_0 is exactly of the order of h: $\bar{d}$ ~ = h. So all points $z(x; p)$ in the range $x \in V$, $p \in [p(z_0); p(z_0) + \bar{d}]$ are points on the trajectories of the flow and therefore from D_n. This is what is required: D_n contains a ball of diameter exactly of the order of h.

In view of (9), points from this sphere with a diameter of the order of h will remain at a distance of the order of h during the time of revolution of the phase circle K_0. And then, as follows from Lemma 4, during this time the differences of their p - coordinates change only by a value of the order of h^2. Hence for the region $D^+_n = \bigcup_{i=0}^{n} \bar{G}^i(D_n)$ for all $\bar{x} \in [0; 2\pi]$ the length of its intersection with a straight line $\{z \mid x(z) = \bar{x}\}$ has exactly the order of h. Reversing time, i.e. in the previous reasoning, everywhere replacing $\bar{G}$ to $\bar{G}^{-1}$, we obtain a region D^-_n, symmetric to D^+_n with respect to K_0 with an accuracy of the order of h^2. Therefore, the domain $\bar{D}_n = D^+_n \cup D^-_n$ forms a neighborhood of the circle K_0 exactly of the order h:

$$z \notin \bar{D}_n \Rightarrow \|z - K_0\| \sim\geq h. \tag{18}$$

2) In accordance with Definition 3, it is required to show that for all $z_0 \in D_n$ and sufficiently large t < T ~ $\frac{1}{h}$ $\left| \frac{1}{N_t} \sum_{i=0}^{N_t} J(\bar{G}^i z_0) - \alpha_n \right| < Bh^2$, where N_t is the number of all iterations in time t.

For considering DS $J(\bar{G}^i z_0) - J(z_0) = a(\Delta p_i) + O(\Delta p_i)^2$, where $\Delta p_i = p(\bar{G}^i z_0) - p(z_0)$, $a = \frac{\partial J}{\partial p}(p(z_0))$. As $p(\bar{G}^i z_0) - p(z_0) \sim h$ for all $i = 1, 2, \ldots, N_t$ in view of (12), then $J(\bar{G}^i z_0) - J(z_0) = a(\Delta p_i) + O(h^2)$. Therefore

$$\frac{1}{N_t} \sum_{i=0}^{N_t} J(\bar{G}^i z) - J(z_0) = \frac{1}{N_t} \sum_{i=0}^{N_t} (J(\bar{G}^i z) - J(z_0)) = \frac{\alpha}{N_t} \sum_{i=0}^{N_t} \Delta p_i + O(h^2). \tag{19}$$

First, let z_0 be contained in the region between the curves K_1 and K_0. Then for $z \in K_0$ we have $J(z_0) - \alpha_n = J(z_0) - J(z) = a(p(z_0) - p(z)) + O(h^2) \sim h^2$ as distance between K_0 and $K_1 = \bar{G}^n K_0$ is of order h^2 by Lemma 1.

We put

$$\beta = \sum_{k=0}^{n-1} \Delta p_k. \quad \gamma_{ij} = p(\bar{G}^{in+j} z_0) - p(\bar{G}^{(i-1)n+j} z_0) = \Delta p_{in+j} - p_{(i-1)n+j}$$

at $1 \le i \le m$, $0 \le j < n$. Let $\delta_i = \sum_{j=0}^{n-1} \gamma_{ij} \sim h^2$ by Lemma 3, the integer part of $\frac{N_t}{n}$ is equal to $[\frac{N_t}{n}] = M$ ~ = $\frac{1}{h}$ with t ~ = $\frac{1}{h}$. In view of (19),

$$\begin{aligned} \frac{1}{N_t} \sum_{k=0}^{N_t} J(\bar{G}^k z_0) - J(z_0) &= \frac{\alpha}{N_t} \sum_{k=0}^{N_t} \Delta p_k + O(h^2) = \\ &= \frac{\alpha}{N_t} \cdot [M\beta + \sum_{i=1}^{M} k\delta_i + \sum_{k=Mn}^{N_t} \Delta p_k] + O(h^2). \end{aligned} \tag{20}$$

But $\frac{a}{N_t}\cdot M\beta \sim h^2$ in view of $\beta = \sum_{k=0}^{n-1}\Delta p_k \sim h$ by (11) Lemma 3 and $\frac{M}{N_t} \le \frac{1}{n} \sim h$. Further $\frac{a}{N_t}\cdot \sum_{i=1}^{M} k\delta_i \le |\frac{aM}{N_t}\sum_{i=1}^{M}\delta_i| \sim h^2$ as $\frac{M}{N_t} \le \frac{1}{n} \sim h$, all $|\delta_i| \le d_2h^2$ by (12) Lemma 3 and $M = [\frac{N_t}{n}] \sim \frac{1}{h}$ at $t \sim \frac{1}{h}$. Finally $\frac{a}{N_t}\cdot\sum_{k=Mn}^{N_t}\Delta p_k \sim h^2$, since everyone $\Delta p_k \sim h$, $N_t - Mn \le n \sim \frac{1}{h}$, and $\frac{a}{N_t} \sim h^2$ at $t \sim \frac{1}{h}$. As a result $\frac{1}{N_t}\sum_{k=0}^{N_t} J(\overline{G}^k z_0) - J(z_0) \sim h^2$, and since $J(z_0) - \alpha_n \sim h^2$, then $\frac{1}{N_t}\sum_{k=0}^{N_t} J(\overline{G}^k z_0) - \alpha_n \sim h^2$, which was required.

Consider now the general case $z_0 \in D_n$. By construction, there is a point $\bar{z}_0$ between the curves K_1 and K_0 such that $\overline{G}^{jn}\bar{z}_0 = z_0$ for some j $(1 \le j \le m)$. Choose a time τ such that $N_\tau = jn$. Then in (20) $M = [\frac{N_\tau}{n}] = j$, $N_\tau - Mn = 0$ and

$$\Phi_1 = \frac{1}{N_\tau}\sum_{i=0}^{N_\tau} H(\overline{G}^i\bar{z}_0) - \alpha_n = \frac{1}{N_\tau}a\cdot[M\beta + \sum_{i=1}^{M}k\delta_i] + O(h^2) =\sim h^2 \tag{21}$$

since $\frac{1}{N_\tau}a\cdot M\beta \sim h^2$ and $\frac{1}{N_\tau}a\sum_{i=1}^{M}k\delta_i \sim h^2$ as there, and the condition $\frac{1}{N_\tau}a\cdot\sum_{k=Mn}^{N_t}\Delta p_k \sim h^2$ from (36), which is true only if $t \sim = \frac{1}{h}$, is not needed here. On the other hand, for sufficiently large $t \sim = \frac{1}{h}$ we obtain

$$\Phi_2 = \frac{1}{N_{t+\tau}}\sum_{i=0}^{N_{t+\tau}} H(\overline{G}^i\bar{z}_0) - \alpha_n \sim h^2,$$

as in (20), in view of $t + \tau \sim = \frac{1}{h}$. Hence $\sum_{i=0}^{N_{t+\tau}} H(\overline{G}^i\bar{z}_0) = N_{t+\tau}\alpha_n + N_{t+\tau}\phi_2$, and from (21) $\sum_{i=0}^{N_\tau} H(\overline{G}^i\bar{z}_0) = N_\tau\alpha_n + N_\tau\phi_1$. Hence

$$\frac{1}{N_t}\sum_{i=0}^{N_t} H(\overline{G}^i z_0) - \alpha_n = \frac{1}{N_t}(N_\tau\Phi_1 + N_{t+\tau}\Phi_2).$$

But $\frac{1}{N_t}(N_\tau\phi_1 + N_{t+\tau}\phi_2) \sim h^2$ as $\phi_1, \phi_2 \sim h^2$ and $0 < \frac{N_\tau}{N_t} < 1$, $1 < \frac{N_{t+\tau}}{N_t} < 2$ for sufficiently large t, as required.

These results extend directly to the domain $\overline{D}_n = D^+_n \cup D^-_n$ so that, in view of (18), either $\frac{1}{N_t}\sum_{i=0}^{N_t} H(\overline{G}^i z_0) - \alpha_n \sim h^2$ or $\| z_0 - K_0\| \sim \ge h$ for all $z_0 \in \Lambda$.

3) Let $z_0 \in \Lambda$, K is a phase circle in Λ, $z_0 \in K$, $z_1 = Gz_0 \in K$. In general, for the ρ - length $2\pi I$ of the circle $K = K_I$, the number $R = 2\pi I - 2\cdot\frac{\pi h}{2}$ is not a multiple of $2\pi h$: integer part $N = [R / 2\pi h] < R / 2\pi h$, fractional part $r = R / 2\pi h - N > 0$. Then $G^N(z_0 + r) = z_0$, $G^N z_1 = z_1 - r = G^{N+1}z_0$, $G^{N+1}(z_0 + r) = z_1$. Hence

$$G^N[z_0+r;z_1]=[z_0;z_1-r]; \quad G^{N=1}[z_0;z_0+r]=[z_1-r;z_1]. \tag{22}$$

Let us define the Poincaré succession function on an arbitrary arc $[z_0; z_1]$ of ρ - length 2πh of the circle K_I. We put $W(z) = \overline{G}^N z - G^N z$ for $z \in [z_0; z_1]$. If $r = 0$ for K_I, then $\overline{G}^N z = z + W(z)$ ($z \in [z_0; z_1]$): this is the case discussed above. In the general case, for each traversal of the circle, the deviation W(z) is added to another point of the segment $[z_0; z_1]$ and the dynamics around K_I turns out to be completely different. Namely, we glue the points z_0 and z_1, thereby transforming the arc $[z_0; z_1]$ into a circle C_I of ρ - length 2πh i.e. of ρ – radius h. Then, according to (38), the operator $Vz = G^N z$ is the rotation of the circle C_I by ρ - length r, that is, by an angle $\varphi = 2\pi \frac{r}{2\pi h} = \frac{r}{h}$ counterclockwise, if clockwise rotation on the phase circle and vice versa. And W (z) is added at the point Vz, the succession function for the disturbed dynamics $Fz = Vz + W(z)$: formally in accordance with (22) $Fz = \overline{G}^N z$ at $z \in [z_0 + r; z_1]$ and $Fz = \overline{G}^{N+1} z$ at $z \in [z_0; z_0 + r]$ on the circle C_I.

For an arbitrary arc $[z_0(t); z_1(t)]$ on K_I of ρ – length 2πh, where $z_i(t)$ is the image of z_i (i = 1, 2) under the action of dynamics in time t, we obtain, as in (22)

$$G^N[z_0(t) + r; z_1(t)] = [z_0(t); z_1(t) - r]; \quad G^{N+1}[z_0(t); z_0(t) + r] = [z_1(t) - r; z_1(t)].$$

Hence, after gluing the points $z_0(t)$ and $z_1(t)$ we also obtain on the circle $C_I(t)$ of the ρ- radius h, the rotation $Vz = G^N z$ through the angle $\varphi = \frac{r}{h}$. Then the succession function on $C_I(t)$ for DQM dynamics is $F_t z = \overline{G}^N z$ at $z \in [z_0(t) + r; z_1(t)]$ and is $Fz = \overline{G}^{N+1} z$ at $z \in [z_0(t); z_0(t) + r]$, $F_t z = Vz + W(z,t)$.

This dynamics is transferred from K_I to C_I: point $z \in C_I$ at time t is transformed into $\overline{F}(z,t) = Vz + W(z, t)$. By construction, the function $W(z, t)$ is 2πI periodic in time t and $W(z, t) \sim h^2$ for all $z \in C_I$ and t according to Lemmas 3 and 4. When decomposing into a segment of a Taylor series for the mapping $\overline{F}(z,t)$ on a circle C_I of ρ- radius h it mean, that Vz is the linear, and $W(z, t)$ is the nonlinear part of $\overline{F}(z,t)$. Therefore linearization of the dynamics on K_I induces uniform rotation in optical time on C_I in accordance with the Floquet theorem [14]. In a complexified form, such dynamics is given by the equation

$$\dot{\zeta} = i\omega\zeta, \tag{23}$$

where ζ is the coordinate on C_I, $|\zeta| = h$, ω is the rotation frequency. Then, taking into account the nonlinear part of $\overline{F}(z,t)$ the dynamics on K_I induces on C_I the dynamics given by the equation

$$\dot{\zeta} = i\omega\zeta + w(\zeta, \overline{\zeta}, t), \tag{24}$$

where $w(\zeta, \overline{\zeta}, t)$ is the nonlinear part of the equation generating the deviation $W(z, t)$ that as smooth as dynamics on K_I and is 2πI periodic in time t. Let's change the time t → t / I: then $W(z, t)$ and $w(\zeta, \overline{\zeta}, t)$ are 2π periodic in time t, and $2\pi\omega = \varphi = \frac{r}{h}$, whence $\omega = \frac{r}{2\pi h}$, i.e., ω is the fraction of the ρ - length r of rotation on the circle C_I under the action of $\overline{F}(z,t)$ to the ρ - the length of the circle C_I. So, (24) is a smooth equation, which is 2π - periodic in t on a two-dimensional phase space. Then [15] for ω ≠ 0 (40) is reduced by a smooth change of variables to the following canonical form:

$$\dot{\zeta} = i\omega\zeta + a\zeta|\zeta|^2 + b\overline{\zeta}^{q-1} + o(\zeta^3), \tag{25}$$

where q is the denominator of the irreducible fraction $\omega = \frac{p}{q}$ and a is a parameter. For irrational ω, here b = 0. For q = 2, b = 0 since the linear part of (25) coincides with (24). In this form, the equation does not depend on t to $O(\zeta^3)$ at least, i.e. with an accuracy of $\sim h^3$.

In 1) the phase circles from Λ, that are not contained in $\overline{D}_n = D^+_n \cup D^-_n$, are located at a distance of the order of ≥ h from this circle according to (18). According to 2) for any $z_0 \in \overline{D}_n$ $\frac{1}{N_t}$

$\sum_{i=0}^{N_t} H(\overline{G}^i z_0) - \alpha_n \sim h^2$ for sufficiently large N_t, i.e. according to definition 3 $\overline{D}_n$ is carrier of DQM eigenvalue α_n. And according to (25), the dynamics on the DQM trajectory outside of $\overline{D} = \bigcup_n \overline{D}_n$ and therefore for $\omega \neq 0$, there is motion along it with a constant velocity with an accuracy of the order of h^3 during the turnover time and of the order of h^2 during the time $T \sim \frac{1}{h}$ (including q = 3 according to [16]). That is, these trajectories remain motionless during time $T \sim \frac{1}{h}$ with an accuracy of the order of h^2. So the trajectories outside $\overline{D}$ with a distance between them in order of magnitude greater than h^2 will remain at a distance greater in order of h^2 during this time. For phase circles outside $\overline{D}$, (24) is reduced to the form (25) by a smooth change of variables different from the identity on only of order $O(\zeta^2) \sim h^2$[15]. So, for two trajectories outside $\overline{D}$ with a distance of an order greater than h^2 between them, the average values of the Hamiltonian on them differ by the same order h^2 over time T from Definition 3. So, outside the domains $\overline{D}_n$, carriers of the eigenvalues α_n, there are no subsets in which (7) is satisfied and which contain a ball of diameter of order h $\Rightarrow$ there are no DQM eigenvalues by Definition 3, QED.

3. Conclusion

Like the human brain, an artificial neural network is a complex nonlinear parallel processor; it is often called a neurocomputer. Accordingly, mathematical models of a neural network are usually continuous and stochastic, naturally associated with fuzzy logic.

Classical systems of artificial intelligence always involve the use of a symbolic language. Their mathematical models are naturally associated with classical logic and discrete mathematics. Thus, the representations and models of knowledge, undeniable at least since Aristotle, do not correspond to the cognitive models that are obtained as a result of studying the human brain.

In view of Niels Bohr, quantization is a phenomenon of a discrete and sequential process, that inherent in continuous, stochastic and parallel systems. However, this idea has not been embodied in a concrete model. The traditional mathematical model of quantum mechanics did not imply generalization to dissipative and stochastic systems.

The corresponding generalization, called the Dynamic quantum model (DQM), was proposed by author. It is defined for any dynamic system, given by ordinary differential equation or by diffeomorphism on any smooth Riemannian manifold, or for dynamic systems that using logical operations: algorithms, theorems, software applications. The neural network is exactly the DQM in the space of input signals.

In this paper DQM is defined and constructed universally for both Hamiltonian systems and systems with the fuzzy logic truth function on phase space. It occurs, that in the second case the point of the DQM spectrum is interpreted exactly as the average value of truth for approximate logical conclusions.

The paper goal is to demonstrate quantization on DQM, i.e. actually on neural networks, and to extend the classical Bohr-Sommerfeld condition to the general case, in particular, to systems with a fuzzy truth function.

4. References

[1] Ojha Varun Kumar, Ajith Abraham, Václav Snášel, "Metaheuristic design of feedforward neural networks: A review of two decades of research." Engineering Applications of Artificial Intelligence 60 (2017): 97–116. doi: 10.1016/j.engappai.2017.01.013.

[2] Guillaume Dubois, Modeling and Simulation, Challenges and Best Practices for Industry. 1st. ed., CRC Press, 2018.

[3] Ian Goodfellow, Yoshua Bengio, Aaron Courville, Deep Learning, MIT Press, 2016.
[4] Alexandra Nagy, "Variational Quantum Monte Carlo Method with a Neural-Network Ansatz for Open Quantum Systems." Physical Review Letters 122 (25) (2019). doi: 10.1103/PhysRevLett.122.250501.
[5] Alan Turing, "Computing Machinery and Intelligence. "Mind LIX(236) (1950): 433–460. https://doi.org/10.1093/mind/LIX.236.433.
[6] Stuart Russell, Peter Norvig, Artificial Intelligence: A Modern Approach, · 4th ed., Pearson, 2020.
[7] Helge Kragh, Niels Bohr and the quantum atom: the Bohr model of atomic structure 1913–1925, Oxford: Oxford University Press, 2012.
[8] Physicist Erwin Schrödinger's Google doodle marks quantum mechanics work, The Guardian, 2013. URL: https://www.theguardian.com/technology/2013/aug/12/erwin-schrodinger-google-doodle.
[9] E. Tesse, Mathematical Principals of Dynamic Systems and the Foundations of Quantum Physic, 2011. URL: https://www.e-booksdirectory.com/details.php?ebook=6083.
[10] A. Weissblut, "Non-Hamiltonian Quantum Mechanics and the Numerical Researches of the Attractor of a Dynamical System." Informational Technologies in Education 11 (2011): 73 – 78.
[11] A. Weissblut, The Computational Modeling: Dynamic Quantum Model Approach, in: ICT in Education, Research and Industrial Applications. Integration, Harmonization and Knowledge Transfer, Kharkiv, Ukraine, 2020, pp. 13 - 28.
[12] A. Weissblut, Computational Modeling and Structural Stability, in: ICT in Education, Research and Industrial Applications. Integration, Harmonization and Knowledge Transfer, Kherson, Ukraine, 2019, pp. 552–567.
[13] David D. Nolte, "Introduction to Modern Dynamics: Chaos, Networks, Space and Time." American Journal of Physics 83, 741 (2015). doi: 10.1119/1.4919244
[14] V. I. Arnold, Mathematical methods of classical mechanics, Springer-Verlag New York, 1978.
[15] A. Katok, B. Hasselblatt, Introduction to the Modern Theory of Dynamical Systems (Encyclopedia of Mathematics and its Applications), Cambridge: Cambridge University Press, 1995. doi:10.1017/CBO9780511809187.
[16] Gerald Teschl, Ordinary Differential Equations and Dynamical Systems, Graduate Studies in Mathematics, Amer. Math. Soc., Providence, 2012.