

STABILITY OF NEURAL ORDINARY DIFFERENTIAL EQUATIONS WITH POWER NONLINEARITIES

Vasily Ye. Belozyorov*, Danylo V. Dantsev†

Abstract. The article presents a study of solutions of ODEs system with a special nonlinear part, which is a continuous analogue of an arbitrary recurrent neural network (neural ODEs). As a nonlinear part of the mentioned system of differential equations, we used sums of piecewise continuous functions, where each term is a power function. (These are activation functions.) The use of power activation functions (PAF) in neural networks is a generalization of well-known the rectified linear units (ReLU). In the present time ReLU are commonly used to increase the depth of trained of a neural network. Therefore, the introduction of PAF into neural networks significantly expands the possibilities of ReLU. Note that the purpose of introducing power activation functions is that they allow one to obtain verifiable Lyapunov stability conditions for solutions of the system differential equations simulating the corresponding dynamic processes. In turn, Lyapunov stability is one of the guarantees of the adequacy of the neural network model for the process under study. In addition, from the global stability (or at least the boundedness) of continuous analog solutions it follows that learning process of the corresponding neural network will not diverge for any training sample.

Key words: ordinary autonomous differential equations, power activation function, neural network, Lyapunov stability.

2010 Mathematics Subject Classification: 13A50, 14L24, 15A72, 93C05.

Communicated by Prof. A. V. Plotnikov

1. Introduction

One of the fundamental goals of machine learning is modeling and understanding real-world phenomena from observations. In the process of modeling two main problems are considered: (i) the problem of minimizing deviations between the trajectory of the described model and the real dynamics that the system under study demonstrates; (ii) the problem of increasing the degree of convergence of the procedure for minimizing deviations, which is solved in the problem (i). However, there is also a third problem: (iii) if a neural network models a certain dynamic process, then how to guarantee the stability or boundedness of solutions of a system of differential equations describing a continuous analog of the aforementioned neural network?

The solution of all three problems substantially depends on the choice of activation functions included in the architecture of a neural network. In this work

*Department of Applied Mathematics, Oles Honchar Dnipro National University, 72, Gagarin's Avenue, 49010, Dnipro, Ukraine, belozvye2017@gmail.com

†Chaotic Dynamics Inc., 201-45793, Luckakuck Way, Chilliwack BC V2R 5S3, Canada, danylo@chaodyna.com

we will primarily study the problem (iii) with respect to the choice of activation functions in the domain of modeling dynamical systems.

A wide variety of naturally occurring phenomena can be approximated by power laws: from frequency distribution of words in natural language [11] to the size distribution of neuronal avalanches in cortical networks [8, 9, 15]. Moreover, it is common for chaotic processes to contain unbounded nonlinearities while describing a bounded and stable process [1, 3, 4, 10, 15]. Therefore, it is only natural to question the fit and expressivity of power functions as activation functions. In order to solve the numerous problems associated with the stability of neural networks, it is necessary to expand the class of activation functions. At present, this set already includes unbounded activation functions [13]. In this regard, the creation of neural networks with unbounded activation functions requires guaranteeing the stability or boundedness conditions of solutions of the equations describing the neural networks [6] – [9].

The key contributions of this article are:

- Section 3 introduces the concept of a piecewise continuous power activation function (PAF) and presents its main properties;
- The conditions for the global stability of neural ODEs with PAFs, found with the help of the well-known method of Lyapunov functions, are given in Sections 4 and 5. Note that checking these conditions is reduced only to checking the negative definiteness of two known matrices.

In the future, these results will be used to construct new types of chaotic neural networks.

1.1. The relationship between neural networks and differential equations

In recent years, an interesting idea has appeared to interpret a system of ordinary differential equations in the form of a suitable neural network (residual network) [3] – [6]. The essence of this idea is as follows.

Consider the following neural network (this is a system of difference equations):

$$\mathbf{x}(t+1) = \mathbf{x}(t) + \mathbf{H}(\mathbf{x}(t), \mathbf{\Omega}), \mathbf{x}(0) = \mathbf{x}_0; t = 1, \dots, N. \quad (1.1)$$

Here $\mathbf{x} \in \mathbb{R}^n$ is a vector of states, $\mathbf{\Omega} \in \mathbb{R}^k$ is a vector of parameters, $\mathbf{H}(\mathbf{x}, \mathbf{\Omega}) : \mathbb{R}^n \times \mathbb{R}^k \rightarrow \mathbb{R}^n$ is a vector field of continuous functions. (The number N in neurodynamics denotes the number of "layers" in the neural network (1.1).)

Now we rewrite relation (1.1) in the following form:

$$\frac{\mathbf{x}(t+1) - \mathbf{x}(t)}{(t+1) - t} = \mathbf{H}(\mathbf{x}(t), \mathbf{\Omega}).$$

If we consider function $\mathbf{x}(t)$ as a function of a continuous argument on some interval $[\mathbf{x}_0, \mathbf{x}_N]$, then the last equation can be rewritten in the following form:

$$\frac{\mathbf{x}(t + \Delta t) - \mathbf{x}(t)}{\Delta t} = \mathbf{H}(\mathbf{x}(t), \mathbf{\Omega}).$$

If now we direct the number of "layers" $N \rightarrow \infty$ and we assume $\Delta t \rightarrow 0$, then we get the following system of ordinary differential equations

$$\dot{\mathbf{x}}(t) = \mathbf{H}(\mathbf{x}(t), \mathbf{\Omega}), \mathbf{x}(0) = \mathbf{x}_0, \quad (1.2)$$

So we can say that neural network (1.1) is the well-known Euler discretization procedure of system (1.2):

$$\mathbf{x}(t + \Delta t) - \mathbf{x}(t) = \Delta t \cdot (\mathbf{H}(\mathbf{x}(t), \mathbf{\Omega})), \mathbf{x}(0) = \mathbf{x}_0, \quad (1.3)$$

where Δt is the discretization step.

It is clear that sequence (1.1) can be viewed as a neural network with $N - 1$ hidden layers, input layer \mathbf{x}_0 and output layer \mathbf{x}_N . The architecture of such neural network is determined by the vector $\mathbf{H}(\mathbf{x}(t), \mathbf{\Omega})$ [4, 5].

Thus, in some cases, we can replace the study of neural network (1.1) with its continuous analog (1.2). In this case, the stability of neural network (1.1) at $N \rightarrow \infty$ will follow from the stability of the system of differential equations (1.2). (If it is known that the process described by equation (1.2) converges, then by increasing the number of layers N of neural network (1.1) or decreasing step Δt in equation (1.3), we can achieve convergence of the learning process of this neural network [4, 5].)

Therefore, in the future, this article will be devoted to the study of the stability of system (1.2), which in neurodynamics is called the system of neural ODEs [4, 5]. (Note that a new application of neural ODEs using invariant theory was proposed in [2].)

2. Mathematical preliminaries

Now we give three well-known results of the nonlinear theory of systems, which will be used later.

Let

$$\dot{\mathbf{x}}(t) = \mathbf{G}(\mathbf{x}(t)), \mathbf{x}(t) \in \mathbb{R}^n, \mathbf{G}(\mathbf{x}) \in \mathbb{R}^n, t \in \mathbb{R} \quad (2.1)$$

be a system of ordinary autonomous differential equations and let $\mathbf{x}(t, \mathbf{x}_0)$ be a trajectory of this system with initial data $\mathbf{x}_0 \in \mathbb{R}^n$. Here $\mathbf{G}(\mathbf{x}) : \mathbb{R}^n \rightarrow \mathbb{R}^n$ is a continuous vector-function; $\mathbf{x}(0, \mathbf{x}_0) = \mathbf{x}_0$.

In the future, we will assume that point $\mathbf{x} = 0$ is the equilibrium point of system (2.1).

Definition 2.1. [9] The equilibrium point $\mathbf{x} = 0$ of system (2.1) is stable if, for each $\epsilon > 0$, there is $\delta = \delta(\epsilon) > 0$ such that $\|\mathbf{x}_0\| < \delta \Rightarrow \|\mathbf{x}(t)\| < \epsilon, \forall t \geq 0$; asymptotically stable if it is stable and δ can be chosen such that $\|\mathbf{x}_0\| < \delta \Rightarrow \lim_{t \rightarrow \infty} \mathbf{x}(t) = 0$.

Theorem 2.1. (Lyapunov's Theorem [9]). Let $\mathbf{x} = 0$ be an equilibrium point for system (2.1) and $\mathbb{D} \subset \mathbb{R}^n$ be a domain containing $\mathbf{x} = 0$. Let $V : \mathbb{R}^n \rightarrow \mathbb{R}$ be a

continuously differentiable function, such that $V(0) = 0$, $V(\mathbf{x}) > 0$ in $\mathbb{D} - \{0\}$, and $\dot{V}(\mathbf{x}) \leq 0$ in \mathbb{D} . Then, $\mathbf{x} = 0$ is stable. Moreover, if $\dot{V}(\mathbf{x}) < 0$ in $\mathbb{D} - 0$, then the equilibrium point $\mathbf{x} = 0$ is asymptotically stable and if it is unique, then $\mathbf{x} = 0$ is globally asymptotically stable.

Definition 2.2. [9] A continuously differentiable function $V(\mathbf{x})$ satisfying the conditions Lyapunov's Theorem is called a Lyapunov function.

A set $\mathbb{M} \subset \mathbb{R}^n$ is said to be a positively invariant set with respect to (2.1) if from $\mathbf{x}_0 \in \mathbb{M}$ it follows that $\mathbf{x}(t, \mathbf{x}_0) \in \mathbb{M} \forall t \geq 0$.

A point $\mathbf{s} \in \mathbb{R}^n$ is said to be a positive limit point of $\mathbf{x}(t, \mathbf{x}_0)$ if there is a sequence $\{t_m\}$, with $t_m \rightarrow \infty$ as $m \rightarrow \infty$, such that $\mathbf{x}(t_m, \mathbf{x}_0) \rightarrow \mathbf{s}$ as $m \rightarrow \infty$. The set \mathbb{L}^+ of all positive limit points of $\mathbf{x}(t, \mathbf{x}_0)$ is called the positive limit set of $\mathbf{x}(t, \mathbf{x}_0)$.

Let $\mathbb{D} \subset \mathbb{R}^n$ be a compact set and let solution $\mathbf{x}(t, \mathbf{x}_0)$ is bounded and belongs to \mathbb{D} . Then its positive limit set \mathbb{L}^+ is a nonempty, compact, invariant set. Moreover, $\mathbf{x}(t, \mathbf{x}_0) \rightarrow \mathbb{L}^+$ as $t \rightarrow \infty$ (see Lemma 3.1 [9]).

From here the following result can be obtained.

Theorem 2.2. (LaSalle's Theorem [9]). Let $\mathbb{H} \subset \mathbb{R}^n$ be a compact set that is positively invariant with respect to (2.1). Let $V : \mathbb{R}^n \rightarrow \mathbb{R}$ be a continuously differentiable function such that $\dot{V}(\mathbf{x}) \leq 0$ in \mathbb{H} . Let \mathbb{E} be the set of all points in \mathbb{H} where $\dot{V}(\mathbf{x}) = 0$. Let \mathbb{M} be the largest invariant set in \mathbb{E} . Then every solution starting in \mathbb{H} approaches \mathbb{M} as $t \rightarrow +\infty$.

In the future, one more important result will be used.

Theorem 2.3. (Comparison Principle [9]). Consider the real scalar differential equation $\dot{v}_t(t) = g(t, v)$, $v(t_0) = v_0$, where the function $g(t, v)$ is continuous in t and locally Lipschitz in v , for all $t \geq 0$ and all $v(t) \in \mathbb{R}$. Let $[t_0, \infty)$ be the interval of existence of the solution $v(t)$. Let $w(t)$ be a differentiable function satisfies the differential inequality $\dot{w}_t(t) \leq g(t, w(t))$, $w(t_0) < v_0$ with $w(t) \in \mathbb{R}$ for all $t \in [t_0, \infty)$. Then, $w(t) < v(t)$ for all $t \in [t_0, \infty)$.

Several special lemmas will also be used in the proof of the main results.

Lemma 2.1. Suppose that real numbers $A, B, \alpha, \beta, p, q, x$, and y satisfy the following conditions: $A > 0$, $B > 0$, $\alpha \geq 1$, $\beta \geq 1$, $0 \leq p \leq A^{\frac{\alpha}{\alpha+\beta}} B^{\frac{\beta}{\alpha+\beta}}$, $0 \leq q \leq A^{\frac{\beta}{\alpha+\beta}} B^{\frac{\alpha}{\alpha+\beta}}$, $x \geq 0$, $y \geq 0$. Then inequality

$$Ax^{\alpha+\beta} + By^{\alpha+\beta} \geq px^{\alpha}y^{\beta} + qy^{\alpha}x^{\beta} \quad (2.2)$$

is true. In addition, if $p = 0$ (or $q = 0$), then inequality (2.2) becomes inequality $Ax^{\alpha+\beta} + By^{\alpha+\beta} \geq 2A^{\frac{\beta}{\alpha+\beta}} B^{\frac{\alpha}{\alpha+\beta}} y^{\alpha}x^{\beta}$ (or $Ax^{\alpha+\beta} + By^{\alpha+\beta} \geq 2A^{\frac{\alpha}{\alpha+\beta}} B^{\frac{\beta}{\alpha+\beta}} x^{\alpha}y^{\beta}$).

Proof. 1. Let $A = B = 1$, $x \neq 0, y \neq 0$, and $p = A^{\frac{\alpha}{\alpha+\beta}} B^{\frac{\beta}{\alpha+\beta}} = 1$, $q = A^{\frac{\beta}{\alpha+\beta}} B^{\frac{\alpha}{\alpha+\beta}} = 1$. Introduce the variable $u = x/y > 0$. Then inequality (2.2) can be rewritten in the following form: $f(u) = u^{\alpha+\beta} + 1 - u^\alpha - u^\beta \geq 0$.

It is easy to check that the function $f(u)$ has only one positive root: $u = 1$. Let's do some simple calculations:

$$\begin{aligned} f'(u) &= (\alpha + \beta)u^{\alpha+\beta-1} - \alpha u^{\alpha-1} - \beta u^{\beta-1}, \\ f''(u) &= (\alpha + \beta)(\alpha + \beta - 1)u^{\alpha+\beta-2} - \alpha(\alpha - 1)u^{\alpha-2} - \beta(\beta - 1)u^{\beta-2}, \\ f(1) &= f'(1) = 0, f''(1) = 2\alpha\beta > 0. \end{aligned}$$

From here it follows that the point $u = 1$ is the single minimum of function $f(u)$ on the interval $(0, \infty)$. The last statement leads to inequality $u^{\alpha+\beta} + 1 \geq u^\alpha + u^\beta \geq pu^\alpha + qu^\beta$ or

$$x^{\alpha+\beta} + y^{\alpha+\beta} \geq p_1 x^\alpha y^\beta + q_1 y^\alpha x^\beta, \quad (2.3)$$

where $0 \leq p_1 \leq 1, 0 \leq q_1 \leq 1$.

2. Now, instead of variables x and y in inequality (2.2), we introduce new variables $x_1^{\alpha+\beta} = Ax^{\alpha+\beta}$ and $y_1^{\alpha+\beta} = By^{\alpha+\beta}$. Then inequality (2.2) transforms into inequality

$$x_1^{\alpha+\beta} + y_1^{\alpha+\beta} \geq \frac{p}{A^{\frac{\alpha}{\alpha+\beta}} B^{\frac{\beta}{\alpha+\beta}}} x_1^\alpha y_1^\beta + \frac{q}{A^{\frac{\beta}{\alpha+\beta}} B^{\frac{\alpha}{\alpha+\beta}}} y_1^\alpha x_1^\beta,$$

which is equivalent to inequality (2.3) at $p \leq A^{\frac{\alpha}{\alpha+\beta}} B^{\frac{\beta}{\alpha+\beta}}$ and $q \leq A^{\frac{\beta}{\alpha+\beta}} B^{\frac{\alpha}{\alpha+\beta}}$. \square

Lemma 2.2. *Suppose that real numbers $A, B, k_1, k_2, m_1, m_2, p, q, x$, and y satisfy the following conditions: $A > 0, B > 0, k_1 \geq 1, k_2 \geq 1, m_1 \geq 1, m_2 \geq 1, k_1 + k_2 = r, m_1 + m_2 = r, 0 \leq p \leq A^{\frac{k_1}{r}} B^{\frac{k_2}{r}}, 0 \leq q \leq A^{\frac{m_1}{r}} B^{\frac{m_2}{r}}, x \geq 0, y \geq 0$. Then inequality*

$$Ax^r + By^r \geq px^{k_1} y^{k_2} + qx^{m_1} y^{m_2} \quad (2.4)$$

is true.

Proof. Let $A = B = 1$, $x \neq 0, y \neq 0$, and $p = A^{\frac{k_1}{r}} B^{\frac{k_2}{r}} = 1$, $q = A^{\frac{m_1}{r}} B^{\frac{m_2}{r}} = 1$. Then inequality (2.4) takes the form

$$x^r + y^r \geq x^{k_1} y^{k_2} + x^{m_1} y^{m_2} \quad (2.5)$$

1. Let $u = x/y \geq 1$. Divide both sides of inequality (2.5) by a positive value y^r . Then, we have

$$\left(\frac{x}{y}\right)^r + 1 \geq \left(\frac{x}{y}\right)^{k_1} + \left(\frac{x}{y}\right)^{m_1}$$

or

$$u^{k_1}(u^{r-k_1} - 1) \geq (u^{m_1} - 1). \quad (2.6)$$

We will consider that $r \geq k_1 + m_1$. Then $u^{r-k_1} - 1 \geq u^{m_1} - 1$ and inequality (2.6) is obvious.

Let $u = x/y \leq 1$. Then, we have

$$u^r + 1 \geq u^{k_1} + u^{m_1} \geq 2u^{k_1/2}u^{m_1/2} = 2u^{k_1/2+m_1/2}.$$

Since $u \leq 1$ and $r \geq k_1 + m_1$, then $u^{k_1/2+m_1/2} \geq u^{r/2}$. Therefore, the inequality $u^r + 1 \geq 2u^{r/2}$ and inequality (2.5) is true.

2. Now let $u = y/x \leq 1$. Divide both sides of inequality (2.5) by a positive value x^r . Then, we have

$$1 + \left(\frac{y}{x}\right)^r \geq \left(\frac{y}{x}\right)^{k_2} + \left(\frac{y}{x}\right)^{m_2}$$

or

$$1 - u^{m_2} \geq u^{k_2}(1 - u^{r-k_2}). \quad (2.7)$$

It is easy to check that if $r \geq k_1 + m_1$, then $r \leq k_2 + m_2$. Since $u = y/x \leq 1$, then $1 - u^{m_2} \geq 1 - u^{r-k_2}$ and inequality (2.7) is obvious. The proof of case $u = y/x \geq 1$ is similar to the proof given in item 1.

3. Let $0 \leq p_1 \leq 1$, $0 \leq q_1 \leq 1$. Then inequality (2.5) implies inequality

$$x^r + y^r \geq p_1 x^{k_1} y^{k_2} + q_1 x^{m_1} y^{m_2}. \quad (2.8)$$

Now, instead of variables x and y in inequality (2.4), we introduce new variables $x_1^r = Ax^r$ and $y_1^r = By^r$. Then inequality (2.8) transforms into inequality

$$x_1^r + y_1^r \geq \frac{p}{A^{\frac{k_1}{r}} B^{\frac{k_2}{r}}} x_1^{k_1} y_1^{k_2} + \frac{q}{A^{\frac{m_1}{r}} B^{\frac{m_2}{r}}} x_1^{m_1} y_1^{m_2},$$

which is equivalent to inequality (2.4) at $p \leq A^{\frac{k_1}{r}} B^{\frac{k_2}{r}}$ and $q \leq A^{\frac{m_1}{r}} B^{\frac{m_2}{r}}$. \square

Lemma 2.1 admits the following generalization.

Lemma 2.3. (Generalization of Lemma 2.1.) Let the real numbers $\gamma_1 \geq 1, \dots, \gamma_n \geq 1$, and $x_1 \geq 0, \dots, x_n \geq 0$. Put $r = \gamma_1 + \dots + \gamma_n$. Then for natural numbers i and j such that $1 \leq i < j \leq n$ the following inequality

$$\frac{r + \gamma_j - \gamma_i}{r} x_i^r + \frac{r + \gamma_i - \gamma_j}{r} x_j^r \geq x_i^{r-\gamma_i} x_j^{\gamma_i} + x_i^{\gamma_j} x_j^{r-\gamma_j} \quad (2.9)$$

is true.

Proof. For simplicity we put $n = 3$, $i = 1$, and $j = 2$. Then inequality (2.9) takes the form

$$\frac{r + \gamma_2 - \gamma_1}{r} x_1^r + \frac{r + \gamma_1 - \gamma_2}{r} x_2^r - x_1^{r-\gamma_1} x_2^{\gamma_1} - x_1^{\gamma_2} x_2^{r-\gamma_2} \geq 0.$$

Let $u = x_1/x_2$. Then the left-hand side of the last inequality can be represented as the following function

$$f(u) = \frac{2\gamma_2 + \gamma_3}{\gamma_1 + \gamma_2 + \gamma_3} u^r + \frac{2\gamma_1 + \gamma_3}{\gamma_1 + \gamma_2 + \gamma_3} - u^{\gamma_2+\gamma_3} - u^{\gamma_2}.$$

Then from here it follows that

$$\begin{aligned} f'(u) &= (2\gamma_2 + \gamma_3)u^{\gamma_1 + \gamma_2 + \gamma_3 - 1} - (\gamma_2 + \gamma_3)u^{\gamma_2 + \gamma_3 - 1} - \gamma_2 u^{\gamma_2 - 1}, \\ f''(u) &= (2\gamma_2 + \gamma_3)(\gamma_1 + \gamma_2 + \gamma_3 - 1)u^{\gamma_1 + \gamma_2 + \gamma_3 - 2} \\ &\quad - (\gamma_2 + \gamma_3)(\gamma_2 + \gamma_3 - 1)u^{\gamma_2 + \gamma_3 - 2} - \gamma_2(\gamma_2 - 1)u^{\gamma_2 - 2}. \end{aligned}$$

Thus, if $x_1 = x_2$, then $f(1) = f'(1) = 0$ and $f''(1) = 2\gamma_1\gamma_2 + (\gamma_1 + \gamma_2)\gamma_3 > 0$. Let $u \neq 0$. Then we can consider the equation

$$f'(u)u^{1-\gamma_2} = p(u) - q(u) = (2\gamma_2 + \gamma_3)u^{\gamma_1 + \gamma_3} - (\gamma_2 + \gamma_3)u^{\gamma_3} - \gamma_2 = 0, \quad (2.10)$$

where $p(u) = (2\gamma_2 + \gamma_3)u^{\gamma_1 + \gamma_3}$, $q(u) = (\gamma_2 + \gamma_3)u^{\gamma_3} + \gamma_2 u^{\gamma_2}$.

Introduce in (2.10) a new variable $v = u^{\gamma_3} \geq 0$. Then we have $p(v) = q(v)$, where $p(v) = (2\gamma_2 + \gamma_3)v^{1+\gamma_1/\gamma_3}$, $\gamma_1/\gamma_3 > 0$ is a concave power function and $q(v) = (\gamma_2 + \gamma_3)v + \gamma_2$, $\gamma_2 > 0$ is a linear function.

Since $p(0) = 0$ and $q(0) = \gamma_2 > 0$, then $v = 1$ is only one positive solution of the equation $p(v) = q(v)$. This means that $u^{\gamma_3} = 1$ and the point $u = 1$ is the single minimum of function $f(u)$ on the interval $(0, \infty)$. The inequality (2.9) for $i = 1, j = 2$ is proved. The proof of the remaining inequalities for $1 \leq i < j \leq n$ is similar. \square

3. New activation functions.

Let

$$\begin{aligned} \alpha &= \frac{2m_r + 1}{2n_r + 1} > 0 \text{ or } \frac{2m_r}{2n_r + 1} > 0 \text{ or } \frac{2m_r + 1}{2n_r} > 0 \\ \text{and } \beta &= \frac{2m_l + 1}{2n_l + 1} > 0 \text{ or } \frac{2m_l}{2n_l + 1} > 0 \text{ or } \frac{2m_l + 1}{2n_l} > 0 \end{aligned}$$

be rational irreducible fractions. (Here m_r, n_r, m_l, n_l are nonnegative integers.)

Consider one of the piecewise continuous real functions:

either

$$f(x, \alpha \vee \beta) = \begin{cases} x^\beta \equiv \sqrt[2n_l+1]{x^{2m_l+1}} < 0, \text{ if } x < 0; \\ x^\alpha \equiv \sqrt[2n_r+1]{x^{2m_r+1}} \geq 0, \text{ if } x \geq 0 \end{cases} \quad (3.1)$$

or

$$\begin{aligned} &f(x, \alpha \vee \beta) \\ &= \begin{cases} x^\beta \equiv \sqrt[2n_l+1]{x^{2m_l}} > 0, \text{ if } x < 0; \\ x^\alpha \equiv \sqrt[2n_r+1]{x^{2m_r+1}} \geq 0 \text{ or } \sqrt[2n_r]{x^{2m_r+1}} \geq 0 \text{ or } \sqrt[2n_r+1]{x^{2m_r}} \geq 0, \text{ if } x \geq 0. \end{cases} \end{aligned} \quad (3.2)$$

Let $r = m/n > 0$ be an irreducible rational fraction; m and n are positive integers.

Definition 3.1. The fraction r is called odd if the numbers m and n are odd. (Otherwise, the fraction r is called even.) The function $f(x, \alpha \vee \beta)$ is called an strictly odd activation function if α and β are odd. (If at least one of the numbers α or β is even, then the function $f(x, \alpha \vee \beta)$ is called an strictly even activation function.)

From an applied point of view, the representation function $f(x, \alpha \vee \beta)$ in the form (3.1) is not very convenient. In order to make this representation more applied, we prove the following lemma.

Lemma 3.1. *The fraction r can be approximated with any preassigned accuracy ϵ by either an odd or even fraction.*

Proof. Let us introduce the rational fraction $r_1 = r_1(k) = (2^k m + 1)/(2^k n + 1)$, where k is a positive integer. Then, we have

$$\lim_{k \rightarrow \infty} |r - r_1| = \lim_{k \rightarrow \infty} \left| \frac{m}{n} - \frac{2^k m + 1}{2^k n + 1} \right| = \lim_{k \rightarrow \infty} \frac{|m - n|}{n(2^k n + 1)} = 0.$$

From here it follows that

$$\frac{|m - n|}{n(2^k n + 1)} < \frac{|m - n|}{n 2^k} < \epsilon \text{ and } \frac{|m - n|}{\epsilon n^2} < 2^k.$$

Thus, we have

$$k = k(\epsilon) > \log_2 \frac{|m - n|}{\epsilon n^2} \text{ and } r \approx r_1(k).$$

Now let either $r_1 = r_1(k) = (2^k m)/(2^k n + 1)$ or $r_1 = r_1(k) = (2^k m + 1)/(2^k n)$. Then, in the case $r_1(k) = (2^k m)/(2^k n + 1)$, we have

$$\lim_{k \rightarrow \infty} |r - r_1| = \lim_{k \rightarrow \infty} \left| \frac{m}{n} - \frac{2^k m}{2^k n + 1} \right| = \lim_{k \rightarrow \infty} \frac{m}{n(2^k n + 1)} = 0.$$

From here it follows that

$$\frac{m}{n(2^k n + 1)} < \frac{m}{n 2^k} < \epsilon \text{ and } \frac{m}{\epsilon n^2} < 2^k.$$

Thus, we again have

$$k = k(\epsilon) > \log_2 \frac{m}{\epsilon n^2} \text{ and } r \approx r_1(k).$$

Further reasoning is obvious. This completes the proof of Lemma 3.1. \square

Thus, any irreducible rational even fraction with any preassigned accuracy can be represented by a rational odd fraction and vice versa.

Since any finite or periodic decimal fraction can be represented as a rational fraction, Lemma 1 allows us to consider odd and even activation functions $f(x, \alpha \vee \beta)$ with powers in the form of decimal fractions.

Thus, in computer simulation, instead of (3.1), we can use the following representation of the activation function:

$$f(x, \alpha \vee \beta) = \begin{cases} -(-x)^\beta & \text{if } (x < 0 \text{ and } \beta > 0); 0 & \text{if } (x < 0 \text{ and } \beta = 0) \\ x^\alpha & \text{if } (x \geq 0 \text{ and } \alpha > 0); 0 & \text{if } (x \geq 0 \text{ and } \alpha = 0), \end{cases} \quad (3.3)$$

where $\alpha \geq 0, \beta \geq 0$ are finite decimal fractions.

Similarly, instead (3.2), we can use the following representation of the activation function:

$$f(x, \alpha \vee \beta) = \begin{cases} (-x)^\beta & \text{if } (x < 0 \text{ and } \beta > 0); 0 & \text{if } (x < 0 \text{ and } \beta = 0) \\ x^\alpha & \text{if } (x \geq 0 \text{ and } \alpha > 0); 0 & \text{if } (x \geq 0 \text{ and } \alpha = 0), \end{cases} \quad (3.4)$$

where $\alpha \geq 0, \beta \geq 0$ are finite decimal fractions.

It is clear that representations (3.3) and (3.4) generalize representations (3.1) and (3.2). Therefore, the following definition is justified.

Definition 3.2. Representation (3.3) ((3.4)) is called an odd (even) activation function.

The last remark allows us to use the tools of well-known mathematical packages to calculate power functions. Below, these packages were used to plot several activation functions. (It should be noted that these graphs are substantially differ from each other at $0 < \alpha < 1, 0 < \beta < 1$ and $\alpha > 1, \beta > 1$. This difference is the main reason for different approaches when proving stability in case $0 < \alpha < 1, 0 < \beta < 1$ and case $\alpha > 1, \beta > 1$.)

Important note. Functions (3.3) and (3.4) may have properties that functions (3.1) and (3.2) do not have. For example, $f(x) = x^2$ is the even function, while the function $f(x) = x^{201/99}$ close to it is odd. The classic view of whether a function $f(x) = x^p$ is even or odd is that the number p must be a positive integer. We have tried to extend these concepts to arbitrary (not integer) positive degrees. In our work, evenness or oddness is defined as follows: any power function $f(x) = x^p$ at $x \geq 0$ is nonnegative. But at $x < 0$, function $f(x) = x^p$ is either positive (even) or negative (odd). A natural question arises: why are such functions the subject of research in this article? In our opinion, the most comprehensive answer is given in [13]. It was said in [13] that ReLU [14] – [16] became a new building block of deep neural networks instead of traditional bounded activation functions. Therefore, we hope that generalization (3.1) (or (3.3)) and (3.2) (or (3.4)) will lead to deeper results than those obtained using ReLU.

3.1. Even and odd activation functions

The concepts of even and odd activation functions introduced above can be extended as follows.

Let $g(x)$ and $h(x)$ be two nonnegative increasing functions continuously differentiable on the interval (r, ∞) . Suppose also that each of these functions has a single root $x^* = r$ ($g(r) = h(r) = 0$) on the interval $[r, \infty)$.

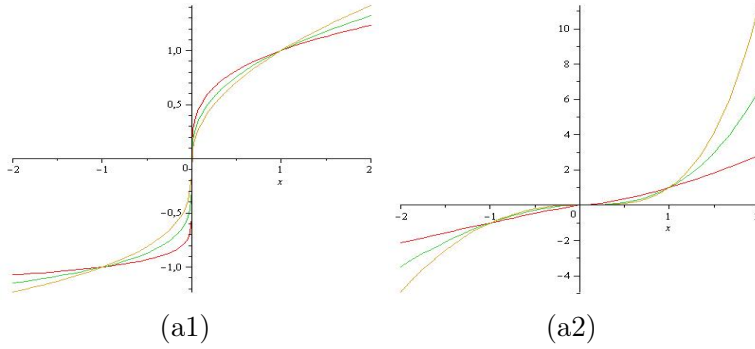


Fig. 3.1. The odd activation functions: (a1) $f_1(x) = \text{piecewise}(x < 0, -(-x)^{1/11}, x \geq 0, x^{3/11})$, $f_2(x) = \text{piecewise}(x < 0, -(-x)^{1/5}, x \geq 0, x^{2/5})$, $f_3(x) = \text{piecewise}(x < 0, -(-x)^{3/11}, x \geq 0, x^{5/11})$; (a2) $f_1(x) = \text{piecewise}(x < 0, -(-x)^{1.1}, x \geq 0, x^{1.5})$, $f_2(x) = \text{piecewise}(x < 0, -(-x)^{9/5}, x \geq 0, x^{23/10})$, $f_3(x) = \text{piecewise}(x < 0, -(-x)^{23/9}, x \geq 0, x^{7/2})$

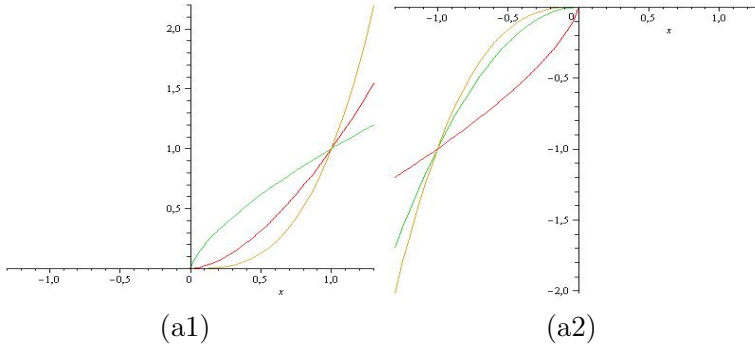


Fig. 3.2. The semi-even (a1) and semi-odd (a2) activation functions: (a1) $f_1(x) = \text{piecewise}(x < 0, 0, x \geq 0, x^{1.67})$, $f_2(x) = \text{piecewise}(x < 0, 0, x \geq 0, x^{0.7})$, $f_3(x) = \text{piecewise}(x < 0, 0, x \geq 0, x^3)$; (a2) $f_1(x) = \text{piecewise}(x < 0, -(-x)^{7/3}, x \geq 0, 0)$, $f_2(x) = \text{piecewise}(x < 0, -(-x)^3, x \geq 0, 0)$, $f_3(x) = \text{piecewise}(x < 0, -(-x)^{0.7}, x \geq 0, 0)$

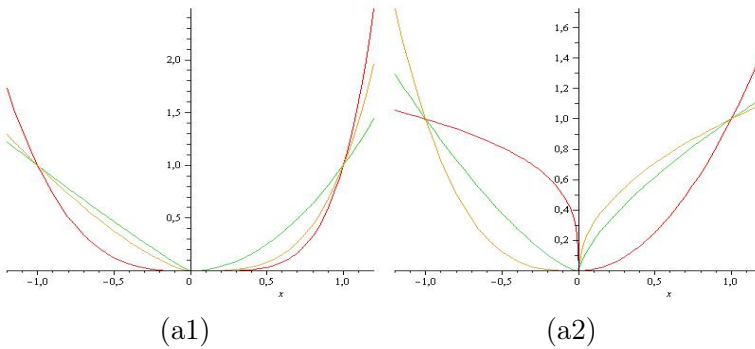


Fig. 3.3. The even activation functions: (a1) $f_1(x) = \text{piecewise}(x < 0, (-x)^{1.1}, x^2)$, $f_2(x) = \text{piecewise}(x < 0, (-x)^{1.4}, x^{3.7})$, $f_3(x) = \text{piecewise}(x < 0, (-x)^3, x^5)$; (a2) $f_1(x) = \text{piecewise}(x < 0, (-x)^{0.3}, x^2)$, $f_2(x) = \text{piecewise}(x < 0, (-x)^{1.4}, x^{0.7})$, $f_3(x) = \text{piecewise}(x < 0, (-x)^3, x^{0.5})$

Consider the following piecewise continuous functions on an open interval $(-\infty, \infty)$:

$$f_o(x) = -h(-x) \text{ if } x < r \text{ and } g(x) \text{ if } x \geq r \quad (3.5)$$

and

$$f_e(x) = h(-x) \text{ if } x < r \text{ and } g(x) \text{ if } x \geq r. \quad (3.6)$$

Definition 3.3. Representation (3.5) ((3.6)) is called an extended odd (extended even) activation function.

Let the root $x^* = r$ be the same for all odd and even functions. Then it is easy to check the following properties of these functions:

(i) the product of two odd or even functions is even.

(ii) the product of even and odd functions is odd.

(iii) the derivative of an even function is odd and vice versa. (Indeed, let $f_o(x)$ be odd. Then $\dot{g}(x) \geq 0, \dot{h}(x) \geq 0$ if $x \geq r$. Hence $\dot{f}_o(x) = -\dot{h}(-x)(-1) = \dot{h}(-x) > 0$ if $x < r$ and $\dot{g}(x) \geq 0$ if $x \geq r$. Thus, the function $\dot{f}_o(x)$ is even.)

In what follows, we will assume that $r = 0$ and $g(x) = ax^\alpha, h(x) = bx^\beta$ are power activation functions. Here $a > 0, b > 0, \alpha > 0$, and $\beta > 0$.

Definition 3.4. Let $\alpha = \beta$ and $a = b$ for the power function $f_o(x)$ or $f_e(x)$. Then the function $f_o(x)$ is called odd antisymmetric, and the function $f_e(x)$ is called even symmetric.

In the future, systems $\dot{\mathbf{x}}(t) = \mathbf{H}(\mathbf{x})$ ($\mathbf{x} \in \mathbb{R}^n, \mathbf{H} : \mathbb{R}^n \rightarrow \mathbb{R}^n$) of neural ODEs of order n of the following two types will be considered.

(T1) Each of the n equations of system $\dot{\mathbf{x}}(t) = \mathbf{H}(\mathbf{x})$ includes no more than n power functions. These functions are the same for each equation of the system.

(T2) Each of the n equations of system $\dot{\mathbf{x}}(t) = \mathbf{H}(\mathbf{x})$ includes only one power function. Moreover, all the functions mentioned are, generally speaking, different.

Thus, in both cases (T1) and (T2), the system $\dot{\mathbf{x}}(t) = \mathbf{H}(\mathbf{x})$ contains no more than n different power functions.

The sense of introducing such types of ODE is as follows.

All neurons of a neural network located in one layer are divided into n groups, each of which contains no more than n neurons.

In the case of (T1), the signals entering the neurons of each group are transformed according to different power laws. At the same time, these laws are the same for different groups.

In the case of (T2), the signals entering the neurons of each group are transformed only according to one power law. At the same time, these laws are, generally speaking, different for different groups.

In the future, two different algorithms for learning neural networks will be designed. One of them will be based on property (T1) and the other on property (T2). Here it is important to find out for which of the algorithms the learning problem is solved more efficiently. (For example, for which of the algorithms a higher rate of convergence can be achieved.)

The rest of the article will be devoted to the study of the stability problem.

4. First use of power functions in neural ODEs: Architecture (T1)

Consider the following system of real differential equations

$$\left\{ \begin{array}{l} \dot{x}_1(t) = c_1 + a_{11}x_1(t) + a_{12}x_2(t) + \dots + a_{1n}x_n(t) \\ \quad + b_{11} \text{piecewise}(x_1(t) < 0, -(-x_1(t))^{\beta_1}, x_1(t)^{\alpha_1}) \\ \quad \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \\ \quad + b_{1n} \text{piecewise}(x_n(t) < 0, -(-x_n(t))^{\beta_n}, x_n(t)^{\alpha_n}), \\ \quad \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \\ \dot{x}_n(t) = c_n + a_{n1}x_1(t) + a_{n2}x_2(t) + \dots + a_{nn}x_n(t) \\ \quad + b_{n1} \text{piecewise}(x_1(t) < 0, -(-x_1(t))^{\beta_1}, x_1(t)^{\alpha_1}) \\ \quad \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \\ \quad + b_{nn} \text{piecewise}(x_n(t) < 0, -(-x_n(t))^{\beta_n}, x_n(t)^{\alpha_n}) \end{array} \right. \quad (4.1)$$

with initial values $x_i(0) = x_{i0}$. Here $\alpha_i \geq 0, \beta_i \geq 0; c_i, a_{ij}, b_{ij} \in \mathbb{R}; i, j = 1, \dots, n$.

Introduce the following designations : $f(x_i, \alpha_i \vee \beta_i) = \text{piecewise}(x_i(t) < 0, -(-x_i(t))^{\beta_i}, x_i(t)^{\alpha_i}), i = 1, \dots, n$;

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \cdots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix} \in \mathbb{R}^{n \times n}, B = \begin{pmatrix} b_{11} & \cdots & b_{1n} \\ \vdots & \cdots & \vdots \\ b_{n1} & \cdots & b_{nn} \end{pmatrix} \in \mathbb{R}^{n \times n},$$

$\mathbf{x} = (x_1, \dots, x_n)^T, \mathbf{f}(\mathbf{x}) = (f(x_1, \alpha_1 \vee \beta_1), \dots, f(x_n, \alpha_n \vee \beta_n))^T, \mathbf{c} = (c_1, \dots, c_n)^T \in \mathbb{R}^n$.

Then the system (4.1) can be rewritten in more compact form

$$\dot{\mathbf{x}}(t) = \mathbf{c} + A \cdot \mathbf{x}(t) + B \cdot \mathbf{f}(\mathbf{x}(t)). \quad (4.2)$$

Theorem 4.1. *Let $\alpha_i > 0$ and $\beta_i > 0$, and the function $f(x_i, \alpha_i \vee \beta_i)$ be odd; $i = 1, \dots, n$. Suppose also that for system (4.1) one of the three conditions:*

(i) $\alpha_i > 1, \beta_i > 1, i = 1, \dots, n$, and the symmetric matrix $B + B^T$ is negative definite;

(ii) $\alpha_i < 1, \beta_i < 1, i = 1, \dots, n$, and the symmetric matrix $A + A^T$ is negative definite;

(iii) $\alpha_i > 1, \beta_i < 1$ or $\alpha_i < 1, \beta_i > 1, i = 1, \dots, n$, and the symmetric matrices $A + A^T$ and $B + B^T$ are negative definite, is satisfied.

Then any solution of system (4.1) is bounded. In addition, if in case (i) $A = 0$ and $\mathbf{c} = 0$ or in case (ii) $B = 0$ and $\mathbf{c} = 0$, then the equilibrium point 0 is globally asymptotically stable.

Proof. Let $(\mathbf{u}, \mathbf{v}) \equiv ((u_1, \dots, u_n), (v_1, \dots, v_n))$ be a scalar product of vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$.

We define the applicant for the role of the Lyapunov function for system (4.1) at $\mathbf{c} = 0$ the real function $V(x_1, \dots, x_n)$ by the following rule:

$$V(x_1, \dots, x_n) = \frac{f(x_1, \gamma_1 + 1)}{\gamma_1 + 1} + \dots + \frac{f(x_n, \gamma_n + 1)}{\gamma_n + 1} = 0.5 \left(\left(\frac{f(x_1, \gamma_1)}{\gamma_1 + 1}, \dots, \frac{f(x_n, \gamma_n)}{\gamma_n + 1} \right), (x_1, \dots, x_n) \right) + 0.5 \left((x_1, \dots, x_n), \left(\frac{f(x_1, \gamma_1)}{\gamma_1 + 1}, \dots, \frac{f(x_n, \gamma_n)}{\gamma_n + 1} \right) \right),$$

where $\forall i \in \{1, \dots, n\}$

$$\gamma_i = \begin{cases} \beta_i, & \text{if } x_i < 0, \\ \alpha_i, & \text{if } x_i \geq 0. \end{cases}$$

(i1) The case of strictly odd function $f(x_i, \alpha_i \vee \beta_i)$ and $\mathbf{c} = 0$; $i = 1, \dots, n$.

Since fractions $\beta_i + 1$ and $\alpha_i + 1$ have an even numerator and odd denominator, then the function $V(x_1, \dots, x_n)$ will be positive definite.

Further, from the definition of function $V(x_1, \dots, x_n)$ and system (4.1) it follows that

$$\begin{aligned} \dot{V}_i(x_1(t), \dots, x_n(t)) &= \left(\frac{f(x_1, \gamma_1 + 1)}{\gamma_1 + 1} + \dots + \frac{f(x_n, \gamma_n + 1)}{\gamma_n + 1} \right)'_t \\ &= f(x_1(t), \alpha_1 \vee \beta_1) \cdot \dot{x}_1(t) + \dots + f(x_n(t), \alpha_n \vee \beta_n) \cdot \dot{x}_n(t) \\ &= 0.5 \left(f(x_1(t), \alpha_1 \vee \beta_1), f(x_2(t), \alpha_2 \vee \beta_2), \dots, f(x_n(t), \alpha_n \vee \beta_n) \right) \cdot A \cdot \begin{pmatrix} x_1(t) \\ x_2(t) \\ \vdots \\ x_n(t) \end{pmatrix} \\ &\quad + 0.5(x_1(t), x_2(t), \dots, x_n(t)) \cdot A^T \cdot \begin{pmatrix} f(x_1(t), \alpha_1 \vee \beta_1) \\ f(x_2(t), \alpha_2 \vee \beta_2) \\ \vdots \\ f(x_n(t), \alpha_n \vee \beta_n) \end{pmatrix} \\ &\quad + \left(f(x_1(t), \alpha_1 \vee \beta_1), f(x_2(t), \alpha_2 \vee \beta_2), \dots, f(x_n(t), \alpha_n \vee \beta_n) \right) S \begin{pmatrix} f(x_1(t), \alpha_1 \vee \beta_1) \\ f(x_2(t), \alpha_2 \vee \beta_2) \\ \vdots \\ f(x_n(t), \alpha_n \vee \beta_n) \end{pmatrix}, \end{aligned} \tag{4.3}$$

where $S := 0.5(B + B^T)$.

Introduce the norm of matrix $Q = \{q_{ij}\} \in \mathbb{R}^n$ by the following formula:

$$\|Q\| = \sum_{1 \leq i, j \leq n} |q_{ij}|.$$

Similarly, we define the norm of vector $\mathbf{u} = (u_1, \dots, u_n)^T$:

$$\|\mathbf{u}\| = \sum_{1 \leq i \leq n} |u_i|.$$

Now we estimate the derivative $\dot{V}_t(x_1(t), \dots, x_n(t))$ of function $V(x_1(t), \dots, x_n(t))$, taking into account the fact that matrix S is negative definite:

$$\begin{aligned} \dot{V}_t(\mathbf{x}(t)) &\leq 0.5\|(f(x_1(t), \alpha_1 \vee \beta_1), \dots, f(x_n(t), \alpha_n \vee \beta_n))\| \cdot (\|A\| + \|A^T\|) \cdot \|\mathbf{x}(t)\| \\ &\quad + \lambda_{\max}(S) \cdot (f^2(x_1(t), \alpha_1 \vee \beta_1) + f^2(x_2(t), \alpha_2 \vee \beta_2) + \dots + f^2(x_n(t), \alpha_n \vee \beta_n)) \\ &= 0.5(|f(x_1(t), \gamma_1)| + \dots + |f(x_n(t), \gamma_n)|) \cdot (|x_1(t)| + \dots + |x_n(t)|) \cdot (\|A\| + \|A^T\|) \\ &\quad + \lambda_{\max}(S) \cdot (f^2(x_1(t), \alpha_1 \vee \beta_1) + f^2(x_2(t), \alpha_2 \vee \beta_2) + \dots + f^2(x_n(t), \alpha_n \vee \beta_n)), \end{aligned}$$

where $\lambda_{\max}(S)$ denotes the maximal eigenvalue of symmetric matrix S .

The last inequality can be rewritten as follows:

$$\frac{d}{dt}V(x_1(t), \dots, x_n(t)) \leq a \cdot W(x_1(t), \dots, x_n(t)) - b \cdot (f(x_1(t), 2\gamma_1) + \dots + f(x_n(t), 2\gamma_n)), \quad (4.4)$$

where

$$W(x_1, \dots, x_n) = \sum_{1 \leq i, j \leq n} |r_{ij}| \cdot |x_i^{\gamma_i}| \cdot |x_j|,$$

and the constants $a = 0.5(\|A\| + \|A^T\|) = \|A\| > 0$, and $b = -\lambda_{\max}(S) > 0$.

The solution of inequality (4.4) can be found by the formula

$$\begin{aligned} &V(x_1(t), \dots, x_n(t)) \quad (4.5) \\ &\leq V_0 \cdot \exp \int_0^t \left[\frac{aW(x_1(\tau), \dots, x_n(\tau)) - b \cdot (f(x_1(\tau), 2\gamma_1) + \dots + f(x_n(\tau), 2\gamma_n))}{V(x_1(\tau), \dots, x_n(\tau))} \right] d\tau, \end{aligned}$$

where the constant $V_0 = V(x_1(0), \dots, x_n(0)) > 0$.

Note that the functions $W(x_1, \dots, x_n)$, $V(x_1, \dots, x_n)$, and the function $H(x_1, \dots, x_n) = f(x_1, 2\gamma_1) + \dots + f(x_n, 2\gamma_n)$ are positive definite power functions. In addition, $\deg W(x_1, \dots, x_n) = \deg V(x_1, \dots, x_n) = \max(\gamma_i + 1)$, and $\deg H(x_1, \dots, x_n) = 2 \max \gamma_i$; $i = 1, \dots, n$.

Now assume that $\alpha_i > 1$ and $\beta_i > 1$. Then, we have $2 \max \gamma_i > \max(\gamma_i + 1)$; $i = 1, \dots, n$. In this case, on the one hand, there exists a moment $T_0 > 0$ such that if $t > T_0$, then $aW(x_1, \dots, x_n) - bH(x_1, \dots, x_n) < 0$, and

$$\lim_{t \rightarrow \infty} \frac{aW(x_1(t), \dots, x_n(t)) - bH(x_1(t), \dots, x_n(t))}{V(x_1(t), \dots, x_n(t))} < 0. \quad (4.6)$$

Thus, we have $V(x_1(t), \dots, x_n(t)) \rightarrow 0$ at $t \rightarrow \infty$. But on the other hand this fact means that if the function $V(x_1(t), \dots, x_n(t))$ is small enough, then there exists the moment $T_1 > T_0 > 0$ such that if $t > T_1$, then $aW(x_1(t), \dots, x_n(t)) - bH(x_1(t), \dots, x_n(t)) > 0$ and the positive function $V(x_1(t), \dots, x_n(t))$ increases, and so on.

Denote by $\mathbb{H} = \{h_1, \dots, h_n\} \subset \mathbb{R}^n$ the set of all points satisfying condition $aW(h_1, \dots, h_n) - bH(h_1, \dots, h_n) \leq 0$. Since function $H(x_1, \dots, x_n)$ is positive definite,

then \mathbb{H} is the compact positively invariant set with respect to (4.2). Therefore, if solution $\mathbf{x}(t)$ of system (4.2) belongs to \mathbb{H} , then it is bounded. It means that the solution $V(x_1, \dots, x_n)$ of equation (4.3) also should be bounded.

Now we assume that there exists a moment T_u such that for $t = T_u$ $\mathbf{x}(T_u) \notin \mathbb{H}$. Then inequality (4.6) holds.

Denote by \mathbb{S} the ball centered at the origin whose surface passes through point $T_u \in \mathbb{S}$. In this case $\mathbb{H} \subset \mathbb{S}$. Therefore, by virtue of (4.6) and according to LaSalle's Theorem, there is a moment $T_s > T_u$ such that $\mathbf{x}(T_s) \in \mathbb{H}$. Again we get that solution $\mathbf{x}(t)$ of system (4.2) starting at \mathbb{S} belongs to \mathbb{S} . In addition, $\mathbf{x}(t)$ is attracted to the boundary of \mathbb{H} as $t \rightarrow +\infty$. Thus, it is bounded.

Now we use Comparison Principle. Then it remain to compare the solution $V(x_1, \dots, x_n)$ of equation (4.3) and a similar solution of inequality (4.4). From here it follows the boundedness of solution $\mathbf{x}(t)$ of system (4.2) for any initial condition $\mathbf{x}_0 \in \mathbb{R}^n$. Finally, if $A = 0$ and $\mathbf{c} = 0$, then $\mathbb{H} = 0$. In this case system (4.2) has only zero equilibrium point and any trajectory of this system is attracted to the origin (see Lyapunov's Theorem). The assertion of Theorem 4. 1 for item (i) become obvious. This completes the proof of case (i1).

(i2) The case of strictly odd function $f(x_i, \alpha_i \vee \beta_i)$ and $\mathbf{c} \neq 0$; $i = 1, \dots, n$.

Now we show that system (4.2) with the strictly odd activation function $f(x, \alpha_i \vee \beta_i)$ can be reduced to case $\mathbf{c} = 0$; $i = 1, \dots, n$.

Let $\alpha = \alpha_i$ and $\beta = \beta_i$. Consider the equation

$$A \cdot \mathbf{x} + B \cdot \mathbf{f}(\mathbf{x}) + \mathbf{c} = 0, \quad (4.7)$$

where $\mathbf{f}(\mathbf{x}) = ((x_1)^{\alpha \vee \beta}, \dots, (x_n)^{\alpha \vee \beta})^T$.

Let $q = \{\text{least common multiply of numbers } 2n_r + 1, 2n_l + 1\}$ (see (3.1)). We introduce new variables u_1, \dots, u_n by the formulas: $x_1 = (u_1/x_0)^q, \dots, x_n = (u_n/x_0)^q$, where $x_0 \neq 0$ is a new variable. Then in the new variables the left-hand sides of equations (4.7) are the sums of monomials of odd degrees: $U_1(x_0, u_1, \dots, u_n) + c_1 x_0^s = 0, \dots, U_n(x_0, u_1, \dots, u_n) + c_n x_0^s = 0$, where the number $s = \{\text{least common multiply of numbers } q, q \cdot \alpha, \text{ and } q \cdot \beta\}$ is odd.

Now it remains to apply Theorems 1 or 3 [12] to the system of polynomial equations

$$U_1(x_0, u_1, \dots, u_n) + c_1 x_0^s = 0, \dots, U_n(x_0, u_1, \dots, u_n) + c_n x_0^s = 0. \quad (4.8)$$

According to [12] system (4.8) has a nontrivial real solution $(x_0^* \neq 0, u_1^*, \dots, u_n^*)$.

Put $x_0^* = 1$ and introduce into system (4.2) new variables v_1, \dots, v_n by the formulas $v_1 = x_1 - (u_1^*)^q, \dots, v_n = x_n - (u_n^*)^q$.

As a result, instead of system (4.2), we obtain the following system

$$\dot{\mathbf{v}}(t) = A \cdot \mathbf{v} + B \cdot \mathbf{f}(\mathbf{v}) + \mathbf{P}(\mathbf{v}), \quad (4.9)$$

where $\mathbf{P}(\mathbf{0}) = \mathbf{0}$ and

$$\text{if } \alpha > 1, \beta > 1, \text{ then } \lim_{\|\mathbf{v}\| \rightarrow \infty} \frac{\|\mathbf{P}(\mathbf{v})\|}{\|\mathbf{f}(\mathbf{v})\|} = \left(\frac{\infty}{\infty} \right) = 0. \quad (4.10)$$

(Here $\|\mathbf{z}\|$ is any norm of vector $\mathbf{z} \in \mathbb{R}^n$.)

Now in system (4.9) we put $\mathbf{v} = \mathbf{x}$ and introduce the same function $V(x_1, \dots, x_n)$ as in Theorem 4.1. From here it follows that

$$\begin{aligned} \dot{V}(x_1(t), \dots, x_n(t)) \leq & \text{right side of formula (4.4)} + \mathbf{P}^T(\mathbf{x}(t)) \begin{pmatrix} f(x_1(t), \alpha_1 \vee \beta_1) \\ f(x_2(t), \alpha_2 \vee \beta_2) \\ \vdots \\ f(x_n(t), \alpha_n \vee \beta_n) \end{pmatrix} \\ & + (f(x_1(t), \alpha_1 \vee \beta_1), \dots, f(x_n(t), \alpha_n \vee \beta_n)) \cdot \mathbf{P}(\mathbf{x}(t)). \end{aligned}$$

Then instead of inequality (4.4) we have the following inequality:

$$\begin{aligned} \frac{d}{dt}V(x_1(t), \dots, x_n(t)) \leq & a \cdot W(x_1(t), \dots, x_n(t)) - b \cdot (f(x_1(t), 2\gamma_1) + \dots + f(x_n(t), 2\gamma_n)) \\ & + 2\|\mathbf{P}(\mathbf{x}(t))\| \cdot \|(f(x_1(t), \alpha_1 \vee \beta_1), \dots, f(x_n(t), \alpha_n \vee \beta_n))\|. \end{aligned} \quad (4.11)$$

To study the solutions of inequality (4.11), we use the same reasonings as in the study of inequality (4.4). Then, taking into account (4.10), in the integrand of formula (4.5), we get

$$\begin{aligned} & \lim_{t \rightarrow \infty} \frac{aW(x_1(t), \dots, x_n(t)) - bH(x_1(t), \dots, x_n(t))}{V(x_1(t), \dots, x_n(t))} \\ & + \lim_{t \rightarrow \infty} \frac{2\|\mathbf{P}(\mathbf{x}(t))\| \cdot \|(f(x_1(t), \alpha_1 \vee \beta_1), \dots, f(x_n(t), \alpha_n \vee \beta_n))\|}{V(x_1(t), \dots, x_n(t))} < 0. \end{aligned}$$

Now the completion of the proof of item **(i)** in case $\mathbf{c} \neq 0$ completely repeats the proof in case $\mathbf{c} = 0$.

(i3) The case of odd function $f(x_i, \alpha_i \vee \beta_i)$ and $\mathbf{c} = 0$; $i \in \{1, \dots, n\}$.

Consider the differential equation $\dot{x}(t) = x^\gamma(t)$, where the function x^γ is odd. Introduce the function $V(x) = x^{\gamma+1} = x^\gamma \cdot x$, where the function x is odd. It is clear that the function $V(x)$ is even.

We have: if $x \geq 0$, then $\dot{V}_t = (\gamma+1) \cdot x^\gamma \cdot \dot{x}(t) = (\gamma+1) \cdot x^\gamma \cdot x^\gamma = (\gamma+1) \cdot x^{2\gamma} \geq 0$; if $x < 0$, then $\dot{V}_t = -(\gamma+1) \cdot (-x)^\gamma \cdot \dot{x}(t) = -(\gamma+1) \cdot (-x)^\gamma \cdot [-(-x)^\gamma] = (\gamma+1) \cdot (-x)^{2\gamma} > 0$.

Thus, the function \dot{V}_t is also even. From here it follows that all the results obtained in item **(i1)** for strictly odd functions remain valid for odd functions.

(i4) The case of odd function $f(x_i, \alpha_i \vee \beta_i)$ and $\mathbf{c} \neq 0$; $i = 1, \dots, n$.

We return to equation (4.7). We assume that the activation functions in this equation are odd. Then, according to Lemma 3.1, any odd function with parameters α_o and β_o can be approximated with any given accuracy $\epsilon > 0$ by a strictly odd function whose parameters are ordinary fractions α_{so} and β_{so} .

Let $\mathbf{x}_{so}^*(\alpha_{so} \vee \beta_{so}) \in \mathbb{R}^n$ be a root of system (4.7) with strictly odd activation functions. Then, by the principle of the continuous dependence of the solution of

system (4.7) on parameters α and β , there must be a solution $\mathbf{x}_o^*(\alpha_o \vee \beta_o) \in \mathbb{R}^n$ of system (4.7) with odd activation functions such that $\|\mathbf{x}_{so}^*(\alpha_{so} \vee \beta_{so}) - \mathbf{x}_o^*(\alpha_o \vee \beta_o)\| < \epsilon$.

This completes the proof of item **(i4)** and the whole item **(i)**.

(ii1) The case of strictly odd function $f(x_i, \alpha_i \vee \beta_i)$ and $\mathbf{c} = 0$; $i = 1, \dots, n$.

Define the applicant for the role of the Lyapunov function for system (4.2) at $\mathbf{c} = 0$ the real function $V(x_1, \dots, x_n)$ by the following rule: $V(x_1, \dots, x_n) = x_1^2 + x_2^2 + \dots + x_n^2 = (\mathbf{x}, \mathbf{x}) = \mathbf{x}^T \cdot \mathbf{x}$. From here it follows that

$$\begin{aligned} \dot{V}_t(x_1(t), \dots, x_n(t)) &= (x_1(t), x_2(t), \dots, x_n(t)) \cdot (A + A^T) \begin{pmatrix} x_1(t) \\ x_2(t) \\ \vdots \\ x_n(t) \end{pmatrix} \\ &+ \left(f(x_1(t), \alpha_1 \vee \beta_1), f(x_2(t), \alpha_2 \vee \beta_2), \dots, f(x_n(t), \alpha_n \vee \beta_n) \right) \cdot B \cdot \begin{pmatrix} x_1(t) \\ x_2(t) \\ \vdots \\ x_n(t) \end{pmatrix} \\ &+ (x_1(t), x_2(t), \dots, x_n(t)) \cdot B^T \cdot \begin{pmatrix} f(x_1(t), \alpha_1 \vee \beta_1) \\ f(x_2(t), \alpha_2 \vee \beta_2) \\ \vdots \\ f(x_n(t), \alpha_n \vee \beta_n) \end{pmatrix}. \end{aligned} \quad (4.12)$$

Thus, from (4.12) we have:

$$\begin{aligned} \dot{V}_t(x_1(t), \dots, x_n(t)) &\leq -\lambda_{\max}(A + A^T) \cdot \|\mathbf{x}(t)\|^2 \\ &+ f(x_1(t), \gamma_1 + 1) + \dots + f(x_n(t), \gamma_n + 1) (\|B\| + \|B^T\|). \end{aligned}$$

Consider the equation

$$\frac{d}{dt} V(x_1(t), \dots, x_n(t)) = -a \cdot V(x_1(t), \dots, x_n(t)) + bW(x_1(t), \dots, x_n(t)), \quad (4.13)$$

where

$$W(x_1, \dots, x_n) = f(x_1, \gamma_1 + 1) + \dots + f(x_n, \gamma_n + 1),$$

and the constants $a = -\lambda_{\max}(A + A^T) > 0$, and $b = \|B\| + \|B^T\| > 0$. Then the solution of equation (4.13) can be found by the formula

$$V(x_1(t), \dots, x_n(t)) = V_0 \cdot \exp \int_0^t \left[\frac{-aV(x_1(\tau), \dots, x_n(\tau)) + bW(x_1(\tau), \dots, x_n(\tau))}{V(x_1(\tau), \dots, x_n(\tau))} \right] d\tau, \quad (4.14)$$

where the constant $V_0 = V(x_1(0), \dots, x_n(0)) > 0$.

As $\alpha_i < 1$, $\beta_i < 1$, then $\gamma_i < 1$ and $2\gamma_i < \gamma_i + 1 < 2$; $i = 1, \dots, n$. Then, in all the arguments given for the proof of Theorem 4.1 (item **(i1)**), it is necessary to replace the function $aW(x_1(t), \dots, x_n(t)) - bH(x_1(t), \dots, x_n(t))$ with the function $-aV(x_1(t), \dots, x_n(t)) + bW(x_1(t), \dots, x_n(t))$. As a result, we again obtain the boundedness of solution $V(x_1, \dots, x_n)$ of equation (4.13).

It is clear that the set of all points $\mathbb{H} \subset \mathbb{R}^n$ satisfying condition $-aV(x_1, \dots, x_n) + bH(x_1, \dots, x_n) \leq 0$ is the compact positively invariant set with respect to (4.2). Therefore, according to LaSalle's Theorem, if a solution $\mathbf{x}(t)$ of system (4.2) belongs to \mathbb{H} , then it is bounded. It means that the solution $V(x_1, \dots, x_n)$ of equation (4.13) also should be bounded.

We use Comparison Principle. Now it again remain to compare the solution $V(x_1, \dots, x_n)$ of equation (4.12) and a similar solution of equation (4.13). From here it follows the boundedness of solution $\mathbf{x}(t)$ of system (4.2) for any initial condition $\mathbf{x}_0 \in \mathbb{R}^n$. Finally, if $B = 0$ and $\mathbf{c} = 0$, then $\mathbb{H} = 0$. In this case system (4.2) has only zero equilibrium point and any trajectory of this system is attracted to the origin (see again Lyapunov's Theorem). This completes the proof of case **(ii1)**.

(ii2) The case of strictly odd function $f(x_i, \alpha_i \vee \beta_i)$ and $\mathbf{c} \neq 0$; $i = 1, \dots, n$.

Repeating the same reasoning as at the beginning of **(ii2)**, as a result we get the system (4.9), in which $\mathbf{P}(\mathbf{0}) = \mathbf{0}$ and

$$\text{if } \alpha_i < 1, \beta_i < 1, \text{ then } \lim_{\|\mathbf{v}\| \rightarrow \infty} \frac{\|\mathbf{P}(\mathbf{v})\|}{\|\mathbf{v}\|} = \left(\frac{\infty}{\infty} \right) = 0. \quad (4.15)$$

(Here $\|\mathbf{z}\|$ is any norm of vector $\mathbf{z} \in \mathbb{R}^n$.)

Let $\alpha_i < 1, \beta_i < 1$. The function $V(x_1, \dots, x_n)$ is introduced as before. Then, we have

$$\begin{aligned} \dot{V}_t(x_1(t), \dots, x_n(t)) &= (f(x_1(t), \gamma_1 + 1) + \dots + f(x_n(t), \gamma_n + 1))'_t \\ &= \text{right side of formula (4.13)} + \mathbf{P}^T(\mathbf{x}(t)) \cdot \begin{pmatrix} x_1(t) \\ x_2(t) \\ \vdots \\ x_n(t) \end{pmatrix} + (x_1(t), \dots, x_n(t)) \cdot \mathbf{P}(\mathbf{x}(t)). \end{aligned}$$

Now, taking into account (4.15), in the integrand of formula (4.14), we get

$$\lim_{t \rightarrow \infty} \frac{-aV(x_1(t), \dots, x_n(t)) + bW(x_1(t), \dots, x_n(t)) + 2\|\mathbf{P}(\mathbf{x}(t))\| \cdot \|\mathbf{x}(t)\|}{V(x_1(t), \dots, x_n(t))} < 0.$$

The further proof for $\mathbf{c} \neq 0$ repeats the proof of item **(i3)** in Theorem 4.1.

(ii3) The case of odd function $f(x_i, \alpha_i \vee \beta_i)$ and $\mathbf{c} = 0$; $i \in \{1, \dots, n\}$.

(ii4) The case of odd function $f(x_i, \alpha_i \vee \beta_i)$ and $\mathbf{c} \neq 0$; $i = 1, \dots, n$.

If we replace $\alpha_i < 1$ and $\beta_i < 1$ with $\alpha_i > 1$ and $\beta_i > 1$, then the proof of these items verbatim repeat the proof of items **(i3)** and **(i4)**.

(iii) In general (with the exception of a few minor details), the proof of item (iii) repeats the proofs of items (i) and (ii)

As before, we again define the applicant for the role of the Lyapunov function for system (4.2) at $\mathbf{c} = 0$ the real function $V(x_1, \dots, x_n)$ by the following rule:

$$V(x_1, \dots, x_n) = V_1(x_1, \dots, x_n) + V_2(x_1, \dots, x_n),$$

where

$$V_1(x_1, \dots, x_n) = x_1^2 + \dots + x_n^2, V_2(x_1, \dots, x_n) = \frac{f(x_1, \gamma_1 + 1)}{\gamma_1 + 1} + \dots + \frac{f(x_n, \gamma_n + 1)}{\gamma_n + 1}.$$

It is clear that the functions $V_1(x_1, \dots, x_n)$ and $V_2(x_1, \dots, x_n)$ are positive definite. Then, as follows from the conditions of Theorem 4.1, the functions $\dot{V}_{1t}(x_1(t), \dots, x_n(t))$ and $\dot{V}_{2t}(x_1(t), \dots, x_n(t))$ are negative definite. Hence the function $\dot{V}_t(x_1(t), \dots, x_n(t))$ will also be negative definite. Therefore, the function $V(x_1, \dots, x_n)$ is the Lyapunov function. Thus, item (iii) of Theorem 4.1 is a simple corollary of items (i) and (ii).

If we follow the logic of the proofs of items (i) and (ii), then the following positive definite function $U(x_1, \dots, x_n)$ could claim the role of the Lyapunov function:

$$U(x_1, \dots, x_n) = W(x_1, \alpha_1 \vee \beta_1) + \dots + W(x_n, \alpha_n \vee \beta_n),$$

where

$$W(x_i, \alpha_i \vee \beta_i) = \text{if } (x_i < 0 \text{ and } 0 < \beta_i < 1) \text{ then } x_i^2 \text{ else if } (x_i < 0 \text{ and } \beta_i > 1)$$

$$\text{then } (-x_i)^{\beta_i+1} \text{ if } (x_i \geq 0 \text{ and } 0 < \alpha_i < 1) \text{ then } x_i^2 \text{ else if } (x_i \geq 0 \text{ and } \alpha_i > 1)$$

$$\text{then } x_i^{\alpha_i+1}; \quad i = 1, \dots, n.$$

It is clear that all n terms of function $U(x_1, \dots, x_n)$ are some terms of function $V(x_1, \dots, x_n)$. However, the negative definiteness of function $\dot{V}_t(x_1(t), \dots, x_n(t))$ does not imply negative definiteness of function $\dot{U}_t(x_1(t), \dots, x_n(t))$. (Only inequality $\dot{V}_t(x_1(t), \dots, x_n(t)) \leq \dot{U}_t(x_1(t), \dots, x_n(t))$ holds, but not the inequality $\dot{V}_t(x_1(t), \dots, x_n(t)) \leq \dot{U}_t(x_1(t), \dots, x_n(t)) \leq 0$.) Therefore, the negative definiteness of function $\dot{U}_t(x_1(t), \dots, x_n(t))$ still needs to be proved. But since the moments of switching t_i^* of variable $x_i(t)$ from condition $x_i(t) < 0$ to condition $x_i(t) \geq 0$, $i \in \{1, \dots, n\}$ are unknown, a constructive verification of the negative definiteness of function $\dot{U}_t(x_1(t), \dots, x_n(t))$ is impossible.

Therefore, the conditions of item (iii) guaranteeing the existence of the Lyapunov function $V(x_1, \dots, x_n)$ are more restrictive than the conditions for the existence of the Lyapunov function $U(x_1, \dots, x_n)$. However, these conditions are easily verified for the coefficients of system (4.2), in contrast to the non-constructive conditions for checking the existence of the Lyapunov function $U(x_1, \dots, x_n)$.

The proof of all parts of Theorem 4.1 is complete. □

4.1. Generalization of Theorem 4.1 for classical neural ODEs

If we construct a continuous analogue for some recurrent neural network (see Subsection 1.1), then we can get neural ODEs of the following form:

$$\begin{cases} \dot{x}_1(t) = f_1(b_{11}x_1(t) + b_{12}x_2(t) + \dots + b_{1n}x_n(t) + c_1), \\ \dots \\ \dot{x}_n(t) = f_n(b_{n1}x_1(t) + b_{n2}x_2(t) + \dots + b_{nn}x_n(t) + c_n), \end{cases} \quad (4.16)$$

where $f_i(\dots)$ is an activation function; $i = 1, \dots, n$. (It is one of the common forms of neural ODE [6, 7].)

Suppose that for system (4.1) $A = 0$ and $\det B \neq 0$. Then, using the change of variables $\mathbf{x}(t) \rightarrow B^{-1}\mathbf{x}(t)$, we can go from system (4.1) to system of type (4.16).

Let the functions $f_i(\dots), i = 1, \dots, n$, be the same as in Theorem 4.1. It is clear that point $\mathbf{x}_e = -B^{-1}\mathbf{c}$ is the only equilibrium point of system (4.16). Then, under the conditions of item (i) of Theorem 4.1, the point \mathbf{x}_e is globally asymptotically stable.

5. Second use of power functions in neural ODEs: Architecture (T2)

Consider the following system of real differential equations

$$\begin{cases} \dot{x}_1(t) = c_1 + a_{11}x_1(t) + a_{12}x_2(t) + \dots + a_{1n}x_n(t) \\ \quad + b_{11}piecewise(x_1(t) < 0, -(-x_1(t))^{\beta_1}, x_1(t)^{\alpha_1}) \\ \quad \dots \\ \quad + b_{1n}piecewise(x_n(t) < 0, -(-x_n(t))^{\beta_1}, x_n(t)^{\alpha_1}), \\ \dots \\ \dot{x}_n(t) = c_n + a_{n1}x_1(t) + a_{n2}x_2(t) + \dots + a_{nn}x_n(t) \\ \quad + b_{n1}piecewise(x_1(t) < 0, -(-x_1(t))^{\beta_n}, x_1(t)^{\alpha_n}) \\ \quad \dots \\ \quad + b_{nn}piecewise(x_n(t) < 0, -(-x_n(t))^{\beta_n}, x_n(t)^{\alpha_n}) \end{cases} \quad (5.1)$$

with initial values $x_i(0) = x_{i0}$. Here $\alpha_i \geq 0, \beta_i \geq 0; c_i, a_{ij}, b_{ij} \in \mathbb{R}; i, j = 1, \dots, n$.

Let $a_{ij} = 0, c_i = 0$. In this case we rewrite system (5.1) in the following form

$$\begin{cases} \dot{x}_1(t) = b_{11}f_1(x_1) + b_{12}f_1(x_2) + \dots + b_{1n}f_1(x_n), \\ \dots \\ \dot{x}_n(t) = b_{n1}f_n(x_1) + b_{n2}f_n(x_2) + \dots + b_{nn}f_n(x_n), \end{cases} \quad (5.2)$$

where $f_i(x_j) = piecewise(x_j(t) < 0, -(-x_j(t))^{\beta_i}, x_j(t)^{\alpha_i}); i, j = 1, \dots, n$. Thus, each right-hand side of the equations of system (5.1) is composed only of odd functions.

Let n be an even number. We will assume that all equations of system (5.2) are composed of antisymmetric odd functions $f_i(x_j): \alpha_i = \beta_i = \gamma_i; i, j = 1, \dots, n$.

Introduce the following function

$$V(x_1, \dots, x_n) = \left[\underbrace{x_1 f_2(x_1) f_3(x_1) \cdot \dots \cdot f_n(x_1)}_n + \underbrace{x_2 f_1(x_2) f_3(x_2) \cdot \dots \cdot f_n(x_2)}_n + \dots + \underbrace{x_n f_1(x_n) f_2(x_n) \cdot \dots \cdot f_{n-1}(x_n)}_n \right]. \quad (5.3)$$

Then, we have the following derivative: $V'_t(x_1, \dots, x_n) =$

$$\begin{aligned} &= \left[(b_{11}f_1(x_1) + b_{12}f_1(x_2) + \dots + b_{1n}f_1(x_n))(1 + \gamma_2 + \dots + \gamma_n)f_2(x_1)f_3(x_1) \cdot \dots \cdot f_n(x_1) \right. \\ &+ (b_{21}f_2(x_1) + b_{22}f_2(x_2) + \dots + b_{2n}f_2(x_n))(1 + \gamma_1 + \gamma_3 + \dots + \gamma_n)f_1(x_2)f_3(x_2) \cdot \dots \cdot f_n(x_2) \\ &+ \dots + (b_{n1}f_n(x_1) + b_{n2}f_n(x_2) + \dots + b_{nn}f_n(x_n))(1 + \gamma_1 + \dots + \gamma_{n-1})f_1(x_n) \cdot \dots \cdot f_{n-1}(x_n) \left. \right] \\ &= \left[b_{11}(1 + \gamma_2 + \dots + \gamma_n)f_1(x_1)f_2(x_1) \cdot \dots \cdot f_n(x_1) \right. \\ &\quad + b_{22}(1 + \gamma_1 + \gamma_3 + \dots + \gamma_n)f_1(x_2)f_2(x_2) \cdot \dots \cdot f_n(x_2) \\ &\quad + \dots + b_{nn}(1 + \gamma_1 + \dots + \gamma_{n-1})f_1(x_n)f_2(x_n) \cdot \dots \cdot f_n(x_n) \\ &\quad \left. + \sum_{1 \leq i_1, \dots, i_n \leq n} h_{i_1, \dots, i_n}(b_{ij}, \gamma_1, \dots, \gamma_n) \cdot f_1(x_{i_1})f_2(x_{i_2}) \cdot \dots \cdot f_n(x_{i_n}) \right], \quad (5.4) \end{aligned}$$

where indices i_1, \dots, i_n take only two values: $i_1 = \dots = i_{n-1} = i$, $i_n = j$; $h_{i_1, \dots, i_n}(b_{ij}, \gamma_1, \dots, \gamma_n) \in \mathbb{R}$; $1 \leq i \leq j \leq n$.

From the method of constructing the function $V'_t(x_1, \dots, x_n)$ it follows that $V'_t(x_1, \dots, x_n)$ is a homogeneous function of degree $\gamma_1 + \gamma_2 + \dots + \gamma_n$.

Theorem 5.1. *Let in system (5.2) $n = 2$ and functions $f_1(x_1)$, $f_1(x_2)$, $f_2(x_1)$, and $f_2(x_2)$ are antisymmetric odd. Assume also that $b_{11} < 0$, $b_{22} < 0$, $\gamma_1 \geq 1$, $\gamma_2 \geq 1$, and*

$$\frac{|b_{12}(1 + \gamma_2) + b_{21}(1 + \gamma_1)|}{[-b_{11}(1 + \gamma_2)]^{\frac{\gamma_2}{\gamma_1 + \gamma_2}} \cdot [-b_{22}(1 + \gamma_1)]^{\frac{\gamma_1}{\gamma_1 + \gamma_2}}} < 2. \quad (5.5)$$

Then system (5.2) is a globally asymptotically stable.

Proof. We introduce a positive definite function

$$V(x_1, x_2) = x_1 f_2(x_1) + x_2 f_1(x_2) = x_1^{1+\gamma_2} + x_2^{1+\gamma_1}.$$

Then from here it follows that

$$\begin{aligned} V'_t(x_1, x_2) &= x_1' f_2(x_1) + x_1 f_2'(x_1) + x_2' f_1(x_2) + x_2 f_1'(x_2) \\ &= b_{11}(1 + \gamma_2)x_1^{\gamma_1 + \gamma_2} + (b_{12}(1 + \gamma_2) + b_{21}(1 + \gamma_1))x_2^{\gamma_1} x_1^{\gamma_2} + b_{22}(1 + \gamma_1)x_2^{\gamma_1 + \gamma_2}. \end{aligned}$$

Let $b_{11} < 0, b_{22} < 0$. Further, we have

$$V'_t(x_1, x_2) \leq b_{11}(1+\gamma_2)x_1^{\gamma_1+\gamma_2} + b_{22}(1+\gamma_1)x_2^{\gamma_1+\gamma_2} + |b_{12}(1+\gamma_2) + b_{21}(1+\gamma_1)| |x_2|^{\gamma_1} |x_1|^{\gamma_2}.$$

We apply Lemmas 2.1 and 2.2 (or 2.3) to the last inequality. Therefore, we put $x = x_1, y = y_1, \alpha = \gamma_1, \beta = \gamma_2, A = -b_{11}(1 + \gamma_2), B = -b_{22}(1 + \gamma_1), p = 0$, and $q = |b_{12}(1 + \gamma_2) + b_{21}(1 + \gamma_1)|$. Then, according to Lemma 2.2, there must be

$$|b_{12}(1+\gamma_2) + b_{21}(1+\gamma_1)| = 2A^{\frac{\beta}{\alpha+\beta}} B^{\frac{\alpha}{\alpha+\beta}} = 2[-b_{11}(1+\gamma_2)]^{\frac{\gamma_2}{\gamma_1+\gamma_2}} \cdot [-b_{22}(1+\gamma_1)]^{\frac{\gamma_1}{\gamma_1+\gamma_2}}.$$

Let

$$s = \frac{|b_{12}(1 + \gamma_2) + b_{21}(1 + \gamma_1)|}{[-b_{11}(1 + \gamma_2)]^{\frac{\gamma_2}{\gamma_1+\gamma_2}} \cdot [-b_{22}(1 + \gamma_1)]^{\frac{\gamma_1}{\gamma_1+\gamma_2}}}.$$

It is easy to verify that if $s \leq 2$, then $V'_t(x_1, x_2)$ is non-positive, and in case $s < 2$ it will be negative definite. Thus, under the condition (5.5), the function $V(x_1, x_2)$ is the Lyapunov function. This statement completes the proof. \square

Corollary. *The assertion of Theorem 5.1 is preserved if condition (5.5) is replaced by the inequality*

$$b_{12}b_{21} < 0.$$

Proof. We introduce new changes of variables in system (5.2) : $x_1 \rightarrow d_1 x_1, d_1 > 0$; $x_2 \rightarrow d_2 x_2, d_2 > 0$. Then the coefficients b_{12} and b_{21} of this system are transformed into:

$$b_{12} \rightarrow \frac{b_{12}d_2^{\gamma_1}}{d_1}, b_{21} \rightarrow \frac{b_{21}d_1^{\gamma_2}}{d_2}.$$

In this case, the coefficient $b_{12}(1 + \gamma_2) + b_{21}(1 + \gamma_1)$ of function $V'_t(x_1, x_2)$ is transformed into

$$b_{12}(1 + \gamma_2) + b_{21}(1 + \gamma_1) \rightarrow \Delta(d_1, d_2) \equiv b_{12}(1 + \gamma_2) \frac{d_2^{\gamma_1}}{d_1} + b_{21}(1 + \gamma_1) \frac{d_1^{\gamma_2}}{d_2}.$$

Since $b_{12}b_{21} < 0$, there will always be numbers $d_1^* > 0$ and $d_2^* > 0$ such that $\Delta(d_1^*, d_2^*) = 0$. This means that function

$$V'_t(x_1, x_2) \rightarrow b_{11}(1 + \gamma_2)(d_1^*)^{\gamma_1-1} x_1^{\gamma_1+\gamma_2} + b_{22}(1 + \gamma_1)(d_2^*)^{\gamma_2-1} x_2^{\gamma_1+\gamma_2}$$

is negative definite and function $V(x_1, x_2) \rightarrow x_1^{1+\gamma_2} + x_2^{1+\gamma_1}$ (it will not change in new variables) is the Lyapunov function. \square

Theorem 5.2. *Let $r = \gamma_1 + \dots + \gamma_n$. Assume that in system (5.2) the number n be even and functions $f_1(x_1), \dots, f_1(x_n), \dots, f_n(x_1), \dots, f_n(x_n)$ be antisymmetric odd. Assume also that $b_{11} < 0, b_{22} < 0, \dots, b_{nn} < 0, \gamma_1 \geq 1, \gamma_2 \geq 1, \dots, \gamma_n \geq 1$, and*

$$\begin{cases} |b_{ij}| < \frac{[-b_{ii}(r+1-\gamma_i)]^{\frac{r-\gamma_i}{r}} [-b_{jj}(r+1-\gamma_j)]^{\frac{\gamma_i}{r}}}{(n-1)(r+1-\gamma_i)}, \\ |b_{ji}| < \frac{[-b_{ii}(r+1-\gamma_i)]^{\frac{\gamma_j}{r}} [-b_{jj}(r+1-\gamma_j)]^{\frac{r-\gamma_j}{r}}}{(n-1)(r+1-\gamma_j)}; \end{cases} \quad (5.6)$$

where $1 \leq i < j \leq n$. Then system (5.2) is a globally asymptotically stable.

Proof. Let $n = 4$. We introduce a positive definite function

$$V(x_1, x_2, x_3, x_4) = x_1 f_2(x_1) f_3(x_1) f_4(x_1) + x_2 f_1(x_2) f_3(x_2) f_4(x_2) \\ + x_3 f_1(x_3) f_2(x_3) f_4(x_4) + x_4 f_1(x_4) f_2(x_4) f_3(x_4).$$

Then

$$V'_t(x_1, x_2, x_3, x_4) \\ = b_{11}(r+1-\gamma_1)x_1^r + b_{12}(r+1-\gamma_1)x_1^{r-\gamma_1}x_2^{\gamma_1} + b_{13}(r+1-\gamma_1)x_1^{r-\gamma_1}x_3^{\gamma_1} + b_{14}(r+1-\gamma_1)x_1^{r-\gamma_1}x_4^{\gamma_1} \\ + b_{21}(r+1-\gamma_2)x_1^{\gamma_2}x_2^{r-\gamma_2} + b_{22}(r+1-\gamma_2)x_2^r + b_{23}(r+1-\gamma_2)x_2^{r-\gamma_2}x_3^{\gamma_2} + b_{24}(r+1-\gamma_2)x_2^{r-\gamma_2}x_4^{\gamma_2} \\ + b_{31}(r+1-\gamma_3)x_1^{\gamma_3}x_3^{r-\gamma_3} + b_{32}(r+1-\gamma_3)x_2^{\gamma_3}x_3^{r-\gamma_3} + b_{33}(r+1-\gamma_3)x_3^r + b_{34}(r+1-\gamma_3)x_3^{r-\gamma_3}x_4^{\gamma_3} \\ + b_{41}(r+1-\gamma_4)x_1^{\gamma_4}x_4^{r-\gamma_4} + b_{42}(r+1-\gamma_4)x_2^{\gamma_4}x_4^{r-\gamma_4} + b_{43}(r+1-\gamma_4)x_3^{\gamma_4}x_4^{r-\gamma_4} + b_{44}(r+1-\gamma_4)x_4^r.$$

We compose the following functions:

$$W_1 = \frac{1}{3}(b_{11}(r+1-\gamma_1)x_1^r + b_{22}(r+1-\gamma_2)x_2^r) + b_{12}(r+1-\gamma_1)x_1^{r-\gamma_1}x_2^{\gamma_1} + b_{21}(r+1-\gamma_2)x_1^{\gamma_2}x_2^{r-\gamma_2}$$

$$W_2 = \frac{1}{3}(b_{11}(r+1-\gamma_1)x_1^r + b_{33}(r+1-\gamma_3)x_3^r) + b_{13}(r+1-\gamma_1)x_1^{r-\gamma_1}x_3^{\gamma_1} + b_{31}(r+1-\gamma_3)x_1^{\gamma_3}x_3^{r-\gamma_3}$$

$$W_3 = \frac{1}{3}(b_{11}(r+1-\gamma_1)x_1^r + b_{44}(r+1-\gamma_4)x_4^r) + b_{14}(r+1-\gamma_1)x_1^{r-\gamma_1}x_4^{\gamma_1} + b_{41}(r+1-\gamma_4)x_1^{\gamma_4}x_4^{r-\gamma_4}$$

$$W_4 = \frac{1}{3}(b_{22}(r+1-\gamma_2)x_2^r + b_{33}(r+1-\gamma_3)x_3^r) + b_{23}(r+1-\gamma_2)x_2^{r-\gamma_2}x_3^{\gamma_2} + b_{32}(r+1-\gamma_3)x_2^{\gamma_3}x_3^{r-\gamma_3}$$

$$W_5 = \frac{1}{3}(b_{22}(r+1-\gamma_2)x_2^r + b_{44}(r+1-\gamma_4)x_4^r) + b_{24}(r+1-\gamma_2)x_2^{r-\gamma_2}x_4^{\gamma_2} + b_{42}(r+1-\gamma_4)x_2^{\gamma_4}x_4^{r-\gamma_4}$$

$$W_6 = \frac{1}{3}(b_{33}(r+1-\gamma_3)x_3^r + b_{44}(r+1-\gamma_4)x_4^r) + b_{34}(r+1-\gamma_3)x_3^{r-\gamma_3}x_4^{\gamma_3} + b_{43}(r+1-\gamma_4)x_3^{\gamma_4}x_4^{r-\gamma_4}.$$

(It's clear that $V'_t(x_1, x_2, x_3, x_4) = W_1 + \dots + W_6$.)

Let $b_{11} < 0, b_{22} < 0$. We assume that $W_1 \leq 0$. Then we have

$$-\frac{1}{3}(b_{11}(r+1-\gamma_1)x_1^r + b_{22}(r+1-\gamma_2)x_2^r) \geq b_{12}(r+1-\gamma_1)x_1^{r-\gamma_1}x_2^{\gamma_1} + b_{21}(r+1-\gamma_2)x_1^{\gamma_2}x_2^{r-\gamma_2}.$$

Let $A = [-b_{11}(r+1-\gamma_1)]^{\frac{1}{r}}, B = [-b_{22}(r+1-\gamma_2)]^{\frac{1}{r}}$. Consider the following inequality:

$$x_1^r + x_2^r \geq \frac{3|b_{12}(r+1-\gamma_1)||x_1^{r-\gamma_1}x_2^{\gamma_1}|}{[-b_{11}(r+1-\gamma_1)]^{\frac{r-\gamma_1}{r}}[-b_{22}(r+1-\gamma_2)]^{\frac{\gamma_1}{r}}} \\ + \frac{3|b_{21}(r+1-\gamma_2)||x_1^{\gamma_2}x_2^{r-\gamma_2}|}{[-b_{11}(r+1-\gamma_1)]^{\frac{\gamma_2}{r}}[-b_{22}(r+1-\gamma_2)]^{\frac{r-\gamma_2}{r}}}. \quad (5.7)$$

It is clear that inequality (5.7) follows from Lemma 2.2 (see (2.8)). In turn, inequalities (5.6) follow from formulas (5.7) and (2.8).

The proof of inequalities $W_i \leq 0, i = 2, \dots, 6$, is carried out in the same way as the proof of inequality $W_1 \leq 0$. Thus, we obtain that under the conditions of Theorem 5.2 the function $V_t'(x_1, x_2, x_3, x_4)$ is negative definite. Hence, V is the Lyapunov function. The proof of Theorem 5.2 for the case of an arbitrary even n now becomes obvious. \square

Theorem 5.3. *Let $r = \gamma_1 + \dots + \gamma_n$. Assume that in system (5.2) the number n be odd and functions $f_1(x_1), \dots, f_1(x_n), \dots, f_n(x_1), \dots, f_n(x_n)$ be antisymmetric odd. Assume also that $b_{11} < 0, b_{22} < 0, \dots, b_{nn} < 0, \gamma_1 \geq 1, \gamma_2 \geq 1, \dots, \gamma_n \geq 1$, and*

$$\begin{cases} |b_{ij}| < \frac{[-b_{ii}(r+2-\gamma_i)]^{\frac{r+1-\gamma_i}{r+1}} [-b_{jj}(r+2-\gamma_j)]^{\frac{\gamma_i}{r+1}}}{(n-1)(r+2-\gamma_i)}, \\ |b_{ji}| < \frac{[-b_{ii}(r+2-\gamma_i)]^{\frac{\gamma_j}{r+1}} [-b_{jj}(r+2-\gamma_j)]^{\frac{r+1-\gamma_j}{r+1}}}{(n-1)(r+2-\gamma_j)}; \end{cases} \quad (5.8)$$

where $1 \leq i < j \leq n$. Then system (5.2) is a globally asymptotically stable.

Proof. Let $n = 3$. We introduce a positive definite function

$$V(x_1, x_2, x_3) = x_1^2 f_2(x_1) f_3(x_1) + x_2^2 f_1(x_2) f_3(x_2) + x_3^2 f_1(x_3) f_2(x_3).$$

Then

$$\begin{aligned} & V_t'(x_1, x_2, x_3) \\ &= b_{11}(r+2-\gamma_1)x_1^{r+1} + b_{12}(r+2-\gamma_1)x_1^{r+1-\gamma_1}x_2^{\gamma_1} + b_{13}(r+2-\gamma_1)x_1^{r+1-\gamma_1}x_3^{\gamma_1} \\ &+ b_{21}(r+2-\gamma_2)x_1^{\gamma_2}x_2^{r+1-\gamma_2} + b_{22}(r+2-\gamma_2)x_2^{r+1} + b_{23}(r+2-\gamma_2)x_2^{r+1-\gamma_2}x_3^{\gamma_2} \\ &+ b_{31}(r+2-\gamma_3)x_1^{\gamma_3}x_3^{r+1-\gamma_3} + b_{32}(r+2-\gamma_3)x_2^{\gamma_3}x_3^{r+1-\gamma_3} + b_{33}(r+2-\gamma_3)x_3^{r+1}. \end{aligned}$$

We compose the following functions:

$$\begin{aligned} W_1 &= \frac{1}{2}(b_{11}(r+2-\gamma_1)x_1^{r+1} + b_{22}(r+2-\gamma_2)x_2^{r+1}) + b_{12}(r+2-\gamma_1)x_1^{r+1-\gamma_1}x_2^{\gamma_1} \\ &+ b_{21}(r+2-\gamma_2)x_1^{\gamma_2}x_2^{r+1-\gamma_2}; W_2 = \frac{1}{2}(b_{11}(r+2-\gamma_1)x_1^{r+1} + b_{33}(r+2-\gamma_3)x_3^{r+1}) \\ &+ b_{13}(r+2-\gamma_1)x_1^{r+1-\gamma_1}x_3^{\gamma_1} + b_{31}(r+2-\gamma_3)x_1^{\gamma_3}x_3^{r+1-\gamma_3}; W_3 = \frac{1}{2}(b_{22}(r+2-\gamma_2)x_2^{r+1} \\ &+ b_{33}(r+2-\gamma_3)x_3^{r+1}) + b_{23}(r+2-\gamma_2)x_2^{r+1-\gamma_2}x_3^{\gamma_2} + b_{32}(r+2-\gamma_3)x_2^{\gamma_3}x_3^{r+1-\gamma_3}. \end{aligned}$$

(It's clear that $V_t'(x_1, x_2, x_3) = W_1 + W_2 + W_3$.)

Let $b_{11} < 0, b_{22} < 0$. We assume that $W_1 \leq 0$. Then we have $-\frac{1}{2}(b_{11}(r+2-\gamma_1)x_1^{r+1} + b_{22}(r+2-\gamma_2)x_2^{r+1}) \geq b_{12}(r+2-\gamma_1)x_1^{r+1-\gamma_1}x_2^{\gamma_1} + b_{21}(r+2-\gamma_2)x_1^{\gamma_2}x_2^{r+1-\gamma_2}$. Now you can use Lemma 2.2.

Let $A = [-b_{11}(r+2-\gamma_1)]^{\frac{1}{r+1}}$, $B = [-b_{22}(r+2-\gamma_2)]^{\frac{1}{r+1}}$. Consider the following inequality:

$$\begin{aligned} x_1^{r+1} + x_2^{r+1} &\geq \frac{2|b_{12}(r+2-\gamma_1)||x_1^{r+1-\gamma_1}x_2^{\gamma_1}|}{[-b_{11}(r+2-\gamma_1)]^{\frac{r+1-\gamma_1}{r+1}}[-b_{22}(r+2-\gamma_2)]^{\frac{\gamma_1}{r+1}}} \\ &+ \frac{2|b_{21}(r+2-\gamma_2)||x_1^{\gamma_2}x_2^{r+1-\gamma_2}|}{[-b_{11}(r+2-\gamma_1)]^{\frac{\gamma_2}{r+1}}[-b_{22}(r+2-\gamma_2)]^{\frac{r+1-\gamma_2}{r+1}}}. \end{aligned} \quad (5.9)$$

Now the proof of inequalities (5.8) for Theorem 5.3 follows from inequalities (5.9) and (2.8). (See the end of the proof of Theorem 5.2.) \square

The following corollary to Theorems 5.2 and 5.3 has a very convenient practical meaning.

Corollary. *Suppose that under the conditions of Theorem 5.3 all diagonal elements b_{ii} of matrix B are approximately equal to each other: $b_{ii} \approx -\Delta$, where $\Delta > 0$. We will also consider that all degrees of activation functions are greater than or equal to 1 and they are also approximately equal to each other. Then system (5.2) is stable. Moreover, we have $|b_{ij}| < \Delta/(n-1)$; $i, j = 1, \dots, n$; $i \neq j$.*

5.1. Boundedness conditions of solutions for system (5.1)

Let $\mathbf{c} = 0$ and $\gamma_1 > 1, \dots, \gamma_n > 1$. Introduce for system (5.1) the positive definite function $Q(x_1, \dots, x_n) = x_1^2 + \dots + x_n^2 + V(x_1, \dots, x_n)$, where $V(x_1, \dots, x_n)$ is defined by the relation (5.3). Then $Q'_t = W(A, x_1, \dots, x_n) + V'_t(x_1, \dots, x_n)$, where $W(A, x_1, \dots, x_n)$ and $V'_t(x_1, \dots, x_n)$ are polynomials of x_1, \dots, x_n ; in addition, the function $V'_t(x_1, \dots, x_n)$ does not depend on A and is defined by formula (5.4).

Let us turn to the inequality (4.4). Note that inequality $\deg W(A, x_1, \dots, x_n) < \deg V'_t(x_1, \dots, x_n) = r$ holds. By the condition of Theorem 5.1, function $V'_t(x_1, \dots, x_n)$ is negative definite. Therefore, inequality (4.4) can be rewritten as

$$\frac{d}{dt}Q(x_1(t), \dots, x_n(t)) \leq |W(A, x_1(t), \dots, x_n(t))| - |V'_t(x_1(t), \dots, x_n(t))|. \quad (5.10)$$

Further investigation of inequality (5.10) repeats a similar study of inequality (4.4) at $t \rightarrow \infty$. In this case, inequality $\deg W(A, x_1, \dots, x_n) < \deg V'_t(x_1, \dots, x_n) = r$ plays a key role in this study.

Thus, under the conditions of Theorems 5.2, 5.3 (with a change $\gamma_i \geq 1 \rightarrow \gamma_i > 1$; $i = 1, \dots, n$), all solutions of system (5.1) in case $A \neq 0$ will be bounded.

6. Conclusion

Two variants of neural ODEs with power nonlinearities corresponding to two different architectures of a recurrent neural network are considered. Sufficient conditions for the global stability of the mentioned neural ODEs are found. Note that the obtained stability conditions can be interesting by themselves. For the

first type of neural ODEs, these conditions are reduced to the study of the positive and negative definiteness of square matrices. For the second type of neural ODEs, there is no analog of these conditions known earlier.

References

1. V. YE. BELOZYOROV, YE. M. KOSARIEV, M. M. PULIN, V. G. SYCHENKO, V. G. ZAYTSEV. *A new mathematical model of dynamic process in direct current traction power supply system*, *Journal of Optimization, Differential Equations and Their Applications (JODEA)*, **27**(1) (2019), 21 – 55.
2. V. YE. BELOZYOROV, D. V. DANTSEV, S. A. VOLKOVA. *On Equivalence of Linear Control Systems and Its Usage to Verification of the Adequacy of Different Models for A Real Dynamic Process*, *Journal of Optimization, Differential Equations and Their Applications (JODEA)*, **28**(1)(2020), 43 – 97.
3. J. JIA, A. R. BENSON, *Neural jump stochastic differential equations*, arXiv preprint arXiv:1905.10403v3[cs.LG], (2020), 1 – 14.
4. B. CHANG, M. CHEN, E. HABER, E.D. CHI, *Antisymmetric RNN: A Dynamical System View on Recurrent Neural Networks*, In conference ICLR, May 6 - May 9, New Orleans, Louisiana, USA, (2019), 1 – 15.
5. R.T.Q. CHEN, Y. RUBANOVA, J. BETTENCOURT, D. DUVENAUD, *Neural ordinary differential equations*, arXiv preprint arXiv:1806.07366v5[cs.LG], (2019), 1 – 18.
6. E. HABER, L. RUTHOTTO, *Stable Architectures for Deep Neural Networks*, arXiv preprint arXiv: 1705.03341v1[cs.LG], (2019), 1 – 23.
7. S. HAYKIN, *Neural Networks. A Comprehensive Foundation*, Second Edition, Pearson Education, Prentice Hall, 2005.
8. E. M. IZHIKEVICH, *Dynamical Systems in Neuroscience: The Geometry of Excitability and Bursting*, The MIT Press Cambridge, Massachusetts, London, England, 2007.
9. H. K. KHALIL, *Nonlinear Systems – 2nd Edition*, (Prentice Hall/New-Jersey), 1996.
10. J. LIANG, W. SONG, M. WANG, *Stock Price Prediction Based on Procedural Neural Networks*, *Advances in Artificial Neural Systems*, **2011**, 2011, Article ID 814769, 11 pages.
11. S. T. PIANTADOSI, *Zipf's word frequency law in natural language: a critical review and future directions*. *Psychonomic bulletin and review*, **21**(5), (2014), 1112 –1130.
12. J. A. SJOGREN. *Homogeneous Systems and Euclidean Topology*, In: *SIAM Conference on Applied Algebraic Geometry (AG17)*, July 31 – August 4, Georgia Tech, Atlanta GA, USA, (2017), 1 – 20.
13. S. SONODA, N. MURATA, *Neural network with unbounded activation functions is universal approximator*, *Applied and Computational Harmonic Analysis*, **43**, 2017, pp. 233 – 268.
14. N.-E. TATAR, *Hopfield neural networks with unbounded monotone activation functions*, *Hindawi Publishing Corporation. Advances in Artificial Neural Systems*, **2012**, 2012, 571358-1 – 5.
15. Z. WANG, J. LIANG, Y. LIU, *Mathematical Problems for Complex Networks*, *Mathematical Problems in Engineering*, **2012**, 2012, Article ID 934680, 5 pages.
16. H. ZHANG, X. GAO, J. UNTERMAN, T. ARODZ, *Approximation capabilities of neural ordinary differential equations*, arXiv preprint arXiv: 1907.12998v1[cs.LG], (2019), 1 – 11.

Received 7.11.2020