JOURNAL OF OPTIMIZATION, DIFFERENTIAL EQUATIONS AND THEIR APPLICATIONS (JODEA)
Volume **33**, Issue **1**, June **2025**, pp. 15–41, DOI 10.15421/142502

ISSN (print) 2617–0108
ISSN (on-line) 2663–6824

HOMOCLINIC AND HETEROCLINIC NEURAL ODES: THEORY AND ITS USE TO CONSTRUCT NEW CHAOTIC ATTRACTORS

Vasiliy Ye. Belozyorov[*], Dmytro M. Moroz[†], Svetlana A. Volkova[‡]

Abstract. New types of neural ordinary differential equations (NODE) with power nonlinearities are considered. For these NODE systems, new conditions for the existence of homoclinic and heteroclinic orbits are found. In the future, the implementation of these conditions guarantees the existence of chaotic attractors in the mentioned NODE systems.

Key words: system of ordinary autonomous differential equations, limit cycle, homoclinic and heteroclinic orbits, chaotic attractor, neural network.

Mathematics Subject Classification MSC2020: 34A34, 37C27, 37C75, 37N40, 68P25.

Communicated by Prof. O.P. Kasyanov

1. Introduction

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

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

Here $\mathbf{x} \in \mathbb{R}^n$ is a vector of states, $\mathbf{\Omega} \in \mathbb{R}^k$ is a vector of parameters, $\mathbf{H}(\mathbf{x}, \mathbf{\Omega}) : \mathbb{R}^n \times \mathbb{R}^k \to \mathbb{R}^n$ 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:

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

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

[†]Department of Computer Systems Software, National Technical University "Dnipro Polytechnic", Dmytra Yavornytskogo avenue, 19, 49005, Dnipro, Ukraine, moroz.d.m@nmu.one

[‡]Department of Computer Information Technologies, Ukrainian State University of Science and Technologies, Academican Lazaryan's street, 2, 49010, Dnipro, Ukraine, svolkovav2017@gmail.com

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

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

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

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

where Δt is the discretization step. (If $\Delta t = 1$, then (1.3) becomes (1.1).)

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 $\mathbf{H}(\mathbf{x}(t), \mathbf{\Omega})$ [8, 9, 16].

Thus, in some cases, we can replace the study of neural network (1.1) with its continuous analog (1.2), which in neurodynamics is called the system of neural ODEs (NODE) [8, 9, 16]. In the future, model (1.2) can be viewed as a control system, the parameter vector $\mathbf{\Omega}$ of which is the control. By adjusting this vector, it is possible to ensure that the trajectory of model (1.2) differs as little as possible from the trajectory of the real process for which the specified model was built.

Note that the process of tuning model (1.2) can be much more successful if its architecture generates some invariant manifolds [1,3,5,10]. The main contribution to the solution of the problem of classification of invariant manifolds was made in [25]. It was indicated that a large class of chaotic systems can be divided into the following four types: chaos of the homoclinic orbit type; chaos of the heteroclinic orbit type; chaos of the hybrid type; i.e. those with both homoclinic and heteroclinic orbits; chaos of other types.

For many decades a chaotic behavior of dynamic systems remains in the focus of mathematicians, physicists and engineers. There are hundreds publications, in which different forms of this phenomenon is considered [1, 3–7, 17, 23, 25]. However, there are only a few publications, in which (from the mathematical point of view) the existence of chaotic dynamics is rigorously proved. For example, the mathematically rigorous proof of the chaos existence in a modified Lorenz systems is presented in papers [19, 20, 23]. (In these papers authors use the theory of Shilnikov bifurcations of homoclinic and heteroclinic orbits.)

This article is devoted to finding the conditions that must be satisfied by system (1.2), which generates periodic orbits, homoclinic and heteroclinic trajectories.

2. Mathematical preliminaries

Definition 2.1. [8–12, 21] A set of continuous real functions $\mathbb{F} \subset \mathbf{C}(\mathbb{X})$ is called separating points of the set $\mathbb{X} \subset \mathbb{R}^n$ if for any different $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{X}$ ($\mathbf{x}_1 \neq \mathbf{x}_2$), there exists a function $f \in \mathbb{F}$ such that $f(\mathbf{x}_1) \neq f(\mathbf{x}_2)$.

Let $f(w) \in \mathbb{F} = \phi(w) \vee \psi(w)$ be the function of one real variable w such that $f(0) = 0$ and

$$\text{either conditions for } f(w) : \begin{cases} \text{if } w < 0 \text{ then } f(w) = -\psi(-w) < 0, \\ \text{if } w > 0 \text{ then } f(w) = \phi(w) > 0 \end{cases} \tag{2.1}$$

$$\text{or conditions for } f(w) : \begin{cases} \text{if } w < 0 \text{ then } f(w) = \psi(-w) > 0, \\ \text{if } w > 0 \text{ then } f(w) = \phi(w) > 0 \end{cases} \tag{2.2}$$

are fulfilled. In addition, $\phi(w)$ and $\psi(w)$ are differentiable increasing functions of one variable w such that $\phi(0) = \psi(0) = 0$ and $\dot{\phi}(0) = \dot{\psi}(0)$.

Definition 2.2. [8–10] Representation (2.1) ((2.2)) is called an odd (even) activation function.

Note that any odd activation function $f(w)$ is the separating points function.

For example, functions $h(u) = u^3 + u$, $h(u) = \tanh(u)$, $h(u) = u^2(u \geq 0) \vee -u^4(u < 0)$, and $h(u) = u^{2m+1}|u| + a_2 u^{2m-1} + \cdots + a_{2m}u$ are odd. (Here $a_2, ..., a_{2m}$ are positive constants.)

Let us consider the ODE system

$$\dot{\mathbf{x}}(t) = \mathbf{G}(\mathbf{x}(t)) \in \mathbb{R}^n, \tag{2.3}$$

where $\mathbf{G}(\mathbf{x})$ is a continuous vector function and $\mathbf{G}(\mathbf{0}) = \mathbf{0}$.

In the future we will need the following well-known theorem.

Theorem 2.1. *(***LaSalle's Theorem***) [18]. Let $\mathbb{H} \subset \mathbb{R}^n$ be a compact set that is positively invariant with respect to (2.3). Let $V : \mathbb{R}^n \to \mathbb{R}$ be a continuously differentiable function such that $\dot{V}(\mathbf{x}) \leq 0$ (or $\dot{V}(\mathbf{x}) \geq 0$) 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 $t \to +\infty$.*

Currently, the following type of NODE (a special case of system (2.3)) is often used to construct the architecture of residual neural networks

$$\dot{\mathbf{x}}(t) = \mathbf{G}(\mathbf{x}(t)) \equiv \sigma(W\mathbf{x}(t) + \mathbf{b}). \tag{2.4}$$

Here the matrix $W = \{w_{ij}\} \in \mathbb{R}^{n \times n}$ and the vector $\mathbf{b} = \{b_i\} \in \mathbb{R}^n$; $\sigma(u)$ is a scalar activation function. (It is usually assumed that $W = G - G^T - \mu I$, where $G \in \mathbb{R}^{n \times n}$, $I \in \mathbb{R}^{n \times n}$ is the identity matrix, $\mu \geq 0$ [10, 11, 13, 21].)

Now let's consider the same system (2.4), but with a real vector activation function on the right-hand side:

$$\sigma(W\mathbf{x}(t) + \mathbf{b}) = \begin{pmatrix} \sigma_1(w_{11}x_1(t) + \cdots + w_{1n}x_n(t) + b_1), \\ \ldots, \\ \sigma_n(w_{n1}x_1(t) + \cdots + w_{nn}x_n(t) + b_n) \end{pmatrix},$$

where $\sigma(u_1, ..., u_n) = (\sigma_1(u_1), ..., \sigma_n(u_n))^T \in \mathbb{R}^n$.

Let us define the function $\sigma_i(u_i), i = 1, ..., n$, by the following rule:

$$\sigma_i(u_i) = f_i(u_i) + \sum_{j=1}^{k} c_j f_{ij}(u_i) = \phi_i(u_i) \vee \psi_i(u_i) + \sum_{j=1}^{k} c_{ij}\phi_{ij}(u_i) \vee \psi_{ij}(u_i), \quad (2.5)$$

where $f_i(u_i)$ and $f_{ij}(u_i)$ are continuous functions of their arguments (these can be functions (2.1), (2.2)); $c_{ij} \in \mathbb{R}$.

Taking into account (2.5) we introduce the following functions:

$$F_i(u) := \int_0^u (\phi_i(\omega) \vee \psi_i(\omega))d\omega + \int_0^u \sum_{j=1}^{k} c_{ij} \cdot (\phi_{ij}(\omega) \vee \psi_{ij}(\omega))d\omega; 1 = 1, ..., n$$

and

$$G_i(u) := \int_{-u}^0 (\phi_i(\omega) \vee \psi_i(\omega))d\omega + \int_{-u}^0 \sum_{j=1}^{k} c_{ij} \cdot (\phi_{ij}(\omega) \vee \psi_{ij}(\omega))d\omega; 1 = 1, ..., n.$$

Theorem 2.2. *Let us assume that for system (2.4) the following conditions*

(a1) $\det W \neq 0$*;*

(a2) the matrix $W + W^T$ *is negative definite;*

(a3) $\forall i \in \{1, ..., n\}$ *the function* $f_i(u_i)$ *is odd;*

(a4) $\forall i \in \{1, ..., n\}$ $\lim\limits_{u\to\infty} F_i(u) > 0$ *and* $\lim\limits_{u\to-\infty} G_i(u) > 0$

are satisfied. Then for system (2.4) there exists nonempty compact invariant set.

Proof. First we will perform the following substitution of the variable: $\mathbf{y} = W\mathbf{x} + \mathbf{b}$. Since $\det W \neq 0$ (see (a1)), then after this substitution system (2.4) transits to the system

$$\dot{\mathbf{y}}(t) = W\sigma(\mathbf{y}(t)), \mathbf{y}(0) = \mathbf{y}_0. \quad (2.6)$$

Now we will use the system (2.6) and introduce the function

$$\dot{V}_t(\mathbf{y}) = \sigma^T(\mathbf{y})\dot{\mathbf{y}} + \dot{\mathbf{y}}^T\sigma(\mathbf{y}) = \sigma^T(\mathbf{y})(W + W^T)\sigma(\mathbf{y}).$$

Then from the condition (a2) of Theorem 2.2 it follows that $\dot{V}_t(\mathbf{y}) \leq 0$.

We define a set $\mathbb{L} \subset \mathbb{R}^n$ such that $\forall \mathbf{y} \in \mathbb{L}$, we have $\dot{V}_t(\mathbf{y}) = 0$. Then from here and conditions (a3), (a4) of Theorem 2.2 it follows that

$$V(\mathbf{y}) = \sum_{i=1}^{n} \int_0^{y_i} (\sigma^T(\mathbf{y})\dot{\mathbf{y}} + \dot{\mathbf{y}}^T\sigma(\mathbf{y}))dt = \sum_{i=1}^{n} F_i(y_i)$$

$$= \sum_{i=1}^{n} \int_0^{y_i} (\phi_i(\omega) \vee \psi_i(\omega))d\omega + \Phi(y_1, ..., y_n) = C_F(\mathbf{y}_0) = const > 0,$$

where $\Phi(y_1, ..., y_n)$ is a continuous function. (Here we use the fact that the integral of the odd activation function is the even activation function [8,9]). Similarly, we have

$$V(\mathbf{y}) = \sum_{i=1}^{n} \int_{-y_i}^{0} (\sigma^T(\mathbf{y})\dot{\mathbf{y}} + \dot{\mathbf{y}}^T\sigma(\mathbf{y}))dt = \sum_{i=1}^{n} G_i(y_i)$$

$$= \sum_{i=1}^{n} \int_{-y_i}^{0} (\phi_i(\omega) \vee \psi_i(\omega))d\omega + \Psi(y_1, ..., y_n) = C_G(\mathbf{y}_0) = const > 0,$$

where $\Psi(y_1, ..., y_n)$ is a continuous function. Thus, according to condition (a4) we have

$$\lim_{t\to\infty} V(\mathbf{y}(t)) > 0.$$

(Note that if $t = t_f < \infty$ is finite, then situation $V(\mathbf{y}(t_f)) < 0$ is not excluded.)

This means that the set $\mathbb{L}$ is compact and it is an invariant set for system (2.6). Then, from here it follows that the set $\mathbb{D} := \{\mathbf{y}(t)|\dot{V}_t(\mathbf{y}) \leq 0\}$ has a boundary $\mathbb{L}$ and therefore is also the compact invariant set for system (2.6) (see Theorem 2.1). If we now return from variable $\mathbf{y}$ to variable $\mathbf{x} = W^{-1}\mathbf{y} - W^{-1}\mathbf{b}$, we obtain the same statement for system (2.4). □

Thus, Theorem 2.2 allows us to significantly expand the class of activation functions that can be used to design neural network architectures.

In what follows we will need the following two auxiliary lemmas

Lemma 2.1. *[14]. Let $H \in \mathbb{R}^{n\times n}$ be the antisymmetric nonsingular matrix and let $D = \mathrm{diag}(d_1, ..., d_n)$ be the diagonal matrix such that $d_1 > 0, ..., d_n > 0$ (or $d_1 < 0, ..., d_n < 0$). Then all the eigenvalues of the matrix $H \cdot D$ are purely imaginary.*

Proof. First of all, we note that matrix H must be a matrix of even order ($n = 2k$). In addition, all eigenvalues of the matrix H are purely imaginary (see [2]).

Now consider the following linear system of differential equations

$$\dot{\mathbf{x}}(t) = H \cdot D \cdot \mathbf{x}, \mathbf{x} \in \mathbb{R}^n. \tag{2.7}$$

Let us construct the first integral of system (2.7). Then we have

$$\mathbf{x}^T(t)D\dot{\mathbf{x}}(t) = \mathbf{x}^T DHD\mathbf{x} = (D\mathbf{x})^T H(D\mathbf{x}) \equiv 0.$$

From here it follows that

$$\frac{1}{2}(d_1x_1^2(t) + \cdots + d_nx_n^2(t)) = C = const \geq 0.$$

Due to the fact that $d_1 > 0, ..., d_n > 0$, the last relation means that under any initial conditions the trajectory of system (2.7) is bounded by an ellipsoid and it lies on a surface of central manifold $\mathbb{M}_c$ of the equilibrium point $\mathbf{0} = (0, ..., 0)^T \in \mathbb{R}^n$ (see [18, 24]).

It is clear that in our case we have $\mathbb{M}_c = \mathbb{R}^n$ and therefore the eigenvalues of the matrix $H \cdot D$ must be only purely imaginary. (The last statement is also obvious in case $d_1 < 0, ..., d_n < 0$.) □

Lemma 2.2. *[14]. Let $H \in \mathbb{R}^{n\times n}$ be the antisymmetric nonsingular matrix and let $D = \mathrm{diag}(d_1, ..., d_n)$ be the diagonal matrix such that $d_1 > 0, d_2 < 0, ..., d_n < 0$. Then the matrix $H \cdot D$ has two real eigenvalues $\lambda_1 = r > 0$ and $\lambda_2 = -r < 0$, so $\lambda_1\lambda_2 = -r^2 < 0$; the remaining $n - 2$ eigenvalues of this matrix are purely imaginary.*

Proof. Denote by $\lambda_1, ..., \lambda_{n=2k}$ the eigenvalues (complex and real) of matrix HD.

(a) Let us use the following well-known result of matrix analysis: the determinant of antisymmetric matrix $H = S - S^T$ of even order $n = 2k$ is the square of a homogeneous polynomial of order k in the products of elements of matrix S (see [2])

$$\det(H) = \Big(\sum_{i_1<j_1,...,i_k<j_k;(i_1,j_1,...,i_k,j_k)\in\{1,...,2k\}} (s_{i_1j_1} \cdot ... \cdot s_{i_kj_k})\Big)^2. \qquad (2.8)$$

Here $S = \{s_{ij}, 1 \le i < j \le n = 2k\}$ and the symbol $\{1, ..., 2k\}$ means the set of all permutations from elements $1, 2, ..., 2k$; $k \ge 1$.

From formula (2.8) it follows that $\det H > 0$ and therefore $\det(HD) = \det H \cdot \det D = \det H \cdot d_1 \cdot d_2 \cdot \dots \cdot d_{n=2k} = \lambda_1 \cdot ... \cdot \lambda_{n=2k} < 0$. This means that at least one eigenvalue (for example λ_2) of the matrix HD is negative.

(b) Denote by $\mathbf{v}_2 \in \mathbb{R}^n$ the eigenvector of matrix HD corresponding to the eigenvalue λ_2.

Let $T = H|_{\mathbb{R}^n/\mathbf{v}_2}$ mean the restriction of the action of matrix HD on the factor subspace $\mathbb{R}^n/\mathbf{v}_2$. It is clear that the matrix $T \in \mathbb{R}^{(n-1)\times(n-1)}$ is a matrix of odd order. Therefore among the eigenvalues $\lambda_1, \lambda_3, ..., \lambda_{n=2k}$ of matrix T there is also one real number (for example λ_1). Since $\lambda_2 < 0$, $\lambda_1\lambda_3 \cdot ... \cdot \lambda_{n=2k} > 0$ and the number $n - 1 = 2k - 1$ is odd, then the number λ_1 must be positive.

Thus, we have $\lambda_1\lambda_2 < 0$ and $\lambda_3 \cdot ... \cdot \lambda_{n=2k} > 0$.

(c) Denote by $\mathbf{v}_1 \in \mathbb{R}^n$ the eigenvector of matrix HD corresponding to the eigenvalue λ_1. Now we introduce an invariant subspace $(\mathbf{v}_1, \mathbf{v}_2) \subset \mathbb{R}^n$ spanned by the vectors $\mathbf{v}_1$ and $\mathbf{v}_2$.

We introduce the following matrices

$$G = HD = \begin{pmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} \end{pmatrix},$$

where $G_{11} \in \mathbb{R}^{2\times 2}, G_{12} \in \mathbb{R}^{2\times(n-2)}, G_{21} \in \mathbb{R}^{(n-2)\times 2}, G_{22} \in \mathbb{R}^{(n-2)\times(n-2)}$;

$$S = \left(\begin{array}{c|c} & \mathbf{0} \\ \mathbf{v}_1, \mathbf{v}_2 & --- \\ & I_{n-2} \end{array}\right) = \left(\begin{array}{c|c} S_{11} & \mathbf{0} \\ --- & --- \\ S_{21} & I_{n-2} \end{array}\right) \in \mathbb{R}^{n\times n},$$

where $\mathbf{0} \in \mathbb{R}^{2\times(n-2)}$ is the zero matrix, $I_{n-2} \in \mathbb{R}^{(n-2)\times(n-2)}$ is the identity matrix, $S_{11} \in \mathbb{R}^{2\times 2}, S_{21} \in \mathbb{R}^{(n-2)\times 2}$.

We also compute the inverse matrix

$$S^{-1} = \left(\begin{array}{c|c} S_{11}^{-1} & \mathbf{0} \\ --- & --- \\ -S_{21}S_{11}^{-1} & I_{n-2} \end{array} \right).$$

(Since vectors $\mathbf{v}_1$ and $\mathbf{v}_2$ are linearly independent, then the matrix S^{-1} always exists.)

With the help of the similarity transformation $G \to S^{-1}GS$, the matrix G can be reduced to the form

$$G_1 = S^{-1}GS = \left(\begin{array}{cc|c} \lambda_1 & 0 & \\ 0 & \lambda_2 & S_{11}^{-1}G_{12} \\ -- & -- & ----------- \\ \mathbf{0} & \mathbf{0} & -S_{21}S_{11}^{-1}G_{12} + G_{22} \end{array} \right),$$

where $\mathbf{0} \in \mathbb{R}^{n-2}$ is the zero vector.

After the similarity transformation, the determinant of the matrix remains the same. So we get

$$\det G_1 = \det G = \det(HD) = (\lambda_1\lambda_2) \cdot (\lambda_3 \cdot ... \cdot \lambda_{2k}) < 0. \tag{2.9}$$

Let us introduce the designation $F = -S_{21}S_{11}^{-1}G_{12} + G_{22}$. Since $\lambda_1\lambda_2 < 0$, then $\lambda_3 \cdot ... \cdot \lambda_{2k} = \det(F) > 0$.

(d) Let $\det(\lambda I_{2k} - H)$ be the characteristic polynomial of the matrix H. Then it is obvious that this polynomial contains only even powers of the variable λ: $\det(\lambda I_{2k} - H) = \lambda^{2k} + h_2\lambda^{2k-2} + \cdots + h_{2k-2}\lambda^2 + h_{2k}$, where $h_1 = h_3 = ... = h_{2k-1} = 0$. All roots of this polynomial are calculated using the formulas: $\lambda^2 = p_j < 0, j = 1, ..., k$. Therefore, we have $\lambda_{1,2} = \pm i\sqrt{|p_1|}, \lambda_{3,4} = \pm i\sqrt{|p_3|}, ..., \lambda_{2k-1,2k} = \pm i\sqrt{|p_{2k-1}|}$; $i = \sqrt{-1}$.

Now consider the characteristic polynomial of matrix HD. It is easy to check that all the principal minors of odd orders of matrix HD are equal to zero. Therefore, polynomial $\det(\lambda I_{2k} - HD)$ has the same structure as the characteristic polynomial of the matrix H: $\det(\lambda I_{2k} - HD) = \lambda^{2k} + q_2\lambda^{2k-2} + \cdots + h_{2k-2}\lambda^2 + q_{2k}$. However, its roots can already be either purely imaginary or real (for example, if $\lambda^2 = r_j > 0$, then $\lambda_{2j-1,2j} = \pm\sqrt{r_j}; j \in \{1, ..., k\}$).

(e) Assume thet $n = 2k = 4$. Then from item (c) it follows that $\det(F) = \lambda_3\lambda_4 > 0$. Taking into account everything that has been said in item (d), we have only a single possibility $\lambda_3\lambda_4 = -(ip)^2 > 0$. In addition, since $\lambda_1\lambda_2 < 0$, we have the following result: $\lambda_{1,2} = \pm r$, $\lambda_{3,4} = \pm i|p|$. The proof of the theorem for $n = 4$ is complete.

The proof of the last item in case $n > 4$ is carried out using the same algorithm as in items (b) - (d). Only in this case the role of matrix HD will be played by matrix F. The proof is carried out by the method of induction on the number $n = 2k, k > 2$.

Indeed, suppose that, for example, $k = 3$. We compute the eigenvalues $\lambda_1 = r, \lambda_2 = -r, \lambda_3 = ip, \lambda_4 = -ip$ and construct an invariant $4D$ subspace $\mathbb{W} \subset \mathbb{R}^6$ spanned by the vectors $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4$. (Here $\mathbf{v}_3$ and $\mathbf{v}_4$ are the basis vectors of invariant $2D$ subspace $\mathbb{V} \subset \mathbb{W} \subset \mathbb{R}^6$ which are not eigenvectors of the operator HD; in this case eigenvectors are $\mathbf{v}_3 - i\mathbf{v}_4, \mathbf{v}_3 + i\mathbf{v}_4 \in \mathbb{V} + i\mathbb{V}$.) Now, we pass to the new basis $(\mathbb{W}, \mathbf{z}_5, \mathbf{z}_6)$ in the space $\mathbb{R}^6$. Then we get $F \in \mathbb{R}^{2\times 2}$ and $\det(F) = \lambda_5\lambda_6 > 0$. From item (e) it follows that $\lambda_1 \cdot ... \cdot \lambda_4 < 0$ and λ_3, λ_4 are imaginary. Now, if we assume that $\lambda_5, \lambda_6 \in \mathbb{R}$ and $\lambda_5\lambda_6 = -c^2 < 0$, then we get the contradiction $\det(HD) > 0$ with formula (2.9). Therefore, it must be $\lambda_5\lambda_6 = -(ic)^2 > 0$. In this case formula (2.9) remains valid. Thus, we have obtained that all eigenvalues $\lambda_3, ..., \lambda_6$ are purely imaginary. □

3. On existence of homoclinic orbits in system (2.6)

For system (2.6) we introduce the following notation:

Let $\mathbf{0} = (0, ..., 0)^T \in \mathbb{R}^n$ be the equilibrium point of system (2.6).

Definition 3.1. [8–12, 14, 21, 24, 25]. A bounded trajectory $\mathbf{y}(t, \mathbf{0}) \in \mathbb{R}^n$ of system (2.6) is called a homoclinic orbit if the trajectory converges to the same equilibrium point $\mathbf{0}$ as $t \to \pm\infty$.

Theorem 3.1. *Assume that the following conditions are true for system (2.6):*

(b1) $\sigma_i(y_i) = y_i \cdot (f_i(y_i) + k_i)$, *where* $k_i = const_i \neq 0$ *and* $f_i(y_i)$ *is a continuous function such that*

$$f_i(0) = 0, \lim_{s\to\infty} \int_{-s}^{s} (f_i(\tau) + k_i) \cdot \tau d\tau > 0, i = 1, ..., n$$

(usually, $f_i(\tau)$ *is the even function (2.2));*

(b2) $k_{i_1} > 0, ..., k_{i_m} > 0$ *and* $k_{i_{m+1}} < 0, ..., k_{i_n} < 0$ *(or* $k_{i_1} < 0, ..., k_{i_m} < 0$ *and* $k_{i_{m+1}} > 0, ..., k_{i_n} > 0$*), where* $m \in \{1, ..., n-1\}$ *and* $(i_1, ..., i_n)$ *is a permutation of elements of the set* $\{1, ..., n\}$*;*

(b3) $W = \{w_{ij}\} = W_a + W_d$, *where* W_a *is the antisymmetric nonsingular matrix,* $W_d = \operatorname{diag}(w_{11}, ..., w_{nn})$ *and* $K = \operatorname{diag}(k_1, ..., k_n)$ *are diagonal matrices such that* $\det WK \neq 0$*;*

(b4) $\delta = w_{11}k_1 + \cdots + w_{nn}k_n \leq 0$ *(system (2.6) is dissipative or conservative) and* $WK + K^TW^T < 0$.

Then there exists the vector of initial data $\mathbf{y}_0^*$ *such that the trajectory* $\mathbf{y}(t, \mathbf{y}_0^*)$ *of system (2.6) is the homoclinic orbit connected at equilibrium point* $\mathbf{0}$.

Proof. First let's put $w_{ii} = 0$ ($\delta = 0$). Suppose, for example, that $k_1 > 0$ and $k_2 < 0, ..., k_n < 0$. In this case, the Jacobian matrix has the form

$$J(\mathbf{0}) = W_a \cdot \begin{pmatrix} k_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & k_n \end{pmatrix}$$

and the equilibrium point $\mathbf{0} \in \mathbb{R}^n$ is a saddle point (see Lemma 2.2).

In this case, the function $V(\mathbf{y})$ takes the following form

$$V(y_1, ..., y_n) = \sum_{i=1}^{n} F_i(y_i) + G_i(y_i) + \frac{k_1 y_1^2 + k_2 y_2^2 + \cdots + k_n y_n^2}{2} = C = const. \quad (3.1)$$

(See the proof of Theorem 2.2.)

Denote by $\mathbb{M}_u \subset \mathbb{R}^n$, $\mathbb{M}_s \subset \mathbb{R}^n$, and $\mathbb{M}_c \subset \mathbb{R}^n$ respectively the unstable, stable, and central manifolds of the point $\mathbf{0} \in \mathbb{R}^n$. Then from Lemma 2.2, we have $\dim \mathbb{M}_u = 1$, $\dim \mathbb{M}_s = 1$, and $\dim \mathbb{M}_c = n - 2$.

It is known that the tangent vector to the manifold $\mathbb{M}_u$ at the point $\mathbf{0}$ is the eigenvector $\mathbf{v}_1 = (v_{11}, ..., v_{n1})^T$ corresponding to the eigenvalue $\lambda_1 > 0$ of matrix $J(\mathbf{0})$. Similarly, there is also a second eigenvector $\mathbf{v}_2 = (v_{12}, ..., v_{n2})^T$ corresponding to the eigenvalue $\lambda_2 = -\lambda_1 < 0$ of matrix $J(\mathbf{0})$ [18] (see also Lemma 2.2, where $HD = J(0)$).

Let $\mathbb{B}(\epsilon)$ be an open ball with center at point $\mathbf{0}$ and radius $\epsilon > 0$ such that the set $\mathbb{L} \bigcap \mathbb{B}(\epsilon)$ does not contain other invariant subsets except $\mathbf{0}$. (The set $\mathbb{L}$ is defined in the proof of Theorem 2.2.)

In formula (3.1) we put the value $C = 0$. Then, by multiplying the eigenvector $\mathbf{v}_1$ by the corresponding parameter $\delta_1 > 0$, we will have $|V(\delta_1 \mathbf{v}_1)| = |V(\delta_1 v_{11}, ..., \delta_1 v_{n1})| < \epsilon$. Let's also choose parameter $\delta_2 > 0$ so that $|V(\delta_2 \mathbf{v}_2)| < \epsilon$.

In what follows we will need one Poincare result.

It is known that for any initial condition $\mathbf{y}_0$ the solution $\mathbf{y}(t)$ of system (2.6) can be uniquely determined by the formula $\mathbf{y}(t) = \mathbf{P}^t(\mathbf{y}_0)$, where $\mathbf{P}^t$ is the evolution operator.

Theorem 3.2. *(Poincare's Recurrence Theorem [22]). Let $\mathcal{H} \subset \mathbb{R}^n$ be a compact set that is positively invariant with respect to (2.6). In other words, $\forall \mathbf{y}_0 \in \mathcal{H}$, we have*

$$(\mathbf{P}^t)^l(\mathbf{y}_0) = \underbrace{\mathbf{P}^t(\mathbf{P}^t(...(\mathbf{P}^t(\mathbf{y}_0))))}_{l} \in \mathcal{H}; \forall l \in \mathbb{N}, \forall t > 0.$$

Then from here it follows that for any $\epsilon > 0$ and almost all points $\mathbf{y}_0 \in \mathcal{H}$ there is a time $\tau > 0$ and a natural number $k \geq 1$ such that $\|(\mathbf{P}^\tau)^k(\mathbf{y}_0) - \mathbf{y}_0\| < \epsilon$.

Thus, Theorem 3.2 states that in the phase space $\mathcal{H}$ of a dissipative system, any trajectory starting from almost any point $\mathbf{y}_0 \in \mathcal{H}$ of this space will pass arbitrarily close to $\mathbf{y}_0$ after some finite time (even if it is very large). (In our case $\mathcal{H}$ is a set of points from $\mathbb{R}^n$ such that $\forall \mathbf{y} \in \mathcal{H}$ we have $\dot{V}(\mathbf{y}) \leq 0$; in addition

$\mathbf{y}_0 = \mathbf{v}_1 \in \mathcal{H}$, $\mathbf{y}_0 \neq \mathbf{0}$, and $\|\mathbf{v}_1\| \to 0$. Then for any sufficiently small $\epsilon > 0$, we have $\|(\mathbf{P}^t)^l(\mathbf{y}_0) - \mathbf{v}_1\| < \epsilon$ or $\lim\limits_{l\to\infty} \|(\mathbf{P}^t)^l(\mathbf{y}_0) - \mathbf{v}_1\| = 0$, where $(\mathbf{P}^t)^l(\mathbf{y}_0) \to \mathbf{v}_2 \in \mathcal{H}$ and $\|\mathbf{v}_2\| \to 0$.)

Let's assume that $\delta_1\mathbf{v}_1 \in \mathbb{B}(\epsilon)$ and $\delta_2\mathbf{v}_2 \in \mathbb{B}(\epsilon)$. The terms under the summation sign in formula (3.1) at $\mathbf{y}(t) \to \mathbf{0}$ are of a higher order of smallness than the quadratic terms in the same formula. Therefore, if $\delta_2 \to 0$, then trajectory $\mathbf{y}(t, \delta_2\mathbf{v}_2)$ will approach the surface of cone $\mathbb{K}$: $k_1y_1^2 + k_2y_2^2 + \cdots + k_ny_n^2 = 0$. (Here one of the parameters $k_i \in \{k_1, ..., k_n\}$ is negative.)

Now assume that $\forall i \in \{1, ..., n\} |w_{ii}| < \epsilon$ and there exist $j_1, ..., j_k \in \{1, ..., n\}$ such that $w_{j_1j_1} \neq 0, ..., w_{j_kj_k} \neq 0$. Let's return to the condition *(b4)* of Theorem 3.1. According to Theorem 2.2, all solutions of system (2.6) are bounded. Therefore, if $w_{ii}k_i \approx 0; i = 1, ..., n$, then the structure of sets $\mathbb{M}_u, \mathbb{M}_s$, and $\mathbb{M}_c$ will remain almost the same. Thus, we have $\mathbb{M}_u \cup \mathbb{M}_s \cup \mathbb{M}_c \subset \mathbb{K}$. This means that if vector $\delta_1\mathbf{v}_1$ (or $\delta_2\mathbf{v}_2$) is close enough to $\mathbb{M}_u$ (or $\mathbb{M}_s$), then trajectory $\mathbf{y}(t, \delta_1\mathbf{v}_1)$ (or $\mathbf{y}(t, \delta_2\mathbf{v}_2)$) will be close enough to cone $\mathbb{K}$. It is clear that if $\delta_1 \to 0, \delta_2 \to 0$, and $t \to \infty$, then the trajectory $\mathbf{y}(t, \delta_1\mathbf{v}_1)$ approaches the trajectory $\mathbf{y}(t, \delta_2\mathbf{v}_2)$. Thus, we obtain that both of these trajectories are attracted to a certain equilibrium point. But in region $\mathbb{B}(\epsilon)$ there is no other equilibrium position except the origin. Therefore, the only possibility remains: trajectories $\mathbf{y}(t, \delta_1\mathbf{v}_1)$ and $\mathbf{y}(t, \delta_2\mathbf{v}_2)$ must pass fairly close to the origin. In the limiting case, the last statement looks like this

$$\lim_{\delta\to 0, t\to\infty} \mathbf{y}(t, \delta\mathbf{v}_1) \to \mathbb{M}_s \to \mathbf{0}, \lim_{\delta\to 0, t\to -\infty} \mathbf{y}(t, \delta\mathbf{v}_2) \to \mathbb{M}_u \to \mathbf{0}.$$

Now let's replace $t \to -t$ in system (2.6). Obviously, in this case $\lambda_1 \to \lambda_2 = -\lambda_1$, $\lambda_2 \to \lambda_1 = -\lambda_2$; $\mathbf{v}_1 \to \overline{\mathbf{v}}_2, \mathbf{v}_2 \to \overline{\mathbf{v}}_1$, $\mathbb{M}_u \to \overline{\mathbb{M}}_s$, $\mathbb{M}_s \to \overline{\mathbb{M}}_u$ ($\dim \overline{\mathbb{M}}_u = \dim \overline{\mathbb{M}}_s = 1$) and the last limit equality is transformed into

$$\lim_{\delta\to 0, t\to -\infty} \mathbf{y}(t, \delta\overline{\mathbf{v}}_2) \to \overline{\mathbb{M}}_u \to \mathbf{0}, \lim_{\delta\to 0, t\to\infty} \mathbf{y}(t, \delta\overline{\mathbf{v}}_1) \to \overline{\mathbb{M}}_s \to \mathbf{0}.$$

This means that $\mathbb{M}_s \cap \mathbb{M}_u \neq \emptyset$ (or $\overline{\mathbb{M}}_s \cap \overline{\mathbb{M}}_u \neq \emptyset$). Therefore, in system (2.6) there exists the homoclinic orbit connected at equilibrium point $\mathbf{0}$.

Theorem 3.1 has been proved for case $\dim \mathbb{M}_u = \dim \mathbb{M}_s = 1$ and $\dim \mathbb{M}_c = n - 2$. It is easy to check that its generalization to case $\dim \mathbb{M}_u = \dim \mathbb{M}_s = l$ and $\dim \mathbb{M}_c = n - 2l, 1 < l < n/2$ is not difficult. The last remark completes the proof. □

If we now return from the variable $\mathbf{y}$ to the variable $\mathbf{x} = W^{-1}\mathbf{y} - W^{-1}\mathbf{b}$, we obtain the same Theorem 3.1 for the system (2.4), but with one clarification: the homoclinic orbit will be connected at the equilibrium point $\mathbf{x}_0 = -W^{-1}\mathbf{b}$.

The application of Theorem 3.1 are demonstrated on Fig.3.1. (Other forms of homoclinic orbits in systems containing module nonlinearities are indicated in [10].)

Theorem 3.1 admits the following generalization.

Let us introduce the piecewise continuous function (see [9, 10, 13]).

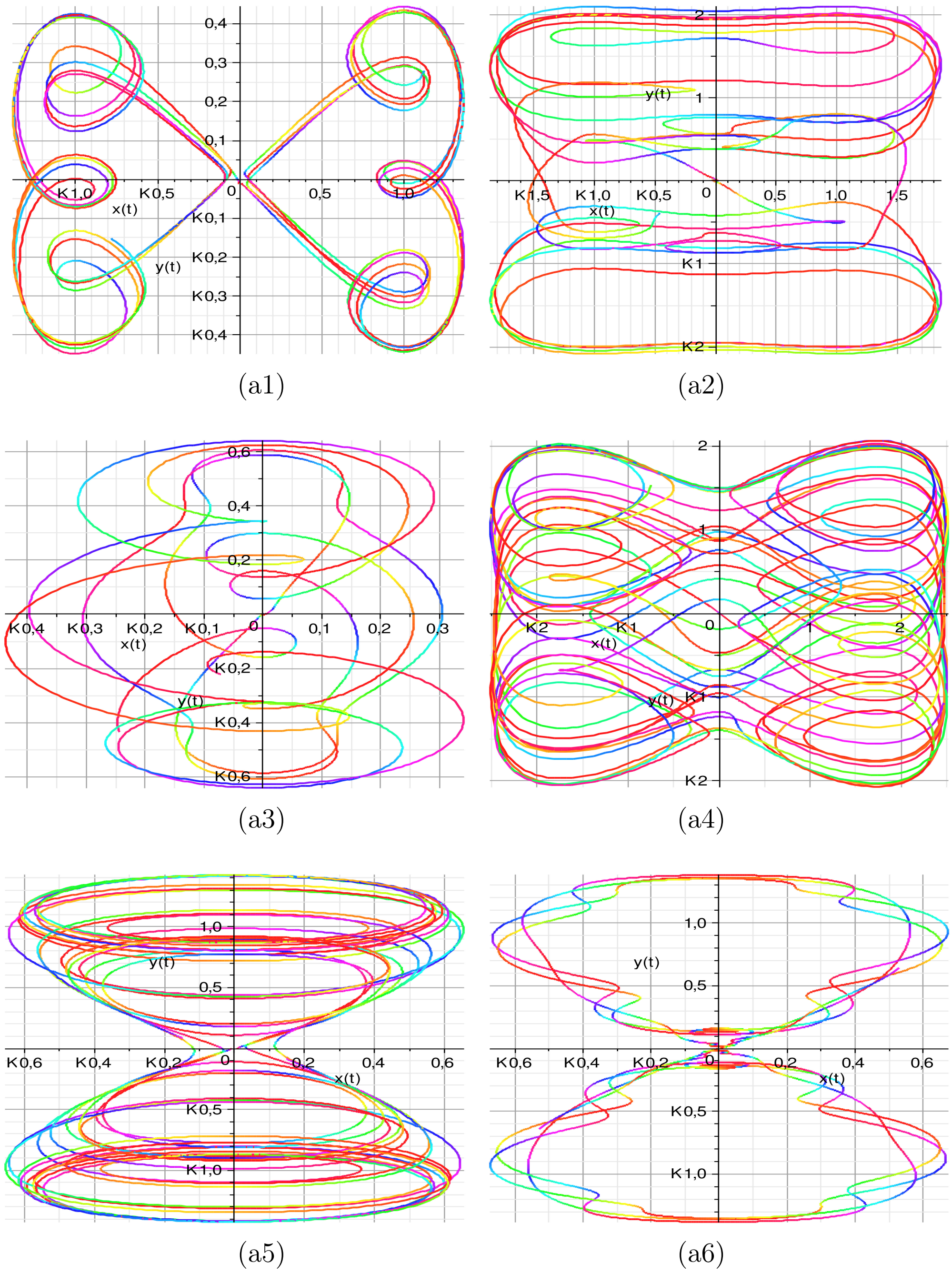


Fig. 3.1. Graphs of 2D projections of homoclinic trajectory of system (2.6) connected at point $\mathbf{0}$ and $\sigma_i(y_i) = y_i^3 + k_i y_i, i = 1, ..., n = 4$, for the following parameter values (see Theorem 3.1): $a_{11} = ... = a_{44} = 0$, $a_{12} = -1, a_{13} = a_{14} = 1, a_{23} = a_{24} = 0, a_{34} = 1$; (a1) $k_1 = -1, k_2 = 2, k_3 = 4, k_4 = 1$; (a2) $k_1 = -1, k_2 = -2, k_3 = 1, k_4 = -1$; (a3) $k_1 = 2, k_2 = 1, k_3 = 1, k_4 = -1$; (a4) $k_1 = -3, k_2 = -1, k_3 = -1, k_4 = 1$; (a5) $k_1 = 1, k_2 = -1, k_3 = -0.2, k_4 = -0.1$; (a6) $k_1 = 0.4, k_2 = -1, k_3 = 3, k_4 = 0.3$.

Definition 3.2. Let $\alpha \neq 1$, $\beta \neq 1$, and $c > 0$ be real constants. The function

$$g(w,\alpha,\beta,c) = \begin{cases} -\frac{\beta-1}{\beta}c^{\frac{\beta}{\beta-1}} - \frac{(-w)^\beta}{\beta}, & \text{if } w < -c^{\frac{1}{\beta-1}} \\ cw, & \text{if } w \leq c^{\frac{1}{\alpha-1}} \\ \frac{\alpha-1}{\alpha}c^{\frac{\alpha}{\alpha-1}} + \frac{w^\alpha}{\alpha}, & \text{otherwise} \end{cases} \tag{3.2}$$

is called an odd generalized power activation function.

Theorem 3.3. *Under the conditions of Theorem 3.1, if we replace the function* $\sigma_i(y_i) = y_i \cdot (f_i(y_i) + k_i)$ *with the function* $\sigma_i(y_i) = g(y_i, \alpha_i, \beta_i, c_i) \cdot (f_i(y_i) + k_i)$ *(see (3.2)),* $i = 1, ..., n$*, then the statement of Theorem 3.1 remains valid.*

Proof. Indeed, the Jacobian matrix of system (2.6) in a sufficiently small neighborhood of the origin for activation functions $\sigma_i(y_i) = y_i \cdot (f_i(y_i) + k_i)$ and $\sigma_i(y_i) = g(y_i, \alpha_i, \beta_i, c_i) \cdot (f_i(y_i) + k_i)$ is the same. Now, the Grobman-Hartman theorem (see Theorem 9.9 [24]) can be applied to system (2.6). After that, it remains to repeat the proof of Theorem 3.1, in which the function $\sigma_i(y_i) = g(y_i, \alpha_i, \beta_i, c_i) \cdot (f_i(y_i) + k_i)$ is used instead of the function $\sigma_i(y_i) = y_i \cdot (f_i(y_i) + k_i)$, $i = 1, ..., n$. □

The application of Theorem 3.3 are demonstrated on Fig.3.2.

As shown in [9,10,13] if $\alpha \to 1$, $\beta \to 1$, and $c = 1$, then we have $g(w, 1, 1, 1) \to w$. Therefore, in this case, in Theorem 3.1, we have $w_i \to y_i, i = 1, ..., n$.

Note that in the above example, the parameters α and β are positive. Nevertheless, the activation function $g(w, \alpha, \beta, c)$ with parameters $\alpha < 0$, $\beta < 0$, and $c > 0$ retained all the main properties that it had with positive parameters. Thus, when designing neural networks, the capabilities of the function $g(w, \alpha, \beta, c)$ can be significantly expanded. In particular, we have:

$$\lim_{w\to\infty,\alpha<0} g(w,\alpha,\beta,c) = \frac{\alpha-1}{\alpha}c^{\frac{\alpha}{\alpha-1}} < \infty,$$

$$\lim_{w\to-\infty,\beta<0} g(w,\alpha,\beta,c) = -\frac{\beta-1}{\beta}c^{\frac{\beta}{\beta-1}} > -\infty.$$

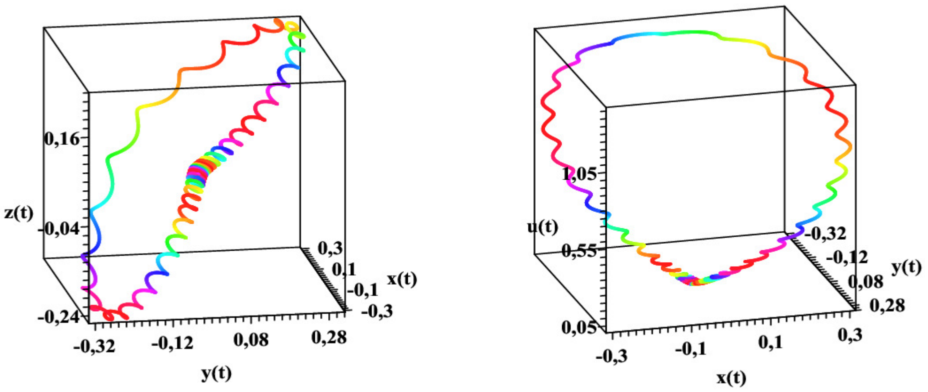
Moreover, if $\alpha < 0$ and $\beta < 0$, then we have

$$-\frac{\beta-1}{\beta}c^{\frac{\beta}{\beta-1}} < g(w,\alpha,\beta,c) < \frac{\alpha-1}{\alpha}c^{\frac{\alpha}{\alpha-1}}$$

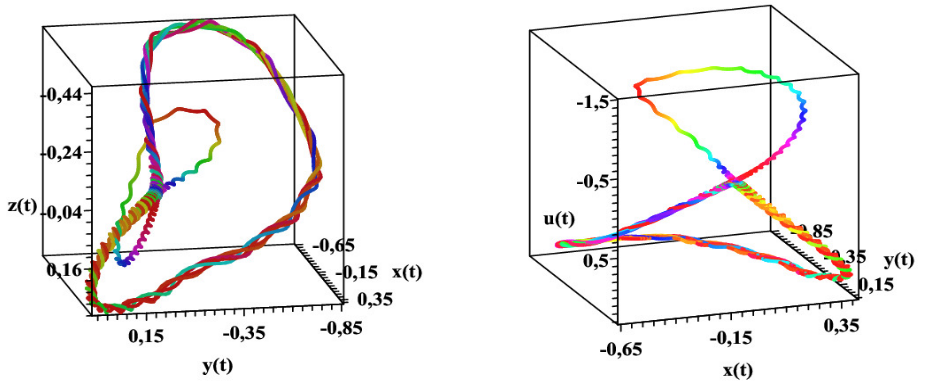
and the function $g(w, \alpha, \beta, c)$ is bounded. In this case, we can say that the function $g(w, \alpha, \beta, c)$ is similar to the function $\tanh(cw)$.

4. On existence of heteroclinic orbits in system (2.6)

Definition 4.1. [23–25]. A bounded trajectory $\mathbf{y}(t, \mathbf{e}_1) \in \mathbb{R}^n$ (or $\mathbf{y}(-t, \mathbf{e}_2) \in \mathbb{R}^n$) of system (2.6) is called heteroclinic if it belongs to the intersection of the stable



1. $\sigma(y_i) = g(y_i, \alpha, \beta, c) \cdot (|y_i| + k_i), i = 1, ..., 4;\ c = 2.5$ (see (3.2))



2. $\sigma(y_i) = g(y_i, \alpha, \beta, c) \cdot (y_i^2 + k_i), i = 1, ..., 4;\ c = 9$ (see (3.2))

Fig. 3.2. Graphs of 3D projections of homoclinic trajectory of system (2.6) connected at point $\mathbf{0}$ at $n = 4$ for the following parameter values: $a_{11} = a_{22} = a_{33} = a_{44} = 0., a_{12} = -3, a_{13} = 2, a_{14} = -1, a_{23} = -2, a_{24} = 0, a_{34} = 1$. In addition: $\alpha = 1.8, \beta = 0.4$; $k_1 = 1, k_2 = 1, k_3 = 2, k_4 = -1$.

invariant manifold of one equilibrium point $\mathbf{e}_1$ (or $\mathbf{e}_2$) with the unstable invariant manifold of another equilibrium point $\mathbf{e}_2$ (or $\mathbf{e}_1$). In addition, in this case the conditions

$$\lim_{t\to\infty} \mathbf{y}(t, \mathbf{e}_1) = \mathbf{e}_2, \lim_{t\to-\infty} \mathbf{y}(t, \mathbf{e}_2) = \mathbf{e}_1.$$

must be satisfied.

Theorem 4.1. *Assume that the following conditions are true for system (2.6):*

(c1) $\sigma_i(y_i) = g_i(y_i) \cdot (f_i(y_i) + k_i)$, *where* $k_i = const_i < 0$, $g_i(y_i)$ *is a differentiable increasing* ($g_i^{'}(y_i) > 0$) *function (2.1), and* $f_i(y_i)$ *is the even function (2.2)*

such that the equation $f_i(y_i) + k_i = 0$ *has exactly two real roots* $m_{i1} = r_i$ *and* $m_{i2} = -r_i$ *of different signs;* $i = 1, ..., n$.

(c2) $W = \{w_{ij}\} = W_a + W_d$, *where* W_a *is the antisymmetric nonsingular matrix,* $W_d = \mathrm{diag}(w_{11}, ..., w_{nn})$ *and* $K = \mathrm{diag}(k_1, ..., k_n)$ *are diagonal matrices such that* $\delta = w_{11}k_1 + \cdots + w_{nn}k_n < 0$ *(system (2.6) is dissipative) and the value* $|\delta|$ *is sufficiently small.*

Then there exists the vector of initial data $\mathbf{y}_0^*$ *such that the trajectory* $\mathbf{y}(t, \mathbf{y}_0^*)$ *of system (2.6) is the heteroclinic orbit connecting points* $(0, ..., 0, y_{j1}, 0, \ ..., 0)^T \in \mathbb{R}^n$ *and* $(0, ..., 0, y_{j2}, 0, ..., 0)^T \in \mathbb{R}^n$; $j \in \{1, ..., n\}$.

Proof. It is obvious that the coordinates of any equilibrium point of system (2.6) are determined by condition

$$g_i(y_i) \cdot (f_i(y_i) + k_i) = g_i(y_i) \cdot (y_i + m_{i1}) \cdot (y_i + m_{i2}) \cdot (\psi_i(y_i) + 1)$$

$$= g_i(y_i) \cdot (y_i + r_i) \cdot (y_i - r_i) \cdot (\psi_i(y_i) + 1) = 0,$$

where $k_i = m_{i1}m_{i2} = -r_i^2 < 0$, $\psi_i(0) = 0$, and $\forall y_i \in \mathbb{R}\ (\psi_i(y_i)+1) \cdot g_i(y_i) \neq 0; i = 1, ..., n$.

Without loss of generality, we can consider that $g_i(y_i)$ is the function (3.2) and $\psi_i(y_i) + 1 = 1; i = 1, ..., n$.

From here it follows that for an arbitrary natural number n, we have 3^n equilibrium points. (Note that if for at least one $j \leq n$, we have $k_j > 0$, then the number of all equilibrium points will be less than 3^n.)

Now for the equilibrium point $\mathbf{e}_0 = (0, ..., 0)^T \in \mathbb{R}^n$, we write the Jacobi matrix

$$J(\mathbf{e}_0) = W \cdot \begin{pmatrix} g_1'(0)k_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & g_n'(0)k_n \end{pmatrix}.$$

We also calculate the Jacobian matrices for the equilibrium points $\mathbf{e}_{i1} = (\underbrace{0, ..., 0, -m_{i1}}_{i}, 0...0)^T \in \mathbb{R}^n$ and $\mathbf{e}_{i2} = (\underbrace{0, ..., 0, -m_{i2}}_{i}, 0...0)^T \in \mathbb{R}^n$; $i = 1, ..., n$:

$$J(\mathbf{e}_{11}) = W \cdot \begin{pmatrix} g_1(-m_{11})f_1'(-m_{11}) & 0 & \dots & 0 \\ 0 & g_2'(0)k_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & g_n'(0)k_n \end{pmatrix},$$

$$J(\mathbf{e}_{12}) = W \cdot \begin{pmatrix} g_1(-m_{12})f_1'(-m_{12}) & 0 & \dots & 0 \\ 0 & g_2'(0)k_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & g_n'(0)k_n \end{pmatrix},$$

....

$$J(\mathbf{e}_{n1}) = W \cdot \begin{pmatrix} g_1^{'}(0)k_1 & \dots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \dots & g_{n-1}^{'}(0)k_{n-1} & 0 \\ 0 & \dots & 0 & g_n(-m_{n1})f_n^{'}(-m_{n1}) \end{pmatrix},$$

$$J(\mathbf{e}_{n2}) = W \cdot \begin{pmatrix} g_1^{'}(0)k_1 & \dots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \dots & g_{n-1}^{'}(0)k_{n-1} & 0 \\ 0 & \dots & 0 & g_n(-m_{n2})f_n^{'}(-m_{n2}) \end{pmatrix}.$$

(a) Let $W_d = 0$. Since $g_1^{'}(0) > 0, ..., g_n^{'}(0) > 0$, then according to Lemma 2.1, the point $\mathbf{e}_0$ is a center (the matrix $J(\mathbf{e}_0) = WK$ has purely imaginary eigenvalues).

First of all, we note that the function $g_i(y_i)$ is odd and the function $f_i(y_i)$ is even. Therefore, the function $g_i^{'}(y_i)$ is even and the function $f_i^{'}(y_i)$ is odd [8]. From here it follows that $g_i(y_i)f_i^{'}(y_i) > 0$, $g_i^{'}(y_i)f_i(y_i) > 0$ and $g_i^{'}(y_i)k_i < 0; i = 1, ..., n$.

Further, the Jacobian matrices $J(\mathbf{e}_{11}), ..., J(\mathbf{e}_{n2})$ in all equilibrium points have the form WD_j, where the diagonal matrix D_j has only one positive diagonal element $g_j(y_j)f_j^{'}(y_j)$; all other diagonal elements $g_i^{'}(y_i)k_i, i \neq j$, are negative. Thus, according to Lemma 2.2, the matrix WD_j has a pair of eigenvalues of different signs; all other eigenvalues will be purely imaginary. From here it follows that all equilibrium points $\mathbf{e}_{11}, ..., \mathbf{e}_{n2}$ are saddle points.

Given that $|m_{i1}| = |m_{i2}|$, we can introduce the notation

$$d_{\min} = \min_{i \in \{1,...,n\}} \|\mathbf{e}_{i1}\| = \min_{i \in \{1,...,n\}} \|\mathbf{e}_{i2}\| = \|\mathbf{e}_{k1}\| = \|\mathbf{e}_{k2}\|.$$

Here $d_{\min}$ is a length of vector $\mathbf{e}_{k1}$ (or vector $\mathbf{e}_{k2}$); $k \in \{1, ..., n\}$.

Now let's use formula (3.1), which in this situation will have the following form:

$$V(y_1, ..., y_n) = \sum_{i=1}^{n} F_i(y_i) + G_i(y_i) - \frac{r_1^2 y_1^2 + \dots + r_n^2 y_n^2}{2} = C = const > 0. \quad (4.1)$$

Let $\mathbf{y}_{b1}, ..., \mathbf{y}_{bk} \in \mathbb{R}^n$ be a sequence of initial vectors such that $\|\mathbf{y}_{b1}\| \leq ... \leq \|\mathbf{y}_{bk}\| = d_{\min}$. Then from the previous equality (4.1) it follows that the solution $\mathbf{y}(t, \mathbf{y}_{bi})$ of system (2.6) is bounded and located around the point $\mathbf{0}$; $i = 1, ..., k$. From here it follows that

$$\min_{t \geq 0} \|\mathbf{y}(t, \mathbf{y}_{b1}) - \mathbf{e}_{k1}\| \geq ... \geq \min_{t \geq 0} \|\mathbf{y}(t, \mathbf{y}_{bk}) - \mathbf{e}_{k1}\| = 0.$$

The point $\mathbf{e}_{k1}$ satisfies equation (4.1) and it is a saddle point. Its unstable manifold $\mathbb{M}_{u,k1}$ and stable manifold $\mathbb{M}_{s,k1}$ are one-dimensional (see Lemma 2.2).

Obviously, if $d_{\min} - \|\mathbf{y}_{bi}\| \to 0$, then $\min_{t \geq 0} \|\mathbf{y}(t, \mathbf{y}_{bi}) - \mathbf{e}_{k1}\| \to 0$ along its stable one-dimensional manifold $\mathbb{M}_{s,k1}$; $i = 1, ..., k$.

Equation (4.1) contains only even functions, so the point $\mathbf{e}_{k2}$ has the same properties as the point $\mathbf{e}_{k1}$. This remark leads to the following conclusion.

Let $\mathbf{y}_{bi} \in \mathbb{M}_{u,k1}$. Then, due to the fact that in the neighborhood of point $\mathbf{0}$, there are no other equilibrium points except $\mathbf{e}_{k1}$ and $\mathbf{e}_{k2}$, the trajectory $\mathbf{y}(t, \mathbf{y}_{bi})$ should be attracted to point $\mathbf{e}_{k2}$ along its stable one-dimensional manifold $\mathbb{M}_{s,k2}$; $i = 1, ..., k$. Thus, we have $\mathbb{M}_{u,k1} \cap \mathbb{M}_{s,k2} \neq \emptyset$. (It is easy to check that the stronger statement $\mathbb{M}_{u,k1} = \mathbb{M}_{s,k2}$ is actually true.) This means that there is a heteroclinic orbit connecting points $\mathbf{e}_{k1}$ and $\mathbf{e}_{k2}$.

(b) Now let $W_d \neq 0$. We choose the elements of the matrix W_d so small that the equilibrium point $\mathbf{0}$ is a stable focus (it can be assumed that $w_{11} < 0, ..., w_{nn} < 0$). By $\mathbb{S} \subset \mathbb{R}^n$ denote the invariant set of points satisfying the equation $\dot{V}_t(y_1, ..., y_n) = \sigma^T(\mathbf{y})(W + W^T)\sigma(\mathbf{y}) = w_{11}(g_1(y_1))^2 \cdot (f_i(y_i) + k_i)^2 + \cdots + w_{nn}(g_n(y_n))^2 \cdot (f_n(y_n) + k_n)^2 = 0$ [18]. (Since $w_{11} < 0, ..., w_{nn} < 0$, then $\mathbb{S}$ is bounded (see Theorem 2.2).)

Obviously, the variety $\mathbb{S}$ contains the points $\mathbf{e}_{k1}, \mathbf{0}$, and $\mathbf{e}_{k2}$.

Now if we start from the point $\mathbf{e}_{k1} \in \mathbb{S}$ along the unstable manifold $\mathbb{M}_{u,k1}$, then the trajectory $\mathbf{y}(t, \mathbf{e}_{k1})$ will be attracted to point $\mathbf{e}_{k2} \in \mathbb{S}$ along the stable manifold $\mathbb{M}_{s,k2}$. A similar situation take places when replacing the time sign: $t \to -t$. In this case, if we start from the point $\mathbf{e}_{k2} \in \mathbb{S}$ along the unstable manifold $\mathbb{M}_{u,k2}$, then the trajectory $\mathbf{y}(t, \mathbf{e}_{k2})$ will be attracted to point $\mathbf{e}_{k1} \in \mathbb{S}$ along the stable manifold $\mathbb{M}_{s,k1}$. (Trajectories $\mathbf{y}(t, \mathbf{e}_{k1})$ and $\mathbf{y}(t, \mathbf{e}_{k2})$ cannot be attracted to the point $\mathbf{0}$ because when the time sign is changed, the point $\mathbf{0}$ becomes repelling. At the same time, the trajectories $\mathbf{y}(t, \mathbf{e}_{k1})$ and $\mathbf{y}(t, \mathbf{e}_{k2})$ only change direction.)

Since the manifolds $\mathbb{M}_{u,k1}$, $\mathbb{M}_{s,k1}$ and $\mathbb{M}_{u,k2}$, $\mathbb{M}_{s,k2}$ are one-dimensional, then according to Lassalle's Theorem 2.1, we have

$$\lim_{t\to\infty} \mathbf{y}(t, \mathbf{e}_{k1}) = \mathbf{e}_{k2}, \lim_{t\to-\infty} \mathbf{y}(t, \mathbf{e}_{k2}) = \mathbf{e}_{k1}.$$

This means that the points $\mathbf{e}_{k1}$ and $\mathbf{e}_{k2}$ are connected by a heteroclinic orbit. □

The application of Theorem 4.1 are demonstrated on Fig.4.1.

Definition 4.2. System (2.6) satisfying the conditions of Theorem (3.1) (Theorem (4.1)) is called a system of neural ordinary differential equations (NODE) of homoclinic (heteroclinic) type.

5. Architecture of homoclinic and heteroclinic neural networks

Consider the following generalization of equation (2.4):

$$\dot{\mathbf{x}}(t) = \sigma_+(W\mathbf{x}(t) + \mathbf{b}) + \sigma_-(W\mathbf{x}(t) + \mathbf{b}). \tag{5.1}$$

Since $\det W \neq 0$, then using the substitution $\mathbf{y} = W\mathbf{x} + \mathbf{b}$ we can reduce equation (5.1) to the following form:

$$\dot{\mathbf{y}}(t) = W \cdot \Big(\sigma_+(\mathbf{y}(t)) + \sigma_-(\mathbf{y}(t))\Big). \tag{5.2}$$

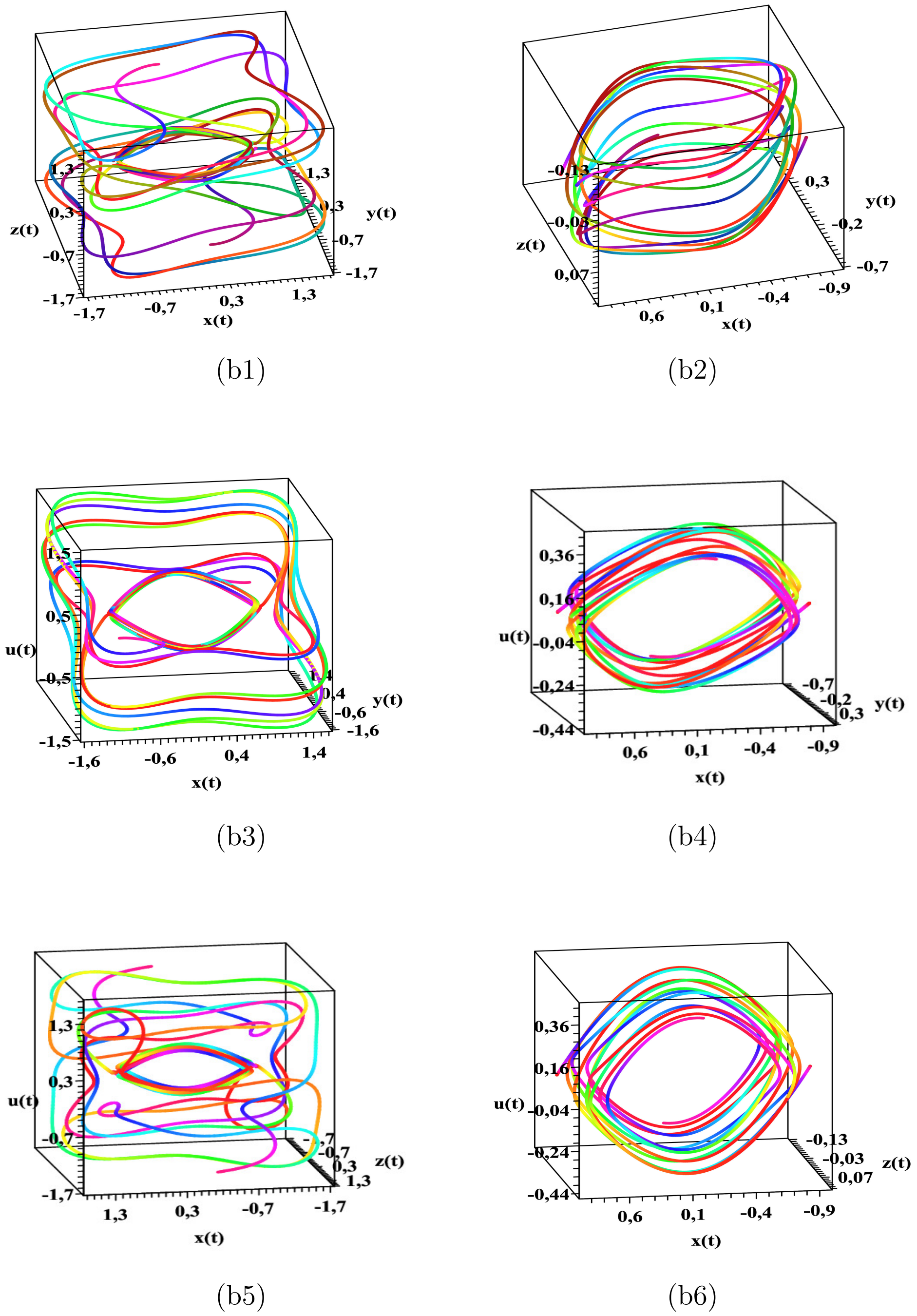


Fig. 4.1. Graphs of 3D projections of heteroclinic trajectory of system (2.6) between points $(1,0,0,0)^T$ and $(-1,0,0,0)^T$ for $\sigma_i(y_i) = y_i^3 + k_i y_i$ (here $g_i(y_i) = y_i$); $i = 1, ..., n = 4$, and at the following parameter values: $a_{11} = ... = a_{44} = 0$, $a_{12} = -3, a_{13} = 0, a_{14} = -2, a_{23} = 1, a_{24} = 0, a_{34} = 1$; $k_1 = k_2 = k_3 = k_4 = -1$. Here on figures (b1), (b3), (b5) $t \to \infty$, and on figures (b2), (b4), (b6) $t \to -\infty$.

Here W is the antisymmetric matrix of even order,

$$\sigma_{+}(\mathbf{y}) = (g(y_1, \alpha_1, \beta_1, c_1) \cdot f(y_1), ..., g(y_n, \alpha_1, \beta_1, c_1) \cdot f(y_n))^T \in \mathbb{R}^n \tag{5.3}$$

(see Definition 3.2) and $f(w)$ is the even function (see (2.2));

$$\sigma_{-}(\mathbf{y}) = (k_1 g(y_1, \alpha_2, \beta_2, c_2), ..., k_i g(y_i, \alpha_2, \beta_2, c_2), ..., k_n g(y_n, \alpha_2, \beta_2, c_2))^T \in \mathbb{R}^n \tag{5.4}$$

and for the vector $\sigma_{-}(\mathbf{y})$ real constants $k_1, ..., k_n$ are assigned arbitrarily.

Let us make some comments on the structure of system (5.2).

We denote by $\mathbb{B}$ a sufficiently small neighborhood of the origin. Then in this neighborhood all equations of system (5.2) have the following form: $\dot{y}_i = y_i \cdot (c_1 f(y_i) + k_i c_2), i = 1, ..., n$. Here a vector of initial conditions $(y_{10}, ..., y_{n0})^T \in \mathbb{B}$.

Since $c_1 > 0$ and $c_2 > 0$, then if $k_i < 0$ the equation $c_1 f(y_i) + k_i c_2 = 0$ has at least 2 real roots, and if $k_i > 0$ the equation $c_1 f(y_i) + k_i c_2 = 0$ has no real roots at all (if $k_i \neq 0$.) A similar situation is described in Theorems 3.1, 3.3, and 4.1. Consequently, for some sets of signs, homoclinic and heteroclinic structures will be encountered among the solutions of system (5.2). This allows us to model dynamic processes with a fairly complex chaotic behavior.

The previous arguments in this section can be given a more rigorous character.

Theorem 5.1. *Assume that for system (5.1) the following conditions:*

(d1) W *is the antisymmetric nonsingular matrix of even order;*

(d2) coordinates of vector (5.3) are the products of odd functions $g(w, \alpha_1, \beta_1, c_1)$ *(3.2) and even functions* $f(w)$ *(2.2), and coordinates of vector (5.4) are odd functions* $g(w, \alpha_2, \beta_2, c_2)$ *(3.2);*

(d3) $\forall i \in \{1, ..., n\}$ $\lim_{u\to\infty} \int\limits_{-u}^{u} [g(w, \alpha_1, \beta_1, c_1) \cdot f(w) + k_i g(w, \alpha_2, \beta_2, c_2)] dw > 0$

are satisfied. Then for system (5.1) there exists nonempty compact invariant set.

Proof. Using functions (5.2), (5.3), and (5.4) we introduce the derivative $\dot{V}_t(\mathbf{y})$ of function $V(\mathbf{y})$ as follows:

$$\dot{V}_t(\mathbf{y}) = (\sigma_{+}(\mathbf{y}) + \sigma_{-}(\mathbf{y}))^T \dot{\mathbf{y}} = (\sigma_{+}(\mathbf{y}) + \sigma_{-}(\mathbf{y}))^T W^T W (\sigma_{+}(\mathbf{y}) + \sigma_{-}(\mathbf{y})).$$

According to condition (d1) we have $\dot{V}_t(\mathbf{y}) = 0$. Now we integrate the last relation taking into account conditions (d2) and (d4). Then we will have

$$V(\mathbf{y}) = \lim_{y_i\to\infty} \int\limits_{-y_i}^{y_i} (\sigma_{+}(y_i) + \sigma_{-}(y_i)) dy_i$$

$$= \lim_{y_i\to\infty} \int\limits_{-y_i}^{y_i} [g(w_i, \alpha_1, \beta_1, c_1) \cdot f(w_i) + k_i g(w_i, \alpha_2, \beta_2, c_2)] dw_i > 0.$$

This means that the set $\mathbb{L} := \{\mathbf{y}(t)|\dot{V}_t(\mathbf{y}) = 0\}$ is compact and it is the invariant set for system (5.2). Then, from here it follows that the set $\mathbb{D} := \{\mathbf{y}(t)$ $|\dot{V}_t(\mathbf{y}) \leq 0\}$ has the boundary $\mathbb{L}$ and therefore it is also the compact invariant set for system (5.2) (see LaSalle's Theorem 2.1). If we now return from variable $\mathbf{y}$ to variable $\mathbf{x} = W^{-1}\mathbf{y} - W^{-1}\mathbf{b}$, we get the same statement for system (5.1). □

It can be shown that if $\alpha \to 1, \beta \to 1$, and $c \to 1$, then the activation function (3.2) is transformed into the function

$$\lim_{\alpha\to1,\beta\to1,c\to1} g(w,\alpha,\beta,c) = w. \tag{5.5}$$

(This is the limiting version of function (3.2).)

In this case, on the one hand, equation (5.1) can be simplified; on the other hand, in this equation the number of adjustable parameters can be expanded by replacing the matrix $W \cdot \operatorname{diag}(k_1, ..., k_n)$ with an arbitrary matrix $A \in \mathbb{R}^{n\times n}$.

Now if we replace in equation (5.1) the activation function (5.4) with its limiting analogue (5.5), we get

$$\dot{\mathbf{x}}(t) = \sigma_+(B\mathbf{x}(t) + \mathbf{b}) + A\mathbf{x}(t) + \mathbf{a}, \tag{5.6}$$

where $(B - \operatorname{diag}(b_{11}, ..., b_{nn}))^T + (B - \operatorname{diag}(b_{11}, ..., b_{nn})) = 0$, $B \in \mathbb{R}^{n\times n}$; $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$, and $\forall \mathbf{z} = (z_1, ..., z_n)^T \in \mathbb{R}^n$

$$\sigma_+(\mathbf{z}) = (g(z_1,\alpha,\beta,c)\cdot f(z_1), ..., g(z_n,\alpha,\beta,c)\cdot f(z_n))^T \in \mathbb{R}^n$$

(see (5.3)).

In the following, in equation (5.6), the function $f(w)$ will be replaced by the function

$$f(w) = |w|^\gamma;\ \text{where } \gamma > 0, \alpha+\gamma > 1, \beta+\gamma > 1.$$

As a result, function (3.2) is transformed into the function

$$\sigma_+(w,\alpha,\beta,\gamma,c) = \begin{cases} -\frac{\beta-1}{\beta}c^{\frac{\beta}{\beta-1}}(-w)^\gamma - \frac{(-w)^{\beta+\gamma}}{\beta}, & \text{if } w < -c^{\frac{1}{\beta-1}} \\ -c(-w)^{1+\gamma}, & \text{if } w < 0 \\ cw^{1+\gamma}, & \text{if } w \leq c^{\frac{1}{\alpha-1}} \\ \frac{\alpha-1}{\alpha}c^{\frac{\alpha}{\alpha-1}}w^\gamma + \frac{w^{\alpha+\gamma}}{\alpha}. & \text{otherwise} \end{cases} \tag{5.7}$$

From the point of view of practical implementation of model (5.6), we will simplify the activation function (5.7).

Let $c = 1$, $\gamma = 1$. Then we get

$$\sigma_+(w,\alpha,\beta) = \begin{cases} -\frac{\beta-1}{\beta}(-w) - \frac{(-w)^{\beta+1}}{\beta}, & \text{if } w < -1 \\ -(-w)^2, & \text{if } w < 0 \\ w^2, & \text{if } w \leq 1 \\ \frac{\alpha-1}{\alpha}w + \frac{w^{\alpha+1}}{\alpha}. & \text{otherwise} \end{cases} \tag{5.8}$$

To adjust the coefficients of equation (5.6), we will use gradient methods [9]. In this case, we will need derivatives of the function $\sigma_+(w,\alpha,\beta)$ (5.8). (Here it is necessary to take into account that w depends on W and $\mathbf{b}$.) Let us represent these derivatives in explicit form:

$$\sigma'_{(+,w)}(w,\alpha,\beta) = \begin{cases} \left(\frac{\beta-1}{\beta} + \frac{\beta+1}{\beta}(-w)^{\beta}\right)\cdot(-w)', & \text{if } w < -1 \\ 2w\cdot w', & \text{if } w \le 1 \\ \left(\frac{\alpha-1}{\alpha} + \frac{\alpha+1}{\alpha}w^{\alpha}\right)\cdot w'. & \text{otherwise} \end{cases} \tag{5.9}$$

$$\sigma'_{(+,\alpha)}(w,\alpha,\beta) = \begin{cases} 0, & \text{if } w \le 1 \\ \frac{w}{\alpha^2} + \frac{w^{\alpha+1}}{\alpha^2}(\ln w^{\alpha} - 1). & \text{otherwise} \end{cases} \tag{5.10}$$

$$\sigma'_{(+,\beta)}(w,\alpha,\beta) = \begin{cases} \frac{(-w)}{\beta^2} + \frac{(-w)^{\beta+1}}{\beta^2}(\ln(-w)^{\beta} - 1), & \text{if } w < -1 \\ 0. & \text{otherwise} \end{cases} \tag{5.11}$$

Now we will show that in formula (5.10) $\forall w > 1$, we have $\sigma'_{(+,\alpha)}(w,\alpha,\beta) > 0$. Let us introduce the notation $w^{\alpha} = z+1$. Then the last inequality can be represented in the form $\forall z > 0$ $(z+1)\ln(z+1) > z$ or $\ln(z+1) > z/(z+1)$. (If $z = 0$, then the last inequality becomes a strict equality $0 = 0$.)

Now we take into account that $z \to \infty$ we have $\ln(z+1) \to \infty$ and $z/(z+1) \to 1$. In addition, $(\ln(z+1))' = 1/(1+z)$ and $(z/(z+1))' = 1/(1+z)^2$. In other words, $\forall z > 0$ we have $(\ln(z+1))' > (z/(z+1))'$ (if $z = 0$, then the last inequality becomes a strict equality $1 = 1$). This relation proves inequality $\ln(z+1) > z/(z+1)$ (see Comparison Lemma 2.5 [18]), and, therefore, inequality $\sigma'_{(+,\alpha)}(w,\alpha,\beta) > 0$. (Inequality $\sigma'_{(+,\beta)}(w,\alpha,\beta) > 0$ in formula (5.11) is proved similarly.)

In the future we will need the following known result.

Theorem 5.2. *(***Bihari's Lemma***) [15]. Let u and f be non-negative continuous functions defined on the half-infinite ray $[0,\infty)$, and let w be a continuous non-decreasing function defined on $[0,\infty)$ and $w(u) > 0$ on $(0,\infty)$. If u satisfies the following integral inequality,*

$$u(t) \le \alpha + \int_0^t f(s)w(u(s))ds, t \in [0,\infty), \tag{5.12}$$

where α is a non-negative constant, then

$$u(t) \le G^{-1}\Big(G(\alpha) + \int_0^t f(s)ds\Big), t \in [0,T], \tag{5.13}$$

where the function G is defined by formula

$$G(x) = \int_{x_0}^{x} \frac{dy}{w(y)}, x \geq 0, x_0 \geq 0,$$

and G^{-1} *is the inverse function of* G *and* T *is chosen so that*

$$\forall t \in [0, T]\ G(\alpha) + \int_0^t f(s)ds \in Dom(G^{-1});$$

here $Dom(G^{-1}) = \{x \in \mathbb{R} | G^{-1}(x) \neq \emptyset\}$.

Corollary 5.1. *Let in Theorem 5.2 be* $w(u) = u^p, 0 < p < 1$. *Then*

$$u(t) \leq \Big(\alpha^{1-p} + (1-p)\int_0^t f(s)ds\Big)^{\frac{1}{1-p}}, t \geq 0. \tag{5.14}$$

If $p > 1$ *and* $\int_0^t f(s)ds < \frac{1}{(p-1)\alpha^{p-1}}$, $t \geq 0$, *then*

$$u(t) \leq \frac{\alpha}{\Big(1 - (p-1)\alpha^{p-1}\int_0^t f(s)ds\Big)^{\frac{1}{p-1}}}. \tag{5.15}$$

Consider the following system of differential equations

$$\dot{\mathbf{x}}(t) = A\mathbf{x}(t) + \mathbf{a} + \sigma(B\mathbf{x}(t) + \mathbf{b}), \mathbf{x}(0) = \mathbf{x}_0, \tag{5.16}$$

where $A, B \in \mathbb{R}^{n\times n}$; $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$, $\sigma(...) = (\sigma_1(...), ..., \sigma_n(...))^T$ (without loss of generality we can assume that $\sigma_i(u_i) = u_i^p$ (if $u_i > 0$) or $-(-u_i)^q$ (if $u_i < 0); i = 1, ..., n)$).

Theorem 5.3. *Let in system (5.16) the activation function* $\sigma(u)$ *be odd and there exist numbers* $c > 0$, $0 < p < 1$, *and* $0 < q < 1$ *such that (a)* $\lim_{u\to\infty}(c + u^p - \sigma(u)) \geq 0$ *and (b)* $\lim_{u\to-\infty}(\sigma(-u) - (-u)^q - c) \leq 0$. *In addition, let us also assume that the matrix* A *is Hurwitz and* $\det B \neq 0$. *Then any solution of this system is bounded.*

Explanations of the proof of Theorem 5.3 (see Fig.5.1).

Proof. Since $\det B \neq 0$, then by replacing the variable $\mathbf{x} \to \mathbf{y} - B^{-1}\mathbf{b}$ we can transform equation (5.16) into an equation of the same structure in which $\mathbf{b} = \mathbf{0}$. Therefore, in the future we can assume that in (5.16) $\mathbf{b} = \mathbf{0}$.

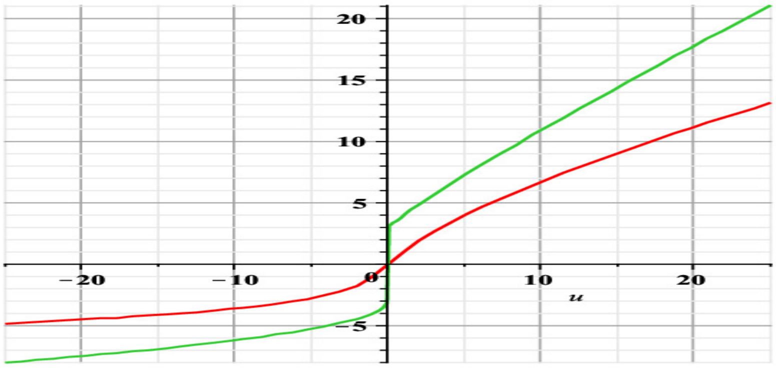


Fig. 5.1. Function $\sigma(u)$ (see (3.2)) with parameters $\alpha = 0.7, \beta = 0.1, c = 1$ (red) and functions $c + u^p, u \geq 0$ and $-c - (-u)^q, u < 0$ with parameters $p = 0.8, q = 0.9, c = 3$ (green)

Now we transform equation (5.16) into the following integral equation

$$\mathbf{x}(t) = \exp(At)\mathbf{x}(0) + \int_0^t \exp(A(t-\tau))[\mathbf{a} + \sigma(B\mathbf{x}(\tau))]d\tau.$$

From here it follows that

$$\|\mathbf{x}(t)\| \leq \|\exp(At)\|\|\mathbf{x}(0)\| + \int_0^t \|\exp(A(t-\tau))\|(\|\mathbf{a}\| + \|\sigma(B\mathbf{x}(\tau))\|)d\tau. \quad (5.17)$$

(a) $u \to \infty$. Since matrix A is Hurwitz, then the relation $\|\exp(At)\| \leq \beta\exp(-\gamma t), \beta > 0$ is valid; here $-\gamma < 0$ is the the maximum real part of all eigenvalues of the matrix A. Besides, $\|\sigma(B\mathbf{x}(\tau))\| < nc + \|\mathbf{x}(\tau)\|^p$, where $p > 0$.

Thus, inequality (5.17) can be rewritten in the form

$$\|\mathbf{x}(t)\| \leq \beta\exp(-\gamma t)\|\mathbf{x}(0)\| + \int_0^t \exp(-\gamma(t-\tau))(nc + \|\mathbf{a}\| + \|\mathbf{x}(\tau)\|^p)d\tau, t > \tau. \quad (5.18)$$

Let $s = \tau$. We note that

$$\int_0^t f(s)ds = \int_0^t \exp(-\gamma\cdot(t-\tau))d\tau = \frac{1-\exp(-\gamma t)}{\gamma}.$$

Then inequality (5.18) can be represented as

$$\|\mathbf{x}(t)\| \leq \beta\|\mathbf{x}(0)\| + \frac{nc + \|\mathbf{a}\|}{\gamma} + \int_0^t \exp(-\gamma(t-\tau))(\|\mathbf{x}(\tau)\|^p)d\tau, t > \tau. \quad (5.19)$$

Let $\alpha = \beta\|\mathbf{x}(0)\| + (nc + \|\mathbf{a}\|)/\gamma$.

Now we use the corollary given above. If $0 < p < 1$, then applying formula (5.14) to inequality (5.19) we get at $t \to \infty$ and $\forall\alpha > 0$, and $\forall\gamma > 0$:

$$\|\mathbf{x}(t)\| \leq \Big(\alpha^{1-p} + (1-p)\int\limits_0^t f(s)ds\Big)^{\frac{1}{1-p}} \leq \Big(\alpha^{1-p} + \frac{1-p}{\gamma}\Big)^{\frac{1}{1-p}} < \infty.$$

(b) $u \to -\infty$. In this case, the proof of Theorem 5.3 almost verbatim repeats the proof of case (a).

Now let $p > 1$. Let us assume that p also satisfies inequality $(p-1)\alpha^{p-1} < \gamma$. In this case, the denominator of formula (5.15) does not turn into 0. Therefore, we get $\forall t \in [0, \infty)$ $\|\mathbf{x}(t)\| < \infty$.

Let us write out all the constraints on the parameters under which system (5.16) has bounded solutions:

1. $0 < p < 1$, $0 < q < 1$; 2. $p > 1$, $(p-1)\alpha^{p-1} < \gamma$, $0 < q < 1$;
3. $0 < p < 1$, $q > 1$, $(q-1)\alpha^{q-1} < \gamma$;
4. $p > 1$, $(p-1)\alpha^{p-1} < \gamma$, $q > 1$, $(q-1)\alpha^{q-1} < \gamma$.

It is clear that among all four groups of constraints, only the first group of constraints does not depend on the initial conditions. This completes the proof of Theorem 5.3. □

6. Practical use of models (5.1) and (5.16)

In this section we will use model (5.1) with activation functions (5.3), (5.4) and model (5.16) whose behavior satisfies the conditions of Theorem 5.3.

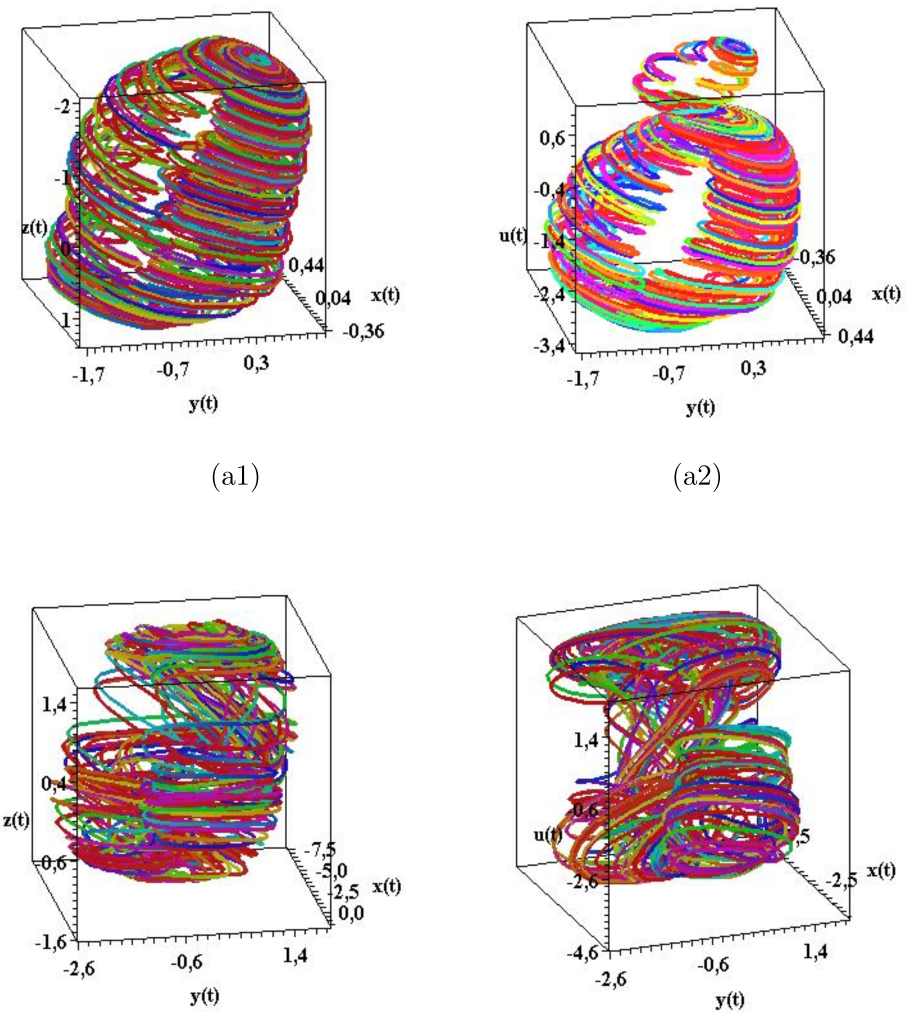
In this case, we will consider not only the situation described in Theorem 5.3, but also other combinations of parameters p and q, which were discussed in the proof of Theorem 5.3.

It is interesting to note that the trajectories of model (5.1) exhibit chaotic behavior at certain parameter values (see Fig.6.1).

At the same time, for model (5.16) the chaotic behavior of trajectories could not be detected. In the presented Fig.6.2 we can see (under the conditions of Theorem 5.3) only quasi-periodic behavior of the trajectories of this system. (To simplify the calculations, we replaced system (5.16) with the equivalent system

$$\dot{\mathbf{y}}(t) = A_e\mathbf{y}(t) + \mathbf{a}_e + B\sigma(\mathbf{y}(t)), \mathbf{y}(0) = \mathbf{y}_0, \tag{6.1}$$

where $\mathbf{y}(t) = B\mathbf{x}(t) + \mathbf{b}, A_e = BAB^{-1}$. The correctness of this transition is proven in [13].)



(a1) (a2)

(a3) (a4)

Fig. 6.1. Graphs of 2D projections of trajectories of system (5.2) of homoclinic and heteroclinic type for $\sigma_i(y_i) = g(y_i, \alpha_{i1}, \beta_{i1}, c_{i1})|y_i| + k_i g(y_i, \alpha_{i2}, \beta_{i2}, c_{i2}), i = 1, ..., n = 4$ (see (5.3), (5.4)) and the following values of parameters (see Theorem 5.1): $a_{10} = ... = a_{40} = a_{11} = ... = a_{44} = 0$; (a1, a2) $a_{12} = -3, a_{13} = -1, a_{14} = 1, a_{23} = 0, a_{24} = 1, a_{34} = 1,\ k_1 = 2.4, k_2 = 0, k_3 = 0, k_4 = -1.2$, $\alpha_1 = 1.8, \beta_1 = 0.4, c_1 = 5,\ \alpha_2 = 0.2, \beta_2 = -1.2, c_2 = 7$; (a3, a4) $a_{12} = -3, a_{13} = 0, a_{14} = 1, a_{23} = 0, a_{24} = 3, a_{34} = 1,\ k_1 = -2, k_2 = 0, k_3 = 3, k_4 = -1.2,\ \alpha_1 = 1.8, \beta_1 = 0.4, c_1 = 3$, $\alpha_2 = -0.1, \beta_2 = 1.2, c_2 = 2$.

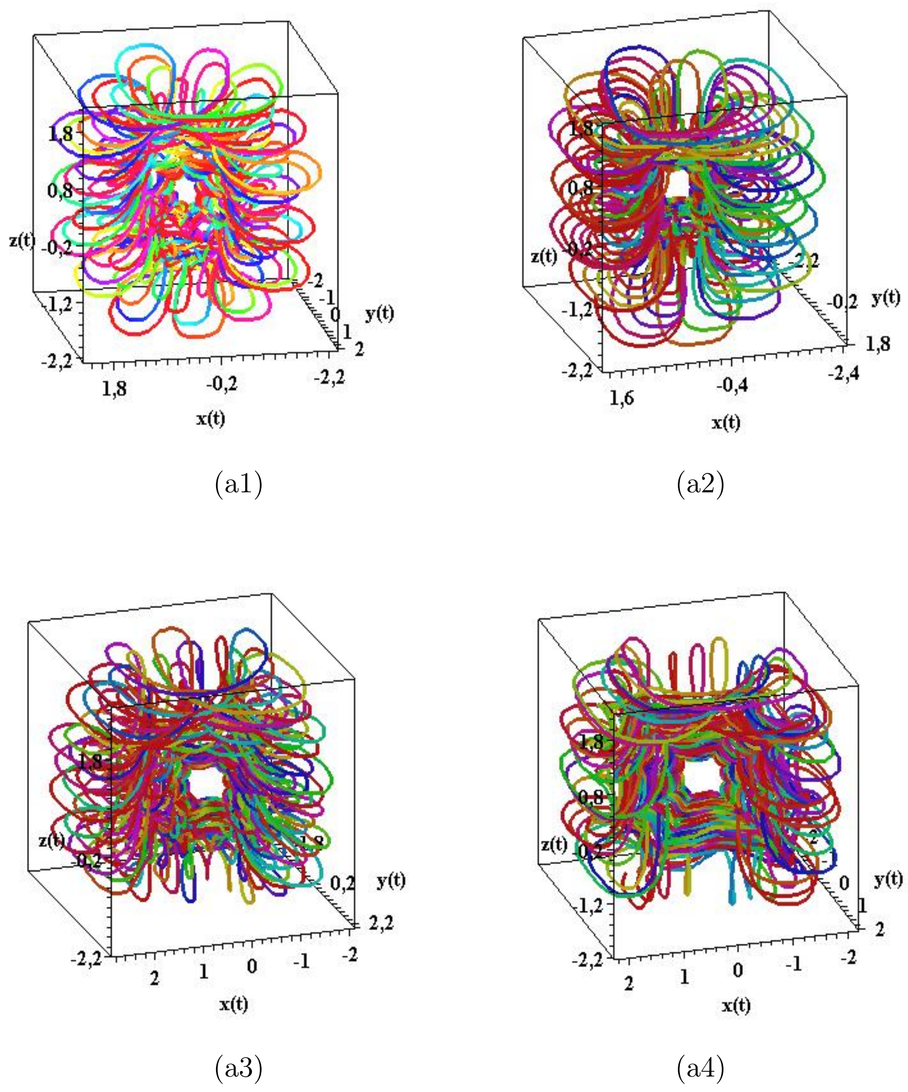


Fig. 6.2. Graphs of 2D projections of trajectories of system (6.1) for $\sigma(y_i) = g(y_i, \alpha_i, \beta_i, c_i)|y_i|, i = 1, ..., n = 4$ and the following values of parameters (see Theorem 5.3): $a_{10} = ... = a_{40} = 0, a_{11} = a_{44} = -0.02$, $a_{22} = a_{33} = -0.01, a_{ij} = a_{ji} = 0$ at $i \neq j$; $b_{10} = ... = b_{40} = 0, b_{12} = -b_{21} = -3, b_{13} = -b_{31} = 0, b_{14} = -b_{41} = 1, b_{23} = -b_{32} = 0, b_{24} = -b_{42} = 3, b_{34} = -b_{43} = -1, b_{11} = ... = b_{44} = 0$; (a1) $\alpha_1 = 0.18, \beta_1 = 0.14, c_1 = 3$; (a2) $\alpha_2 = 1.18, \beta_2 = 0.14, c_2 = 3$; (a3) $\alpha_3 = 0.03, \beta_3 = 3, c_3 = 3$; (a4) $\alpha_4 = 3, \beta_4 = 4, c_4 = 3$.

References

1. P. G. Barrientos, S. Ibanez, A. A. Rodrigues, J. A. Rodrigues, *Emergence of Chaotic Dynamics from Singularities*, (IMPA, Published in Brazil), 2019.
2. R. Bellman, *Introduction to Matrix Analysis – 2nd edition*, (Society for Industrial and Applied Mathematics 3600 University City Science Center/ Philadelphia, PA 19104–2688), 1997.
3. V. Ye. Belozyorov, *On existence of homoclinic orbits for some types of autonomous quadratic systems of differential equations*, Applied Mathematics and Computation, **217**(9)(2011), 4582–4595.
4. V. Ye. Belozyorov, S. A. Volkova, *Role of logistic and Ricker's maps in appearance of chaos in autonomous quadratic dynamical systems*, Nonlinear Dynamics, **83**(1)(2016), 719–729.
5. V. Ye. Belozyorov, *Invariant approach to existence problem of chaos in 3D autonomous quadratic dynamical systems*, International Journal of Bifurcation and Chaos, **26**, (2016), Article ID 1650012, 14 pages.
6. V. Ye. Belozyorov, *On novel conditions of chaotic attractors existence in autonomous polynomial dynamical systems*, Nonlinear Dynamics, **91**(4) (2018), P. 2435–2452.
7. V. Ye. Belozyorov, *Universal approach to the problem of emergence of chaos in autonomous dynamical systems*, Nonlinear Dynamics, **95**(1)(2019), 579–595.
8. V. Ye. Belozyorov, D. V. Dantsev, *Stability of neural ordinary differential equations with power nonlinearities*, Journal of Optimization, Differential Equations and Their Applications (JODEA), **28**(2)(2020), 21–46.
9. V. Ye. Belozyorov, D. V. Dantsev, *Modeling of chaotic processes by means of antisymmetric neural ODEs*, Journal of Optimization, Differential Equations and Their Applications (JODEA), **30**(1)(2022), 1–41.
10. V. Ye. Belozyorov, S. A. Volkova, *Odd and even functions in the design problem of new chaotic attractors*, International Journal of Bifurcation and Chaos, **32**, (2022), Article ID 2250218, 26 pages.
11. V. Ye. Belozyorov, S. A. Volkova, V. G. Zaytsev, *Singular differential equations and their applications for modeling strongly oscillating processes*, Journal of Optimization, Differential Equations and Their Applications (JODEA), **31**(1)(2023), 22–52.
12. V. Ye. Belozyorov, O.A. Inkin, *Systems of singular differential equations as the basis for neural network modeling of chaotic processes*, Journal of Optimization, Differential Equations and Their Applications (JODEA), **31**(2)(2023), 24–48.
13. V. Ye. Belozyorov, Y. V. Koshel, *On systems of neural ODEs with generalized power activation functions*, Journal of Optimization, Differential Equations and Their Applications (JODEA), **32**(2)(2024), 56–91.
14. V. Ye. Belozyorov, N. A. Guk, D. I. Yehoshkin, *Encryption of color images based on chaotic attractors generated by ODE systems containing module nonlinearities*, Journal of Optimization, Differential Equations and Their Applications (JODEA), **32**(2)(2024), 92–117.
15. I. Bihari, *A generalization of a lemma of Bellman and its application to uniqueness problems of differential equations*, Acta Mathematica Hungarica, **7**(1)(1956), 81–94.
16. R. T. Q. Chen, Y. Rubanova, J. Bettencourt, D. Duvenaud, *Neural ordinary differential equations*, arXiv preprint *arXiv:1806.07366v5[cs.LG]*, (2019), 1–18.
17. B. Feng, R. Hu, *A survey on homoclinic and heteroclinic orbits*, Applied Mathematics E-Notes, **3** (2003), 16–37.

18. H. K. Khalil, *Nonlinear Systems – 2nd Edition*, (Prentice Hall/New-Jersy), 1996.
19. Y. C. Li, *Existence of chaos in evolution equations*, Mathematical and Computer Modelling, **36** (2003), 1211–1219
20. Z. Li, G. Chen, W. A. Halang, *Homoclinic and heteroclinic orbits in a modified Lorenz system*, Information Sciences, **165**(2004), 235–245.
21. Y. V. Koshel, V. Ye. Belozyorov, *Univariate time series analysis with hyper neural ODE*, Journal of Optimization, Differential Equations and Their Applications (JODEA), **31**(2)(2023), 50–66.
22. H. Poincare, *Sur la probleme des trois corps et les equations de la dynamique*, Acta Mathematica, **13**(1890), 1–270.
23. D. Shang, M. Han, *The existence of homoclinic orbits to saddle-focus*, Applied Mathematics and Computation, **163**(2005), 621–631.
24. G. Teschi, *Ordinary Differential Equations and Dynamical Systems*, Graduate Studies in Mathematics, **140** (AMS, Providence/ Rhode Island), 2012.
25. T. Zhou, G. Chen, *Classification of chaos in 3D autonomous quadratic systems - I: Basic framework and methods*, International Journal of Bifurcation and Chaos, **16**(9) (2006), 2459–2479.

Received 30.4.2025