PERGAMON

Neural Networks 11 (1998) 495–508

Neural Networks

Contributed article

Multiple cusp bifurcations[1,2,3]

Eugene M. Izhikevich*

Center for Systems Science and Engineering, Arizona State University, Tempe, AZ 85287-7606 U.S.A.

Received 20 June 1996; revised 12 September 1997; accepted 12 September 1997

Abstract

The cusp bifurcation provides one of the simplest routes leading to bistability and hysteresis in neuron dynamics. We show that weakly connected networks of neurons near cusp bifurcations that satisfy a certain adaptation condition have quite interesting and complicated dynamics. First, we prove that any such network can be transformed into a canonical model by an appropriate continuous change of variables. Then we show that the canonical model can operate as a multiple attractor neural network or as a globally asymptotically stable neural network depending on the choice of parameters.

Keywords: Weakly connected neural networks; Multiple cusp bifurcations; Multiple pitchfork bifurcations; Canonical models; Hebbian learning; Bistability of perception

1. Introduction

1.1. Weakly connected neural networks

A useful assumption in mathematical neuroscience is that neurons are weakly connected (Hoppensteadt and Izhikevich, 1997). It is based on the observation that the averaged size of postsynaptic potential (PSP) is less than 1 mV, which is small in comparison with the mean size necessary to discharge a cell (around 20 mV) or the averaged size of the action potential (around 100 mV). For example, PSPs in hippocampus cells are as small as 0.1 ± 0.03 mV (McNaughton et al., 1981; Sayer et al., 1990). The majority of PSPs in pyramidal neurons of the rat visual cortex are less than 0.5 mV, with the range 0.05–2.08 mV (Mason et al., 1991). As was pointed out by Mason et al. (1991), there is an underestimate of the true range because PSPs smaller than 0.03 mV would have gone undetected.

We see that it is reasonable to consider weakly connected neural networks (WCNNs). Any WCNN can be written in the ''weakly connected'' form

$$\dot{X}_i = F_i(X_i, \lambda) + \varepsilon G_i(X_1, \ldots, X_n, \rho, \lambda, \varepsilon), \tag{1}$$

where vector $X_i \in \mathbb{R}^m$ denotes activity of the ith neuron or the ith local population of neurons, which we call the ''averaged'' neuron. We can think of X_i as a continuous variable representing the number of action potentials per unit of time, or an amount of chemical neurotransmitters released by synaptic terminals, or any other physiological observable. The vector $\lambda \in \mathbb{R}^l$ denotes the set of internal parameters of the network, and the vector $\rho \in \mathbb{R}^r$ denotes input from external receptors to the network. Parameter $\varepsilon \ll 1$ is small reflecting the strength of connections between neurons. Hoppensteadt and Izhikevich (1997) obtained the estimation

$$0.004 < \varepsilon < 0.008,$$

using experimental data for the hippocampus.

Since little is known about neurophysiological processes taking place inside biological neurons, we do not have detailed information about the functions F_i and G_i above. Nevertheless, using the condition $\varepsilon \ll 1$ we could analyze behavior of networks of the form (1) without even telling how F_i and G_i should look like. Much progress has been achieved in the case when each neuron in the uncoupled network

$$\dot{X}_i = F_i(X_i, \lambda), \quad i = 1, \ldots, n \tag{2}$$

is a pacemaker; that is, each X_i converges to a limit cycle attractor in $\mathbb{R}^m$ (see reviews by Ermentrout, 1994;

* Email: eugene.izhikevich@asu.edu.

[1] This paper received the SIAM Award as the best student paper in applied mathematics in 1995. It was written while the author was a graduate student at department of Mathematics, Michigan State University, and was supported in part by NSF grant DMS 9206677.

[2] This work could not be accomplished without Dr. Frank Hoppensteadt. I am very grateful for all kinds of investments he made in me and my research.

[3] This paper is dedicated to Frank C. Hoppensteadt on the occasion of his 60th birthday.

0893–6080/98/$19.00
PII: S0893-6080(97)00117-2

Kopell, 1995; and the book by Hoppensteadt and Izhikevich, 1997).

In this paper we consider the case when each neuron in (2) is quiescent; that is, each X_i is in a neighborhood of an equilibrium in $\mathbb{R}^m$. It is erroneously assumed that the system (1) cannot have interesting behavior in this case since ε is small. Below we show that this is not true. System (1) may have quite interesting and complicated behavior for any $\varepsilon \ll 1$ provided that neurons are near thresholds; that is, the equilibria are nonhyperbolic. In particular, we consider the case when the nonhyperbolic equilibria correspond to cusp bifurcations, which are often referred to as cusp singularities or cusp catastrophes.

1.2. Cusp bifurcations

Cusp bifurcation is one of the simplest bifurcations leading to local bistability and hysteresis. It may be observed in many neural models, such as the additive neuron or Wilson–Cowan neural oscillator (Borisyuk and Kirillov, 1992; Hoppensteadt and Izhikevich, 1997); see Fig. 1. The latter provides an example of cusp bifurcation in a model exhibiting prominent oscillatory behavior. The exact definition of the bifurcation is provided, e.g., by Kuznetsov (1995) or Hoppensteadt and Izhikevich (1997), see also Section 2 below. A typical system at the cusp bifurcation is

$$\dot{x} = a + bx + \sigma x^3, \quad \sigma = \pm 1, \tag{3}$$

at the origin $a = b = x = 0$; see Fig. 2. Incidentally, this system is a local canonical model for the cusp bifurcation in the sense that any other dynamical system of the form $\dot{X} = F(X, \lambda)$ at the bifurcation can be transformed to the form (3) by a suitable continuous (possibly non-invertible) local change of variables $x = h(X)$. Peculiarities of the function F do not affect the form of the canonical model (3), but affect only the values of the parameters a, b and σ. When $\sigma = -1$ ($\sigma = +1$), the cusp bifurcation is called

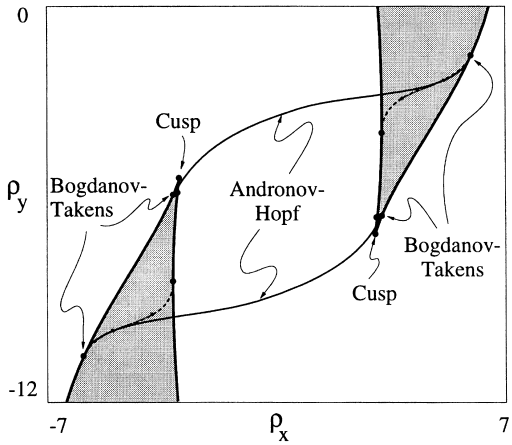

Fig. 1. Bifurcation sets of Wilson–Cowan neuron oscillator model $\dot{x} = -x + S(ax + by + \rho_x)$, $\dot{y} = -y + S(x + dy + \rho_y)$, where $S(\rho) = 1/(1 + e^{-\rho})$, $a = b = c = 10$ and $d = -2$ (from Hoppensteadt and Izhikevich, 1997).

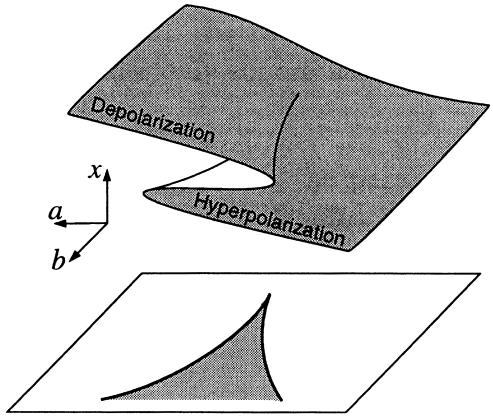

Fig. 2. Cusp surface (from Hoppensteadt and Izhikevich, 1997).

supercritical (*subcritical*). Below we consider only supercritical cusp bifurcations unless we state otherwise.

When (3) describes activity of a neuron, the parameter a denotes the input to the neuron from the network, and b denotes a feedback parameter. When $b < 0$, the neuron always has one equilibrium for any input a, which is approximately $\sqrt[3]{a}$ (for large $|a|$). The dynamics of the canonical model (3) is relatively simple in this case. When $b > 0$, the canonical model acquires a new property—*bistability*. It exhibits hyperpolarization for inhibitory inputs ($a \ll -1$, see Fig. 3A), depolarization for excitatory inputs ($a \gg 1$, see Fig. 3C) and co-existence of the hyperpolarized and depolarized activities for intermediate values of a (see Fig. 3B). When the external input a is changed periodically, the neuron exhibits hysteresis behavior. The canonical model (3) is one of the simplest systems exhibiting these behaviors since it has only one nonlinear term, namely x^3.

The equilibria corresponding to hyperpolarized and depolarized states appear and disappear via saddle–node bifurcations. The simplest way to see this is to fix $b > 0$ and vary a as a bifurcation parameter. The bifurcation diagrams in this case are depicted at the top of Fig. 4. Many other bifurcation diagrams can be obtained by assuming that the bifurcation parameters a and b represent a curve in the ab-plane parametrized by s (that is, $a = a(s)$ and $b = b(s)$). A summary of some special cases for $\sigma = -1$ is depicted in Fig. 4 showing that there can be many interesting dynamical regimes despite the simplicity of the canonical model (3).

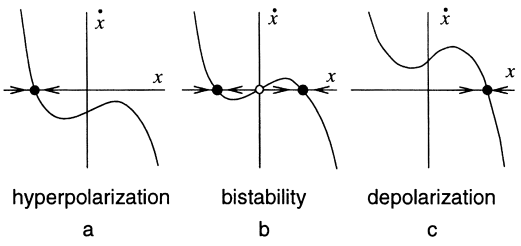

Fig. 3. Dynamic behavior of a neuron governed by $\dot{x} = a + bx - x^3$. The neuron can exhibit (A) hyperpolarization (for $a \ll -1$), (B) co-existence of hyperpolarization and depolarization and (C) depolarization (for $a \gg 1$).

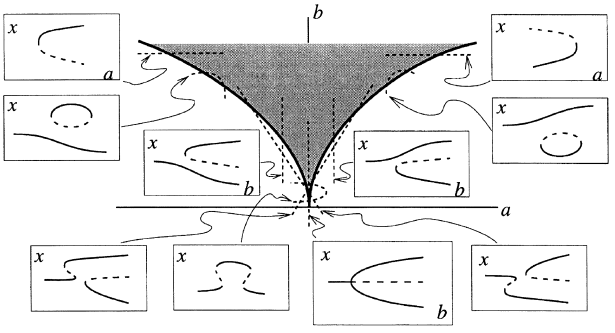

Fig. 4. Summary of special cases for the supercritical cusp bifurcation. Dotted segments and curves crossing the bifurcation sets represent one-dimensional bifurcation parameters. Bifurcation diagrams are depicted in boxes: Continuous curves represent stable solutions, dashed curves represent unstable solutions (from Hoppensteadt and Izhikevich, 1997).

1.3. Multiple cusp bifurcations

A WCNN (1) is said to be near multiple cusp bifurcation when some or all of its neurons are near the bifurcation. In Section 2 we show how a general WCNN of the form (1) near multiple cusp bifurcation can be transformed to a canonical model

$$\dot{x}_i = r_i \pm x_i^3 + \sum_{j=1}^{n} c_{ij} x_j, \tag{4}$$

by a continuous change of variables. Here $x_i \in \mathbb{R}$ is the canonical variable denoting activity of the ith neuron, parameters $r_i \in \mathbb{R}$ are the inputs from external receptors, $c_{ij} \in \mathbb{R}$, $i \neq j$ are synaptic coefficients, and $c_{ii} \in \mathbb{R}$ are bifurcation parameters. Notice that each equation in the canonical model (4) can be written in the form (3), where

$$a = r_i + \sum_{j \neq i} c_{ij} x_j \quad \text{and} \quad b = c_{ii}.$$

We start to analyze the behavior of the canonical model (4) in Section 3, where we show that the most interesting case is multiple supercritical cusp bifurcation, which corresponds to the choice $-x_i^3$. We show in Section 4 that the canonical model can behave as globally asymptotically stable (GAS) or multiple attractor (MA) neural network depending on the choice of parameters.

In Section 5 we derive rigorously a learning rule for the canonical model. It turns out to be Hebbian. Thus, instead of been postulated, the Hebbian learning rule is a consequence of the assumption that neurons are weakly connected and each of them is near cusp bifurcation. Dynamics of the canonical model with the Hebbian learning rule is scrutinized in subsequent sections. In Section 8 we use classical example in perception to illustrate bistability of the canonical model dynamics.

2. Derivation of the canonical model

Consider a weakly connected neural network (WCNN) of the form

$$\dot{X}_i = F_i(X_i, \lambda) + \varepsilon G_i(X_1, \ldots, X_n, \rho, \lambda, \varepsilon), \tag{5}$$

and suppose that each equation in the uncoupled system

$$\dot{X}_i = F_i(X_i, \lambda), \quad i = 1, \ldots, n, \tag{6}$$

is at equilibrium, say $X_i = 0$ for $\lambda = 0$; that is,

$$F_i(0, 0) = 0, \quad i = 1, \ldots, n.$$

Since $\lambda \in \mathbb{R}^l$ is a vector, this condition may be satisfied for all i simultaneously provided that $l \geq n$. If the vector X_i describes activity of a single cell, we may assume that $X_i = 0$ denotes the quiescent state corresponding to the rest membrane potential. If X_i describes activity of a local population of neurons, which we refer to as being the ''averaged'' neuron, then $X_i = 0$ may also mean that the neurons from the population fire repeatedly with a constant mean firing rate.

2.1. Fundamental theorem of WCNN theory

Obviously, the uncoupled system (6) is not interesting as a model of the brain. Does the WCNN (5) acquire any new properties that make it more interesting? The answer is given by the fundamental theorem of WCNN theory (Izhikevich, 1996; Hoppensteadt and Izhikevich, 1997) that says that behavior of (5) and (6) is essentially the same (locally linear and uncoupled) unless some or all neurons are near a bifurcation point. Moreover, those neurons that are not near a bifurcation point can be removed from the network without changing its behavior. If X_i describes activity of a single cell, then to be near a bifurcation means that the rest potential is near a threshold potential. This is not necessarily valid when X_i describes activity of the ''averaged'' neuron. Finally, we notice that the fundamental theorem does not preclude the WCNN (5) from having interesting *global* properties even when each equation in (6) has an hyperbolic equilibrium.

Without loss of generality we assume that each equation in (6) is at a bifurcation. The WCNN (5) is said to be near *multiple* bifurcation in this case. In the present paper we study multiple cusp bifurcation. First, we seek a change of variables that transforms the WCNN (5) to a simpler model, which captures its non-linear behavior. Such a simpler model is called the *canonical model* and it is derived rigorously by Izhikevich (1996), see also Hoppensteadt and Izhikevich (1997). Below we outline the derivation, which is done in two steps.

2.1.1. Step 1: Center manifold reduction

A general WCNN of the form (5) near multiple cusp bifurcation can be reduced to its center manifold by a continuous change of variables $x = h(X,\varepsilon)$. The mapping h is

non-invertible when each X_i is a vector. On the center manifold system (5) is governed by a dynamical system of the form

$$\dot{x}_i = f_i(x_i, \lambda) + \varepsilon g_i(x, \rho, \lambda, \varepsilon), \tag{7}$$

where $x_i \in \mathbb{R}$ is a scalar and functions f_i satisfy

$$f_i(0,0) = 0 \quad \text{and} \quad \frac{\partial f_i(0,0)}{\partial x_i} = 0 \tag{8}$$

for all i. Such a reduction is a consequence of center manifold reduction for weakly connected systems, see Theorem 5.2 in Hoppensteadt and Izhikevich (1997).

Since the reduced system (7) is near multiple cusp bifurcation too, we may deduce some additional information, namely,

$$\frac{\partial^2 f_i}{\partial x_i^2} = 0 \quad \text{but} \quad q_i = \frac{1}{6}\frac{\partial^3 f_i}{\partial x_i^3} \neq 0$$

at the equilibrium $x_i = 0$ for $\lambda = 0$. This is equivalent to say that the Taylor series of f_i at the origin starts from the cubic term, i.e.

$$f_i(x_i, 0) = q_i x_i^3 + \text{higher order terms.}$$

There are also some transversality conditions imposed on $\lambda \in \mathbb{R}^l$ and function f_i which we do not use below.

2.1.2. Step 2: Simplifying change of variables

Let us assume that

$$\lambda = \lambda(\varepsilon) = 0 + \varepsilon\lambda_1 + O(\varepsilon^2),$$

$$\rho = \rho(\varepsilon) = \rho_0 + \sqrt{\varepsilon}\rho_{1/2} + O(\varepsilon),$$

for some $\lambda_1 \in \mathbb{R}^l$ and $\rho_0, \rho_{1/2} \in \mathbb{R}^r$. Then we can rewrite the initial portion of the Taylor series of (7) as

$$\dot{x}_i = \varepsilon a_i + \varepsilon\sqrt{\varepsilon} d_i + \varepsilon \sum_{j=1}^{n} s_{ij} x_j + q_i x_i^3 + \text{h.o.t.}, \tag{9}$$

where

$$a_i = D_\lambda f_i \cdot \lambda_1 + g_i,$$

$$d_i = D_\rho g_i \cdot \rho_{1/2},$$

$$s_{ij} = \frac{\partial g_i}{\partial x_j}, \quad \text{for } i \neq j,$$

$$s_{ii} = D_\lambda \frac{\partial f_i}{\partial x_i} \cdot \lambda_1 + \frac{\partial g_i}{\partial x_i}$$

are some parameters evaluated at the equilibrium $x_i = 0$ for $\lambda = 0$, $\rho = \rho_0$ and $\varepsilon = 0$.

2.1.3. Adaptation condition

The condition $a_i = 0$, $i = 1,\ldots,n$ is called the adaptation condition. Behavior of the system (9) depends crucially on whether or not the adaptation condition is satisfied. If it is not, then the rescaling $x_i = \sqrt[3]{\varepsilon} y_i$ transforms (9) into another weakly connected system

$$y_i' = a_i + q_i x_i^3 + O(\sqrt[3]{\varepsilon}),$$

where $' = \mathrm{d}/\mathrm{d}\tau$ and $\tau = \sqrt{\varepsilon} t$ is the slow time. Since all $a_i \neq 0$, we can apply the fundamental theorem of WCNN theory to this system to conclude that its behavior is trivial and uninteresting (Hoppensteadt and Izhikevich, 1997, Section 5.3.8).

The adaptation condition

$$a_i = D_\lambda f_i(0,0) \cdot \lambda_1 + g_i(0,0,\rho_0,0) = 0 \tag{10}$$

can be explained in ordinary language as follows: The internal parameter λ counterbalances up to order ε the steady state input, $g_i(0,0,\rho_0,0)$, from the entire network onto the ith neuron. An important case when the adaptation condition is always satisfied is when $g_i(0,0,\rho_0,0) = 0$ and $\lambda = O(\varepsilon^2)$. The former can be interpreted as absence of any input when the neurons are silent. The latter means that the distance to the bifurcation, λ, is much smaller than the strength of connections ε.

When the adaptation condition is satisfied we use the change of variables

$$y_i = \varepsilon^{-1/2} |q_i|^{1/2} x_i, \qquad \text{(activity of neurons)},$$

$$r_i = |q_i|^{1/2} d_i, \qquad \text{(external inputs)},$$

$$\sigma_i = \operatorname{sign} q_i = \pm 1, \qquad \text{(criticality parameters)},$$

$$c_{ij} = |q_i|^{1/2} s_{ij} |q_j|^{-1/2}, \qquad \text{(synaptic coefficients)},$$

$$\tau = \varepsilon t, \qquad \text{(slow time)}$$

to transform (9) into the system (11) below. Now we may combine the transformation h from Step 1 (center manifold reduction) and the change of variables above to state the following result (Izhikevich, 1996; Hoppensteadt and Izhikevich, 1997).

Theorem 1 (Canonical Model for Multiple Cusp Bifurcations in WCNNs). *There is a continuous change of variables that transforms a WCNN*

$$\dot{X}_i = F_i(X_i, \lambda) + \varepsilon G_i(X_1, \ldots, X_n, \rho, \lambda, \varepsilon),$$

satisfying adaptation condition near multiple cusp bifurcation to the canonical model

$$y_i' = r_i + \sigma_i y_i^3 + \sum_{j=1}^{n} c_{ij} y_j + O(\sqrt{\varepsilon}).$$

The choice $\sigma_i = -1$ ($\sigma_i = +1$) corresponds to the supercritical (subcritical) multiple cusp bifurcation.

Notice that the canonical model is not weakly connected anymore. We remark that *all* WCNNs near multiple cusp bifurcation are governed by the *same* canonical model (11). Particulars of the functions F_i and G_i and parameters λ and ρ in the WCNN above do not affect the form of the canonical

model, but affect only the values of the parameters r_i and c_{ij}. This is the advantage of studying the canonical model. The disadvantage is that it describes WCNN only in a narrow parameter regimes, viz. near the multiple cusp bifurcation.

We analyze the canonical model for $\varepsilon \to 0$. Thus, we study the system

$$\dot{y}_i = r_i + \sigma_i y_i^3 + \sum_{j=1}^{n} c_{ij} y_j, \tag{12}$$

which is an interesting neural network model in itself without any connection to the WCNN theory. It exhibits useful behavior from a computational point of view and deserves to be studied per se.

3. Supercritical vs. subcritical cusp

Below we show that supercritical multiple cusp bifurcations in WCNNs have certain advantage over the subcritical ones.

Let $B_0(R) \subset \mathbb{R}^n$ denote a ball with center at the origin and radius R. A dynamical system is *bounded* if there is $R > 0$ such that $B_0(R)$ attracts all trajectories; i.e., for any initial condition $x(0)$ there exists t_0 such that $x(t) \in B_0(R)$ for all $t \geq t_0$. Obviously, all attractors of such a system lie inside $B_0(R)$.

Theorem 2. *A necessary and sufficient condition for the system (12) to be bounded is that* $\sigma_1 = \ldots = \sigma_n = -1$; *i.e., (12) must be*

$$y_i' = r_i - y_i^3 + \sum_{j=1}^{n} c_{ij} y_j, \tag{13}$$

which corresponds to a multiple supercritical cusp bifurcation.

Proof. We are interested in the flow structure of (12) outside some ball $B_0(R)$ with sufficiently large radius R. Let $\delta = R^{-2}$ be a small parameter. After the rescaling $x = \sqrt{\delta} y$, $t = \delta^{-1}\tau$, (12) can be rewritten as

$$\dot{x}_i = \sigma_i x_i^3 + \delta\left(\sum_{j=1}^{n} c_{ij} x_j + \sqrt{\delta} r_i\right),$$

which must be studied outside the unit ball $B_0(1)$. Note that this is a δ-perturbation of the uncoupled system

$$\dot{x}_i = \sigma_i x_i^3, \quad i = 1, \ldots, n. \tag{14}$$

Obviously, the unit ball $B_0(1)$ attracts all trajectories of (14) if and only if all $\sigma_i < 0$. Any δ-perturbation of (14) has the same property, provided that δ is small enough. □

So, the flow structure of (12) outside a ball with sufficiently large radius looks like that of (14), as depicted in Fig. 5 for the cases $(\sigma_1, \sigma_2) = (-1, \mp 1)$.

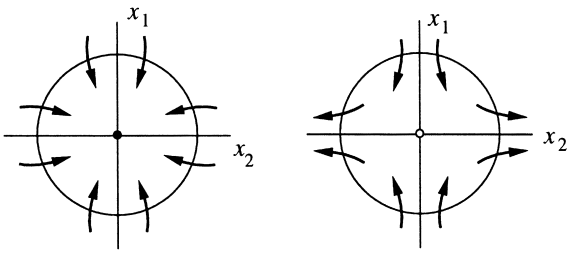

Fig. 5. Global flow structures for the canonical model (12) and (14) for $n = 2$. When $\sigma_1 = \sigma_2 = -1$ (left), the activity is bounded. It is not bounded when $\sigma_2 = +1$ (right).

Bounded dynamics is a desirable property in applications. Any initial condition $x(0)$ lies in a domain of attraction of some attractor that lies somewhere inside $B_0(R)$. Hence, for any $x(0)$ we have at least a hope of finding the asymptotic dynamics. From now on we will consider only (13) as the canonical model of the WCNN near a multiple cusp bifurcation point; i.e., we study the supercritical cusp bifurcation.

4. Neural network types

Most of the neural networks (NNs) dealing with static input patterns can be divided into two groups according to the way in which the pattern is presented:

- *MA-type* (*Multiple Attractor NN*). The key pattern is given as an initial state of the network, and the network converges to one of many attractors (Hopfield, 1982; Grossberg, 1988).
- *GAS-type* (*Globally Asymptotically Stable NN*). The key pattern is given as a parameter, which controls the location and shape of a unique attractor (Hirsch, 1989).

All attractors usually considered in these NNs are equilibrium points, although there are many attempts to understand the role of limit cycle and chaotic attractors in brain functioning, see Hoppensteadt and Izhikevich (1997) and Elbert et al. (1994) for reviews.

4.1. Canonical model as GAS-type NN

Systems that have only one asymptotically stable equilibrium point and do not have any other attractors are called *globally asymptotically stable* systems (Hirsch, 1989).

Theorem 3. *The dynamical system*

$$\dot{y}_i = r_i - y_i^3 + \sum_{j=1}^{n} c_{ij} y_j$$

is globally asymptotically stable if

$$c_{ii} < -\frac{1}{2}\sum_{j \neq i} |c_{ij} + c_{ji}|, \tag{15}$$

for all $i = 1, \ldots, n$.

Proof. We prove this using Hirsch' theorem (Hirsch, 1989). Let L be the Jacobian of the system above at a point $y = (y_1,\ldots,y_n)$. Hirsch' theorem claims that if there is a constant $-\delta < 0$ such that

$$\langle L\xi, \xi\rangle \le -\delta\langle \xi, \xi\rangle,$$

for all $\xi = (\xi_1,\ldots,\xi_n) \in \mathbb{R}^n$ then the system is globally asymptotically stable. Here $\langle \xi,\eta\rangle$ denote the inner (dot) product of vectors ξ and η. It is easy to check that

$$\begin{aligned}\langle L\xi, \xi\rangle &= -3\sum_{i=1}^{n} y_i^2\xi_i^2 + \sum_{i,j=1}^{n} c_{ij}\xi_i\xi_j = \sum_{i=1}^{n}(c_{ii} - 3y_i^2)\xi_i^2 \\ &\quad + \frac{1}{2}\sum_{\substack{i,j=1\\ i\neq j}}^{n}(c_{ij} + c_{ji})\xi_i\xi_j \le \sum_{i=1}^{n} c_{ii}\xi_i^2 \\ &\quad + \frac{1}{4}\sum_{\substack{i,j=1\\ i\neq j}}^{n}|c_{ij} + c_{ji}|(\xi_i^2 + \xi_j^2) \\ &= \sum_{i=1}^{n}\left(c_{ii} + \frac{1}{2}\sum_{j\neq i}|c_{ij} + c_{ji}|\right)\xi_i^2 < -\delta\sum_{i=1}^{n}\xi_i^2,\end{aligned}$$

where $-\delta = \max_i(c_{ii} + 1/2\Sigma_{j\neq i}|c_{ij} + c_{ji}|)$. We used here the fact that

$$\xi_i\xi_j \le \tfrac{1}{2}(\xi_i^2 + \xi_j^2).$$

The inequality (15) guarantees that $-\delta < 0$ and, hence, all the conditions of Hirsch's theorem are satisfied. This completes the proof. □

Remark 4. The external input $r \in \mathbb{R}^n$ does not come into the condition (15). What it does affect is the location of the unique attractor of the network. Therefore, the canonical model can operate as a GAS-type NN.

4.2. Canonical model as MA-type NN

The following theorem is due to Cohen and Grossberg (1983). It was generalized for oscillatory neural networks (complex valued y_i and c_{ij}) by Hoppensteadt and Izhikevich (1996a, b, 1997).

Theorem 5 (Cohen and Grossberg, 1983). *If the matrix of synaptic connections $C = (c_{ij})$ is symmetric, then the neural network*

$$\dot{y}_i = r_i - y_i^3 + \sum_{j=1}^{n} c_{ij}y_j \tag{16}$$

is a gradient system.

Proof. To prove the theorem it suffices to present a function $U : \mathbb{R}^n \to \mathbb{R}$ satisfying

$$\dot{y}_i = -\frac{\partial U}{\partial y_i}.$$

It is easy to check that

$$U(y) = -\sum_{i=1}^{n}\left(r_i y_i - \frac{1}{4}y_i^4\right) - \frac{1}{2}\sum_{i,j=1}^{n} c_{ij}y_i y_j$$

is such a function for (16). Note that far away from the origin $U(y)$ behaves like $1/4\Sigma_{i=1}^{n} y_i^4$, hence $U(y)$ is bounded below. □

Being a gradient system imposes many restrictions on the possible dynamics of (16). For example, its dynamics cannot be oscillatory or chaotic. Nevertheless, this property is considered to be very useful from neuro-computational point of view. In particular, it implies that the canonical model (16) can operate as an MA-type NN.

The symmetry of $C = (c_{ij})$ is a strong requirement, but it arises naturally if one considers the Hebbian learning rules, which we discuss in the next section.

5. Hebbian learning rule

Not much is known about learning in the central nervous system, so our assumptions, which have already been used by Hoppensteadt and Izhikevich (1996b, 1997), are reasonably nonrestrictive. We assume that

- Learning results from modifying synaptic connections between neurons (Hebb, 1949).
- Learning is local; i.e., the modification depends upon activities of pre- and postsynaptic neurons and does not depend upon activities of the other neurons.
- The modification of synapses is slow compared with characteristic times of neuron dynamics.
- If either pre- or postsynaptic neurons or both are silent, then no synaptic changes take place except for exponential decay, which corresponds to forgetting.

These assumptions in terms of the WCNN

$$\dot{x}_i = f_i(x_i, \lambda) + \varepsilon g_i(x_1, \ldots, x_n, \lambda, \rho, \varepsilon)$$

have the following implications: The first hypothesis states that learning is described by modification of the coefficients $c_{ij} = \partial g_i/\partial x_j$. Recall that the actual synaptic connections have order ε. We denote them temporary by $w_{ij} = \varepsilon c_{ij}$. The second hypothesis says that for fixed i and j the coefficient w_{ij} is modified according to equations of the form

$$w_{ij}' = H(w_{ij}, x_i, x_j).$$

We introduce the slow time $\tau = \varepsilon t$ to account for the third hypothesis. We say that a neuron is silent if its activity is at an equilibrium point. Then the fourth hypothesis says that

$$H(w_{ij}, 0, x_j) = H(w_{ij}, x_i, 0) = H(w_{ij}, 0, 0) = -\gamma w_{ij} + \alpha w_{ij}^2 + \cdots$$

for all x_i,x_j, so that H has the form

$$H(w_{ij}, x_i, x_j) = -\gamma w_{ij} + \beta x_i x_j + \delta_1 w_{ij}x_i + \delta_2 w_{ij}x_j + \alpha w_{ij}^2 + \cdots$$

Now we use the fact that the synaptic coefficient w_{ij} is of order ε, and rewrite this in the original variables $c_{ij} = w_{ij}/\varepsilon$, $y_i = x_i/\sqrt{\varepsilon}$, to obtain the learning rule

$$c_{ij}' = -\gamma c_{ij} + \beta y_i y_j + O(\sqrt{\varepsilon}), \tag{17}$$

which we refer to as the *Hebbian* synaptic modification rule. Note that although we consider general functions H, after the rescaling only two constants, γ and β, are significant to leading order. They are the rate of memory fading and the rate of synaptic plasticity, respectively.

If the neural network activity $y(t) = \xi \in \mathbb{R}^n$ is a constant, then

$$c_{ij} \to \frac{\beta}{\gamma}\xi_i\xi_j.$$

This leads to the following well known definition of the Hebbian learning rule for synaptic matrix C.

5.1. Learning rule for synaptic matrix C

Suppose $\xi^1,\ldots,\xi^m \subset \mathbb{R}^n$ are m key patterns to be memorized and reproduced by a neural network. Here each $\xi^s = (\xi_1^s, \ldots, \xi_n^s)^{\mathrm{T}} \in \mathbb{R}^n$, $s = 1,\ldots,m$, is a vector, and $m \le n$. Let constants β_s measure ''strength'', or ''quality'', of the memory about the patterns ξ^s; Then the matrix C is said to be formed according to a Hebbian learning rule if

$$c_{ij} = \frac{1}{n}\sum_{s=1}^{m} \beta_s \xi_i^s \xi_j^s, \quad 1 \le i,j \le n. \tag{18}$$

The Hebbian learning rule (18) can be rewritten in the more convenient form

$$C = \frac{1}{n}\sum_{s=1}^{m} \beta_s \xi^s (\xi^s)^{\mathrm{T}}, \tag{19}$$

where superscript T means transpose.

It is easy to see that the synaptic matrix C constructed according to (19) is symmetric. It is also true that any symmetric matrix C can be represented as (19) for some, possibly non-unique, choice of the orthogonal vectors $\xi^1,\ldots,\xi^m$. Thus, we can state the following result (Hoppensteadt and Izhikevich, 1997).

Proposition 6. *The matrix of synaptic connections C is symmetric if and only if there is a set of orthogonal patterns $\xi^1,\ldots,\xi^m$ such that C is constructed according to the Hebbian learning rule (19).*

Note that the orthogonal vectors $\xi^1,\ldots,\xi^m$ are eigenvectors of C. If we normalize them such that

$$|\xi^s|^2 = \langle \xi^s, \xi^s \rangle = \sum_{i=1}^{n} (\xi_i^s)^2 = n,$$

then the constants β_s are eigenvalues of C and (19) is the spectral decomposition of C. We assume that $\beta_1 \ge \cdots \ge \beta_m > 0$. If $m < n$, then there is an $n - m$ dimensional eigenspace $\ker C \subset \mathbb{R}^n$ corresponding to the zero eigenvalue. We denote this eigenvalue by $\beta_{m+1} = 0$.

To summarize we can say that the Hebbian learning rule for orthogonal patterns gives a way of constructing the matrix of synaptic connections such that each pattern is an eigenvector of the matrix corresponding to a positive eigenvalue.

In order to analyze the canonical model without resorting to computer simulations we also assume that $|\xi_i^s| = 1$ for all i, i.e.

$$\xi^s = (\pm 1, \ldots, \pm 1), \quad s = 1, \ldots, m.$$

For these purposes we introduce the set

$$\Xi^n = \{\xi \in \mathbb{R}^n,\ \xi = (\pm 1, \ldots, \pm 1)\} \subset \mathbb{R}^n. \tag{20}$$

We will also need for our analysis an orthogonal basis for $\mathbb{R}^n$ that contains vectors only from Ξ^n. This basis always exists if $n = 2^k$ for some integer $k > 0$. The assumption that $\xi^s \in \Xi^n$ might look artificial, but it is very important in neurocomputer applications and in digital circuit design.

6. Multiple pitchfork bifurcation ($r = 0$)

We start the bifurcation analysis of the canonical model (13) for the special case when there are no receptor inputs, i.e. when $r_1 = \cdots = r_n = 0$. We introduce a new bifurcation parameter b and subtract it from c_{ii}, so that the canonical model has the form

$$y_i' = by_i - y_i^3 + \sum_{j=1}^{n} c_{ij} y_j. \tag{21}$$

It is easy to check that (21) describes a WCNN with Z_2 symmetry $y \to -y$ near a *multiple pitchfork* bifurcation point. Symmetry means that for $b > 0$ each neuron is essentially a bistable element with two states: depolarization ($y_i = \sqrt{b}$) and hyperpolarization ($y_i = -\sqrt{b}$).

6.1. Stability of the origin: General C

Note that (21) always has an equilibrium point $y_1 = \cdots = y_n = 0$. It follows from Theorem 3 that the origin is the only equilibrium point, which is a stable node, for sufficiently small b. What happens while b is increasing? How does the origin lose its stability? How many and of what type are the new equilibrium points? What is the relationship between them and the synaptic matrix $C = (c_{ij})$? These and other questions are studied in this section.

Let L be the Jacobian of the right-hand side of (21) at the origin. It is easy to see that

$$L = bE + C,$$

where E is the unit matrix. Let $\beta_1,\ldots,\beta_m$ be the eigenvalues of C ordered such that

$$\mathrm{Re}\,\beta_1 \ge \cdots \ge \mathrm{Re}\,\beta_m.$$

Obviously, L has m eigenvalues

$$\lambda_s = b + \beta_s, \quad s = 1, \ldots, m$$

with the same eigenvectors as those of C. The matrix L has all eigenvalues with negative real parts, and, hence, the origin is a stable equilibrium point for (21) if and only if $b < -\mathrm{Re}\,\beta_1$ (see Fig. 6A).

6.1.1. Bifurcations of the origin

If β_1 is a real eigenvalue with multiplicity one, then (21) undergoes a pitchfork bifurcation when b crosses $-\beta_1$. For b slightly larger than $-\beta_1$ the origin is a saddle surrounded by two sinks (see Fig. 6b) and those are the only equilibrium points for (21).

If (β_1,β_2) is a pair of complex conjugate eigenvalues with multiplicity one, then we can observe the Andronov–Hopf bifurcation for $b = -\mathrm{Re}\,\beta_1$.

For $b > -\beta_1$ it is possible to observe the birth of a pair of saddles or an unstable limit cycle every time b crosses $-\mathrm{Re}\,\beta_s$, where β_s is an eigenvalue with multiplicity one.

For the eigenvalues with multiplicity more than one bifurcations could be more complicated. Nevertheless, we will consider some of them later.

6.1.2. Self-ignition

Recall that each neuron is bistable only for $b > 0$. For negative b there is only one stable state $y_i = 0$ and, hence, it is ''passive''. But when the neurons are connected they acquire a new property: bistability for $-\beta_1 < b < 0$. This is the property that each neuron alone cannot have. Thus a network of ''passive'' elements can exhibit ''active'' properties. This effect, which can be called self-ignition (Rapaport, 1952), is discussed in details by Kowalski et al. (1992). Smale (1974) interpreted this phenomena in terms of live and dead cells: Two non-oscillating (dead) cells start to oscillate (become alive) when they are coupled together. We encounter this phenomena frequently in our analysis of brain function.

6.2. Stability of the other equilibria

It is noteworthy that we have not restricted the synaptic matrix C yet. All the bifurcations discussed above take place for any C. In return for this generality, we cannot trace the new equilibria and study their stability. Fortunately, we can do it if we assume that the synaptic matrix C is constructed according to the Hebbian learning rule,

$$C = \frac{1}{n}\sum_{s=1}^{m} \beta_s \xi^s (\xi^s)^{\mathrm{T}}, \quad \beta_1 \geq \cdots \geq \beta_m > 0,$$

and that the memorized images $\xi^1,\ldots,\xi^m \in \Xi^n$ are orthogonal. For simplicity we assume that all β_s are different. At the end of this section we will discuss the case $\beta_1 = \cdots = \beta_m$. Let $y_i = x_s \xi_i^s$ for $i = 1,\ldots,n$. Then

$$y_i' = x_s' \xi_i^s = b x_s \xi_i^s - x_s^3 (\xi_i^s)^3 + \beta_s x_s \xi_i^s.$$

After (dot) multiplication by ξ_i^s, we have

$$x_s' = (b + \beta_s) x_s - x_s^3. \tag{22}$$

This equation has only one equilibrium point $x_s = 0$ for $b < -\beta_s$. For $b > -\beta_s$ there are three points $x_s = 0$, $x_s = \pm\sqrt{b+\beta_s}$. Hence, the original system (21) has two new equilibrium points $y = \pm\sqrt{b+\beta_s}\,\xi^s$ after b crosses $-\beta_s$.

Note that the pair of new equilibrium points lies on the line spanned by the memorized pattern ξ^s. Every attractor lying on or near $\mathrm{span}(\xi^s)$ is called an attractor corresponding to the pattern ξ^s. When the network activity $y(t)$ approaches such an attractor, we say that it has recognized the memorized image ξ^s.

Since the origin is unstable for $b > -\beta_1$, only the pair $y = \pm\sqrt{b+\beta_1}\,\xi^1$ is a pair of stable nodes, whereas the others $y = \pm\sqrt{b+\beta_s}\,\xi^s$, $s \geq 2$ are pairs of saddles, at least when b is near $-\beta_s$ (see Fig. 6C).

Let us study the stability of the equilibrium $y = \pm\sqrt{b+\beta_k}\,\xi^k$ for some $1 \leq k \leq m$. The matrix of linearization L at the equilibrium is

$$L = (b - 3(b + \beta_k))E + C.$$

It has eigenvalues

$$\lambda_s = -2(b + \beta_k) + \beta_s - \beta_k, \quad 1 \leq s \leq m+1,$$

where $\beta_{m+1} = 0$ corresponds to $\ker C$. Note that $\lambda_1 \geq \cdots \geq \lambda_m > \lambda_{m+1}$. The maximum eigenvalue $\lambda_1 = -2(b + \beta_k) +$

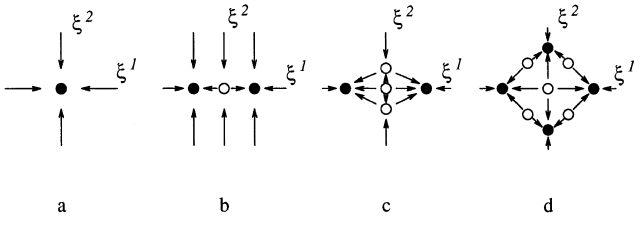

Fig. 6. Phase portrait of the canonical model (21) of weakly connected neural network near multiple pitchfork bifurcation point for different values of the bifurcation parameter b. (A) $b < -\beta_1$; (B) $-\beta_1 < b < -\beta_2$; (C) $-\beta_2 < b < -\beta_2 + (\beta_1 - \beta_2)/2$; (D) $-\beta_2 + (\beta_1 - \beta_2)/2 < b$.

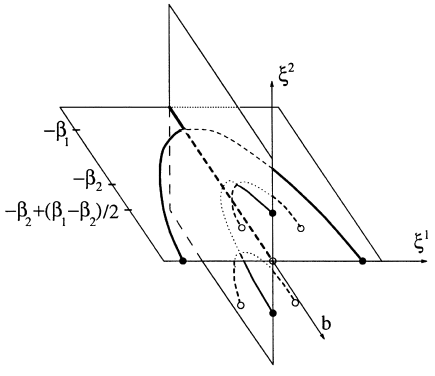

Fig. 7. Bifurcation diagram.

$\beta_1 - \beta_k$ is always negative only for $k = 1$. For $k \geq 2$ the inequality $\lambda_1 < 0$ gives us the condition

$$b > -\beta_k + \frac{\beta_1 - \beta_k}{2}. \tag{23}$$

One could say that in the life of the equilibrium point $\sqrt{b+\beta_k}\xi^k$ there are two major events: Birth (when $b = -\beta_k$) and maturation ($b = -\beta_k + (\beta_1 - \beta_k)/2$) when the point becomes a stable node. For $k = 2$ see Fig. 6D and Fig. 7.

It is easy to see that when $b = -\beta_k$ the eigenvalues of L are

$$\lambda_1 \geq \cdots \geq \lambda_{k-1} \geq 0 = \lambda_k \geq \lambda_{k+1} \geq \cdots \geq \lambda_m.$$

So, $\sqrt{b+\beta_k}\xi^k$ is the saddle such that $k - 1$ directions corresponding to $\xi^1,\ldots,\xi^{k-1}$ are unstable (see Fig. 8A). Every time b crosses $-\beta_k + (\beta_s - \beta_k)/2$, $s < k$ there is a pitchfork bifurcation. As a result, the direction corresponding to ξ^s becomes stable and there appears a new pair of saddles lying in span(ξ^s,ξ^k) (see Fig. 8B–D).

To summarize, we can say that for $b < -\beta_1$ the only equilibrium point is the origin, but for $b > -\beta_m + (\beta_1 - \beta_m)/2$ there are m pairs of stable nodes corresponding to the memorized images $\xi^1,\ldots,\xi^m$ and many saddles lying in between these nodes.

6.2.1. Spurious memory

It is customary to refer to the attractors that do not correspond to any of the memorized images as being spurious memory. Is there any spurious memory in (21)? The answer is *yes*. Fortunately, it happens for large b. Indeed, when $b > 0$ all eigenvalues of the matrix of linearization at the origin are positive, not only $\lambda_1,\ldots,\lambda_m$. The new unstable directions correspond to ker C (of course, if $m < n$). It is easy to check that for $0 < b < \beta_1/2$ all the equilibrium points lying in ker C are saddles (except the origin, which is an unstable node), whereas for $b > \beta_1/2$ there are $2(n - m)$ stable nodes among them.

In order to avoid spurious memory, one should keep the bifurcation parameter below the critical value $\beta_1/2$. Actually, it is more reasonable to demand that b be negative. By this means we guarantee that nothing interesting is going on in the directions orthogonal to the all memorized images. But we must be cautious because not all equilibrium points corresponding to memorized patterns are stable for $b < 0$. Indeed, the stability condition (23) for $b < 0$ can be satisfied only if

$$\beta_k > \frac{\beta_1}{3}.$$

Thus, we have proved the following

Proposition 7. *All memorized images are stable nodes and successful recognition is possible if the weight β_m of the weakest image is greater than one third of that of the strongest one.*

Obviously, we do not have this kind of problem when $\beta_1 = \cdots = \beta_m = \beta > 0$. For $b < -\beta$ the neural network is globally asymptotically stable. For $b = -\beta$ there is a multiple pitchfork bifurcation with the birth of $2m$ stable nodes corresponding to the memorized images. For $-\beta < b < 0$ these nodes are the only attractors[3] and the behavior of (21) is very simple in the directions orthogonal to the span($\xi^1,\ldots,\xi^m$). Thus, the system (21) can work as a typical MA-type NN. If the initial condition $y(0)$ is an input from receptors, then the activity $y(t)$ of (21) approaches the closest attractor, which corresponds to one of the previously memorized images. It is believed that this simple procedure is a basis for neurocomputing.

Nevertheless, our opinion is that this is too far from the basic principles of how a real brain functions (despite the fact that we know almost nothing about these principles). In the next section we explore another, more realistic, approach.

7. Bifurcations for $r \neq 0$

Comprehensive analysis of the canonical model

$$y_i' = r_i + by_i - y_i^3 + \sum_{j=1}^{n} c_{ij}y_j \tag{24}$$

is formidable, even for a symmetric synaptic matrix $C = (c_{ij})$, although there is a Liapunov function in that case. Hence, every piece of information about (24) obtained by analytical tools is precious. The next step in studying (24) is to assume that the number of memorized images $m \leq 2$. The key result in this direction is the Reduction Lemma that enables us to reduce the number of independent variables to 2.

7.1. The reduction lemma

Suppose that $\xi^1,\ldots,\xi^n \in \Xi^n$ form an orthogonal basis for $\mathbb{R}^n$ and that ξ^1 and ξ^2 coincide with memorized images, where Ξ is defined in (20). Let

$$x_s = \frac{1}{n}\langle y, \xi^s\rangle = \frac{1}{n}\sum_{i=1}^{n} y_i\xi_i^s$$

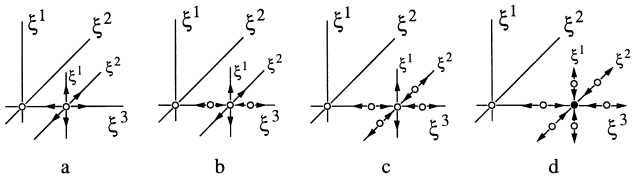

Fig. 8. Every equilibrium point $\pm\sqrt{b+\beta_k}\xi^k$ becomes an attractor after the sequence of the pitchfork bifurcations. Every time b crosses $-\beta_k + (\beta_s - \beta_k)/2$, $s < k$, the ξ^s-direction becomes stable.

[3] Actually, the correct statement is that these are the only attractors that bifurcated from the origin. Whether there are other attractors or not is still an open question.

be the projection of $y \in \mathbb{R}^n$ onto $(1/n)\xi^s$. Obviously, y can be represented as the sum

$$y = \sum_{s=1}^{n} x_s \xi^s. \tag{25}$$

A similar decomposition is possible for any input $r \in \mathbb{R}^n$: Let

$$a_s = \frac{1}{n}\langle r, \xi^s \rangle,$$

then

$$r = \sum_{s=1}^{n} a_s \xi^s. \tag{26}$$

Let us prove the following:

Lemma 8 (Reduction Lemma). *If*

$$C = \frac{1}{n}(\beta_1 \xi^1 (\xi^1)^{\mathrm{T}} + \beta_2 \xi^2 (\xi^2)^{\mathrm{T}})$$

for orthogonal $\xi^1, \xi^2 \in \Xi^n$, $b < 0$ *and* $a_3 = \cdots = a_n = 0$, *then the plane* $x_3 = \cdots = x_n = 0$ *is a stable invariant manifold for (24) and dynamics on the manifold are governed by the system*

$$\begin{cases} x_1' = a_1 + (b+\beta_1)x_1 - 3x_1 x_2^2 - x_1^3, \\ x_2' = a_2 + (b+\beta_2)x_2 - 3x_2 x_1^2 - x_2^3, \end{cases} \tag{27}$$

see Fig. 9.

The condition $a_3 = \cdots = a_n = 0$ means that the external input r is in the span(ξ^1, ξ^2). For example, if we studied the olfactory system, then with this restriction only two odors are inhaled and recognizable.

Proof. Substituting (25) and (26) into (24) gives

$$\sum_{s=1}^{n} x_s' \xi_i^s = \sum_{s=1}^{n} a_s \xi_i^s + b \sum_{s=1}^{n} x_s \xi_i^s - \sum_{s=1}^{n}\sum_{p=1}^{n}\sum_{q=1}^{n} x_s x_p x_q \xi_i^s \xi_i^p \xi_i^q + \sum_{s=1}^{n} x_s \beta_s \xi_i^s,$$

where $\beta_3 = \cdots = \beta_n = 0$. Projecting both sides onto $(1/n)\xi^k$ gives

$$x_k' = a_k + (b+\beta_k)x_k - \sum_{s=1}^{n}\sum_{p=1}^{n}\sum_{q=1}^{n} x_s x_p x_q \left(\frac{1}{n}\sum_{i=1}^{n} \xi_i^k \xi_i^s \xi_i^p \xi_i^q\right), \tag{28}$$

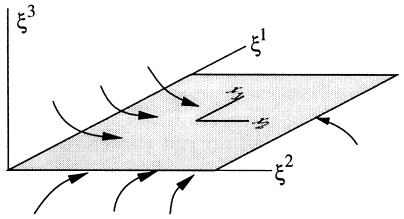

Fig. 9. The plane spanned by ξ^1 and ξ^2 is stable and invariant when the canonical model lives in a world having two images.

Note that

$$\frac{1}{n}\sum_{i=1}^{n} \xi_i^k \xi_i^s \xi_i^p \xi_i^q = \begin{cases} 1, & \text{if } k=s,\ p=q \\ & \text{or } k=p,\ s=q \\ & \text{or } k=q,\ s=p, \\ d_{kspq}, & \text{all indices are different,} \\ 0, & \text{otherwise,} \end{cases}$$

where $d_{kspq} \in \mathbb{R}$ are some constants. We used the assumption that $\xi_i^s = \pm 1$ for any s and i.

If the number of the memorized images m were greater than 2, then all equations in (28) would contain the constants d_{kspq}. It is possible to eliminate them if we consider (28) on the plane $x_3 = \cdots = x_n = 0$ for $m \le 2$. Indeed, the product $x_s x_p x_q$ is always zero unless $1 \le s,p,q \le 2$. The inequality guarantees that at least two indices coincide. Hence, the sum $1/n\Sigma_{i=1}^{n} \xi_i^k \xi_i^s \xi_i^p \xi_i^q$ is either 1 or 0. It is 1 when all the indices are equal (this gives x_k^3) or when k is equal to only one of the three indices s,p,q (there are three such possibilities and, hence, the term $3x_k x_{3-k}^2$). Thus, the system (28) on the plane can be rewritten as (27).

We still must show that the plane is a stable invariant manifold. From the lemma's conditions we know that $a_3 = \cdots = a_n = 0$ and $\beta_3 = \cdots = \beta_n = 0$. Let us fix x_1 and x_2 and consider them as parameters. Keeping only linear terms, we can rewrite (28) for $k \ge 3$ as

$$x_k' = bx_k - 3x_k(x_1^2 + x_2^2) + \text{h.o.t.}, \quad k = 3, \ldots, n.$$

The plane is invariant because $x_3' = \cdots = x_n' = 0$ on it. It is stable because $b - 3(x_1^2 + x_2^2) < b < 0$, which follows from the lemma's condition that $b < 0$. □

It is still an open question whether the invariant plane is globally asymptotically stable or not. Our conjecture is that for $b < 0$ it is true, but we do not need this for our analysis below.

If $\beta_1 = \beta_2 = \beta$, then it is easy to check that (27) can be rewritten as

$$\begin{cases} u' = s + (b+\beta)u - u^3, \\ v' = c + (b+\beta)v - v^3, \end{cases} \tag{29}$$

where $u = x_1 + x_2$, $v = x_1 - x_2$, $s = a_1 + a_2$ and $c = a_1 - a_2$. The advantage of (29) is that it is uncoupled and each equation can be studied independently.

7.2. Recognition: Only one image is presented

Without loss of generality we may assume that it is ξ^1, i.e.

$$r_i = a\xi_i^1.$$

Assuming that all the conditions of the Reduction Lemma are satisfied and that $\beta_1 = \beta_2 = \beta$ we can rewrite (24) as

(29). Note that $s = c = a$. Thus, the dynamics on the (u,v) plane is the direct product of two identical equations

$$\dot{z} = a + (b+\beta)z - z^3.$$

If $b + \beta < 0$, then there is only one equilibrium point for any a. The dynamics on the (u,v) plane is qualitatively the same as that of the canonical model of WCNN near multiple pitchfork bifurcation point (21) for $b + \beta_1 < 0$ which is depicted in Fig. 6A.

7.2.1. Weak input

Suppose $b + \beta > 0$. There are three equilibrium points when $|a| < a^*$, where

$$a^* = 2\left(\frac{b+\beta}{3}\right)^{3/2},$$

see Fig. 10. Hence, (29) has nine equilibrium points. Again, there is not any qualitative distinction between the phase portrait of (21) depicted in Fig. 6D and that of (29), which we depict in Fig. 11A for $a > 0$. We see that $|a| < a^*$ is too weak to produce any qualitative changes in the dynamics of the canonical model (24) in comparison with (21). Nevertheless, it is easy to see that the domain of attraction of the equilibrium point corresponding to the presented image ξ^1 is much bigger than the attraction domains of the other equilibrium points. By the term attraction domain *size* we mean here the distance from the attractor to the closest saddle. We use this definition in order to be able to compare domains that have infinite volumes.

7.2.2. Strong input

When the parameter a crosses $\pm a^*$ one can observe two saddle–node bifurcations and one co-dimension-2 bifurcation (see Fig. 11B). All of them take place simultaneously due to the fact that (29) is a direct product of two identical equations. We consider these bifurcations elsewhere when we study the canonical model for WCNNs near multiple saddle–node bifurcation point (Hoppensteadt and Izhikevich, 1997).

If the input $r = a\xi^1$ is sufficiently strong (i.e. if $|a| > a^*$), then there is only one equilibrium point, which is a stable node (see Fig. 11C). The equilibrium point is globally asymptotically stable in this case.

We see that the canonical model (24) can operate as

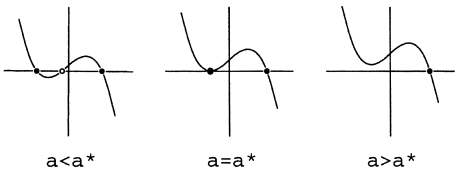

Fig. 10. The graph of $a + (b+\beta)z - z^3$ for $b + \beta > 0$ and various values of a.

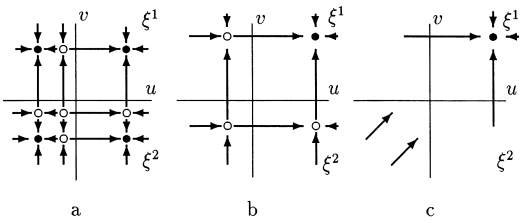

Fig. 11. Phase portrait of the canonical model on the stable invariant plane span(ξ^1,ξ^2). The first image is presented as an input onto the network. (A) Input is weak, i.e. $|a| < a^*$; (B) For $|a| = a^*$ there are fold bifurcations; (C) For $|a| > a^*$ the canonical model is globally asymptotically stable.

GAS-type NN when the input strength a is strong enough, viz.

$$|a| > 2\left(\frac{b+\beta}{3}\right)^{3/2}.$$

We performed all the analyses above for the case of one presented and two memorized images.

7.3. Recognition: Two images are presented

Without loss of generality we may assume in this case that

$$r = a_1\xi^1 + a_2\xi^2$$

for a_1,$a_2 > 0$. If $\beta_1 = \beta_2 = \beta$ and all the conditions of the Reduction Lemma are satisfied, then the canonical model (24) can be reduced to the two-dimensional system (29)

$$\begin{cases} \dot{u} = s + (b+\beta)u - u^3, \\ \dot{v} = c + (b+\beta)v - v^3. \end{cases}$$

We cannot reduce (29) to a one-dimensional system because in general $s \neq c$. The constant $s = a_1 + a_2$ has obvious meaning of overall *strength* of the input from receptors, whereas $c = a_1 - a_2$ is the *contrast* of the input. When $c > 0$ ($c < 0$) we say that ξ^1 (ξ^2) is dominant.

In order to determine the qualitative behavior of (29) we have to compare s and c with the bifurcation value a^*. When both s and c are less than a^*, the qualitative phase portrait of (29) depicted in Fig. 12A coincides with that of (21) depicted in Fig. 6D provided $b + \beta > 0$.

Very interesting behavior arises when the overall input from receptors s is strong enough, i.e. when $s > a^*$. Then,

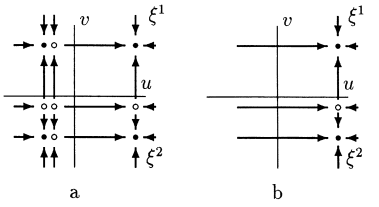

Fig. 12. Phase portrait of the canonical model on the stable invariant plane span(ξ^1,ξ^2). The input is a mixture of two images ξ^1 and ξ^2. (A) Overall input is weak; (B) Strong input and weak contrast. There is a co-existence of two attractors.

(29) generically has either one or three equilibrium points (see Fig. 12B). Its behavior is determined by the equation

$$\dot{v} = c + (b+\beta)v - v^3, \quad v \in \mathbb{R}. \tag{30}$$

Obviously, the dynamics of (30) depends crucially not only upon which image is dominant but also upon how dominant it is. If $|c| < a^*$, then there is a co-existence between these two images (see Fig. 12B). Both equilibria points are stable. If $|c| > a^*$, then only one image survives, viz. the dominant image.

One possibility to explain the co-existence of two attractors corresponding to two different images is that the NN cannot distinguish between them when the contrast $|c|$ is small. One could say that the two-attractor state corresponds to the ''*I do not know*'' answer. We prefer another explanation suggested by the psychological experiment described in the next section.

8. Bistability of perception

In the previous sections we showed that if the conditions of the Reduction Lemma are satisfied and the overall input from receptors is strong ($s > a^*$) then the entire canonical model behaves qualitatively like Eq. (30)

$$\dot{v} = c + bv - v^3, \quad v \in \mathbb{R},$$

where $c = a_1 - a_2$ is the contrast between two images $a_1\xi^1$ and $a_2\xi^2$ and b is a real parameter (we incorporated β into b, so b can be positive or negative). We have already mentioned that if the contrast is weak ($|c| < a^*$) then (30) has two attractors corresponding to the previously memorized images ξ^1 and ξ^2. First of all, note that the co-existence of two attractors contradicts the GAS-type NN paradigm which requires that the NN have only one attractor. We must accept the fact that the brain is a very complicated system having many attractors. Its dynamic behavior depends not only upon the input r, the synaptic memory C and the psychological state b but also upon a short-term past activity (which sometimes is called a short-term memory, Grossberg, 1988). In our case this is the initial condition $x(0)$. Obviously, which attractor will be selected by the NN depends upon the initial state. Simultaneous existence of several attractors for the input that is a mixture of images suggests the following hypothesis: The NN perceives the ambiguous input according to the network's past short-term activity $x(0)$.

The behavior of the *artificial* NN (24) is similar to the behavior of the real human brain in the following psychological experiment (Attneave, 1971): The fourth figure from the left in the top row depicted in Fig. 13 was shown to be perceived with equal probability as a human face or a body. If the figure is included in a sequence, then its perception depends upon the direction in which the sequence is viewed.

This phenomenon was studied from catastrophe theory point of view (Poston and Stewart, 1978; Stewart and Peregoy, 1983) and it was shown that there is a one-dimensional section of a cusp catastrophe in the human perception of the figures.

The remarkable fact is that the WCNN approximated by (30) also exhibits the cusp catastrophe. Suppose ξ^1 and ξ^2 represent human body and face images, respectively. If we fix $b > 0$ and vary the image contrast $c = a_1 - a_2$, then the artificial NN also has the same bistable perception of the presented images $a_1\xi^1$ and $a_2\xi^2$ (see bottom row in Fig. 13).

What we have not explained yet is the switching of our attention (say, from the body to the face and back) while we observe an ambiguous picture. These oscillations in our perception cannot be explained by the catastrophe theory. We can tackle this problem by allowing the internal parameter λ to be a variable (see Izhikevich, 1996). This was used by Ditzinger and Haken (1989).

As we can see, the canonical model (24) can operate as the MA and GAS-type NNs simultaneously. Indeed, its dynamics crucially depends upon the input $r \in \mathbb{R}^n$ from receptors. If the input is strong enough and there is no ambiguity, then (24) has only one attractor and, hence, works as the GAS-type NN. If the input is weak or ambiguous, then (24) can have many attractors and, hence, can work as the MA-type neural network.

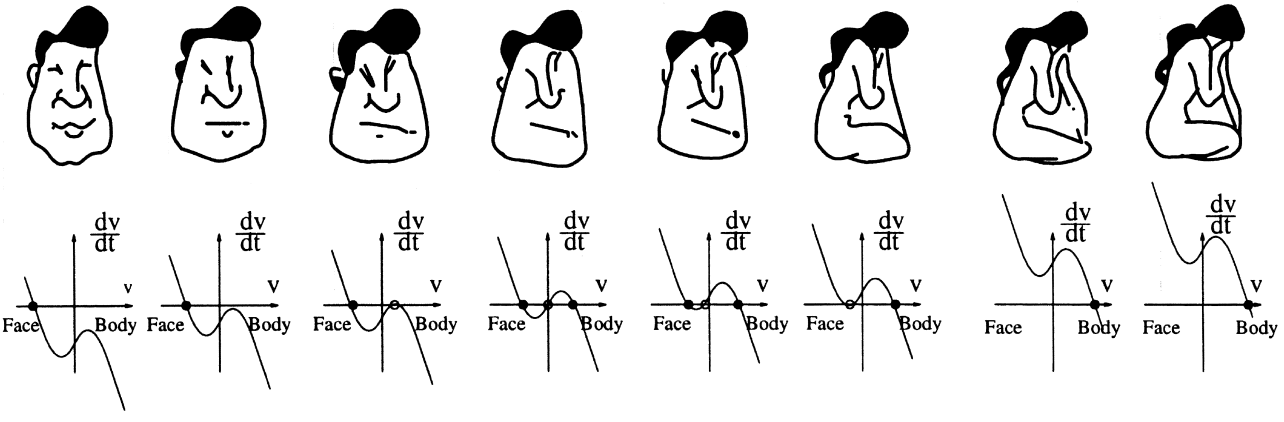

Fig. 13. Bistability of perception. Biasing sequence for the Fisher (1967) figure, due to Attneave (1971).

We think that the real brain might use similar principles. Consider, for example, the olfactory system (Baird, 1986; Erdi et al., 1993; Li and Hopfield, 1989; Skarda and Freeman, 1987). It is believed that each inhaled odor has its own attractor—a stable limit cycle. The analysis of the canonical model (24) suggests that when an animal inhales a mixture of the odors, the appropriate limit cycles become stable so that there is a one-to-one correspondence between the inhaled odors and the attractors. Similar results were obtained by studying another NN (Izhikevich and Malinetskii, 1993), but the attractors there were chaotic.

9. Discussion

The main purpose of the present paper is (1) to show how an arbitrary WCNN

$$\dot{X}_i = F_i(X_i, \lambda) + \varepsilon G_i(X_1, \ldots, X_n, \lambda, \rho, \varepsilon) \tag{31}$$

at a multiple supercritical cusp bifurcation can be transformed into the canonical model

$$y_i' = r_i - y_i^3 + \sum_{j=1}^{n} c_{ij} y_j \tag{32}$$

by a continuous possibly non-invertible change of variables, and (2) to prove that the canonical model can operate as MA-type and/or GAS-type neural network.

The canonical model (32) consists of linearly coupled cusp or pitchfork (when all $r_i = 0$) neurons. Moreover, (32) is probably the simplest continuous-time neural network model since each neuron has only one nonlinear term, namely y_i^3. Nevertheless, its behavior is far from being understood, except in some special cases, such as symmetric $C = (c_{ij})$.

This is especially frustrating after we take into account the universality of the canonical model (32): It captures nonlinear behavior of all weakly connected systems (31) near multiple cusp bifurcation regardless of the equations that describe dynamics of each cell X_i. Thus, the system (31) may be a weakly connected network of Wilson–Cowan oscillators, Hodgkin–Huxley neurons, Traub–Miles neurons, or any other know or still unknown neuron models. As soon as its neurons are near cusp bifurcations, the weakly connected network can be transformed into the canonical model. Particulars of the functions F_i and G_i do not affect the form of the canonical model (32), but affect only the values of the parameters r_i and c_{ij}.

The universality of the canonical model (32) should not be overemphasized: It describes activity of WCNN (31) only *near* multiple cusp bifurcation. When (31) is near multiple saddle–node bifurcation, Andronov–Hopf bifurcation, or has an invariant torus, it can also be reduced to a canonical model, which differs from (32). Derivation of many such canonical models can be found in Hoppensteadt and Izhikevich (1997).

The reason we are interested in canonical models is the following: We do not know and probably will never know all details of neuron dynamics. Therefore, we will never be able to determine exact functions F_i and G_i that describe exactly the dynamics of neurons in central nervous system. Moreover, we do not even know the dimension of each vector X_i denoting activity of the ith neuron or local population of neurons. Nevertheless, if such a system operates near a critical regime, such as multiple bifurcations, then it is possible to prove that there is a continuous change of variables that transforms the system into a simpler canonical model. Derivation of such canonical models is a daunting task, which is accomplished only partially. But once identified, the canonical models provide an invaluable information about behavior of many neural systems, even those that have not been discovered yet.

References

Attneave F. (1971). Multistability in perception. *Scientific American*, *225*, 63–71.

Baird B. (1986). Nonlinear dynamics of pattern formation and pattern recognition in the rabbit olfactory bulb. *Physica D*, *22*, 150–175.

Borisyuk R.M., & Kirillov A.B. (1992). Bifurcation analysis of a neural network model. *Biological Cybernetics*, *66*, 319–325.

Cohen M.A., & Grossberg S. (1983). Absolute stability of global pattern formation and parallel memory storage by competitive neural networks. *IEEE Transactions SMC*, *13*, 815–826.

Ditzinger T., & Haken H. (1989). Oscillations in the perception of ambiguous patterns. *Biological Cybernetics*, *61*, 279–287.

Elbert T., Ray W.J., Kowalik Z.J., Skinner J.E., Graf K.E., & Birbaumer N. (1994). Chaos and physiology: Deterministic chaos in excitable cell assemblies. *Physiological Reviews*, *74*, 1–47.

Erdi P., Grobler T., Barna G., & Kaski K. (1993). Dynamics of the olfactory bulb: Bifurcations, learning, and memory. *Biological Cybernetics*, *69*, 57–66.

Ermentrout, G.B. (1994). An Introduction to Neural Oscillators. In F Ventriglia (Ed.), Neural Modeling and Neural Networks. Pergamon Press, Oxford, pp. 79–110.

Fisher G.H. (1967). Measuring ambiguity. *American Journal of Psychology*, *80*, 541–547.

Grossberg S. (1988). Nonlinear neural networks: Principles, mechanisms, and architectures. *Neural Networks*, *1*, 17–61.

Hebb, D.O. (1949). The Organization of Behavior. Wiley, New York.

Hirsch M.W. (1989). Convergent activation dynamics in continuous time networks. *Neural Networks*, *2*, 331–349.

Hopfield J.J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of National Academy of Sciences U.S.A.*, *79*, 2554–2558.

Hoppensteadt F.C., & Izhikevich E.M. (1996). Synaptic organizations and dynamical properties of weakly connected neural oscillators: I. Analysis of canonical model. *Biological Cybernetics*, *75*, 117–127.

Hoppensteadt F.C., & Izhikevich E.M. (1996). Synaptic organizations and dynamical properties of weakly connected neural oscillators: II. Learning of phase information. *Biological Cybernetics*, *75*, 129–135.

Hoppensteadt, F.C. and Izhikevich, E.M. (1997). Weakly Connected Neural Networks. Springer-Verlag, New York.

Izhikevich, E.M. (1996). Bifurcations in brain dynamics. Ph.D. Thesis, Department of Mathematics, Michigan State University.

Izhikevich, E.M., and Malinetskii, G.G. (1993). A neural network with chaotic behavior. Preprint #17, Institute of Applied Mathematics, Russian Academy of Sciences (in Russian).

Kopell, N. (1995). Chains of coupled oscillators. In Arbib, M.A. (Ed.), Brain Theory and Neural Networks. The MIT press, Cambridge, MA.

Kowalski J.M., Albert G.L., Rhoades B.K., & Gross G.W. (1992). Neuronal networks with spontaneous, correlated bursting activity: Theory and simulations. *Neural Networks*, *5*, 805–822.

Kuznetsov, Yu. (1995). Elements of Applied Bifurcation Theory. Springer-Verlag, New York.

Li Z., & Hopfield J.J. (1989). Modeling the olfactory bulb and its neural oscillatory processings. *Biological Cybernetics*, *61*, 379–392.

Mason A., Nicoll A., & Stratford K. (1991). Synaptic transmission between individual pyramidal neurons of the rat visual cortex in vitro. *The Journal of Neuroscience*, *11*, 72–84.

McNaughton B.L., Barnes C.A., & Andersen P. (1981). Synaptic efficacy and EPSP summation in granule cells of rat fascia dentata studied in vitro. *Journal of Neurophysiology*, *46*, 952–966.

Poston T., & Stewart I. (1978). Nonlinear modeling of multistable perception. *Behavioral Science*, *23*, 318–334.

Rapaport A. (1952). Ignition phenomenon in random nets. *Bulletin of Mathematical Biophysics*, *14*, 35–44.

Sayer R.J., Friedlander M.J., & Redman S.J. (1990). The time course and amplitude of EPSPs evoked at synapses between pairs of CA3/CA1 neurons in the hippocampal slice. *Journal of Neuroscience*, *10*, 826–836.

Skarda C.A., & Freeman W.J. (1987). How brain makes chaos in order to make sense of the world. *Behavioral Brain Sciences*, *10*, 161–195.

Smale S. (1974). A mathematical model of two cells via Turing's equation. *Lectures in Applied Mathematics*, *6*, 15–26.

Stewart I.N., & Peregoy P.L. (1983). Catastrophe theory modeling in psychology. *Psychological Bulletin*, *94*, 336–362.