Attractive Periodic Sets in Discrete Time Recurrent Networks (with Emphasis on Fixed Point Stability and Bifurcations in Two–Neuron Networks)

Peter Tiňo*

Aston University

Aston Triangle, Birmingham B4 7ET, UK

Bill G. Horne and **C. Lee Giles**†

NEC Research Institute

4 Independence Way, Princeton, NJ 08540

Neural Computation, Vol. 13, No. 6, pp. 1379–1414, 2001.

Abstract

We perform a detailed fixed-point analysis of two-unit recurrent neural networks with sigmoid-shaped transfer functions. Using geometrical arguments in the space of transfer function derivatives, we partition the network state space into distinct regions corresponding to stability types of the fixed points. Unlike in the previous studies, we do not assume any special form of connectivity pattern between the neurons, and all free parameters are allowed to vary. We also prove that when both neurons have excitatory self-connections and the mutual interaction pattern is the same (i.e. the neurons either mutually inhibit or excite themselves), new attractive fixed points are created through the saddle-node bifurcation. Finally, for an N-neuron recurrent network, we give lower bounds on the rate of convergence of attractive periodic points towards the saturation values of neuron activations, as the absolute values of connection weights grow.

*also with *the Department of Computer Science and Engineering, Slovak University of Technology, Ilkovicova 3, 812 19 Bratislava, Slovakia*

†Also with School of Information Sciences and Technology, Pennsylvania State University, University Park, PA 16801

1 Introduction

Discrete-time recurrent neural networks offer a wide range of dynamical behavior. They can generate steady-state, periodic, quasi-periodic and chaotic orbits.

Attractive fixed points have been proposed as robust representations of prototype vectors in associative memories (Amit, 1989; Hopfield, 1984; Hui & Zak, 1992). Indeed, a great deal of work has focused on the question of how to constrain the weights in the recurrent network so that it exhibits only attractive steady states (Casey, 1995; Jin, Nikiforuk & Gupta, 1994; Sevrani & Abe, 2000). Jin, Nikiforuk and Gupta (1994) give the conditions on the weight matrix under which *all* fixed points of the network are attractive.

Saddle fixed points were suggested to play an important role in working memories operating in non-stationary environments where both long-term maintenance and quick transitions are desirable (McAuley & Stampfli, 1994; Nakahara & Doya, 1998). Moreover, saddle points often mimic stack-like behavior in recurrent networks trained on context-free languages (Rodriguez, Wiles & Elman, 1999). Also, they were found to be part of a mechanism by which recurrent networks induce non-stable representations of large cycles in finite state machines (Tiňo et al., 1998).

There are many applications where oscillatory dynamics of recurrent networks is desirable. For example, when trained to act as a finite state machine (Cleeremans, Servan-Schreiber & McClelland, 1989; Giles et al., 1992; Tiňo & Sajda, 1995; Watrous & Kuhn, 1992) the network has to induce a stable representation of state transitions associated with each input symbol σ of the machine. Of special importance are period-n cycles driven by σ: given the current state S, when repeatedly presenting the input σ, we return after n steps to S. Cycles in the machine often induce in the recurrent network state space stable periodic orbits corresponding to the input σ (Casey, 1996; Manolios & Fanelli, 1994; Tiňo et al., 1998). In particular, loops, i.e.period-1 cycles, often induce attractive fixed points (see e.g. (Casey, 1996; Tiňo et al., 1998))

Finally, chaotic behavior of recurrent networks has been a focus of a vivid research activity (Botelho, 1999; Klotz & Brauer, 1999; Pasemann, 1995a), especially after Wang (1991) rigorously showed that a simple two-neuron recurrent network is capable of producing chaos.

There is a considerable amount of literature on two-unit recurrent networks (Beer, 1995; Borisyuk & Kirillov, 1992; Botelho, 1999; Klotz & Brauer, 1999; Pakdamann et al., 1998; Pasemann, 1993; Wang, 1991; Zhou, 1996). This is partly due to the lack of mathematical tools for a detailed analysis of higher dimensional dynamical systems, and partly due to the high expressive power of such simple networks, in which many generic properties of larger networks are already present (Botelho, 1999). Moreover, two-neuron networks are sometimes thought of as systems of two modules, where each module represents the mean activity of a spatially localized neural population (Borisyuk & Kirillov, 1992; Pasemann, 1995a; Tonnelier et al., 1999).

Typically, studies of the asymptotic behavior of two-unit or larger networks assume some form of structure in the weight matrix describing the connectivity pattern among recurrent neurons. We mention few examples:

- *Symmetric connectivity and absence of self-interactions* enabled Hopfield (1984) to interpret the network as a physical system having energy minima in the attractive fixed points. These rather strict conditions were weakened in (Casey, 1995). Blum and Wang (1992) globally analyzed networks with non-symmetrical connectivity patterns of special types. In particular, they formulated results for two-neuron networks with the logistic sigmoid transfer function $g(\ell) = 1/(1+e^{-\ell})$, in the *absence of neuron self-connections*.
- Deep mathematical studies were presented for severely restricted topologies, such as the *"ring network"* (Pasemann, 1995b) or the *chain oscillators* (Wang, 1996).
- Often, rather detailed bifurcation studies are performed with respect to only one or two network parameters, the remaining parameters are kept *fixed* to "appropriate" values (Borisyuk & Kirillov, 1992; Nakahara & Doya, 1998).

Also, the assumption of *saturated* sigmoidal or linear transfer functions (like in the Brain-State-in-a-Box model (Anderson, 1993)) makes the study of asymptotically stable equilibrium points more feasible (Botelho, 1999; Hui & Zak, 1992; Sevrani & Abe, 2000).

In this paper we impose no conditions on connection weights or external inputs, apart from the fact that the weights are assumed to be non-zero. To our knowledge, a thorough fixed point analysis of two-unit neural networks with sigmoid transfer functions is still missing[1]. We offer such an analysis in sections 4 and 5. In section 6 we rigorously explain the mechanism by which new attractive fixed points are created - the saddle-node bifurcation. This confirms empirical findings of Tonnelier et al. (1999). When studying such networks, they empirically observed that a new fixed point appears through the saddle-node bifurcation.

Hirsch proved (Hirsch, 1994) that when all the weights in an N-neuron recurrent network with exclusively self-exciting (or exclusively self-inhibiting) neurons are multiplied by an increasing neural gain, attractive fixed points tend towards the saturated activation values. In section 7, we give a lower bound on the rate of convergence of attractive periodic points[2] towards the saturation values. The results hold for "sigmoid-shaped" neuron transfer functions. This investigation was motivated by the observation of many from the recurrent network grammar/automata induction community that the activations of recurrent neurons often cluster away from the center of the network state space and close to the saturated activation values (e.g. (Manolios & Fanelli, 1994; Tiňo et al., 1998)).

[1] The most complete fixed point analysis of continuous time two-neuron networks can be found in (Beer, 1995). We determine the fixed point positions using a similar "null-cline" technique.

[2] fixed points can be considered periodic points of period 1

In fact, this was exploited by Zeng, Goodman and Smyth (1993) in a heuristic to stabilize the induced automata representations in the recurrent network.

In the next two sections we recall some basic concepts from the theory of dynamical systems and introduce the recurrent network model.

2 Basic definitions

A discrete-time dynamical system can be represented as the iteration of a (differentiable) map $f : X \to X$, $X \subseteq \Re^d$:

$$x_{n+1} = f(x_n), \quad n \in \mathbf{N}. \tag{1}$$

Here, $\mathbf{N}$ denotes the set of all natural numbers. For each $x = x_0 \in X$, the iteration (1) generates a sequence of points defining the orbit, or trajectory of x under the map f. In other words, the orbit of x under f is the sequence $\{f^n(x)\}_{n \geq 0}$. For $n \geq 1$, f^n is the composition of the map f with itself n times. f^0 is defined to be the identity map on X.

A point $x \in X$ is called a *fixed point* of the map f, if $f(x) = x$. It is called a *periodic point* of period P, if it is a fixed point of f^P.

Fixed points can be classified according to orbit behavior of points in their vicinity. A fixed point x is said to be asymptotically stable (or an *attractive point* of f), if there exists a neighborhood $O(x)$ of x, such that $\lim_{n\to\infty} f^n(y) = x$, for all $y \in O(x)$. As n increases, trajectories of points near to an asymptotically stable fixed point tend towards it.

A fixed point x of f is asymptotically stable if only if for each eigenvalue λ of $\mathcal{J}(x)$, the Jacobian of f at x, $|\lambda| < 1$ holds. The eigenvalues of the Jacobian $\mathcal{J}(x)$ govern the contracting/expanding directions of the map f in a vicinity of x. Eigenvalues larger in absolute value than one lead to expansions, whereas eigenvalues smaller than one correspond to contractions. If all the eigenvalues of $\mathcal{J}(x)$ are outside the unit circle, x is a *repulsive point*. As the time index n increases, the trajectories of points from a neighborhood of a repulsive point move away from it. If some eigenvalues of $\mathcal{J}(x)$ are inside and some are outside the unit circle, x is said to be a *saddle point*.

3 The model

In this paper we study recurrent neural networks with transfer functions from a general class of "sigmoid-shaped" maps

$$g_{A,B,\mu}(\ell) = \frac{A}{1 + e^{-\mu\ell}} + B \tag{2}$$

transforming $\Re$ into the interval $(B, B + A)$, $B \in \Re$, $A \in (0, \infty)$. The so called "neural gain", $\mu > 0$, controls "steepness" of the transfer function. As $\mu \to \infty$, $g_{A,B,\mu}$ tends to the step function $g_{A,B,\mu}(\ell) = B$, for $\ell < 0$, and $g_{A,B,\mu}(\ell) = B + A$, for $\ell \geq 0$. The commonly used unipolar and bipolar logistic transfer functions can be expressed as $g_{1,0,1}$ and

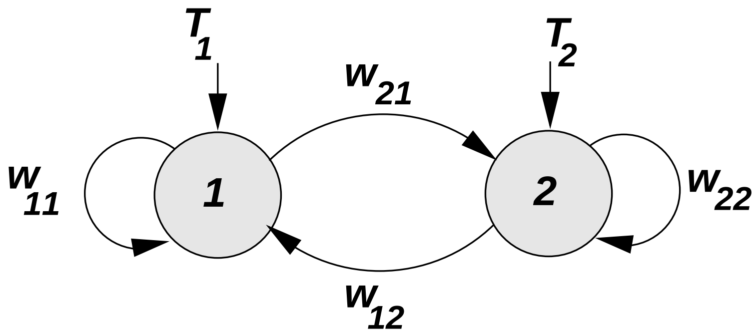

Figure 1: A two dimensional recurrent neural network.

$g_{2,-1,1}$, respectively. Another commonly used transfer function, the hyperbolic tangent, corresponds to $g_{2,-1,2}$.

For recurrent networks consisting of two neurons (see figure 1), the iterative map (1) can be written as follows:

$$\begin{bmatrix} x_{n+1} \\ y_{n+1} \end{bmatrix} = \begin{bmatrix} g_{A,B,\mu}(w_{11}x_n + w_{12}y_n + T_1) \\ g_{A,B,\mu}(w_{21}x_n + w_{22}y_n + T_2) \end{bmatrix}. \tag{3}$$

The neuron outputs (activations) $(x_n, y_n) \in (B, B + A)^2$ form the state of the network at time step n. The connection weights $w_{ij} \in \Re \setminus \{0\}$ and bias/external input terms $T_i \in \Re$ are the adjustable parameters of the network and determine the dynamical behavior of the system (3).

For the purpose of fixed point analysis of (3), it is convenient to include the neural gain μ into the adjustable parameters. We rewrite (3) as

$$\begin{bmatrix} x_{n+1} \\ y_{n+1} \end{bmatrix} = \begin{bmatrix} g(ax_n + by_n + t_1) \\ g(cx_n + dy_n + t_2) \end{bmatrix}, \tag{4}$$

where $g = g_{A,B,1}$, $a = \mu w_{11}$, $b = \mu w_{12}$, $c = \mu w_{21}$, $d = \mu w_{22}$, $t_1 = \mu T_1$ and $t_2 = \mu T_2$.

4 Analysis in the space of transfer function derivatives

In this section, we work with the derivatives

$$G_1(x, y) = g'(\ell)|_{\ell=ax+by+t_1} = g'(ax + by + t_1) \tag{5}$$

$$G_2(x, y) = g'(\ell)|_{\ell=cx+dy+t_2} = g'(cx + dy + t_2) \tag{6}$$

of transfer functions corresponding to the two neurons in the network described by (4). The derivatives G_1 and G_2 are always positive. Our aim is to partition the space of derivatives $(G_1, G_2) \in (0, \frac{A}{4})^2$ into regions corresponding to stability types of the neural network fixed points.

First, we introduce two auxiliary maps $\psi : (B, B+A) \to (0, A/4]$,

$$\psi(u) = \frac{1}{A}(u-B)(B+A-u), \tag{7}$$

and $\phi : (B, B+A)^2 \to (0, A/4]^2$,

$$\phi(x,y) = (\psi(x), \psi(y)). \tag{8}$$

It is easy to show that

$$g'(\ell) = \frac{1}{A}(g(\ell)-B)(B+A-g(\ell)) = \psi(g(\ell)). \tag{9}$$

For each fixed point (x,y) of (4),

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} g(ax+by+t_1) \\ g(cx+dy+t_2) \end{bmatrix}, \tag{10}$$

and so by (5), (6), (9) and (10)

$$G_1(x,y) = \psi(g(ax+by+t_1)) = \psi(x), \tag{11}$$

$$G_2(x,y) = \psi(g(cx+dy+t_2)) = \psi(y), \tag{12}$$

which implies (see (8))

$$(G_1(x,y), G_2(x,y)) = \phi(x,y). \tag{13}$$

Consider a function $F(u,v)$, $F : \mathcal{A} \subseteq \Re^2 \to \Re$. The function F induces a partition of the set $\mathcal{A}$ into three regions

$$F^- = \{(u,v) \in \mathcal{A} | F(u,v) < 0\}, \tag{14}$$

$$F^+ = \{(u,v) \in \mathcal{A} | F(u,v) > 0\}, \tag{15}$$

$$F^0 = \{(u,v) \in \mathcal{A} | F(u,v) = 0\}. \tag{16}$$

Now we are ready to formulate and prove three lemmas that define the correspondence between the derivatives G_1 and G_2 of the neuron transfer functions (see (5), (6)) and the fixed-point stability of the network (4).

Lemma 1: *If $bc > 0$, then all attractive fixed points (x,y) of (4) satisfy*

$$\phi(x,y) \in \left(0, \frac{1}{|a|}\right) \times \left(0, \frac{1}{|d|}\right).$$

Proof: Consider a fixed point (x,y) of (4). By (5) and (6), the Jacobian $\mathcal{J}(x,y)$ of (4) in (x,y) is given by[3]

$$\mathcal{J} = \begin{pmatrix} aG_1(x,y) & bG_1(x,y) \\ cG_2(x,y) & dG_2(x,y) \end{pmatrix}.$$

[3]to simplify the notation, the identification (x,y) of the fixed point in which (4) is linearized is omitted

The eigenvalues of $\mathcal{J}$ are

$$\lambda_{1,2} = \frac{aG_1 + dG_2 \pm \sqrt{\mathcal{D}(G_1, G_2)}}{2}, \tag{17}$$

where

$$\mathcal{D}(G_1, G_2) = (aG_1 - dG_2)^2 + 4G_1G_2bc. \tag{18}$$

The proof continues with an analysis of three separate cases, corresponding to sign patterns of weights associated with the neuron self-connections.

- **Assume $\mathbf{a}, \mathbf{d} > \mathbf{0}$, i.e the weights of self-connections on both neurons are positive.**

  Define

  $$\alpha(G_1, G_2) = aG_1 + dG_2. \tag{19}$$

  Since the derivatives G_1, G_2 can only take on values from $(0, A/4)$, and $a, d > 0$, we have $\mathcal{D}(G_1, G_2) > 0$ and $\alpha(G_1, G_2) > 0$, for all $(G_1, G_2) \in (0, A/4)^2$. In our terminology (see eq. (14–16)), $\mathcal{D}^+, \alpha^+ = (0, A/4)^2 \subseteq (0, \infty)^2$. To identify possible values of G_1 and G_2 so that $|\lambda_{1,2}| < 1$, it is sufficient to solve the inequality $aG_1 + dG_2 + \sqrt{\mathcal{D}(G_1, G_2)} < 2$, or equivalently

  $$2 - aG_1 - dG_2 > \sqrt{\mathcal{D}(G_1, G_2)}. \tag{20}$$

  Consider only (G_1, G_2) such that

  $$\rho_1(G_1, G_2) = aG_1 + dG_2 - 2 < 0, \tag{21}$$

  i.e. (G_1, G_2) lying under the line $\rho_1^0 : aG_1 + dG_2 = 2$. All (G_1, G_2) such that $aG_1 + dG_2 - 2 > 0$, i.e. $(G_1, G_2) \in \rho_1^+$ (above ρ_1^0), lead to at least one eigenvalue of $\mathcal{J}$ greater in absolute value than 1. Squaring both sides of (20) we arrive at

  $$\kappa_1(G_1, G_2) = (ad - bc)G_1G_2 - aG_1 - dG_2 + 1 > 0. \tag{22}$$

  If $ad \neq bc$, the set κ_1^0 is a hyperbola

  $$G_2 = \frac{1}{\tilde{d}} + \frac{C}{G_1 - \frac{1}{\tilde{a}}}, \tag{23}$$

  with

  $$\tilde{a} = a - \frac{bc}{d}, \tag{24}$$

  $$\tilde{d} = d - \frac{bc}{a}, \tag{25}$$

  $$C = \frac{bc}{(ad - bc)^2}. \tag{26}$$

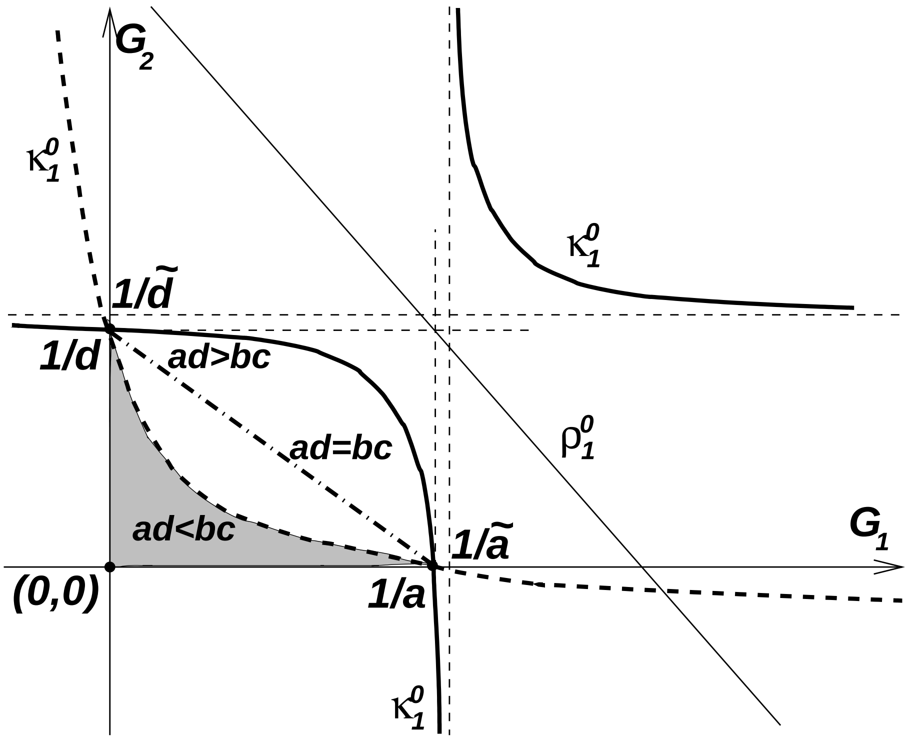

Figure 2: An illustration for the proof of Lemma 1. The network parameters satisfy $a, d > 0$ and $bc > 0$. All $(G_1, G_2) \in (0, A/4]^2$ below the left branch of κ_1^0 (if $ad \geq bc$), or between the branches of κ_1^0 (if $ad < bc$) correspond to the attractive fixed points. Different line styles for κ_1^0 are associated with the cases $ad > bc$, $ad = bc$ and $ad < bc$, namely, the solid, dashed-dotted and dashed lines, respectively.

It is easy to check that the points $(1/a, 0)$ and $(0, 1/d)$ lie on the curve κ_1^0.

If $ad = bc$, the set κ_1^0 is a line passing through the points $(0, 1/d)$ and $(1/a, 0)$. (see figure 2).

To summarize, when $a, d > 0$, by (13), (21) and (22), the fixed point (x, y) of (4) is attractive only if

$$(G_1, G_2) = \phi(x, y) \in \kappa_1^+ \cap \rho_1^-,$$

where the map ϕ is defined by (8).

A necessary (not sufficient) condition for (x, y) to be attractive reads[4]

$$\phi(x, y) \in \left(0, \frac{1}{a}\right) \times \left(0, \frac{1}{d}\right).$$

- **Consider now the case of negative weights on neuron self-connections, i.e. $\mathbf{a}, \mathbf{d} < \mathbf{0}$.**
  Since in this case $\alpha(G_1, G_2)$ (eq. 19) is negative for all values of the transfer function derivatives G_1, G_2, we have $\alpha^- = (0, A/4)^2 \subseteq (0, \infty)^2$. In order to identify possible values of (G_1, G_2) such that $|\lambda_{1,2}| < 1$, it is sufficient to solve the inequality $aG_1 + dG_2 - \sqrt{\mathcal{D}(G_1, G_2)} > -2$, or equivalently

$$2 + aG_1 + dG_2 > \sqrt{\mathcal{D}(G_1, G_2)}. \tag{27}$$

As in the previous case, we shall consider only (G_1, G_2) such that

$$\rho_2(G_1, G_2) = aG_1 + dG_2 + 2 > 0, \tag{28}$$

since all $(G_1, G_2) \in \rho_2^-$ lead to at least one eigenvalue of $\mathcal{J}$ greater in absolute value than 1.

Squaring both sides of (27) we arrive at

$$\kappa_2(G_1, G_2) = (ad - bc)G_1G_2 + aG_1 + dG_2 + 1 > 0, \tag{29}$$

which is equivalent to

$$((-a)(-d) - bc)G_1G_2 - (-a)G_1 - (-d)G_2 + 1 > 0. \tag{30}$$

Further analysis is exactly the same as the analysis from the previous case $(a, d > 0)$ with a, $\tilde{a}$, d and $\tilde{d}$ replaced by $|a|$, $|a| - bc/|d|$, $|d|$ and $|d| - bc/|a|$, respectively.

[4] If $ad > bc$, then $0 < \tilde{a} < a$ and $0 < \tilde{d} < d$ (see eq. (24) and (25)). The derivatives $(G_1, G_2) \in \kappa_1^+$ lie under the "left branch" and above the "right branch" of κ_1^0 (figure 2). It is easy to see that since we are confined to the half-plane ρ_1^- (below the line ρ_1^0), only (G_1, G_2) under the "left branch" of κ_1^0 will be considered. Indeed, ρ_1^0 is a decreasing line going through the point $(1/a, 1/d)$ and so it never intersects the right branch of κ_1^0. If $ad < bc$, then $\tilde{a}, \tilde{d} < 0$ and $(G_1, G_2) \in \kappa_1^+$ lie between the two branches of κ_1^0.

If $ad \neq bc$, the set κ_2^0 is a hyperbola

$$G_2 = \frac{-1}{\tilde{d}} + \frac{C}{G_1 + \frac{1}{\tilde{a}}} \tag{31}$$

passing through the points $(-1/a, 0)$ and $(0, -1/d)$.

If $ad = bc$, κ_2^0 is the line defined by the points $(0, -1/d)$ and $(-1/a, 0)$.

By (13), (28) and (29), the fixed point (x, y) of (4) is attractive only if

$$(G_1, G_2) = \phi(x, y) \in \kappa_2^+ \cap \rho_2^+.$$

In this case, the derivatives $(G_1, G_2) = \phi(x, y)$ must satisfy

$$\phi(x, y) \in \left(0, \frac{1}{|a|}\right) \times \left(0, \frac{1}{|d|}\right).$$

- **Finally, consider the case when the weights on neuron self-connections differ in sign.**

  Without loss of generality assume $a > 0$ and $d < 0$.

  Assume further that $aG_1 + dG_2 \geq 0$, i.e. $(G_1, G_2) \in \alpha^+ \cup \alpha^0$ (see eq. (19)) lie under or on the line

  $$\alpha^0 : \quad G_2 = \frac{a}{|d|} G_1.$$

  It is sufficient to solve the inequality (20). Using (22), (21) and arguments developed earlier in this proof, we conclude that the derivatives (G_1, G_2) lying in

  $$\kappa_1^+ \cap \rho_1^- \cap (\alpha^+ \cup \alpha^0).$$

  correspond to attractive fixed points of (4).

  We refer the reader to figure 3.

  Since $a > 0$ and $d < 0$, the term $ad - bc$ is negative, and so $0 < a < \tilde{a}$, $\tilde{d} < d < 0$. The "transformed" parameters $\tilde{a}$, $\tilde{d}$ are defined in (24) and (25).

  For $(G_1, G_2) \in \alpha^-$, by (29), (28) and earlier arguments in this proof, the derivatives (G_1, G_2) lying in

  $$\kappa_2^+ \cap \rho_2^+ \cap \alpha^-$$

  correspond to attractive fixed points of (4).

  It can be easily shown that the sets κ_1^0 and κ_2^0 intersect on the line α^0 in the open interval (see figure 3).

  $$\left(\frac{1}{\tilde{a}}, \frac{1}{a}\right) \times \left(\frac{1}{|\tilde{d}|}, \frac{1}{|d|}\right)$$

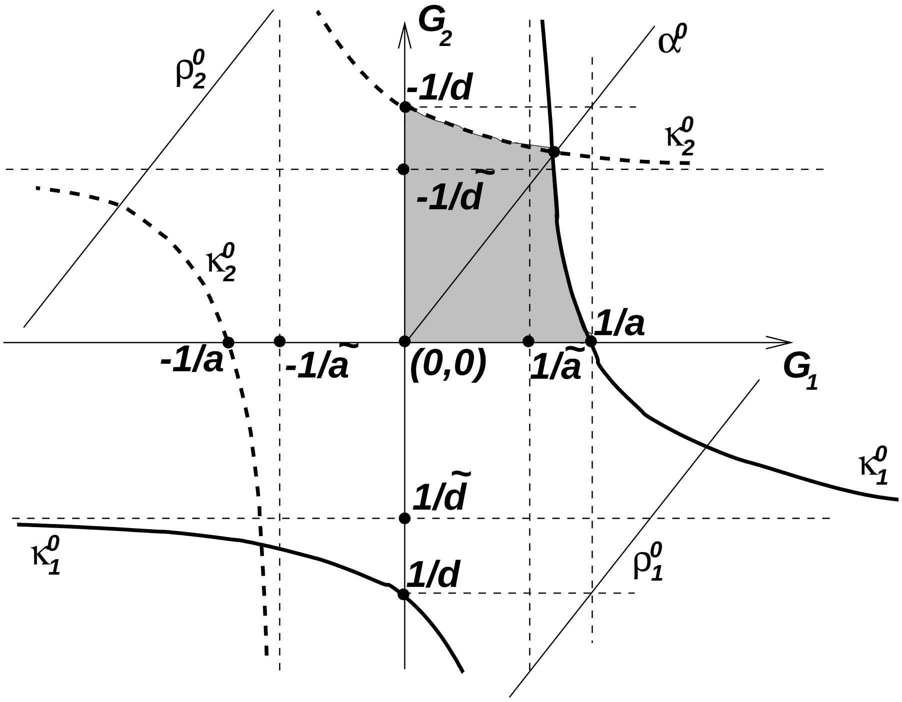

Figure 3: An illustration for the proof of Lemma 1. The network parameters are constrained as follows: $a > 0$, $d < 0$ and $bc > 0$. All $(G_1, G_2) \in (0, A/4]^2$ below and on the line α^0, and between the two branches of κ_1^0 (solid line) correspond to the attractive fixed points. So do all $(G_1, G_2) \in (0, A/4]^2$ above α^0 and between the two branches of κ_2^0 (dashed line).

We conclude that the fixed point (x, y) of (4) is attractive only if

$$\phi(x,y) \in [\kappa_1^+ \cap \rho_1^- \cap (\alpha^+ \cup \alpha^0)] \cup [\kappa_2^+ \cap \rho_2^+ \cap \alpha^-].$$

In particular, if (x, y) is attractive, then the derivatives $(G_1, G_2) = \phi(x, y)$ must lie in

$$\left(0, \frac{1}{a}\right) \times \left(0, \frac{1}{|d|}\right).$$

Examination of the case $a < 0,\ d > 0$ in the same way leads to a conclusion that all attractive fixed points of (4) have their corresponding derivatives (G_1, G_2) in

$$\left(0, \frac{1}{|a|}\right) \times \left(0, \frac{1}{d}\right).$$

■

We remind the reader that the "transformed" parameters $\tilde{a}$ and $\tilde{d}$ are defined in (24) and (25).

Lemma 2: *Assume $bc < 0$. Suppose $ad > 0$, or $ad < 0$ with $|ad| \leq |bc|/2$. Then each fixed point (x, y) of (4) such that*

$$\phi(x,y) \in \left(0, \frac{1}{|a|}\right) \times \left(0, \frac{1}{|\tilde{d}|}\right) \cup \left(0, \frac{1}{|\tilde{a}|}\right) \times \left(0, \frac{1}{|d|}\right)$$

is attractive. In particular, all fixed points (x, y) for which

$$\phi(x,y) \in \left(0, \frac{1}{|\tilde{a}|}\right) \times \left(0, \frac{1}{|\tilde{d}|}\right)$$

are attractive.

Proof: The discriminant $\mathcal{D}(G_1, G_2)$ in (18) is no longer exclusively positive. It follows from analytic geometry (see for example (Anton, 1980)) that $\mathcal{D}(G_1, G_2) = 0$ defines either a single point or two increasing lines (that can collide into one, or disappear). Furthermore, $\mathcal{D}(0, 0) = 0$. Hence, the set $\mathcal{D}^0$ is either a single point – the origin, or a pair of increasing lines (that may be the same) passing through the origin.

As in the proof of the first lemma, we proceed in three steps.

- **Assume $\mathbf{a, d > 0}$.**

  Since

  $$\mathcal{D}\left(\frac{1}{a}, \frac{1}{d}\right) = \frac{4bc}{ad} < 0$$

  and

  $$\mathcal{D}\left(\frac{1}{a}, 0\right) = \mathcal{D}\left(0, \frac{1}{d}\right) = 1 > 0,$$

  the point $(1/a, 1/d)$ is always in the set $\mathcal{D}^-$, while $(1/a, 0), (0, 1/d) \in \mathcal{D}^+$.

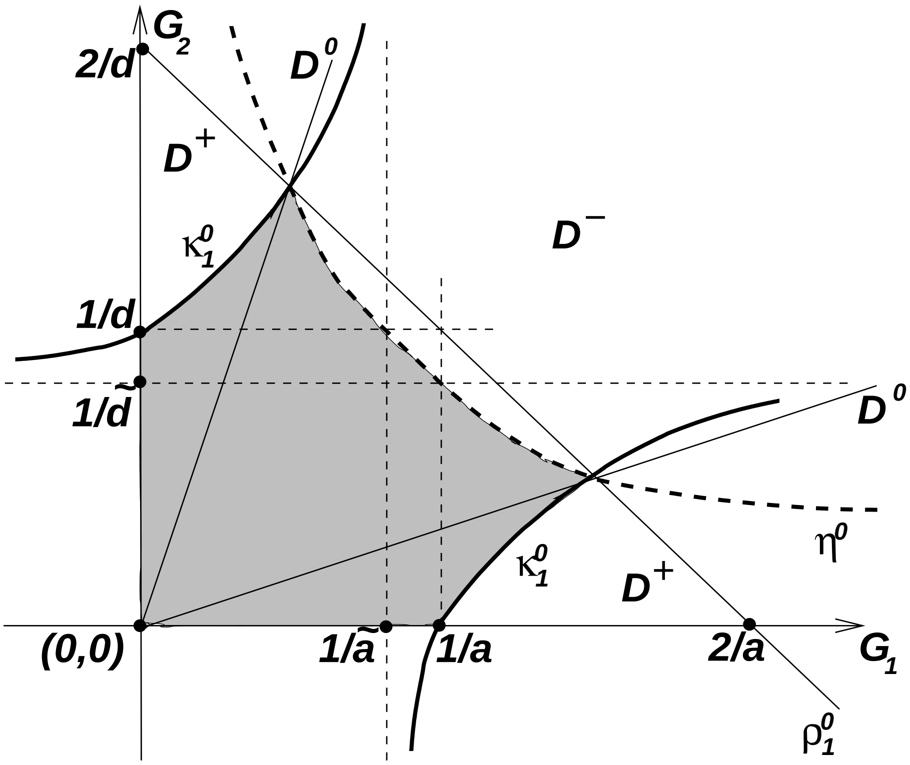

Figure 4: An illustration for the proof of Lemma 2. The parameters satisfy $a, d > 0$ and $bc < 0$. All $(G_1, G_2) \in \mathcal{D}^-$ and below the right branch of η^0 (dashed line) correspond to the attractive fixed points. So do $(G_1, G_2) \in (0, A/4]^2$ in $\mathcal{D}^+$ between the two branches of κ_1^0.

– First, we shall examine the case when $\mathcal{D}(G_1, G_2)$ is negative. From

$$|\lambda_{1,2}|^2 = \frac{(aG_1 + dG_2)^2 + |\mathcal{D}|}{4} = G_1 G_2 (ad - bc)$$

it follows, that $(G_1, G_2) \in \mathcal{D}^-$, for which $|\lambda_{1,2}| < 1$, lie in (figure 4)

$$\mathcal{D}^- \cap \eta^-,$$

where

$$\eta(G_1, G_2) = (ad - bc) G_1 G_2 - 1. \tag{32}$$

It is easy to show that

$$\left(\frac{1}{a}, \frac{1}{\tilde{d}}\right), \left(\frac{1}{\tilde{a}}, \frac{1}{d}\right) \in \eta^0, \quad \frac{1}{\tilde{a}} < \frac{1}{a}, \quad \frac{1}{\tilde{d}} < \frac{1}{d}, \quad \text{and} \quad \left(\frac{1}{a}, \frac{1}{\tilde{d}}\right), \left(\frac{1}{\tilde{a}}, \frac{1}{d}\right) \in \mathcal{D}^-.$$

– Turn now to the case $\mathcal{D}(G_1, G_2) > 0$. Using the technique from the previous proof we conclude that the derivatives (G_1, G_2) corresponding to attractive fixed points of (4) lie in

$$\mathcal{D}^+ \cap \rho_1^- \cap \kappa_1^+,$$

i.e under the line ρ_1^0 and between the two branches of κ_1^0 (see equations (18), (21) and (22)).
The sets $\mathcal{D}^0$, ρ_1^0, κ_1^0 and η^0 intersect in two points as suggested in figure 4. To see this, note that for points on η^0, it holds

$$G_1G_2 = \frac{1}{ad - bc},$$

and for all $(G_1, G_2) \in \mathcal{D}^0 \cap \eta^0$ we have

$$(aG_1 + dG_2)^2 = 4. \tag{33}$$

For $G_1, G_2 > 0$, (33) defines the line ρ_1^0.
Similarly, for (G_1, G_2) in the sets κ_1^0 and η^0, it holds

$$aG_1 + dG_2 = 2,$$

which is the definition of the line ρ_1^0. The curves κ_1^0 and η^0 are monotonically increasing and decreasing, respectively, and there is exactly one intersection point of the right branch of η^0 with each of the two branches of κ_1^0.

- For $(G_1, G_2) \in \mathcal{D}^0$,

$$|\lambda_{1,2}| = \frac{aG_1 + dG_2}{2},$$

and (G_1, G_2) corresponding to attractive fixed points of (4) are from

$$\mathcal{D}^0 \cap \rho_1^-.$$

In summary, when $a, d > 0$, each fixed point (x, y) of (4) such that

$$(G_1, G_2) = \phi(x, y) \in \left(0, \frac{1}{a}\right) \times \left(0, \frac{1}{\tilde{d}}\right) \cup \left(0, \frac{1}{\tilde{a}}\right) \times \left(0, \frac{1}{d}\right)$$

is attractive.

- **Assume $\mathbf{a, d < 0}$.**
This case is identical to the case $a, d > 0$ examined above, with a, $\tilde{a}$, d, $\tilde{d}$, ρ_1^- and κ_1^+ replaced by $|a|$, $|\tilde{a}|$, $|d|$, $|\tilde{d}|$, ρ_2^+ and κ_2^+, respectively.

  - First, note that the set $\mathcal{D}^0$ is the same as before, since

$$(aG_1 - dG_2)^2 = (|a|G_1 - |d|G_2)^2.$$

  - Furthermore, $ad - bc = |a||d| - bc$, and so $(G_1, G_2) \in \mathcal{D}^-$, for which $|\lambda_{1,2}| < 1$, lie in

$$\mathcal{D}^- \cap \eta^-.$$

Again, it directly follows that

$$\left(\frac{1}{|a|},\frac{1}{|\tilde{d}|}\right),\left(\frac{1}{|\tilde{a}|},\frac{1}{|d|}\right)\in\eta^0,\quad \frac{1}{|\tilde{a}|}<\frac{1}{|a|},\quad \frac{1}{|\tilde{d}|}<\frac{1}{|d|}$$

and

$$\left(\frac{1}{|a|},\frac{1}{|\tilde{d}|}\right),\left(\frac{1}{|\tilde{a}|},\frac{1}{|d|}\right)\in\mathcal{D}^-.$$

– For $\mathcal{D}^+$ the derivatives (G_1,G_2) corresponding to attractive fixed points of (4) lie in

$$\mathcal{D}^+\cap\rho_2^+\cap\kappa_2^+.$$

All $(G_1,G_2)\in\mathcal{D}^0\cap\rho_2^+$ lead to $|\lambda_{1,2}|<1$. Hence, when $a,d<0$, every fixed point (x,y) of (4) such that

$$\phi(x,y)\in\left(0,\frac{1}{|a|}\right)\times\left(0,\frac{1}{|\tilde{d}|}\right)\cup\left(0,\frac{1}{|\tilde{a}|}\right)\times\left(0,\frac{1}{|d|}\right)$$

is attractive.

- **Finally, consider the case $\mathbf{a>0,d<0}$.**
  The case $a<0,d>0$ would be treated in exactly the same way.

  – Assume $\mathcal{D}^-$ is a nonempty region. Then, $ad>bc$ must hold and

  $$\left(\frac{1}{a},\frac{1}{|d|}\right)\in\mathcal{D}^-.$$

  This can be easily seen, since for $ad<bc$ we would have for all (G_1,G_2)

  $$\mathcal{D}(G_1G_2)=(aG_1-dG_2)^2+4G_1G_2bc=(aG_1+dG_2)^2+4G_1G_2(bc-ad)\geq 0.$$

  The sign of

  $$\mathcal{D}\left(\frac{1}{a},\frac{1}{|d|}\right)=4\left(1+\frac{bc}{a|d|}\right)$$

  is equal to the sign of $a|d|+bc=bc-ad<0$.
  The derivatives $(G_1,G_2)\in\mathcal{D}^-$, for which $|\lambda_{1,2}|<1$, lie in

  $$\mathcal{D}^-\cap\eta^-.$$

  Also,

  $$\left(\frac{1}{a},\frac{1}{\tilde{d}}\right),\left(\frac{1}{|\tilde{a}|},\frac{1}{|d|}\right)\in\eta^0.$$

  Note that $\tilde{d}\geq|d|$ and $|\tilde{a}|\geq a$, only if $2a|d|\leq|bc|$.

  – Only those $(G_1,G_2)\in\mathcal{D}^0$ are taken into account for which $|aG_1+dG_2|<2$. This is true for all (G_1,G_2) in

  $$\mathcal{D}^0\cap\rho_1^-\cap\rho_2^+.$$

– If $\mathcal{D}(G_1, G_2) > 0$, the inequalities to be solved depend on the sign of $aG_1 + dG_2$. Following the same reasoning as in the proof of Lemma 1, we conclude that the derivatives (G_1, G_2) corresponding to attractive fixed points of (4) lie in

$$\mathcal{D}^+ \cup \left([\kappa_1^+ \cap \rho_1^- \cap (\alpha^+ \cup \alpha^0)] \cup [\kappa_2^+ \cap \rho_2^+ \cap \alpha^-] \right).$$

∎

We saw in the proof of Lemma 1, that if $a > 0$, $d < 0$ and $bc > 0$, then all $(G_1, G_2) \in (0, 1/\tilde{a}) \times (0, 1/|\tilde{d}|)$ potentially correspond to attractive fixed points of (4) (figure 3). In the proof of the last Lemma, it was shown that when $a > 0$, $d < 0$, $bc < 0$, if $2a|d| \geq |bc|$, then $(1/a, 1/|d|)$ is on or under the right branch of η^0 and each $(G_1, G_2) \in (0, 1/a) \times (0, 1/|d|)$ potentially corresponds to an attractive fixed point of (4). Hence, the following Lemma can be formulated:

Lemma 3: *If* $ad < 0$ *and*

- $bc > 0$, *then every fixed point* (x, y) *of (4) such that*

$$\phi(x, y) \in \left(0, \frac{1}{|\tilde{a}|}\right) \times \left(0, \frac{1}{|\tilde{d}|}\right)$$

  is attractive.

- $bc < 0$ *with* $|ad| \geq |bc|/2$, *then each fixed point* (x, y) *of (4) satisfying*

$$\phi(x, y) \in \left(0, \frac{1}{|a|}\right) \times \left(0, \frac{1}{|d|}\right)$$

  is attractive.

5 Transforming the results to the network state space

Lemmas 1, 2 and 3 introduce a structure reflecting stability types of fixed points of (4) into the space of transfer function derivatives (G_1, G_2). In this section, we transform our results from the (G_1, G_2)-space into the space of neural activations (x, y).

For $u > \frac{4}{A}$, define

$$\Delta(u) = \frac{A}{2}\sqrt{1 - \frac{4}{A}\frac{1}{u}}. \tag{34}$$

The interval $(0, A/4]^2$ of all transfer function derivatives in the (G_1, G_2)-plane corresponds to four intervals partitioning the (x, y)-space. Namely,

$$\left(B, B + \frac{A}{2}\right]^2, \tag{35}$$

$$\left(B, B + \frac{A}{2}\right] \times \left[B + \frac{A}{2}, B + A\right), \tag{36}$$

$$\left[B + \frac{A}{2}, B + A\right) \times \left(B, B + \frac{A}{2}\right], \tag{37}$$

$$\left[B + \frac{A}{2}, B + A\right)^2. \tag{38}$$

Recall that according to (13), the transfer function derivatives $(G_1(x, y), G_2(x, y)) \in (0, A/4]^2$ in a fixed point $(x, y) \in (B, B + A)^2$ of (4) are equal to $\phi(x, y)$, where the map ϕ is defined in (7) and (8). Now, for each couple (G_1, G_2), there are four preimages (x, y) under the map ϕ,

$$\phi^{-1}(G_1, G_2) = \left\{ \left(B + \frac{A}{2} \pm \Delta\left(\frac{1}{G_1}\right), B + \frac{A}{2} \pm \Delta\left(\frac{1}{G_2}\right) \right) \right\}, \tag{39}$$

where $\Delta(\cdot)$ is defined in (34).

Before stating the main results, we introduce three types of regions in the neuron activation space that are of special importance. The regions are parametrized by $\alpha, \delta > A/4$, and correspond to stability types of fixed points[5] of (4) (see figure 5):

$$R_{00}^A(\alpha, \delta) = \left(B, B + \frac{A}{2} - \Delta(\alpha)\right) \times \left(B, B + \frac{A}{2} - \Delta(\delta)\right), \tag{40}$$

$$\begin{aligned} R_{00}^S(\alpha, \delta) = \; & \left(B + \tfrac{A}{2} - \Delta(\alpha), B + \tfrac{A}{2}\right] \times \left(B, B + \tfrac{A}{2} - \Delta(\delta)\right) \cup \\ & \left(B, B + \tfrac{A}{2} - \Delta(\alpha)\right) \times \left(B + \tfrac{A}{2} - \Delta(\delta), B + \tfrac{A}{2}\right], \end{aligned} \tag{41}$$

$$R_{00}^R(\alpha, \delta) = \left(B + \frac{A}{2} - \Delta(\alpha), B + \frac{A}{2}\right] \times \left(B + \frac{A}{2} - \Delta(\delta), B + \frac{A}{2}\right]. \tag{42}$$

Having partitioned the interval $(B, B + A/2]^2$ (35) of the (x, y)-space into the sets $R_{00}^A(\alpha, \delta)$, $R_{00}^S(\alpha, \delta)$ and $R_{00}^R(\alpha, \delta)$, we partition the remaining intervals (36), (37), (38) in the same manner. The resulting partition of the network state space, $(B, B + A)^2$, can be seen in figure 5. Regions symmetrical to $R_{00}^A(\alpha, \delta)$, $R_{00}^S(\alpha, \delta)$ and $R_{00}^R(\alpha, \delta)$ with respect to the line $x = B + A/2$ are denoted by $R_{10}^A(\alpha, \delta)$, $R_{10}^S(\alpha, \delta)$ and $R_{10}^R(\alpha, \delta)$, respectively. Similarly, $R_{01}^A(\alpha, \delta)$, $R_{01}^S(\alpha, \delta)$ and $R_{01}^R(\alpha, \delta)$ denote the regions symmetrical to $R_{00}^A(\alpha, \delta)$, $R_{00}^S(\alpha, \delta)$ and $R_{00}^R(\alpha, \delta)$, respectively, with respect to the line $y = B + A/2$. Finally, we denote by $R_{11}^A(\alpha, \delta)$, $R_{11}^S(\alpha, \delta)$ and $R_{11}^R(\alpha, \delta)$ the regions that are symmetrical to $R_{01}^A(\alpha, \delta)$, $R_{01}^S(\alpha, \delta)$ and $R_{01}^R(\alpha, \delta)$ with respect to the line $x = B + A/2$.

We are now ready to translate the results formulated in Lemmas 1, 2 and 3 into the (x, y)-space of neural activations.

[5] superscripts A, S and R indicate attractive, saddle and repulsive points, respectively

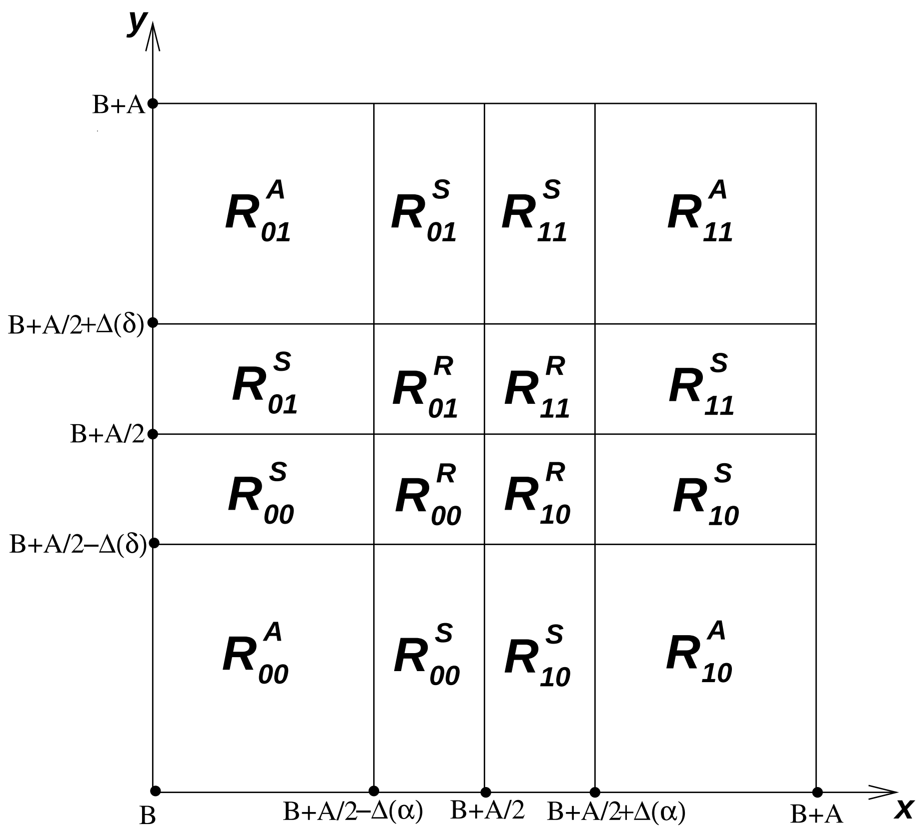

Figure 5: Partitioning of the network state space according to stability types of the fixed points.

Theorem 1: *If* $bc > 0$, $|a| > 4/A$ *and* $|d| > 4/A$, *then all attractive fixed points of (4) lie in*

$$\bigcup_{i \in \mathcal{I}} R_i^A(|a|, |d|),$$

where $\mathcal{I}$ *is the index set* $\mathcal{I} = \{00, 10, 01, 11\}$.

Theorem 2: *If* $bc < 0$, $ad < 0$, $|a| > 4/A$, $|d| > 4/A$ *and* $|ad| \geq |bc|/2$, *then all fixed points of (4) lying in*

$$\bigcup_{i \in \mathcal{I}} R_i^A(|a|, |d|), \quad \mathcal{I} = \{00, 10, 01, 11\}$$

are attractive.

Theorem 3: *If* $|\tilde{a}|, |\tilde{d}| > 4/A$ *(eqs. (24),(25)) and one of the following conditions is satisfied*

- $bc > 0$ *and* $ad < 0$
- $bc < 0$ *and* $ad > 0$
- $bc < 0, ad < 0$ *and* $|ad| \leq |bc|/2$

then all fixed points of (4) lying in

$$\bigcup_{i \in \mathcal{I}} R_i^A(|\tilde{a}|, |\tilde{d}|), \quad \mathcal{I} = \{00, 10, 01, 11\}$$

are attractive.

For an insight into the bifurcation mechanism (explored in the next section) by which attractive fixed points of (4) are created (or dismissed), it is useful to have an idea where other types of fixed points can lie. For the case when the signs of weights on neuron self-connections are equal $(ab > 0)$, as are the signs of weights on the inter-neuron connections $(bc > 0)$, we have the following theorem:

Theorem 4: *Suppose* $ad > 0$, $bc > 0$, $|a| > 4/A$ *and* $|d| > 4/A$. *Then the following can be said about the fixed points of (4):*

- *attractive points can lie only in*

$$\bigcup_{i \in \mathcal{I}} R_i^A(|a|, |d|), \quad \mathcal{I} = \{00, 10, 01, 11\}.$$

- *if* $ad \geq bc/2$, *then all fixed points in*

$$\bigcup_{i\in\mathcal{I}} R_i^S(|a|,|d|)$$

are saddle points; repulsive points can lie only in

$$\bigcup_{i\in\mathcal{I}} R_i^R(|a|,|d|).$$

- *the system (4) has repulsive points only when*

$$|ad-bc| \geq \frac{4}{A} min\{|a|,|d|\}.$$

Proof: Regions for attractive fixed points follow from Theorem 1.

- **Consider first the case** $\mathbf{a}, \mathbf{d} > \mathbf{0}$.

  A fixed point (x,y) of (4) is a saddle if $|\lambda_2| < 1$ and $|\lambda_1| = \lambda_1 > 1$ (see eq. (17) and (18)).

  - Assume $ad > bc$. Then

    $$0 < \sqrt{(aG_1+dG_2)^2 - 4G_1G_2(ad-bc)} = \sqrt{\mathcal{D}(G_1,G_2)} < aG_1 + dG_2.$$

    It follows that if $aG_1 + dG_2 < 2$, (i.e. if $(G_1,G_2) \in \rho_1^-$, see eq. (21)), then

    $$0 < aG_1 + dG_2 - \sqrt{\mathcal{D}(G_1,G_2)} < 2$$

    and so $0 < \lambda_2 < 1$.
    For $(G_1,G_2) \in \rho_1^0 \cup \rho_1^+$, we solve the inequality

    $$aG_1 + dG_2 - \sqrt{\mathcal{D}(G_1,G_2)} < 2,$$

    that is satisfied by (G_1,G_2) from $\kappa_1^- \cap (\rho_1^0 \cup \rho_1^+)$. For a definition of κ_1, see (22). It can be seen (figure 2) that in all fixed points (x,y) of (4) with

    $$\phi(x,y) \in \left(0,\frac{A}{4}\right] \times \left(0, min\left\{\frac{1}{\tilde{d}},\frac{A}{4}\right\}\right] \cup \left(0, min\left\{\frac{1}{\tilde{a}},\frac{A}{4}\right\}\right] \times \left(0,\frac{A}{4}\right],$$

    the eigenvalue $\lambda_2 > 0$ is less than 1. This is certainly true for all (x,y) such that

    $$\phi(x,y) \in (0,A/4] \times (0,1/d) \cup (0,1/a) \times (0,A/4].$$

    In particular, the preimages under the map ϕ of

    $$(G_1,G_2) \in (1/a, A/4] \times (0,1/d) \cup (0,1/a) \times (1/d, A/4]$$

define the region

$$\bigcup_{i\in\mathcal{I}} R_i^S(a,d), \quad \mathcal{I}=\{00,10,01,11\},$$

where only saddle fixed points of (4) can lie.

Fixed points (x,y) whose images under ϕ lie in $\kappa_1^+\cap\rho_1^+$ are repellors. No (G_1,G_2) can lie in that region, if $\tilde{a},\tilde{d}\le 4/A$, that is, if $d(a-4/A)\le bc$ and $a(d-4/A)\le bc$, which is equivalent to

$$max\{a(d-4/A),d(a-4/A)\}\le bc.$$

– When $ad=bc$, we have (see eq. (18))

$$\sqrt{\mathcal{D}(G_1,G_2)}=aG_1+dG_2,$$

and so $\lambda_2=0$. Hence, there are no repelling points if $ad=bc$.

– Assume $ad<bc$. Then

$$\sqrt{\mathcal{D}(G_1,G_2)}>aG_1+dG_2,$$

which implies that λ_2 is negative. It follows that the inequality to be solved is

$$aG_1+dG_2-\sqrt{\mathcal{D}(G_1,G_2)}>-2.$$

It is satisfied by (G_1,G_2) from κ_2^+ (eq. (29)). If $2ad\ge bc$, then $|\tilde{a}|\le a$ and $|\tilde{d}|\le d$.

Fixed points (x,y) with

$$\phi(x,y)\in\left(0,\frac{A}{4}\right]\times\left(0,min\left\{\frac{1}{|\tilde{d}|},\frac{A}{4}\right\}\right]\cup\left(0,min\left\{\frac{1}{|\tilde{a}|},\frac{A}{4}\right\}\right]\times\left(0,\frac{A}{4}\right],$$

have $|\lambda_2|$ less than 1. If $2ad\ge bc$, this is true for all (x,y) such that

$$\phi(x,y)\in(0,A/4]\times(0,1/d)\cup(0,1/a)\times(0,A/4]$$

and the preimages under ϕ of

$$(G_1,G_2)\in(1/a,A/4]\times(0,1/d)\cup(0,1/a)\times(1/d,A/4]$$

define the region $\bigcup_{i\in\mathcal{I}} R_i^S(a,d)$ where only saddle fixed points of (4) can lie.

There are no repelling points, if $|\tilde{a}|,|\tilde{d}|\le 4/A$, that is, if

$$min\{a(d+4/A),d(a+4/A)\}\ge bc.$$

- **The case $\mathbf{a,d<0}$ is analogous to the case $\mathbf{a,d>0}$.** We conclude that

– if $ad > bc$, in all fixed points (x, y) of (4) with

$$\phi(x,y) \in \left(0, \frac{A}{4}\right] \times \left(0, min\left\{\frac{1}{|\tilde{d}|}, \frac{A}{4}\right\}\right] \cup \left(0, min\left\{\frac{1}{|\tilde{a}|}, \frac{A}{4}\right\}\right] \times \left(0, \frac{A}{4}\right],$$

$|\lambda_1| < 1$. Surely, this is true for all (x, y) such that

$$\phi(x,y) \in (0, A/4] \times (0, 1/|d|) \cup (0, 1/|a|) \times (0, A/4].$$

The preimages under ϕ of

$$(G_1, G_2) \in (1/|a|, A/4] \times (0, 1/|d|) \cup (0, 1/|a|) \times (1/|d|, A/4]$$

define the region $\bigcup_{i\in\mathcal{I}} R_i^S(|a|, |d|)$ where only saddle fixed points of (4) can lie. There are no repelling points if $|\tilde{a}|, |\tilde{d}| \leq 4/A$, that is, if $|d|(|a| - 4/A) \leq bc$ and $|a|(|d| - 4/A) \leq bc$, which is equivalent to

$$max\{|a|(|d| - 4/A), |d|(|a| - 4/A)\} \leq bc.$$

– in the case $ad = bc$, we have

$$\sqrt{\mathcal{D}(G_1, G_2)} = |aG_1 + dG_2|,$$

and so $\lambda_1 = 0$. Hence, there are no repelling points.

– if $ad < bc$, in all fixed points (x, y) with

$$\phi(x,y) \in \left(0, \frac{A}{4}\right] \times \left(0, min\left\{\frac{1}{\tilde{d}}, \frac{A}{4}\right\}\right] \cup \left(0, min\left\{\frac{1}{\tilde{a}}, \frac{A}{4}\right\}\right] \times \left(0, \frac{A}{4}\right],$$

$\lambda_1 > 0$ is less than 1. If $2ad \geq bc$, this is true for all (x, y) such that

$$\phi(x,y) \in (0, A/4] \times (0, 1/|d|) \cup (0, 1/|a|) \times (0, A/4]$$

and the preimages under ϕ of

$$(G_1, G_2) \in (1/|a|, A/4] \times (0, 1/|d|) \cup (0, 1/|a|) \times (1/|d|, A/4]$$

define the region $\bigcup_{i\in\mathcal{I}} R_i^S(|a|, |d|)$ where only saddle fixed points of (4) can lie. There are no repelling points if $\tilde{a}, \tilde{d} \leq 4/A$, that is, if

$$min\{|a|(|d| + 4/A), |d|(|a| + 4/A)\} \geq bc.$$

In general, we have shown that if

- $ad < bc$ and $ad + 4min\{|a|, |d|\}/A \geq bc$, or
- $ad = bc$, or
- $ad > bc$ and $ad - 4min\{|a|, |d|\}/A \leq bc$,

then there are no repelling points. ■

6 Creation of a new attractive fixed point through saddle node bifurcation

In this section we are concerned with the actual position of fixed points of (4). We study, how the coefficients a, b, t_1, c, d and t_2 effect the number and position of the fixed points. It is illustrative first to concentrate only on a single neuron $\mathcal{N}$ from the pair of neurons forming the neural network.

Denote the values of the weights associated with the self–loop on the neuron $\mathcal{N}$ and with the interconnection link from the other neuron to the neuron $\mathcal{N}$ by s and r, respectively. The constant input to the neuron $\mathcal{N}$ is denoted by t. If the activations of the neuron $\mathcal{N}$ and the other neuron are u and v, respectively, then the activation of the neuron $\mathcal{N}$ at the next time step is $g(su + rv + t)$.

If the activation of the neuron $\mathcal{N}$ is not to change, (u, v) should lie on the curve $f_{s,r,t}$:

$$v = f_{s,r,t}(u) = \frac{1}{r}\left(-t - su + ln\frac{u-B}{B+A-u}\right). \tag{43}$$

The function

$$ln((u-B)/(B+A-u)) : (B, B+A) \to \Re,$$

is monotonically increasing with

$$\lim_{u\to B^+} ln\frac{u-B}{B+A-u} = -\infty \quad \text{and} \quad \lim_{u\to(B+A)^-} ln\frac{u-B}{B+A-u} = \infty.$$

Even though the linear function $-su + t$ cannot influence the asymptotic properties of $f_{s,r,t}$, it can locally influence its "shape". In particular, while the effect of the constant term $-t$ is just a vertical shift of the whole function, $-su$ (if decreasing, i.e. if $s > 0$, and "sufficiently large") has the power to overcome for a while the increasing tendencies of $ln((u-B)/(B+A-u))$. More precisely, if $s > 4/A$, then the term $-su$ causes the function $-su - t + ln((u-B)/(B+A-u))$ to "bend" so that on

$$\left[B + \frac{A}{2} - \Delta(s), B + \frac{A}{2} + \Delta(s)\right]$$

it is decreasing, while it still increases on

$$\left(B, B + \frac{A}{2} - \Delta(s)\right) \cup \left(B + \frac{A}{2} + \Delta(s), B + A\right).$$

The function

$$-su - t + ln((u-B)/(B+A-u))$$

is always concave and convex on $(B, B+A/2)$ and $(B+A/2, B+A)$, respectively.

Finally, the coefficient r scales the whole function and flips it around the u–axis, if $r < 0$. A graph of $f_{s,r,t}(u)$ is presented in figure 6.

Each fixed point of (4) lies on the intersection of two curves $y = f_{a,b,t_1}(x)$ and $x = f_{d,c,t_2}(y)$ (see eq.(43)).

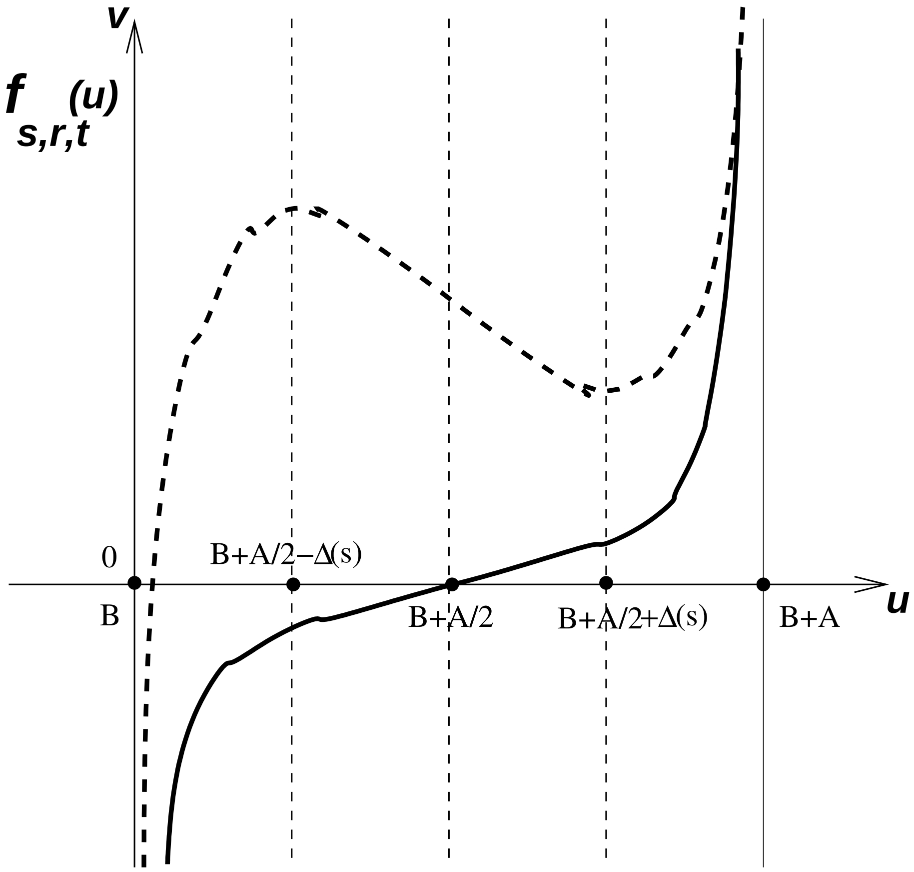

Figure 6: A graph of $f_{s,r,t}(u)$. Solid line represents the case $t, s = 0, r > 0$. Dashed line shows the graph when $t < 0, s > 4/A$ and $r > 0$. Negative external input t shifts the bended part into $v > 0$.

There are $\binom{4}{2}+4=10$ possible cases of coexistence of the two neurons in the network. Based on the results from the previous section, in some cases we are able to predict stability types of the fixed points of (4) according to their position in the neuron activation space. More interestingly, we have structured the network state space $(B, B+A)^2$ into areas corresponding to stability types of the fixed points and, in some cases, these areas directly correspond to monotonicity intervals of the functions f_{a,b,t_1} and f_{d,c,t_2} defining the position of the fixed points.

The results of the last section will be particularly useful when the weights a, b, c, d and external inputs t_1, t_2 are such that the functions f_{a,b,t_1} and f_{d,c,t_2} "bend", thus possibly creating a complex intersection pattern in $(B, B+A)^2$.

For $a > 4/A$, the set

$$f^{\#0}_{a,b,t_1} = \left\{ (x, f_{a,b,t_1}(x)) |\ x \in \left(B, B + \frac{A}{2} - \Delta(a) \right) \right\} \tag{44}$$

contains the points lying on the "first outer branch" of $f_{a,b,t_1}(x)$

Analogously, the set

$$f^{\#1}_{a,b,t_1} = \left\{ (x, f_{a,b,t_1}(x)) |\ x \in \left(B + \frac{A}{2} + \Delta(a), B + A \right) \right\} \tag{45}$$

contains the points in the "second outer branch" of $f_{a,b,t_1}(x)$.

Finally,

$$f^{*}_{a,b,t_1} = \left\{ (x, f_{a,b,t_1}(x)) |\ x \in \left(B + \frac{A}{2} - \Delta(a), B + \frac{A}{2} + \Delta(a) \right) \right\} \tag{46}$$

is the set of points on the "middle branch" of $f_{a,b,t_1}(x)$.

Similarly, for $d > 4/A$, we define the sets

$$f^{\#0}_{d,c,t_2} = \left\{ (f_{d,c,t_2}(y), y) |\ y \in \left(B, B + \frac{A}{2} - \Delta(d) \right) \right\}, \tag{47}$$

$$f^{\#1}_{d,c,t_2} = \left\{ (f_{d,c,t_2}(y), y) |\ y \in \left(B + \frac{A}{2} + \Delta(d), B + A \right) \right\} \tag{48}$$

and

$$f^{*}_{d,c,t_2} = \left\{ (f_{d,c,t_2}(y), y) |\ y \in \left(B + \frac{A}{2} - \Delta(d), B + \frac{A}{2} + \Delta(d) \right) \right\}. \tag{49}$$

containing the points on the "first outer branch", the "second outer branch" and the "middle branch", respectively, of $f_{d,c,t_2}(y)$.

Using Theorem 4 we state the following corollary:

Corollary 1: *Assume $a > 4/A$, $d > 4/A$, $bc > 0$ and $ad \geq bc/2$. Then, attractive fixed points of (4) can lie only on the intersection of the outer branches of f_{a,b,t_1} and f_{d,c,t_2}.*

Whenever the middle branch of f_{a,b,t_1} intersects with an outer branch of f_{d,c,t_2} (or vice-versa), it corresponds to a saddle point of (4). In other words, all attractive fixed points of (4) are from

$$\bigcup_{i,j=0,1} f^{\#i}_{a,b,t_1} \cap f^{\#j}_{d,c,t_2}.$$

Every point from

$$f^{*}_{a,b,t_1} \cap \bigcup_{i=0,1} f^{\#i}_{d,c,t_2},$$

or

$$f^{*}_{d,c,t_2} \cap \bigcup_{i=0,1} f^{\#i}_{a,b,t_1}$$

is a saddle point of (4).

Corollary 1 suggests the saddle-node bifurcation as the usual scenario of creation of a new attractive fixed point. In the saddle-node bifurcation, a pair of new fixed points, of which one is attractive and the other one is a saddle, is created. Attractive fixed points disappear in a reverse manner: an attractive point coalesces with a saddle and they are annihilated. This is illustrated in figure 7.

The curve $f_{d,c,t_2}(y)$ (dashed line) intersects with $f_{a,b,t_1}(x)$ in three points. By increasing d, f_{d,c,t_2} continues to bend (solid curve) and intersects with f_{a,b,t_1} in five points[6]. Saddle and attractive points are marked with squares and stars, respectively. Note that as d increases attractive fixed points move closer to vertices $\{B, B+A\}^2$ of the network state space. Next section rigorously analyzes this tendency in the context of recurrent networks with an arbitrary, finite number of neurons.

7 Attractive Periodic Sets in Recurrent Neural Networks with N neurons

From now on, we study fully connected recurrent neural networks with N neurons. As usual, the weight on a connection from neuron j to neuron i is denoted by w_{ij}. The output of the i-th neuron at time step m is denoted by $x_i^{(m)}$. The external input to neuron i is T_i. Each neuron has a "sigmoid-shaped" transfer function (2).

The network evolves in the N-dimensional state space $\mathcal{O} = (B, B+A)^N$ according to

$$x_i^{(m+1)} = g_{A,B,\mu_i}\left(\sum_{n=1}^{N} w_{in} x_n^{(m)} + T_i\right) \qquad i = 1, 2, ..., N. \tag{50}$$

Note that we allow different neural gains across the neuron population.

[6] At the same time, $|c|$ has to be also appropriately increased so as to compensate for the increase in d so that the "bended" part of f_{d,c,t_2} does not move radically to higher values of x.

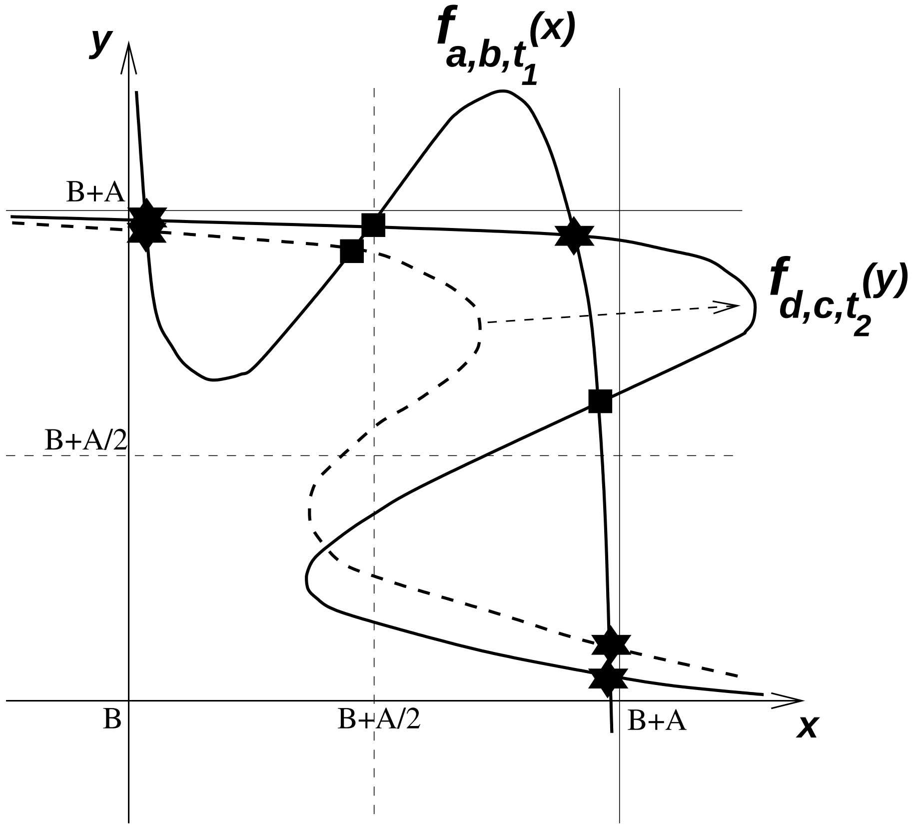

Figure 7: Geometrical illustration of the saddle-node bifurcation in a neural network with two recurrent neurons. $f_{d,c,t_2}(y)$ shown as dashed curve intersects with $f_{a,b,t_1}(x)$ in three points. By increasing d, f_{d,c,t_2} bends further (solid curve) and intersects with f_{a,b,t_1} in five points. Saddle and attractive points are marked with squares and stars, respectively. The parameter settings is: $a, d > 0,\ b, c < 0$.

In vector notation, we represent the state $x^{(m)}$ at time step m as[7]

$$x^{(m)} = (x_1^{(m)}, x_2^{(m)}, ..., x_N^{(m)})^T$$

and (50) becomes

$$x^{(m+1)} = F(x^{(m)}). \tag{51}$$

It is convenient to represent the weights w_{ij} as a weight matrix $\mathcal{W} = (w_{ij})$. We assume that $\mathcal{W}$ is non-singular, i.e. $\det(\mathcal{W}) \neq 0$.

Assuming an external input $T = (T_1, T_2, ..., T_N)^T$, the Jacobian matrix $\mathcal{J}(x)$ of the map F in $x \in \mathcal{O}$ can be written as

$$\mathcal{J}(x) = \mathcal{M}\mathcal{G}(x)\mathcal{W}, \tag{52}$$

where $\mathcal{M}$ and $\mathcal{G}(x)$ are N-dimensional diagonal matrices

$$\mathcal{M} = \mathrm{diag}(\mu_1, \mu_2, ..., \mu_N) \tag{53}$$

and

$$\mathcal{G}(x) = \mathrm{diag}(G_1(x), G_2(x), ..., G_N(x)), \tag{54}$$

respectively, with[8]

$$G_i(x) = g'\left(\mu_i \sum_{n=1}^{N} w_{in} x_n + \mu_i T_i\right), \quad i = 1, 2, ..., N. \tag{55}$$

We denote absolute value of the determinant of the weight matrix $\mathcal{W}$ by W, i.e.

$$W = |\det(\mathcal{W})|. \tag{56}$$

Then

$$|\det(\mathcal{J}(x))| = W \prod_{n=1}^{N} \mu_n G_n(x). \tag{57}$$

Lemma 4 formulates a useful relation between the value of the transfer function and its derivative. It is formulated in a more general setting than that used in sections 4 and 5.

Lemma 4: *Let $u \in (0, A/4]$. Then the set*

$$\{g(\ell)| \;\; g'(\ell) \geq u\}$$

is a closed interval of length $2\Delta(1/u)$ centered at $B + A/2$, where $\Delta(\cdot)$ is defined in (34).

[7] superscript T means the transpose operator

[8] remember that for the sake of simplicity, we denote $g_{A,B,1}$ by g

Proof: For a particular value u of $g'(\ell) = \psi(g(\ell))$, the corresponding values[9] $\psi^{-1}(u)$ of $g(\ell)$ are

$$B + \frac{A}{2} \pm \Delta\left(\frac{1}{u}\right).$$

The transfer function $g(\ell)$ is monotonically increasing with increasing and decreasing derivative on $(-\infty, 0)$ and $(0, \infty)$ respectively. The maximal derivative $g'(0) = A/4$ occurs at $\ell = 0$ with $g(0) = B + A/2$. ∎

Denote the geometrical mean of neural gains in the neuron population by $\tilde{\mu}$

$$\tilde{\mu} = \left[\prod_{n=1}^{N} \mu_n\right]^{\frac{1}{N}}. \tag{58}$$

For $W^{1/N} \geq 4/(A\tilde{\mu})$, define

$$c_- = B + \frac{A}{2} - \Delta\left(\tilde{\mu} W^{\frac{1}{N}}\right), \tag{59}$$

$$c_+ = B + \frac{A}{2} + \Delta\left(\tilde{\mu} W^{\frac{1}{N}}\right). \tag{60}$$

Denote the hypercube $[c_-, c_+]^N$ by $\mathcal{H}$. The next theorem states that increasing neural gains μ_i push the attractive sets away from the center $\{B + A/2\}^N$ of the network state space towards the faces of the activation hypercube $\mathcal{O} = (B, B + A)^N$.

Theorem 5: *Assume*

$$W \geq \left(\frac{4}{A\tilde{\mu}}\right)^N.$$

Then, attractive periodic sets of (51) cannot lie in the hypercube $\mathcal{H}$ with sides of length

$$2\Delta\left(\tilde{\mu} W^{\frac{1}{N}}\right),$$

centered at $\{B + A/2\}^N$, the center of the state space.

Proof: Suppose there is an attractive periodic orbit $X = \{x^{(1)}, ..., x^{(Q)}\} \subset \mathcal{H} = [c_-, c_+]^N$, with $F(x^{(1)}) = x^{(2)}, F(x^{(2)}) = x^{(3)}, ..., F(x^{(Q)}) = x^{(Q+1)} = x^{(1)}$.

Determinant of the Jacobian matrix of the map F^Q in $x^{(j)}$, $j = 1, ..., Q$, is

$$\det(\mathcal{J}_{F^Q}(x^{(j)})) = \prod_{k=1}^{Q} \det(\mathcal{J}(x^{(k)}))$$

and by (57)

$$|\det(\mathcal{J}_{F^Q}(x^{(j)}))| = W^Q \prod_{k=1}^{Q} \prod_{n=1}^{N} \mu_n G_n(x^{(k)}).$$

[9] ψ^{-1} is the set theoretic inverse of ψ

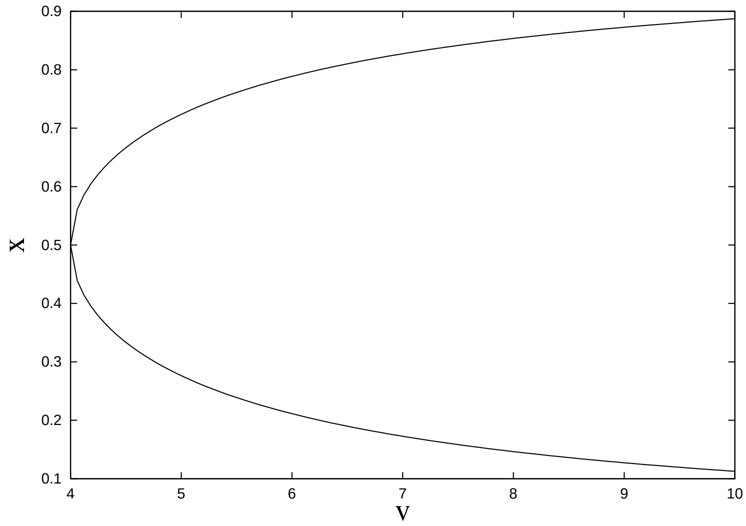

Figure 8: An illustration of the growth of the hypercube $\mathcal{H}$ with increasing $v = \tilde{\mu}W^{1/N}$. Parameters of the transfer function are set to $B = 0, A = 1$. The lower and upper curves correspond to functions $1/2 - \Delta(v)$ and $1/2 + \Delta(v)$, respectively. For a particular value of v, the hypercube $\mathcal{H}$ has each side centered at $1/2$ and spanned between the two curves.

From

$$G_n(x^{(k)}) = \psi(x_n^{(k+1)}),\ n = 1, ..., N,\ k = 1, ..., Q,$$

where ψ is defined in (7), it follows that if all $x^{(k)}$ are from the hypercube $\mathcal{H} = [c_-, c_+]^N$ (see eq. (59), (60)), then by lemma 4 we have

$$W^Q \prod_{k=1}^{Q} \prod_{n=1}^{N} \mu_n G_n(x^{(k)}) \geq W^Q \prod_{k=1}^{Q} \prod_{n=1}^{N} \frac{\mu_n}{\tilde{\mu} W^{\frac{1}{N}}} = 1.$$

But

$$\det(\mathcal{J}_{F^Q}(x^{(1)})) = \det(\mathcal{J}_{F^Q}(x^{(2)})) = ... = \det(\mathcal{J}_{F^Q}(x^{(Q)}))$$

is the product of eigenvalues of F^Q in $x^{(j)}$, $j = 1, ..., Q$, and so the absolute value of at least one of them has to be equal to, or greater than 1. This contradicts the assumption that X is attractive. ∎

The plots in figure 8 illustrate the growth of the hypercube $\mathcal{H}$ with growing $v = \tilde{\mu}W^{1/N}$ in case of $B = 0$ and $A = 1$. The lower and upper curves correspond to functions $\frac{1}{2} - \Delta(v)$ and $\frac{1}{2} + \Delta(v)$, respectively. For a particular value of v, $\mathcal{H}$ is the hypercube with each side centered at $\frac{1}{2}$ and spanned between the two curves.

7.1 The effect of growing neural gain

In (Hirsch, 1994), Hirsch studies recurrent networks with non-decreasing transfer functions f from a broader class than the "sigmoid-shaped" class (eq. (2)) considered here.

Transfer functions can differ from neuron to neuron. He shows that attractive fixed point coordinates corresponding to neurons with a steadily increasing neural gain tend to saturation values[10] as the gain grows without a bound. It is assumed that all the neurons with increasing gain are either self-exciting, or self-inhibiting[11].

To investigate the effect of growing neural gain in our setting, we assume that the neurons are split into two groups. Initially, all the neurons have the same transfer function g_{A,B,μ_*}. The neural gain on neurons from the first group does not change, while the gain on neurons from the second group is allowed to grow. We investigate how the growth of neural gain in the second subset of neurons effects regions for periodic attractive sets of the network. In particular, we answer the following question: Given the weight matrix $\mathcal{W}$ and a "small neighborhood factor" $\epsilon > 0$, for how big a neural gain μ on neurons from the second group, all the attractive fixed points and at least some attractive periodic points must lie within ϵ-neighborhood of the faces of the activation hypercube $\mathcal{O} = (B, B+A)^N$?

Theorem 6: *Let $\epsilon > 0$ be a "small" positive constant. Assume M $(M \leq N)$ and $N - M$ neurons have the neural gain equal to μ and μ_*, respectively. Then, for*

$$\mu > \mu(\epsilon) = \frac{\mu_*^{1-\frac{N}{M}}}{W^{\frac{1}{M}}} \frac{1}{\left[\epsilon\left(1 - \frac{\epsilon}{A}\right)\right]^{\frac{N}{M}}},$$

all attractive fixed points and at least some points from each attractive periodic orbit of the network lie in the ϵ-neighborhood of the faces of the hypercube $\mathcal{O} = (B, B+A)^N$.

Proof: By theorem 5,

$$\epsilon = \frac{A}{2} - \Delta\left(\mu_*^{\frac{N-M}{N}} \mu(\epsilon)^{\frac{M}{N}} W^{\frac{1}{N}}\right) = \frac{A}{2}\left[1 - \sqrt{1 - \frac{4}{A} \frac{1}{\mu_*^{\frac{N-M}{N}} \mu(\epsilon)^{\frac{M}{N}} W^{\frac{1}{N}}}}\right]. \tag{61}$$

Solving (61) for $\mu(\epsilon)$, we arrive at the expression presented in the theorem. ∎

Note that for very close ϵ-neighborhoods of the faces of $\mathcal{O}$, the growth of μ obeys

$$\mu \propto \frac{1}{\epsilon^{\frac{N}{M}}} \tag{62}$$

and if the gain is allowed to grow on each neuron,

$$\mu \propto \frac{1}{\epsilon}. \tag{63}$$

Equation (61) constitutes a lower bound on the rate of convergence of the attractive fixed points towards the faces of $\mathcal{O}$, as the neural gain μ on M neurons grows.

[10]In fact, it is shown that each of such coordinates tend to either a saturation value $f(\pm\infty)$, which is assumed to be finite, or to a critical value $f(v)$, with $f'(v) = 0$. The critical values are assumed to be finite in number. There are no critical values in the class of transfer functions considered in this paper.

[11]self-exciting and self-inhibiting neurons have positive and negative weights, respectively, on the feedback self-loops

7.2 Using alternative bounds on spectral radius of Jacobian Matrix

Results in the previous subsection are based on the bound

$$\textit{if} \ \ |\det(\mathcal{J}(x))| \geq 1, \ \ \textit{then} \ \ \rho(\mathcal{J}(x)) \geq 1,$$

where $\rho(\mathcal{J}(x))$ is the spectral radius[12] of the Jacobian $\mathcal{J}(x)$. Although one can think of other bounds on spectral radius of $\mathcal{J}(x)$, usually the expression describing the conditions on terms $G_n(x)$, $n = 1, ..., N$, such that $\rho(\mathcal{J}(x)) \geq 1$, is very complex (if it exists in a closed form at all). This prevents us from obtaining a relatively simple analytical approximation of the set

$$\Upsilon = \{x \in \mathcal{O} | \ \ \rho(\mathcal{J}(x)) \geq 1\}, \tag{64}$$

where no attractive fixed points of (51) can lie. Furthermore, in general, if two matrices have their spectral radii equal to or greater than one, the same does not necessarily hold for their product. As a consequence, one cannot directly reason about regions with no periodic attractive sets.

In this subsection, we shall use a simple bound on the spectral radius of a square matrix stated in the following lemma.

Lemma 5: For an M-dimensional square matrix $\mathcal{A}$,

$$\rho(\mathcal{A}) \geq \frac{|\mathrm{trace}(\mathcal{A})|}{M}.$$

Proof: Let λ_i, $i = 1, 2, ..., M$, be the eigenvalues of $\mathcal{A}$. From

$$\mathrm{trace}(\mathcal{A}) = \sum_i \lambda_i,$$

it follows that

$$\rho(\mathcal{A}) \geq \frac{1}{M} \sum_i |\lambda_i| \geq \frac{1}{M} \left| \sum_i \lambda_i \right| = \frac{|\mathrm{trace}(\mathcal{A})|}{M}. \quad \blacksquare$$

Theorems 7 and 8 are analogous to theorems 5 and 6. We assume the same transfer function on all neurons. Generalization to transfer functions differing in neural gains is straightforward.

Theorem 7: *Assume all the neurons have the same transfer function (2) and the weights on neural self-loops are exclusively positive, or exclusively negative, i.e. either* $w_{nn} >$

[12]the maximal absolute value of eigenvalues of $\mathcal{J}(x)$

$0,\ n=1,2,...,N,$ *or* $w_{nn}<0,\ n=1,2,...,N$. *If attractive fixed points of (51) lie in the hypercube with sides of length*

$$2\Delta\left(\frac{\mu|\mathrm{trace}(\mathcal{W})|}{N}\right),$$

centered at $\{B+A/2\}^N$*, the center of the state space, then*

$$|\mathrm{trace}(\mathcal{W})|<\frac{4N}{\mu A}.$$

Proof: By lemma 5,

$$\mathcal{K}=\left\{x\in\mathcal{O}\,\middle|\,\frac{|\mathrm{trace}(\mathcal{J}(x))|}{N}\geq 1\right\}\subset\Upsilon.$$

The set $\mathcal{K}$ contains all states x such that the corresponding N-tuple

$$G(x)=(G_1(x),G_2(x),...,G_N(x))$$

lies in the half space Ξ not containing the point $\{0\}^N$, with border defined by the hyper plane

$$\sigma:\ \sum_{n=1}^{N}|w_{nn}|G_n=\frac{N}{\mu}.$$

Since all the coefficients $|w_{nn}|$ are positive, the intersection of the line

$$G_1=G_2=...=G_N$$

with σ exists and is equal to

$$G_*=\frac{N}{\mu\sum_{n=1}^{N}|w_{nn}|}=\frac{N}{\mu|\mathrm{trace}(\mathcal{W})|}.$$

Moreover, since $|w_{nn}|>0$, if

$$|\mathrm{trace}(\mathcal{W})|\geq 4N/(\mu A),$$

then $[G_*,A/4]^N\subset\Xi$.

The hypercube $[G_*,A/4]^N$ in the space of transfer function derivatives corresponds to the hypercube

$$\left[B+\frac{A}{2}-\Delta\left(\frac{1}{G_*}\right),B+\frac{A}{2}+\Delta\left(\frac{1}{G_*}\right)\right]^N$$

in the network state space $\mathcal{O}$. ∎

Theorem 8: *Let* $\epsilon>0$ *be a "small" positive constant. Assume all the neurons have the same transfer function (2). If either* $w_{nn}>0,\ n=1,2,...,N,$ *or* $w_{nn}<0,\ n=1,2,...,N,$ *then for*

$$\mu>\mu(\epsilon)=\frac{N}{\mu|\mathrm{trace}(\mathcal{W})|}\frac{1}{\left[\epsilon\left(1-\frac{\epsilon}{A}\right)\right]}$$

all attractive fixed points of (51) lie in the ϵ-neighborhood of the faces of the hypercube $\mathcal{O}$.

Proof: By theorem 7,

$$\epsilon = \frac{A}{2} - \Delta\left(\mu(\epsilon)\frac{|\mathrm{trace}(\mathcal{W})|}{N}\right). \tag{65}$$

Solving (65) for $\mu(\epsilon)$, we arrive at the expression in the theorem. ∎

8 Relation to continuous-time networks

There is a large amount of literature devoted to dynamical analysis of continuous-time networks. While discrete-time networks are more appropriate for dealing with discrete and symbolic data, continuous-time networks seem more natural in many "physical" models and applications.

Relations between the dynamics of discrete-time and continuous-time networks are considered, for example, in (Blum & Wang, 1992) and (Tonnelier et al., 1999). Hirsch (1994) gives a simple example illustrating that there is no fixed step size for discretizing continuous-time dynamics that would yield a discrete-time dynamics accurately reflecting its continuous-time origin.

It has been pointed out in (Hirsch, 1994) that while there is a saturation result for stable limit cycles of continuous-time networks – for sufficiently high gain, the output along a stable limit cycle is saturated almost all the time – there is no known analog of this for stable periodic orbits of discrete-time networks. Theorems 5 and 6 offer such an analog for discrete-time networks with sigmoid-shaped transfer functions studied in this paper.

Vidyasagar (1993) studied the number, location and stability of high-gain equilibria in continuous-time networks with sigmoidal transfer functions. He concludes that all equilibria in the corners of the activation hypercube are stable. Theorems 6 and 7 formulate analogous results for discrete-time networks.

9 Conclusion

By performing a detailed fixed point analysis of two-neuron recurrent networks, we partitioned the network state space into regions corresponding to the fixed point stability types. The results are intuitive and hold for a large class of sigmoid-shaped neuron transfer functions. Attractive fixed points cluster around the vertices of the activation square $[L, H]^2$, where L and H are the low and high transfer function saturation levels, respectively. Repelling fixed points are concentrated close to the center $\{\frac{L+H}{2}\}^2$ of the activation square. Saddle fixed points appear in the neighborhood of the four sides of the activation square. Unlike in the previous studies (e.g. (Blum & Wang, 1992; Borisyuk & Kirillov, 1992;

Pasemann, 1993; Tonnelier et al., 1999)), we allowed all free parameters of the network to vary.

We have rigorously shown that when the neurons self-excite themselves and have the same mutual-interaction pattern, a new attractive fixed point is created through the saddle-node bifurcation. This is in accordance with recent empirical findings (Tonnelier et al., 1999).

Next, we studied recurrent networks of arbitrary finite number of neurons with sigmoid-shaped transfer functions. Inspired by the result of Hirsch (1994) concerning equilibrium points in high-neural-gain networks, we analyzed the tendency of attractive periodic points to approach saturation faces of the activation hypercube as the neural gain increases. Their distance from the saturation faces is approximately reciprocal to the neural gain.

References

Amit, D. (1989). *Modeling brain function.* Cambridge: Cambridge University Press.

Anderson, J. (1993). The BSB model: A simple nonlinear autoassociative neural network. In M.H. Hassoun (Eds.), *Associative neural memories: Theory and implementation* (pp. 77–103). Oxford: Oxford University Press.

Anton, H. (1980). *Calculus with analytic geometry.* New York: John Wiley and Sons.

Beer, R.D. (1995). On the dynamics of small continuous–time recurrent networks. *Adaptive Behavior*, 3(4), 471–511.

Blum, K.L., & Wang, X. (1992). Stability of fixed points and periodic orbits and bifurcations in analog neural networks. *Neural Networks*, 5, 577–587.

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

Botelho, F. (1999). Dynamical features simulated by recurrent neural networks. *Neural Networks*, 12, 609–615.

Casey, M.P. (1995). *Relaxing the symmetric weight condition for convergent dynamics in discrete-time recurrent networks* (Tech. Rep. INC-9504). La Jolla, CA: Institute for Neural Computation, University of California, San Diego.

Casey, M.P. (1996). The dynamics of discrete-time computation, with application to recurrent neural networks and finite state machine extraction. *Neural Computation*, 8(6), 1135–1178.

Cleeremans, A., Servan-Schreiber, D., & McClelland, J.L. (1992). Finite state automata and simple recurrent networks. *Neural Computation*, 1(3), 372–381.

Giles, C.L., Miller, C.B., Chen, D., Chen, H.H., Sun, G.Z., & Lee, Y.C. (1992) Learning and extracting finite state automata with second–order recurrent neural networks. *Neural Computation*, 4(3), 393–405.

Hirsch, M.W. (1994). Saturation at high gain in discrete time recurrent networks. *Neural Networks*, 7(3), 449–453.

Hopfield, J.J. (1984). Neurons with a graded response have collective computational properties like those of two–state neurons. *Proceedings of the National Academy of Science USA*, 81, 3088–3092.

Hui, S., & Zak, S.H. (1992). Dynamical analysis of the brain-state-in-a-box neural models. *IEEE Transactions on Neural Networks*, 1, 86–94.

Jin, L., Nikiforuk, P.N., & Gupta, M.M. (1994). Absolute stability conditions for discrete-time recurrent neural networks. *IEEE Transactions on Neural Networks*, 6, 954–963.

Klotz, A., & Brauer, K. (1999). A small-size neural network for computing with strange attractors. *Neural Networks*, 12, 601–607.

Manolios, P., & Fanelli, R. (1994). First order recurrent neural networks and deterministic finite state automata. *Neural Computation*, 6(6), 1155–1173.

McAuley, J.D., & Stampfli, J. (1994). Analysis of the effects of noise on a model for the neural mechanism of short-term active memory. *Neural Computation*, 6(4), 668–678.

Nakahara, H., & Doya, K. (1998). Near-saddle-node biffurcation behavior as dynamics in working memory for goal-directed behavior. *Neural Computation*, 10(1), 113–132.

Pakdamann, K., Grotta-Ragazzo, C., Malta, C.P., Arino, O., & Vibert, J.F. (1998). Effect of delay on the boundary of the basin of attraction in a system of two neurons. *Neural Networks*, 11, 509–519.

Pasemann, F. (1993). Discrete dynamics of two neuron networks. *Open Systems & Information Dynamics*, 2, 49–66.

Pasemann, F. (1995a). Neuromodules: a dynamical systems approach to brain modelling. In H.J. Hermann, D.E. Wolf, & E. Poppel (Eds.), *Supercomputing in Brain Research* (pp. 331–348). Singapore: World Scientific.

Pasemann, F. (1995b). Characterization of periodic attractors in neural ring network. *Neural Networks*, 8, 421–429.

Rodriguez, P., Wiles, J., & Elman, J.L. (1999). A recurrent neural network that learns to count. *Connection Science*, 11, 5–40.

Sevrani, F., & Abe, K. (2000). On the synthesis of brain-state-in-a-box neural models with aplication to associative memory. *Neural Computation*, 12(2), 451–472.

Tiňo, P., & Sajda, J. (1995). Learning and extracting initial mealy machines with a modular neural network model. *Neural Computation*, 7(4), 822–844.

Tiňo, P., Horne, B.G., Giles, C.L., & Collingwood, P.C. (1998). Finite state machines and recurrent neural networks – automata and dynamical systems approaches. In J.E. Dayhoff, & O. Omidvar (Eds.), *Neural Networks and Pattern Recognition* (pp. 171–220). Academic Press.

Tonnelier, A., Meignen, S., Bosh, H., & Demongeot, J. (1999). Synchronization and desynchronization of neural oscilators. *Neural Networks*, 12, 1213–1228.

Vidyasagar, M. (1993). Location and Stability of the High-Gain Equilibria of Nonlinear Neural Networks. *Neural Networks*, 4, 660–672.

Wang, D.L. (1996). Synchronous oscillations based on lateral connections. In J. Sirosh, R. Miikkulainen, & Y. Choe (Eds.), *Lateral connections in the cortex: structure and function.* Austin, TX: UTCS Neural Networks Research Group.

Wang, X. (1991). Period-doublings to chaos in a simple neural network: An analytical proof. *Complex Systems*, 5, 425–441.

Watrous, R.L., & Kuhn, G.M. (1992). Induction of finite–state languages using second–order recurrent networks. *Neural Computation*, 4(3), 406–414.

Zeng, Z., Goodman, R.M., & Smyth, P. (1993) Learning finite state machines with self–clustering recurrent networks. *Neural Computation*, 5(6), 976–990.

Zhou, M. (1996). *Fault-tolerance for two neuron networks.* Master's Thesis. Memphis, TN, The University of Memphis.