

# **Dynamical stability conditions for recurrent neural networks with unsaturating piecewise linear transfer functions**

**Heiko Wersing, Wolf-Jürgen Beyn, Helge Ritter**

**2001**

**Preprint:**

This is an accepted article published in Neural Computation. The final authenticated version is available online at: [https://doi.org/\[DOI not available\]](https://doi.org/[DOI not available])

# Dynamical Stability Conditions for Recurrent Neural Networks with Unsaturing Piecewise Linear Transfer Functions

Heiko Wersing<sup>1\*</sup>, Wolf-Jürgen Beyn<sup>2</sup>, and Helge Ritter<sup>1</sup>

<sup>1</sup>Faculty of Technology, <sup>2</sup>Faculty of Mathematics,  
University of Bielefeld  
P.O. Box 100131, D-33501 Bielefeld, Germany

## Abstract

We establish two conditions which ensure the non-divergence of additive recurrent networks with unsaturating piecewise linear transfer functions, also called linear threshold or semilinear transfer functions. As was recently shown by Hahnloser et al. (2000), networks of this type can be efficiently built in silicon and exhibit the coexistence of digital selection and analogue amplification in a single circuit. To obtain this behaviour, the network must be multistable and non-divergent and our conditions allow to determine the regimes where this can be achieved with maximal recurrent amplification. The first condition can be applied to nonsymmetric networks and has a simple interpretation of requiring that the strength of local inhibition must match the sum over excitatory weights converging onto a neuron. The second condition is restricted to symmetric networks, but can also take into account the stabilizing effect of non-local inhibitory interactions. We demonstrate the application of the conditions on a simple example and the orientation-selectivity model of Ben-Yishai et al. (1995). We show that the conditions can be used to identify in their model regions of maximal orientation-selective amplification and symmetry breaking.

## 1 Introduction

A prominent model of the coupled dynamics of networks of neurons is the additive recurrent model

$$\dot{x}_i = -x_i + \sigma_i \left( \sum_j w_{ij} x_j + h_i \right), \quad (1)$$

where  $x_i$  denotes the activity of neuron  $i$ . The synaptic weights  $w_{ij}$  denote the strength of the synaptic connection from neuron  $j$  onto neuron  $i$ . The weights can be positive or negative corresponding to an excitatory or inhibitory synapse respectively. The function  $\sigma_i(\cdot)$  is called the transfer or gain function of the neuron  $i$ , and can be interpreted as computing the cell's average spike rate depending on its membrane potential. The constants  $h_i$  denote external inputs.

The form of the transfer function  $\sigma(\cdot)$  (see Figure 1) plays an important role for bounding the dynamics (1). If the transfer function has lower and upper saturation limits according

---

\*New address: HONDA R&D Europe (Germany) GmbH, Future Technology Research, Carl-Legien-Str.30, 63073 Offenbach/Main, Germany, email: heiko.wersing@hre-ftr.f.rd.honda.co.jp

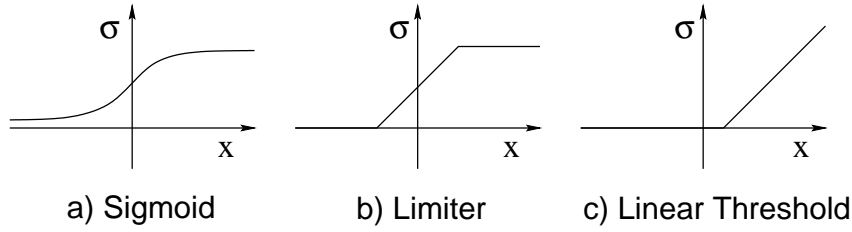


Figure 1: Transfer functions. The sigmoidal function a) is continuous and saturates for  $x \rightarrow \infty$ . The limiter function b) is linear between its saturation points. The linear threshold or semilinear transfer function does not saturate as  $x \rightarrow \infty$ .

to  $a \leq \sigma(x) \leq b$ , then the dynamics is obviously bounded, since  $\dot{x}_i \geq 0$  for  $x_i \leq a$  and  $\dot{x}_i \leq 0$  for  $x_i \geq b$ . Examples for such saturating transfer functions are the logistic or Fermi function

$$\sigma(x) = \frac{1}{1 + e^{-2\beta x}} \quad (2)$$

and the limiter function (also referred to as the saturating linear function or the piecewise linear sigmoidal function)

$$\sigma(x) = \begin{cases} a, & \text{if } x < a, \\ x, & \text{if } a \leq x \leq b, \\ b, & \text{if } x > b. \end{cases} \quad (3)$$

In recent years there has been a growing amount of literature on biologically-based models (Hartline & Ratliff 1958; von der Malsburg 1973; Douglas, Mahowald, & Martin 1994; Ben-Yishai, Lev Bar-Or, & Sompolinsky 1995; Salinas & Abbott 1996; Adorjan, Levitt, Lund, & Obermayer 1999; Bauer, Scholz, Levitt, Obermayer, & Lund 1999), where the transfer function, denoted as semilinear or linear threshold (LT) function, is non-saturating and of the form

$$\sigma(x) = \begin{cases} 0 & \text{if } x < \theta, \\ k(x - \theta) & \text{if } x \geq \theta, \end{cases} \quad (4)$$

where  $\theta$  is the threshold and  $k > 0$  is the gain of the transfer function. It has been argued (Douglas, Koch, Mahowald, Martin, & Suarez 1995) that this transfer function is more appropriate, because cortical neurons rarely operate close to saturation, despite strong recurrent excitation. This indicates that the upper saturation may not be involved in the actual computations of the recurrent network, since the activation is dynamically bounded due to the effective net interactions. Another important field for non-saturating linear threshold transfer functions are winner-take-all (WTA) networks (Lippmann 1987; Hahnloser 1998; Ritter 1990; Wersing, Steil, & Ritter 2000), where the lack of an upper saturation is necessary to exclude spurious ambiguous states.

Some properties of neural networks with piecewise linear transfer functions have been investigated since. Feng & Haderl (1996) and Kuhn & Loehwen (1987) proved conditions which ensure the uniqueness of a stationary state. Feng, Pan, & Roychowdhury (1996) analyzed the general fixed point structure of networks with piecewise linear sigmoidal functions and a convergence condition for asynchronous update procedures was given in (Feng 1997).

Hahnloser (1998) considered non-saturating linear threshold (LT) neurons for winner-take-all (WTA) networks and discussed their piecewise linear analysis. Wersing, Steil, & Ritter (2000) showed that complex perceptual grouping tasks can be accomplished in an LT network composed of competitive layers, where the overall contextual modulation can be enhanced by operating the model close to the stability limits. Recently Hahnloser, Sarpeshkar, Mahowald, Douglas, & Seung (2000) demonstrated an efficient silicon design for LT networks and discussed the coexistence of analogue amplification and digital selection in their circuit. They showed that the stability analysis for symmetric networks can be reduced to the discussion of sets which contain the active neurons, where an active superset of an unstable attractor is also unstable and a subset of a stable attractor is also stable. They also stated a condition for non-divergence of symmetric systems, which, however, depends on eigenvalues of all possible subsystems, and is therefore difficult to evaluate for complex systems.

To summarize, in spite of their wide application in biological and cortically inspired circuit models, there has been paid little attention so far to the general analysis of the conditions which ensure non-divergence in linear threshold networks. Therefore, this paper gives an in-depth analysis of this issue.

## 2 Monostable and Multistable Dynamics

In the following we only consider the case where  $k_i = 1$  and  $\theta_i = 0$ , since any network can be reduced to this form by an affine transformation of the activities that leaves the boundedness property of the dynamics invariant<sup>1</sup>. The dynamics is then given by

$$\dot{x}_i = -x_i + \sigma\left(\sum_j w_{ij}x_j + h_i\right), \quad (5)$$

where the transfer function is of the form  $\sigma(x) = \max(0, x)$ .

The standard approach to obtain a non-diverging dynamics (5) in the general case of nonsymmetric weights is to choose the combined gains of the transfer function and weights sufficiently small (see Steil 1999 for a review). A simple example is the condition given by Hirsch (1989) which is based on the property that all eigenvalues of the symmetrical parts of the Jacobians of the vector field in (5) must be negative. This gives for (5) the conditions

$$w_{ii} + \frac{1}{2} \sum_{j \neq i} |w_{ij}| + |w_{ji}| < 1 \quad \text{for all } i, \quad (6)$$

which poses a limit on the sum of the magnitudes of the incoming and outgoing weights of each neuron  $i$ , compared to its self-coupling weight  $w_{ii}$ . These conditions, however, imply *global* asymptotic stability, i.e. convergence to a single and unique attractor, and can be therefore only employed if the application requires a monostable dynamical behavior (see Figure 2). Nevertheless, there are many applications where *multistable* dynamics is essential to the model's operation. One example are WTA networks, where, depending on the external input (or the initial value) only the neuron with strongest input (or highest initial value) should remain active. If arbitrarily small differences can be amplified by this competitive process, then the system is said to exhibit symmetry-breaking behavior. In

<sup>1</sup>If  $\sigma_i(x_i) = k_i \max(0, x_i - \theta_i)$ , the transformation is given by  $x'_i = \sqrt{k_i}(x_i - \theta_i)$  and the corresponding coefficients are  $w'_{ij} = \sqrt{k_i k_j} w_{ij}$  and  $h'_i = (h_i - \theta_i)/\sqrt{k_i}$ . Note also that for  $w_{ij}$  being invertible, (5) is dynamically equivalent to  $\dot{m}_i = -m_i + \sum_j w_{ij} \sigma(m_j) + h_i$ , where  $m_i = \sum_j w_{ij} x_j + h_i$ .

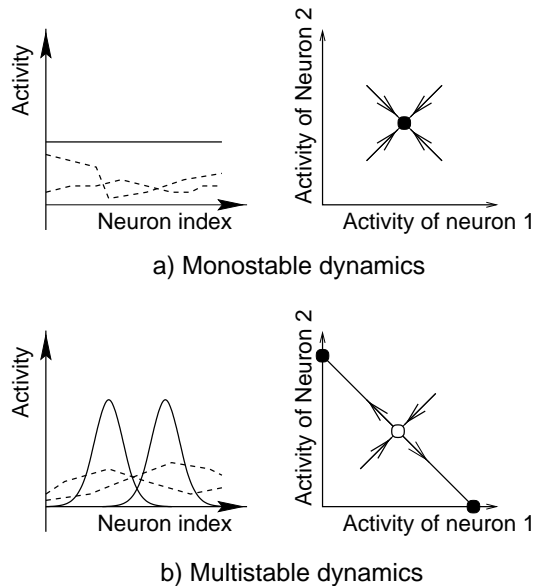


Figure 2: Difference between monostable and multistable dynamics. In a monostable system a) different initial activations (dashed lines) converge to the same and unique attractor (straight line on the left and filled circle on the right). In a multistable system b) many attractors can coexist and the dynamics also has unstable fixed points (open circle) which allow for unstable modes and symmetry breaking.

the piecewise linear LT systems this is only possible if there are unstable directions in the state space, corresponding to local Jacobians with eigenvalues with positive real parts. The conditions that we present in the following allow for this multistable behavior.

### 3 Conditions for Non-Divergence

#### 3.1 Nonsymmetric Networks

Hahnloser (1998) has proved a condition for non-divergence of LT networks that is based on a principle of global inhibition. For a given excitatory weight matrix with components  $a_{ij} \geq 0$ , he assumes a global inhibition with a strength equal to the sum over all excitatory weights diverging from a neuron. This leads to an effective weight matrix  $W$  of the form

$$w_{ij} = a_{ij} - \sum_k a_{kj}, \quad \text{for all } i, j. \quad (7)$$

Although the result is rather restrictive, and global all-to-all interactions are prohibitive in large networks, it is an example for a condition which does not imply monostability<sup>2</sup> and can be applied to some WTA networks (Hahnloser 1998).

In the following we prove a more general criterion, which shows that in fact local inhibition is sufficient to achieve boundedness of the LT dynamics. The condition is based on a

<sup>2</sup>Suppose the matrix  $A$  with components  $a_{ij}$  is symmetric positive definite and has an eigenvector of  $1$ 's. Since the column-constant shift in (7) affects only this eigenvalue, the resulting matrix  $W$  still has positive eigenvalues, causing unstable dynamical modes.

reduced connection matrix  $W^+$  with entries

$$w_{ij}^+ = \delta_{ij}w_{ij} + (1 - \delta_{ij}) \max(0, w_{ij}) \quad (8)$$

which is constructed from the diagonal and the positive off-diagonal entries of  $W$ , and where the Kronecker delta is defined by  $\delta_{ij} = 1$  if  $i = j$  and  $\delta_{ij} = 0$  if  $i \neq j$ .

**Theorem 1.** *If there exists a vector  $\hat{\mathbf{v}}$  with positive components  $\hat{v}_i > 0$  satisfying*

$$\sum_j w_{ij}^+ \hat{v}_j < \hat{v}_i \quad \text{for all } i, \quad (9)$$

*then the LT dynamics  $\dot{x}_i = -x_i + \sigma(\sum_j w_{ij}x_j + h_i)$  is bounded: If  $0 \leq x_i(0) \leq c\hat{v}_i$  for all  $i$ , where  $c > 0$  is sufficiently large to satisfy  $c(\hat{v}_i - \sum_j w_{ij}^+ \hat{v}_j) > h_i$  for all  $i$ , then  $0 \leq x_i(t) \leq c\hat{v}_i$  for all  $i$  and  $t > 0$ .*

*Proof.* First note that initially nonnegative activities remain nonnegative, since  $\dot{x}_i \geq 0$  for  $x_i = 0$ , and therefore the activity variables cannot cross zero to obtain negative values (see footnote below). The boundedness is proved by constructing a ‘‘box’’ in the positive domain which cannot be left by the dynamics. The  $i$ -th face of the box is defined by

$$x_i = c\hat{v}_i, \quad x_{j \neq i} \leq c\hat{v}_j. \quad (10)$$

Now we have either  $\dot{x}_i = -x_i \leq 0$  or

$$\dot{x}_i = -x_i + \sum_j w_{ij}x_j + h_i \quad (11)$$

$$\leq -x_i + \sum_j w_{ij}^+ x_j + h_i. \quad (12)$$

Inserting (10) and due to the positiveness of both activities and  $\hat{\mathbf{v}}$  we obtain (note that for the diagonal term  $w_{ii}x_i = w_{ii}c\hat{v}_i = w_{ii}^+c\hat{v}_i$ , since we sit on face  $i$ ):

$$\dot{x}_i \leq -x_i + \sum_j w_{ij}^+ c\hat{v}_j + h_i \quad (13)$$

$$= c\left(\sum_j w_{ij}^+ \hat{v}_j - \hat{v}_i\right) + h_i. \quad (14)$$

Therefore, if we choose  $c > 0$  sufficiently large such that  $c > \max_i(h_i/(\hat{v}_i - \sum_j w_{ij}^+ \hat{v}_j))$  for all  $i$ , then  $\dot{x}_i \leq 0$ . Thus the vector field defining the dynamics is pointing inwards on all sides of the box, and the neural activities are confined to this bounded region of the state space<sup>3</sup>.  $\square$

Using the theory of nonnegative matrices<sup>4</sup> (Berman & Plemmons 1979), condition (9) may be expressed in terms of eigenvalues of the matrix  $W^+$ . It is known that the maximal real part  $\text{Re}(\lambda_{\max})$  of the eigenvalues of  $W^+$  is an eigenvalue itself with corresponding nonnegative eigenvector  $\hat{\mathbf{v}}$ ,  $\hat{v}_i \geq 0$  for all  $i$ . Condition (9) is then equivalent to  $\lambda_{\max}\{W^+\} < 1$ . This shows that condition (9) is sharp if the eigenvector  $\hat{\mathbf{v}}$  is positive and  $W = W^+$  (no off-diagonal inhibitory weights). In fact, if  $W$  has an eigenvalue  $\lambda > 1$  with positive

<sup>3</sup>In dynamical systems theory this is called a *subtangent condition* (see Amann 1990).

<sup>4</sup>The results can be applied regardless of the sign of the diagonal matrix elements.

eigenvector then the dynamics obviously tends to infinity after initialization with a suitable multiple of this eigenvector.

Another well known characterization of condition (9) (see Berman & Plemmons 1979) states that the linear system

$$\hat{v}_i - \sum_j w_{ij}^+ \hat{v}_j = 1 \quad (15)$$

has a unique solution which has positive components  $\hat{v}_i > 0$ . Thus condition (9) can be tested by solving this system. Finally we notice that the special case  $\hat{v}_i = 1$  for all  $i$  in Theorem 1 yields an especially simple criterion

**Corollary 1.** *If the self-coupling weights satisfy  $w_{ii} < 1 - \sum_{j \neq i} w_{ij}^+$  for all  $i$  then the LT dynamics is bounded.*

If we compare this to the global inhibition principle of Hahnloser, expressed in (7), we see that starting from the same  $a_{ij}$ , local inhibition mediated by a self-inhibitory term is sufficient for non-divergence with

$$w_{ij} = a_{ij} - \delta_{ij} \sum_k a_{ik}. \quad (16)$$

Since self-interacting synapses (autapses) are rare in cortical circuits, the local inhibition should be mediated by local interneurons in a biologically more realistic setting. This inhibition, however, must be sufficiently fast to avoid a qualitative change in the dynamics (Li & Dayan 1999).

The conditions of Corollary 1 can be interpreted as a measure of *inhibitory* diagonal dominance, which is weaker than the condition of equation (6) and thus allows multistable dynamics. This is complementary to networks with saturating nonlinearities, where *excitatory* diagonal dominance (Greenberg 1988) can be used to enforce multistability by a locally divergent dynamics which drives the system into the corners of the bounded state space.

### 3.2 Symmetric Networks

In the following we assume that the weights are symmetric with  $w_{ij} = w_{ji}$  for all  $i, j$ . In this case, an additional criterion can be given that allows to improve the stability margins by taking into account also the non-local inhibitory contribution. The result is based on the energy function of the system which can be constructed for symmetric weights.

**Theorem 2.** *The LT dynamics is bounded if there exists a symmetric matrix  $\hat{W}$  with  $\hat{w}_{ij} \geq w_{ij}$  for all  $i, j$ , for which  $\lambda_{\max}\{\hat{W}\} < 1$ .*

*Proof.* The LT dynamics has an energy function of the form

$$E = -\frac{1}{2} \sum_{ij} (w_{ij} - \delta_{ij}) x_i x_j - \sum_j h_j x_j. \quad (17)$$

If we take  $E_i = \partial E / \partial x_i = \sum_j w_{ij} x_j - x_i + h_i$  then we see that  $E$  is nonincreasing under the dynamics

$$\dot{E} = \sum_j E_j \dot{x}_j = \sum_j E_j (-x_j + \sigma(-E_j + x_j)) \leq 0. \quad (18)$$

Since  $x_i \geq 0$  we simply obtain

$$E = -\frac{1}{2} \sum_{ij} (w_{ij} - \delta_{ij}) x_i x_j - \sum_j h_j x_j \quad (19)$$

$$\geq -\frac{1}{2} \sum_{ij} (\hat{w}_{ij} - \delta_{ij}) x_i x_j - \sum_j h_j x_j \quad (20)$$

$$\geq -\frac{1}{2} (\lambda_{\max}\{\hat{W}\} - 1) \|\mathbf{x}\|^2 - \sum_j h_j x_j \rightarrow \infty, \quad (21)$$

for  $\mathbf{x} \rightarrow \infty$ . Since  $E$  is bounded from below and  $E \rightarrow \infty$  for  $\mathbf{x} \rightarrow \infty$  according to (21),  $E$  is a global Lyapunov function and the dynamics must be bounded and converges to the set of equilibria (Hirsch 1989).  $\square$

The matrix  $W^+$  used in Theorem 1 provides an example for a possible choice of  $\hat{W}$  with nonnegative off-diagonal elements. The components of  $\hat{W}$ , however, need not be positive. In the following we give an argument based on linear perturbation theory, which shows that negative entries can be used to infer larger stability margins.

Suppose the weight matrix  $W$  is decomposed into the excitatory part  $W^+$  and an inhibitory off-diagonal part  $W^-$  according to  $w_{ij} = w_{ij}^+ + \alpha w_{ij}^-$ , where  $\alpha$  controls the strength of the off-diagonal inhibitory part. The condition of Theorem 1 amounts to choosing  $\alpha = 0$ , taking  $\hat{W} = W^+$  and neglecting the off-diagonal inhibition. The eigenvector of  $W^+$  with largest eigenvalue  $\lambda_{\max}$  has nonnegative components  $\hat{v}_i \geq 0$ . According to linear perturbation theory, the eigenvalue of the dominant eigenvector changes to first order in  $\alpha$  by  $\lambda'_{\max} = \lambda_{\max} + \alpha \hat{v}^T W^- \hat{v}$ . Since  $W^-$  has only nonpositive entries,  $\lambda'_{\max} < \lambda_{\max}$ , and by raising  $\alpha$  from zero the stability margin is increased by a decrease of the maximal eigenvalue. Since this, however, may in turn increase other eigenvalues for which then  $\lambda_i > 1$ , the increase in stability can be inferred only for small values of  $\alpha$ . Note that although  $\hat{W}$  corresponds to a monostable network,  $W$  can correspond to a multistable network, since  $\lambda_{\max}\{W\} > \lambda_{\max}\{\hat{W}\}$  for small  $\alpha$ .

What happens if the strength of the non-local inhibition is increased to large values of  $\alpha$ ? For such a strong inhibition, only pools of neurons can remain active, which share only excitatory interactions. Since then only these pools contribute to the dynamics, we can reconsider Theorem 1, applied to all possible active pools. Since now the number of excitatory interactions which must be considered is smaller, strong lateral inhibition has the expected effect of decreasing the local self-inhibition that is necessary to avoid runaway excitation.

## 4 Examples

### 4.1 A Simple Example

Let us consider an example network consisting of four neurons, which is simple enough to allow a complete analytic discussion of the different parameter regimes of the presented conditions for non-divergence. The interaction of neural activities  $x_i$ ,  $i = 1, 2, 3, 4$  (see



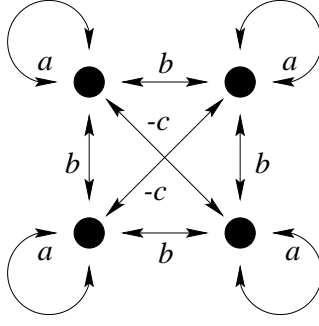


Figure 3: A simple network consisting of four neurons, which is a simple example for a lateral inhibition network with local cooperation.

Figure 3) is given by

$$W = \begin{pmatrix} a & b & -c & b \\ b & a & b & -c \\ -c & b & a & b \\ b & -c & b & a \end{pmatrix}. \quad (22)$$

The network is a simplified version of the lateral inhibition network with closed boundary conditions considered by Ben-Yishai et al. which is discussed in the next section. The parameter  $b \geq 0$  characterizes the cooperation between neighbouring neurons,  $c \geq 0$  implements inhibition between diagonally opposing neurons, and  $a$  is the self-coupling weight. The circulant matrix  $W$  has the corresponding eigenvectors  $\mathbf{v}_i$  and eigenvalues  $\lambda_i$ :

$$\begin{aligned} \mathbf{v}_1 &= (1, 1, 1, 1), & \lambda_1 &= a + 2b - c, \\ \mathbf{v}_2 &= (1, 0, -1, 0), & \lambda_2 &= a + c, \\ \mathbf{v}_3 &= (0, 1, 0, -1), & \lambda_3 &= a + c, \\ \mathbf{v}_4 &= (1, -1, 1, -1), & \lambda_4 &= a - 2b - c. \end{aligned} \quad (23)$$

We now address the following question: Given  $b \geq 0$  and  $c \geq 0$ , how small must be the self-coupling  $a$  to keep non-divergence of the dynamics?

If we apply Corollary 1, which neglects all off-diagonal inhibitory contributions, we obtain the non-divergence condition

$$a < 1 - 2b, \quad (24)$$

which is not dependent on  $c$ .

According to Theorem 2 we can infer non-divergence of a system with parameters  $(a, b, c)$  from the non-divergence of a system with parameters  $(a', b', c')$ , where  $a < a'$ ,  $b < b'$ , and  $c > c'$ . The maximum eigenvalue is given by

$$\lambda_{\max} = \begin{cases} a + 2b - c & \text{if } b \geq c, \\ a + c & \text{if } b < c. \end{cases} \quad (25)$$

The condition  $\lambda_{\max} < 1$  then results in

$$i) \quad a < 1 - 2b + c \quad \text{for } b > c, \quad (26)$$

$$ii) \quad a < 1 - c \quad \text{for } b \leq c. \quad (27)$$

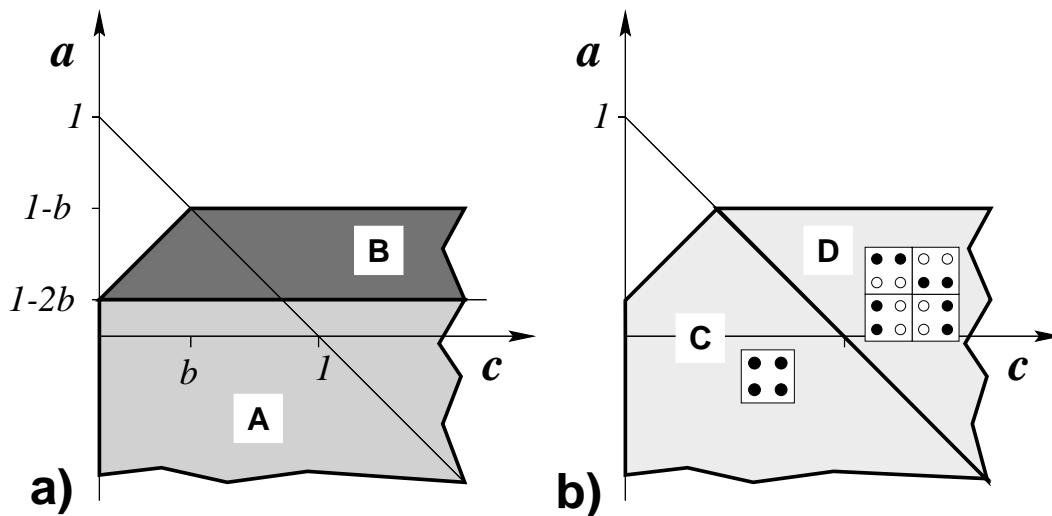


Figure 4: Non-divergence domains and regions of mono- and multistable dynamics. a) shows the non-divergent domains depending on parameters  $a$  and inhibition strength  $c > 0$ . Region A is bounded according to Theorem 1. Theorem 2 covers boundedness in the additional region B. Note that due to Th. 2 the non-divergence for a pair of parameters  $(a, c)$  implies non-divergence for all  $(a', c')$  with  $a' < a$  and  $c' > c$ . b) In region C the monostable attractor has (approximately) equal activity in all neurons. In region D the system is multistable and four attractors of pairwise active neurons coexist.

Case *i*) improves the stability margin towards larger values of  $a$  with a maximum when  $c = b$  to  $a < 1 - b$ . Case *ii*) offers no advantage since  $b \leq c$  implies that  $1 - c < 1 - 2b$ . The domains of non-divergence for large  $c$  and small  $a$  which can be obtained from these inequalities are shown in Figure 4. The diagrams demonstrate, that by taking into account off-diagonal inhibitory contributions Theorem 2 covers a larger area of the parameter space.

Figure 4b shows the domains of mono- and multistable dynamics, if the system obtains a weak input  $h_i = m + \epsilon_i$ , where  $m$  is constant and  $\epsilon_i \ll m$  is a small random modulation. Only in the multistable regime the system is capable of symmetry breaking, where the dynamics causes a nonlinear amplification of small input differences.

## 4.2 A Model for Orientation Selectivity

Ben-Yishai, Lev Bar-Or, & Sompolinsky (1995) have considered a one-dimensional continuum model for orientation selectivity which is of the form

$$\dot{x}(\phi) = -x(\phi) + \sigma(F(\phi)), \quad -\pi/2 \leq \phi \leq \pi/2 \quad (28)$$

where

$$F(\phi) = \frac{1}{\pi} \int_{-\pi/2}^{\pi/2} f(\phi - \phi') x(\phi') d\phi' + h^{\text{ext}}(\phi - \phi_0) \quad (29)$$

with  $f(\psi) = -J_0 + J_2 \cos(2\psi)$  and

$$\sigma(x) = \begin{cases} 0 & \text{if } x \leq 1, \\ k(x-1) & \text{if } 1 < x < k^{-1} + 1, \\ 1 & \text{if } x \geq k^{-1} + 1. \end{cases} \quad (30)$$

The parameters  $J_0 > 0$  and  $J_2 > 0$  characterize on ‘‘On-Center-Off-Surround’’ interaction between a continuum of neurons in a 1-D orientation space with closed boundary conditions. The external input is of the form  $h^{\text{ext}}(\phi) = c(1 - \eta + \eta \cos(2\phi))$ , where  $c$  denotes the contrast and  $\eta \ll 1$  is the amplitude of a weakly biased input stimulus. The parameter  $k$  is the gain of the transfer function. Of particular interest is a region in the  $J_0 \times J_2$  space, where the model exhibits a so-called marginal phase. In this phase the model spontaneously generates orientation selectivity, by recurrently amplifying arbitrarily small anisotropies in the input for  $\eta \rightarrow 0$ . As the results of Ben-Yishai et.al. demonstrate, this occurs in a regime where the upper saturation of the transfer function is not reached, which corresponds to the case of non-divergent linear threshold dynamics. Therefore, we use in the following the presented conditions for non-divergence with multi-stable dynamics to identify this regime in the parameter space.

The linear integral operator in (29) has two continuous eigenfunctions  $x_0, x_2$  with eigenvalues  $\lambda_0, \lambda_2$ , where

$$\begin{aligned} x_0(\phi) &= 1, & \lambda_0 &= -1 - J_0 k, \\ x_2(\phi) &= \cos(2\phi), & \lambda_2 &= -1 + \frac{k}{2} J_2. \end{aligned} \quad (31)$$

As Ben-Yishai et.al. have stated, a necessary condition for the system being in the marginal phase is that  $J_2 > 2k^{-1}$ , which in turn makes the  $x_2$ -mode unstable with  $\lambda_2 > 0$ .

We can now use Corollary 1 to infer a condition which ensures the non-divergence of the system, even if unstable modes are present. The straightforward generalization to the continuum case is given by the condition

$$0 > -1 + \frac{k}{\pi} \int_{-\pi/2}^{\pi/2} \max(0, f(\phi)) d\phi. \quad (32)$$

This can be geometrically interpreted as the condition that the area under the positive part of  $f$  is smaller than  $\pi/k$ . If  $J_0 > J_2$  the lateral interaction is completely negative, and the condition is trivially satisfied. Due to the absence of excitatory interactions the activity settles down to zero for a weak unspecific input. For  $J_0^c < J_0 < J_2$  the area can be estimated from above by the larger rectangle with height  $J_2 - J_0$  and width between the zeroes of  $f$  (see Figure 5). Using the approximations  $J_0 = J_2 - \epsilon$  with  $\epsilon$  small and  $\epsilon \arccos(1 - a\epsilon) \approx \sqrt{2a}\epsilon^{3/2}$  we obtain the condition

$$J_0 > J_0^c = J_2 - \frac{\pi^{2/3}}{2^{1/3}} \frac{1}{k^{2/3}} J_2^{1/3}. \quad (33)$$

This condition can be used to effectively identify the region in the  $J_0 \times J_2$  space, where the lateral interaction is capable of recurrently amplifying small fluctuations in the activity distribution, without saturating at the upper saturation limit (see Figure 5). This leads to

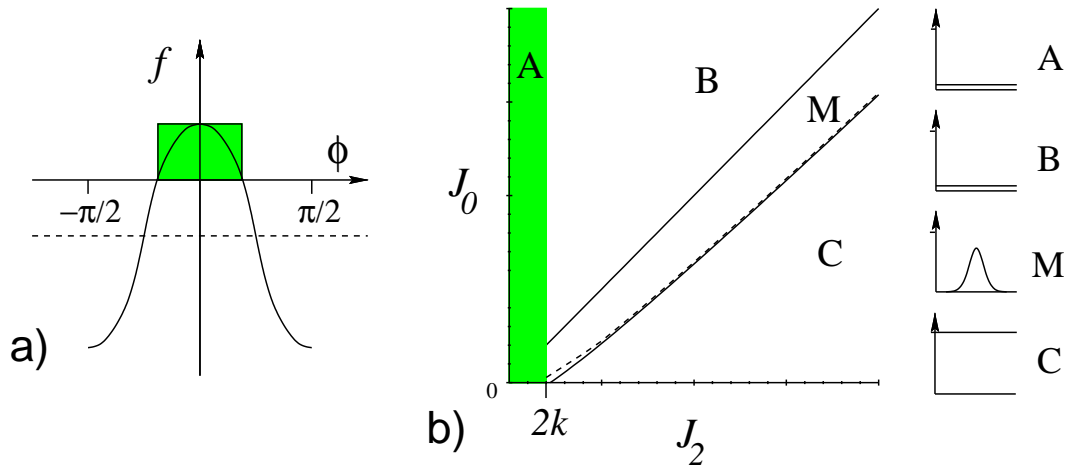


Figure 5: Angular On-Center-Off-Surround interaction and phase diagram of the orientation selectivity model by Ben-Yishai et.al. The shaded rectangle in a) is an estimate to the sum over excitatory interactions. With this estimate a phase diagram shown in b) can be obtained. In region A (shaded area) all eigenvalues are negative and the system always converges to an isotropic state. In region B the activity remains small due to large inhibition with  $J_0 > J_2$  with the same unique isotropic attractor. Region M is the marginal phase, where a sharp and peaked response can be generated from isotropic input. In region C the network saturates due to strong excitation. The border between M and C is given by the non-divergence condition, the bold straight line gives the theoretical prediction for the critical inhibition  $J_0^c$ , while the dashed line was obtained from a simulation of the system. Within region M there is a gradual increase of peak height towards this boundary.

localized peaks with increasing amplification for  $J_0 \rightarrow J_0^c$ . For  $J_0 > J_0^c$  the network runs into saturation. A comparison with a numerical simulation of the dynamics shows that (33) captures the border between maximal amplification and saturation rather precisely.

## 5 Conclusion

The two presented conditions allow to ensure the non-divergence of linear threshold networks without restricting them to be monostable. Although, at least for symmetric systems, the second condition covers a larger parameter regime, it requires the knowledge of eigenvalues which may not be readily available. On the contrary, the first condition can be easily applied by choosing appropriate local (self-inhibitory) interactions matching the sum over excitatory weights. Although the conditions are strictly speaking only sufficient and not necessary, our discussion of the orientation-selectivity model shows that they can be used to effectively approach the regime of critical amplification in LT networks, where small stimulus contrasts can be largely enhanced through dynamical symmetry breaking.

Linear threshold networks provide a promising step towards neuromorphic analogue circuit design and were shown to accomplish complex tasks like visual perceptual grouping. Non-divergence is a crucial property for the coexistence of analogue amplification and digital selection in these circuits. Our results largely extend earlier conditions, which required either global inhibitory interactions or symmetry of the weights, and open the door for more complex models and circuit designs.

## Acknowledgements

The authors would like to thank J. J. Steil for helpful discussions. We also thank the anonymous referees for comments which improved the clarity of the presentation. H. Wersing was supported by DFG grant GK-231.

## References

- Adorjan, P., Levitt, J. B., Lund, J. S., & Obermayer, K. (1999). A model for the intracortical origin of orientation preference and tuning in macaque striate cortex. *Visual Neuroscience*, 16, 303–318.
- Amann, H. (1990). *Ordinary Differential Equations*. De Gruyter.
- Bauer, U., Scholz, M., Levitt, J. B., Obermayer, K., & Lund, J. S. (1999). A biologically-based neural network model for geniculocortical information transfer in the primate visual system. *Vision Research*, 39, 613–629.
- Ben-Yishai, R., Lev Bar-Or, R., & Sompolinsky, H. (1995). Theory of orientation tuning in visual cortex. *Proc. Nat. Acad. Sci. USA*, 92, 3844–3848.
- Berman, A. & Plemmons, R. J. (1979). *Nonnegative matrices in the mathematical sciences*. Academic Press.
- Douglas, R., Koch, C., Mahowald, M., Martin, K., & Suarez, H. (1995). Recurrent excitation in neocortical circuits. *Science*, 269, 981–985.
- Douglas, R., Mahowald, M., & Martin, K. (1994). Hybrid analog-digital architectures for neuromorphic systems. In *IEEE International Conference on Neural Networks*, Volume 3, Orlando, pp. 1848–1853.
- Feng, J. (1997). Lyapunov functions for neural nets with nondifferentiable input-output characteristics. *Neural Computation* 9(1), 43–49.
- Feng, J. & Hadel, K. P. (1996). Qualitative behaviour of some simple networks. *Journal of Physics A*, 29, 5019–5033.
- Feng, J., Pan, H., & Roychowdhury, V. P. (1996). On neurodynamics with limiter function and Linsker's development model. *Neural Computation*, 8, 1003–1019.
- Greenberg, H. J. (1988). Equilibria of the Brain-State-in-a-Box neural model. *Neur. Netw.*, 1, 323–324.
- Hahnloser, R., Sarpeshkar, R., Mahowald, M. A., Douglas, R. J., & Seung, H. S. (2000). Digital selection and analogue amplification coexist in a cortex-inspired silicon circuit. *Nature*, 405, 947–951.
- Hahnloser, R. L. T. (1998). On the piecewise analysis of linear threshold neurons. *Neur. Netw.*, 11, 691–697.
- Hartline, H. K. & Ratliff, F. (1958). Spatial summation of inhibitory influence in the eye of limulus and the mutual interaction of receptor units. *Journal of General Physiology*, 41, 1049–1066.
- Hirsch, M. W. (1989). Convergent activation dynamics in continuous time networks. *Neur. Netw.*, 2, 331–349.

- Kuhn, D. & Loehwen, R. (1987). Piecewise affine bijections of  $R^n$  and the equation  $Sx_+ - Tx_- = y$ . *Linear Algebra and its Appl.*, 96, 109–129.
- Li, Z. & Dayan, P. (1999). Computational differences between asymmetrical and symmetrical networks. *Network*, 10, 59–77.
- Lippmann, R. (1987). An introduction to computing with neural nets. *IEEE ASSP Mag.*, 4, 4–22.
- Ritter, H. (1990). A spatial approach to feature linking. In *Proc. Int. Neur. Netw. Conf. Paris Vol.2*, pp. 898–901.
- Salinas, A. & Abbott, L. F. (1996). A model of multiplicative responses in parietal cortex. *Proc. Nat. Acad. Sci. USA*, 93, 11956–11961.
- Steil, J. J. (1999). *Input-Output Stability of Recurrent Neural Networks*. Ph. D. thesis, University of Bielefeld, Faculty of Technology.
- von der Malsburg, C. (1973). Self-organization of orientation sensitive cells in the striate cortex. *Kybernetik*, 14, 85–100.
- Wersing, H., Steil, J. J., & Ritter, H. (2000). A competitive layer model for feature binding and sensory segmentation. *Neural Computation* (in press).