# Stable and unstable gradients near identical recurrent computation

**Anonymous Author(s)**
Affiliation
Address
email

## 1 Bifurcation analysis of the line attractors

### 1.1 Unbounded line attractor

The parameters:

$$W_{\text{in}} = \alpha \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix}, \; W = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \; W_{\text{out}} = \frac{1}{2\alpha} \begin{pmatrix} 1 & 1 \end{pmatrix}, \; b_{\text{out}} = -\frac{\beta}{\alpha}. \tag{1}$$

The bias to the recurrent units is zero.

#### 1.1.1 Stabilty of the fixed points

We investigate how perturbations to the bounded line affect the Lyapunov spectrum. We calculate the eigenspectrum of the Jacobian:

$$\begin{aligned} \det[W' - (1+\lambda)\mathbb{I}] &= (\epsilon_{11} - 1 - \lambda)(\epsilon_{22} - 1 - \lambda) - (\epsilon_{12} + 1)(\epsilon_{21} + 1) \\ &= \lambda^2 - (2 + \epsilon_{11} + \epsilon_{22})\lambda - \epsilon_{11} - \epsilon_{22} + \epsilon_{11}\epsilon_{22} - \epsilon_{12} - \epsilon_{21} - \epsilon_{12}\epsilon_{21} \end{aligned}$$

Let $u = -(2 + \epsilon_{11} + \epsilon_{22})$ and $v = -\epsilon_{11} - \epsilon_{22} + \epsilon_{11}\epsilon_{22} - \epsilon_{12} - \epsilon_{21} - \epsilon_{12}\epsilon_{21}$

There are only two types of invariant set for the perturbations of the line attractor. Both have as invariant set a fixed point at the origin. What distinguishes them is that one type of perturbations lead to this fixed point being stable while the other one makes it unstable.

### 1.2 Bounded line attractor

Another implementation of a perfect integrator: one which is bounded.

**Input** Parameter that determines step size along line attractor $\delta << 1$. The size determines the maximum number of clicks as the difference between the two channels.

This pushes the input along the line “attractor" in two opposite directions, see below.

The parameters:

$$W_{\text{in}} = \alpha \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix}, \; W = \begin{pmatrix} 0 & -1 \\ -1 & 0 \end{pmatrix}, \; W_{\text{out}} = \frac{1}{2\alpha} \begin{pmatrix} 1 & -1 \end{pmatrix}, \; b = \beta \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \; b_{\text{out}} = 0. \tag{2}$$

Needs to be initialized at $\frac{\beta}{2}(1,1)$ for correct decoding, i.e., output projection

### 1.3 Stabilty of the fixed points

The Jacobian

$$J = -\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \tag{3}$$

We apply the perturbation

$$W' = \begin{pmatrix} 0 & -1 \\ -1 & 0 \end{pmatrix} + \epsilon \tag{4}$$

with

$$\epsilon = \begin{pmatrix} \epsilon_{11} & \epsilon_{12} \\ \epsilon_{21} & \epsilon_{22} \end{pmatrix} \tag{5}$$

The eigenvalues are computed as

$$\begin{aligned} \det[W' - (1+\lambda)\mathbb{I}] &= (\epsilon_{11} - 1 - \lambda)(\epsilon_{22} - 1 - \lambda) - (\epsilon_{12} - 1)(\epsilon_{21} - 1) \\ &= \lambda^2 - (2 + \epsilon_{11} + \epsilon_{22})\lambda - \epsilon_{11} - \epsilon_{22} + \epsilon_{11}\epsilon_{22} + \epsilon_{12} + \epsilon_{21} - \epsilon_{12}\epsilon_{21} \end{aligned}$$

Let $u = -(2 + \epsilon_{11} + \epsilon_{22})$ and $v = -\epsilon_{11} - \epsilon_{22} + \epsilon_{11}\epsilon_{22} + \epsilon_{12} + \epsilon_{21} - \epsilon_{12}\epsilon_{21}$

$$\lambda = \frac{-u \pm \sqrt{u^2 - 4v}}{2} \tag{6}$$

Case 1: $\mathrm{Re}(\sqrt{u^2 - 4v}) < u$, then $\lambda_{1,2} < 0$

Case 2: $\mathrm{Re}(\sqrt{u^2 - 4v}) > u$, then $\lambda_1 < 0$ and $\lambda_2 > 0$

Case 3: $v = 0$, then $\lambda = \frac{1}{2}(-u \pm u)$, i.e., $\lambda_1 = 0$ and $\lambda_2 = -u$

$$\epsilon_{11} = -\epsilon_{22} + \epsilon_{11}\epsilon_{22} + \epsilon_{12} + \epsilon_{21} - \epsilon_{12}\epsilon_{21} \tag{7}$$

We give some examples of the different types of perturbations to the bounded line attractor. The first type is when the invariant set is composed of a single fixed point, for example for the perturbation:

$$\epsilon = \frac{1}{10}\begin{pmatrix} -2 & 1 \\ 1 & -2 \end{pmatrix} \tag{8}$$

See Figure **??**, left upper.

The second type is when the invariant set is composed of three fixed points:

$$\epsilon = \frac{1}{10}\begin{pmatrix} 1 & -2 \\ -2 & 1 \end{pmatrix} \tag{9}$$

The third type is when the invariant set is composed of two fixed points, both with partial support.

$$b' = \frac{1}{10}\begin{pmatrix} 1 & -1 \end{pmatrix} \tag{10}$$

The fourth and final type is when the line attractor is maintained but rotated:

$$\epsilon = \frac{1}{20}\begin{pmatrix} 1 & 10 \\ 10 & 1 \end{pmatrix} \tag{11}$$

**Theorem 1.** *All perturbations of the bounded line attractor are of the types as listed above.*

*Proof.* We enumerate all possibilities for the dynamics of a ReLU activation network with two units. First of all, note that there can be no limit cycle or chaotic orbits.

Now, we look at the different possible systems with fixed points. There can be at most three fixed points [**?** , Corollary 5.3]. There has to be at least one fixed point, because the bias is non-zero.

General form (example):

$$\epsilon = \frac{1}{10}\begin{pmatrix} -2 & 1 \\ 1 & -2 \end{pmatrix} \tag{12}$$

One fixed point with full support:

In this case we can assume $W$ to be full rank.

$$\dot{x} = \mathrm{ReLU}\left[\begin{pmatrix} \epsilon_{11} & \epsilon_{12} \\ \epsilon_{21} & \epsilon_{22} \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{pmatrix} 1 \\ 1 \end{pmatrix}\right] - \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = 0$$

Note that $x > 0$ iff $z_1 \coloneqq \epsilon_{11}x_1 + (\epsilon_{12} - 1)x_2 - 1 > 0$. Similarly for $x_2 > 0$.

So for a fixed point with full support, we have

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = A^{-1}\begin{pmatrix} -1 \\ -1 \end{pmatrix} \tag{13}$$

with

$$A \coloneqq \begin{pmatrix} \epsilon_{11} - 1 & \epsilon_{12} - 1 \\ \epsilon_{21} - 1 & \epsilon_{22} - 1 \end{pmatrix}.$$

Note that it is not possible that $x_1 = 0 = x_2$.

Now define

$$B \coloneqq A^{-1} = \frac{1}{\det A}\begin{pmatrix} \epsilon_{22} - 1 & 1 - \epsilon_{12} \\ 1 - \epsilon_{21} & \epsilon_{11} - 1 \end{pmatrix}$$

with

$$\det A = \epsilon_{11}\epsilon_{22} - \epsilon_{11} - \epsilon_{22} - \epsilon_{12}\epsilon_{21} + \epsilon_{12} + \epsilon_{21}.$$

Hence, we have that $x_1, x_2 > 0$ if $B_{11} + B_{12} > 0$, $B_{21} + B_{22} > 0$ and $\det A > 0$ and $B_{11} + B_{12} < 0$, $B_{21} + B_{22} < 0$ and $\det A < 0$.

This can be satisfied in two ways, If $\det A > 0$, this is satisfied if $\epsilon_{22} > \epsilon_{12}$ and $\epsilon_{11} > \epsilon_{21}$, while if $\det A > 0$, this is satisfied if $\epsilon_{22} < \epsilon_{12}$ and $\epsilon_{11} < \epsilon_{21}$. This gives condition 1.

Finally, we investigate the condition that specify that there are fixed points with partial support. If $x_1 = 0$ then $(\epsilon_{22} - 1)x_2 + 1 = 0$ and $z_1 < 0$. From the equality, we get that $x_2 = \frac{1}{1-\epsilon_{22}}$. From the inequality, we get $(\epsilon_{12} - 1)x_2 + 1 \geq 0$, i.e. $\frac{1}{1-\epsilon_{12}} \geq x_2$. Hence,

$$\frac{1}{1 - \epsilon_{12}} \geq \frac{1}{1 - \epsilon_{22}}$$

and thus

$$\epsilon_{22} \leq \epsilon_{12}. \tag{14}$$

Similiarly to have a fixed point $x^*$ such that $x_2^* = 0$, we must have that

$$\epsilon_{11} \leq \epsilon_{21}. \tag{15}$$

Equation 14 and 15 together form condition 2.

If condition 1 is violated, but condition 2 is satisfied, there are two fixed points on the boundary of the admissible quadrant.

If condition 1 is violated, and only one of the subconditions of condition 2 is satisfied, there is a single fixed point on one of the axes.

If condition 2 is violated, there are three fixed points.

We now look at the possibility of the line attractor being preserved. This is the case if $v = 0$. It is not possible to have a line attractor with a fixed point off it for as there cannot be disjoint fixed points that are linearly dependent [1, Lemma 5.2]. $\square$

### 1.4 Structure of the parameter space

We check what proportion of the bifurcation parameter space is constituted with bifurcations of the type that result in three fixed points.

The conditions are

$$\begin{aligned} 0 &< \epsilon_{11}\epsilon_{22} - \epsilon_{11} - \epsilon_{22} - \epsilon_{12}\epsilon_{21} - \epsilon_{12} - \epsilon_{21}, \\ \epsilon_{22} &\leq \epsilon_{12}, \\ \epsilon_{11} &\leq \epsilon_{21}. \end{aligned}$$

We show that if

$$\begin{aligned} \epsilon_{22} &\leq \epsilon_{12}, \\ \epsilon_{11} &\leq \epsilon_{21}. \end{aligned}$$

then always

$$0 < \epsilon_{11}\epsilon_{22} - \epsilon_{11} - \epsilon_{22} - \epsilon_{12}\epsilon_{21} - \epsilon_{12} - \epsilon_{21}.$$

The only nontrivial cases are

## References

[1] Katherine Morrison, Anda Degeratu, Vladimir Itskov, and Carina Curto. Diversity of emergent dynamics in competitive threshold-linear networks.