Towards a Pseudo-Reaction-Diffusion Model for Turing Instability in Adversarial Learning

Litu Rout, Space Applications Centre, Indian Space Research Organisation

1 Objectives

Long after Turing's seminal **Reaction-Diffusion (RD)** model, the elegance of his fundamental equations alleviated much of the skepticism surrounding pattern formation. Interestingly, we observe Turing-like patterns in a system of neurons with adversarial interaction. In this study, we establish the following:

1. Involvement of Turing instability.
2. A *Pseudo-Reaction-Diffusion* model.
3. Symmetry and homogeneity.
4. Breakdown of symmetry and homogeneity.

2 Introduction

In this paper, we intend to demystify an interesting phenomenon: adversarial interaction between generator and discriminator creates non-homogeneous equilibrium by inducing Turing instability in a Pseudo-Reaction-Diffusion (PRD) model. This is in stark contrast to sole supervision. Thus we state our key observation:

A system in which a generator and a discriminator adversarially interact with each other exhibits Turing-like patterns in the hidden layer and top layer of the two layer generator network.

3 Preliminaries

Supervised Training:

$$\mathcal{L}_{sup}\left(\boldsymbol{U},\boldsymbol{V}\right)=\frac{1}{2}\sum_{p=1}^{n}\left\|\frac{1}{\sqrt{d_{out}m}}\boldsymbol{V}\sigma\left(\boldsymbol{U}\boldsymbol{x}_p\right)-\boldsymbol{y}_p\right\|_2^2=\frac{1}{2}\left\|\frac{1}{\sqrt{d_{out}m}}\boldsymbol{V}\sigma\left(\boldsymbol{U}\boldsymbol{X}\right)-\boldsymbol{Y}\right\|_F^2.$$

Regularized Adversarial Training:

$$\mathcal{L}_{aug}\left(\boldsymbol{U},\boldsymbol{V},\boldsymbol{W},\boldsymbol{a}\right)=\underbrace{\frac{1}{2}\left\|\frac{1}{\sqrt{d_{out}m}}\boldsymbol{V}\sigma\left(\boldsymbol{U}\boldsymbol{X}\right)-\boldsymbol{Y}\right\|_F^2}_{\mathcal{L}_{sup}}-\underbrace{\frac{1}{m\sqrt{d_{out}}}\sum_{p=1}^{n}\boldsymbol{a}^T\sigma\left(\boldsymbol{W}\boldsymbol{V}\sigma\left(\boldsymbol{U}\boldsymbol{x}_p\right)\right)}_{\mathcal{L}_{adv}}.$$

Learning Algorithm:

$$\frac{du_{jk}}{dt}=-\frac{\partial\mathcal{L}_{aug}\left(\boldsymbol{U}(t),\boldsymbol{V}(t),\boldsymbol{W}(t),\boldsymbol{a}(t)\right)}{\partial u_{jk}(t)},$$
$$\frac{dv_{ij}}{dt}=-\frac{\partial\mathcal{L}_{aug}\left(\boldsymbol{U}(t),\boldsymbol{V}(t),\boldsymbol{W}(t),\boldsymbol{a}(t)\right)}{\partial v_{ij}(t)}.$$

Pseudo-Reaction-Diffusion Model[1]:

$$\frac{d\boldsymbol{u}_j}{dt}=\mathfrak{R}_j^{\boldsymbol{u}}\left(\boldsymbol{u}_j,\boldsymbol{v}_j\right)+\mathfrak{D}_j^{\boldsymbol{u}}\left(\nabla^2\boldsymbol{u}_j\right),$$
$$\frac{d\boldsymbol{v}_j}{dt}=\mathfrak{R}_j^{\boldsymbol{v}}\left(\boldsymbol{u}_j,\boldsymbol{v}_j\right)+\mathfrak{D}_j^{\boldsymbol{v}}\left(\nabla^2\boldsymbol{v}_j\right).$$

4 Theoretical Analysis

(Informal) Theorem 1. (Symmetry and Homogeneity) *Suppose the necessary assumptions hold. We obtain the following with probability at least $1-\delta$:*

$$\left\|\boldsymbol{u}_j(t)-\boldsymbol{u}_j(0)\right\|_2\leq\mathcal{O}\left(\frac{n^{3/2}}{m^{1/2}\lambda_0\delta}\left(1-\exp\left(-\frac{\lambda_0}{2}t\right)\right)\right).$$

(Informal) Theorem 2. (Breakdown of Symmetry and Homogeneity) *If the required conditions are satisfied, then with probability at least $1-\delta$, we get*

$$\left\|\boldsymbol{u}_j(t)-\boldsymbol{u}_j(0)\right\|_2\leq\mathcal{O}\left(\frac{n^{3/2}}{\sqrt{m}\lambda_0\delta}\left(1-\exp\left(-\frac{\lambda_0}{2}t\right)\right)+\left(\frac{\mu\left(1+\kappa\sqrt{n}\right)}{\sqrt{m}}\right)t\right).$$

Analogous Bernoulli Differential Equation: Modeling Population Growth,

$$\frac{dP}{dt}=rP\left(1-\frac{P}{K}\right). \quad (1)$$

Modeling Regularized Adversarial Training,

$$\frac{d\psi}{dt}\leq r\psi^{1/2}\left(1-\frac{\psi^{1/2}}{K}\right). \quad (2)$$

5 Experimental Results

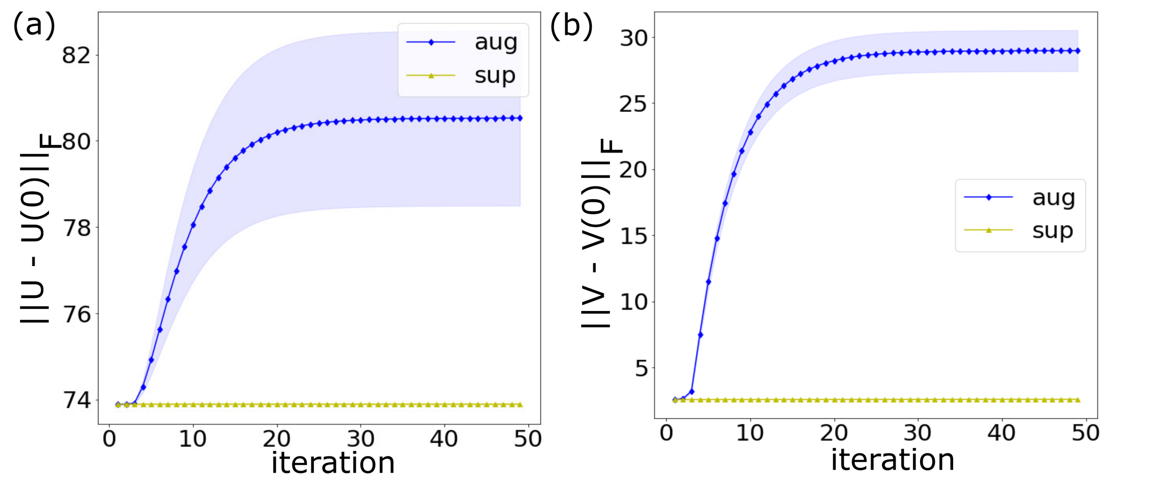

Figure 1: Distance from multiple initialization in the (a) hidden layer and (b) top layer on MNIST.

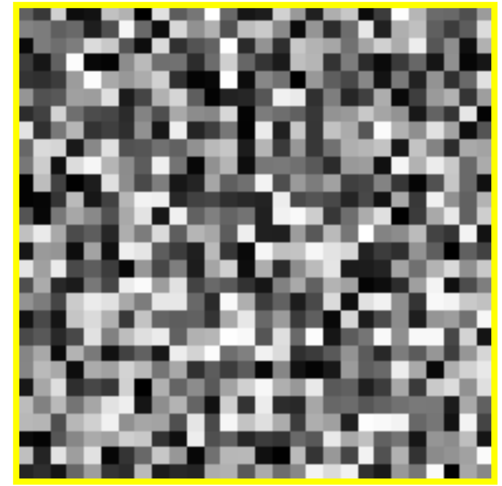

Figure 2: Input image used for the visualization of features in the hidden layer.

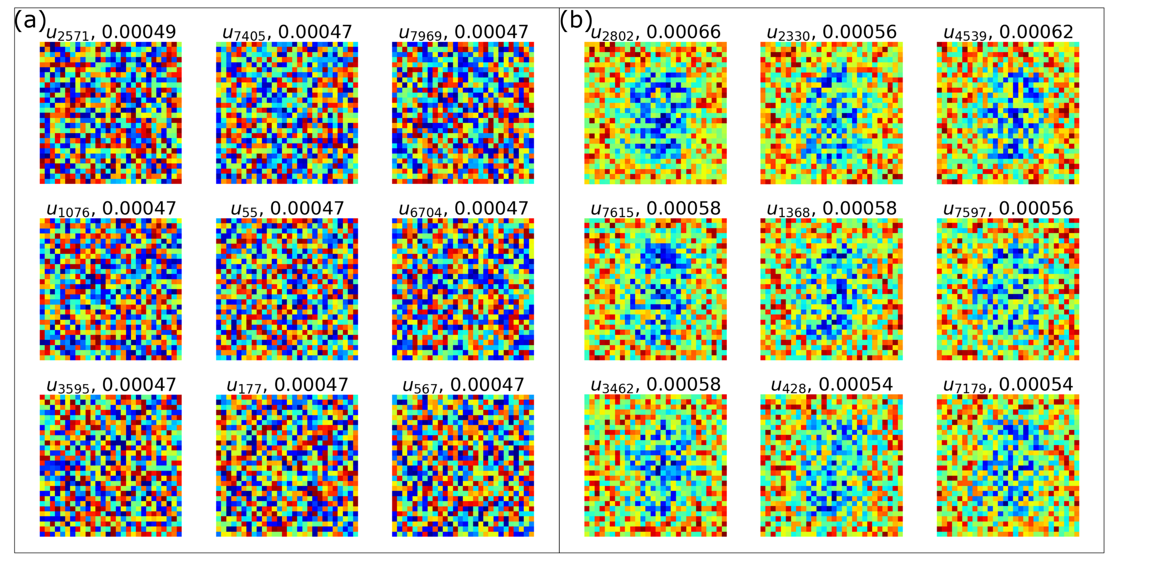

Figure 3: Hidden layer filters on MNIST. (a) Without Diffusion. (b) With Diffusion.

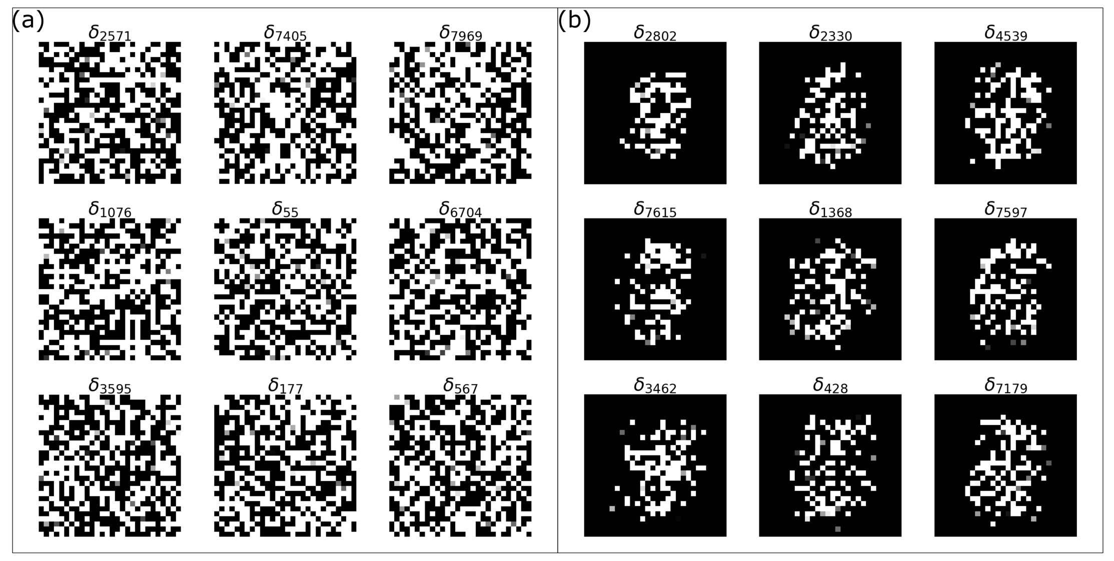

Figure 4: Visualization of features on MNIST. (a) Without Diffusion. (b) With Diffusion.

6 Turing Instability in Adversarial Learning

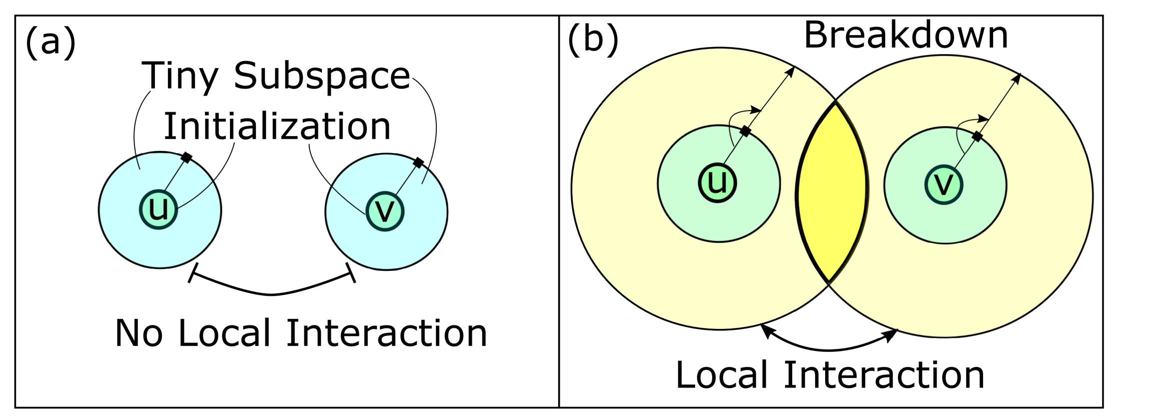

Figure 5: Breakdown of symmetry and homogeneity. (a) Without Diffusion. (b) With Diffusion.

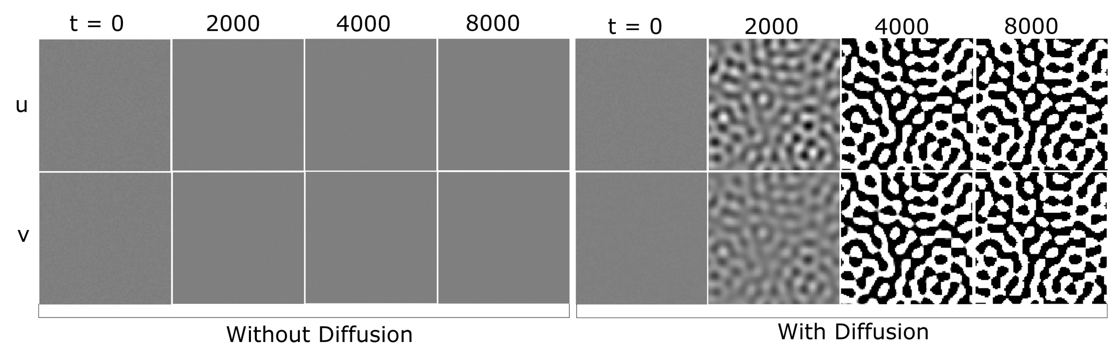

Figure 6: Turing pattern formation. The diffusible factors help break the symmetry and homogeneity.

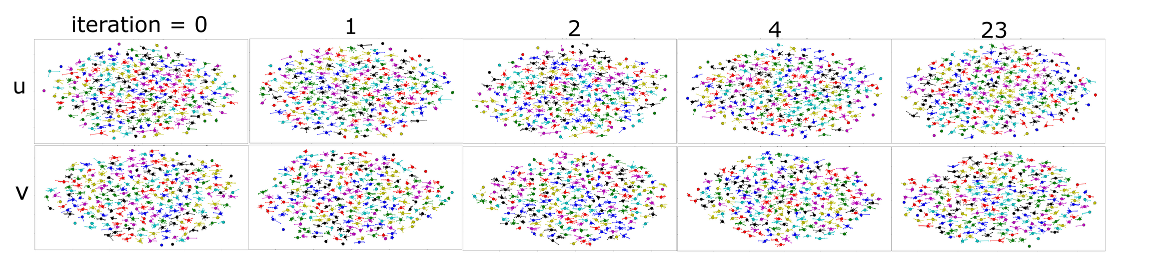

Figure 7: Pattern formation on synthetic data, $d_{in}=784$ **without Diffusion**.

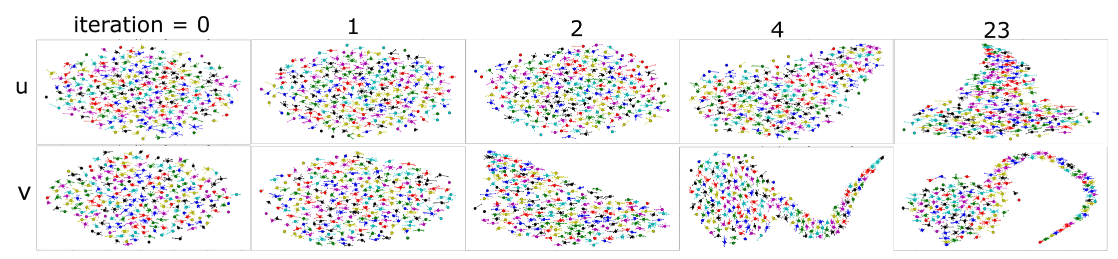

Figure 8: Pattern formation on synthetic data, $d_{in}=784$ **with Diffusion**.

Reference

[1] A.M. Turing. The chemical basis of morphogenesis. *Phil. Trans. of the Royal Soc. of London*, 1952.

7 Future Scope

Though diffusibility ensures more local interaction, it will certainly be interesting to synchronize neurons based on breakdown of symmetry and homogeneity in the future.

Contact Information

Website: https://liturout.github.io/
Email: liturout1997@gmail.com, lr@sac.isro.gov.in