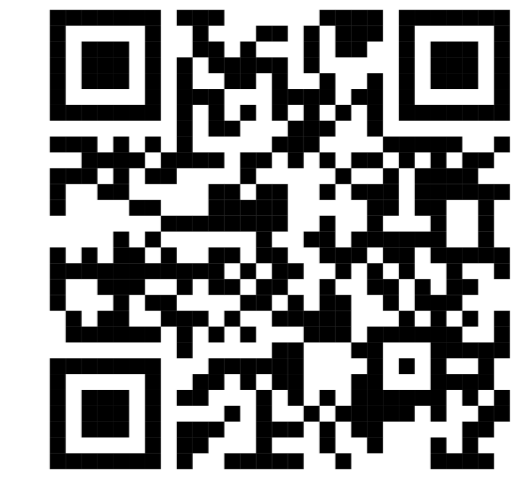


# Complex-valued Neurons Can Learn More but Slower than Real-valued Neurons via Gradient Descent

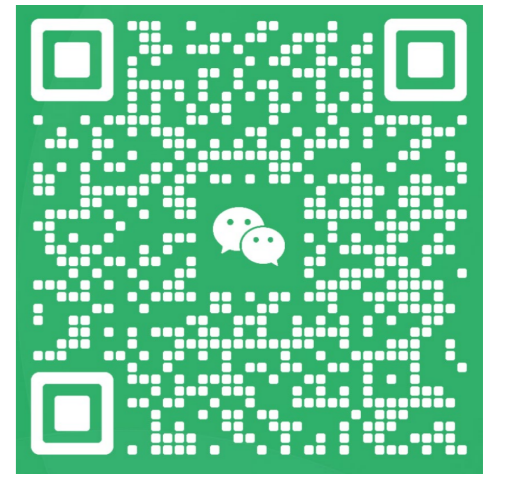
Jin-Hui Wu, Shao-Qun Zhang, Yuan Jiang, and Zhi-Hua Zhou  
 { wujh, zhangsq, jiangy, zhouzh }@lamda.nju.edu.cn



Paper



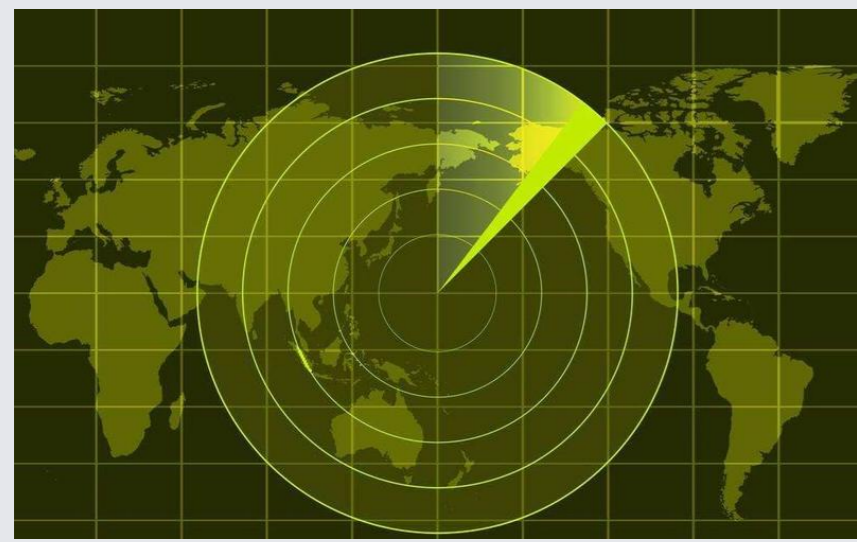
Homepage



WeChat  
(Jin-Hui Wu)

## Motivation

- Complex-valued neural networks (CVNNs) can outperform real-valued neural networks (RVNNs) in some signal processing tasks.
- But CVNNs cannot always outperform RVNNs.



Radar Signals



Audio Signals

- Two important questions:
  - When CVNNs outperform RVNNs via gradient descent (GD)?
  - Can we learn everything with CVNNs without paying additional price?
- We answer from the aspect of neuron learning, i.e., learning a single neuron using another neuron.
  - It is a special case of neural network learning.
  - Its analysis is tractable.
  - It is sufficient to tell us the difference between RVNNs and CVNNs.

## Formulation

- In this paper, neuron learning minimizes the expected square loss via GD

$$L(\mathbf{w}, \psi_w) = \mathbb{E}_{\mathbf{x} \sim \mathcal{D}} \left[ \left( \sigma_{\psi_w}(\mathbf{w}^\top \bar{\mathbf{x}}) - \sigma_{\psi_v}(v^\top \bar{\mathbf{x}}) \right)^2 \right].$$

- Here:
  - $\mathcal{D} = \mathcal{N}(\mathbf{0}, \mathbf{I})$  is the distribution of input  $\mathbf{x} \in \mathbb{C}^d$ .
  - $(\psi_v, v)$  is the fixed target neuron.
  - $(\psi_w, \mathbf{w})$  is the learnable neuron with random initialization.
  - $\sigma_{\psi}$  is the symmetric zReLU activation function

$$\sigma_{\psi}(z) = \begin{cases} \text{Re}(z), & \theta_z \in [-\psi, \psi], \\ 0, & \text{otherwise.} \end{cases}$$

- For a complex-valued neuron (CVN), both  $\psi_w$  and  $\mathbf{w}$  are learnable.
- For a real-valued neuron (RVN),  $\psi_w$  is fixed as  $\frac{\pi}{2}$ , only  $\mathbf{w}$  is learnable, and  $\sigma_{\psi}$  degenerates to the ReLU activation function.

## CVNs Can Learn More than RVNs

Two positive learning results for CVNs.

**Theorem 1** (informal). Let  $d = 1$ , and  $L_{cr}$  is the expected loss of learning an RVN using a CVN via GD. Under random initialization, if the step size of GD satisfies  $\eta_t = \eta \in (0, 1/(12\pi))$ , then we have

$$\Pr[L_{cr} = O(t^{-3})] > 0.$$

A CVN learns an RVN at rate  $O(t^{-3})$ .

**Theorem 2** (informal). Let  $d = 1$ , and  $L_{cc}$  is the expected loss of learning a CVN using a CVN via GD. Under random initialization, if the step size of GD satisfies  $\eta_t = \min\{c_1, c_2/t\}$  with  $c_1 \leq 1/3000$  and  $c_2 \geq 20$ , then we have

$$\Pr[L_{cc} = O(t^{-1})] > 0.$$

A CVN learns a CVN at rate  $O(t^{-1})$ .

One negative learning results for RVNs.

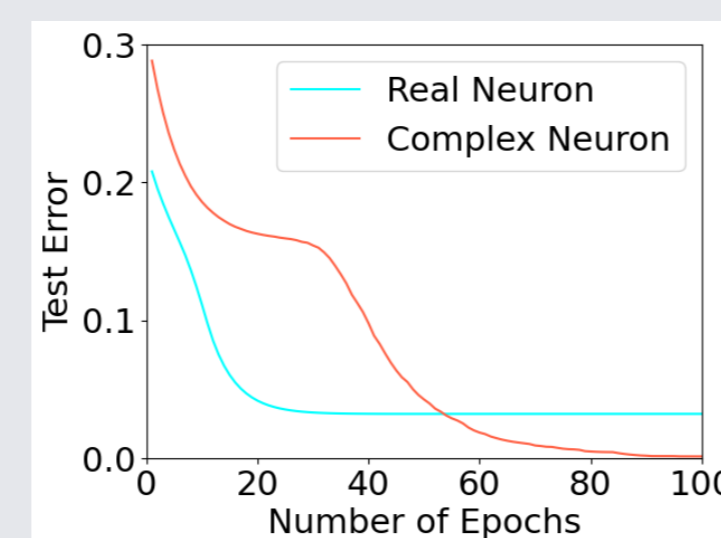
- The phase  $\psi$  of an RVN is fixed as  $\pi/2$ .
- An RVN has less learnable parameters than a CVN.
- The negative result considers learning a CVN using a two-layer RVNN for fairness.

**Theorem 4** (informal). Let  $d = 1$ , and  $L_{rc}$  is the expected loss of learning a CVN using a two-layer RVNN with  $n$  hidden neurons. If the CVN is non-degenerate, i.e.,  $\psi_v \notin \{0, \pi/2\}$  and  $v \neq 0$ , then we have

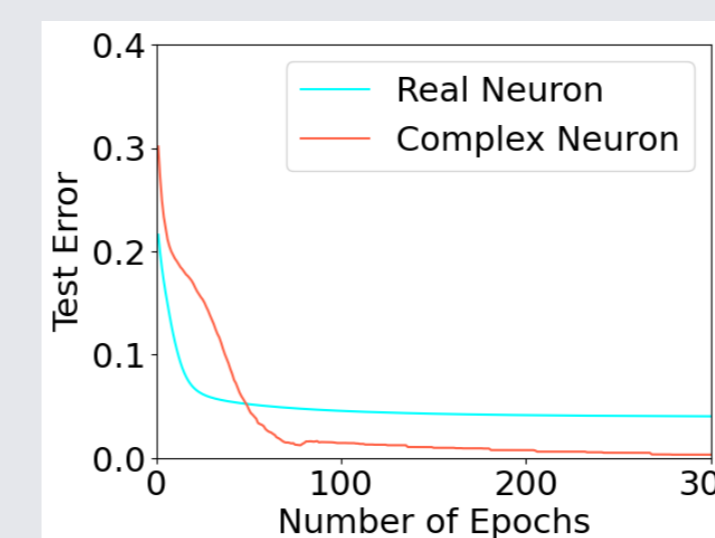
$$L_{rc} \geq \frac{\|v\|^2 \min\{2\psi, \pi - 2\psi\}^3}{24\pi(n+2)^2} > 0.$$

RVNNs with fixed width cannot learn a non-degenerate CVN.

Simulation experiments.



$d = 1$  and no bias term.  
(theoretical setting)



$d = 5$  and with bias term.  
(general setting)

## CVNs Learn Slower than RVNs

**Lemma 5** [Yehudai and Shamir, 2020] (informal). Let  $L_{rr}$  be the expected loss of learning an RVN using an RVN via GD. Under random initialization and suitable step size, we have

$$\Pr[L_{rr} = O(e^{-ct})] > 0.$$

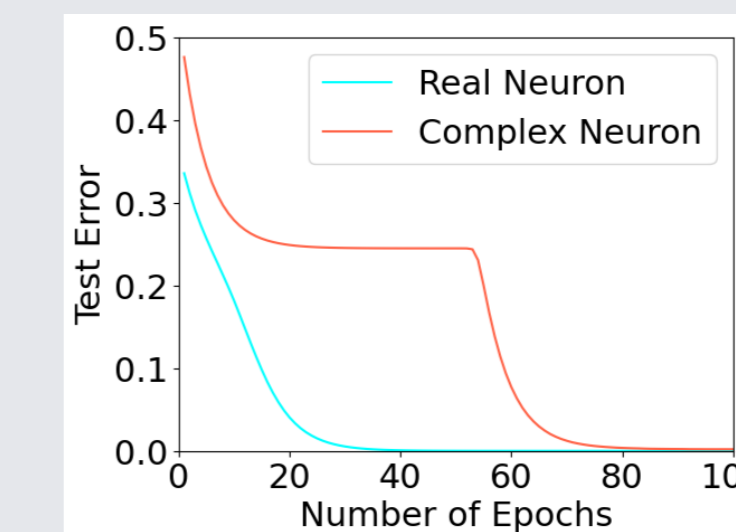
An RVN learns an RVN with exponentially small loss.

**Theorem 6** (informal). Let  $d = 1$ , and  $L_{cr}$  is the expected loss of learning an RVN using a CVN via GD. If the initialization is around the global minimum, and the step size of GD satisfies  $\eta_t = \eta \in (0, 1/(12\pi))$ , then we have

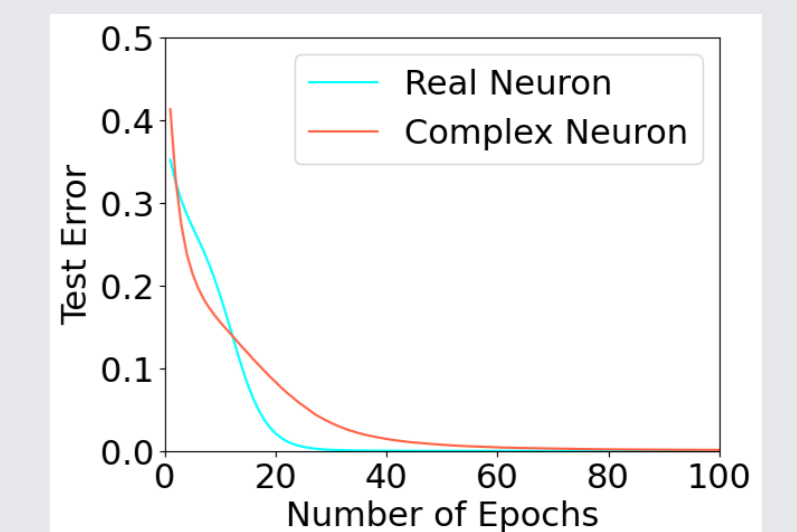
$$L_{cr} = \Omega(t^{-3}).$$

A CVN learns an RVN with polynomially large loss.

Simulation experiments.



$d = 1$  and no bias term.  
(theoretical setting)



$d = 5$  and with bias term.  
(general setting)

## Summary

- CVNs can learn **more** than RVNs
  - A CVN learns an RVN at rate  $O(t^{-3})$ .
  - A CVN learns a CVN at rate  $O(t^{-1})$ .
  - RVNNs with fixed width cannot learn a non-degenerate CVN.
- CVNs learn **slower** than RVNs.
  - An RVN learns an RVN at rate  $O(e^{-ct})$ .
  - A CVN learns an RVN at rate  $\Omega(t^{-3})$ .

Target	RVN	CVN
RVN	$O(e^{-ct})$	$\Theta(t^{-3})$
CVN	Cannot Learn	$O(t^{-1})$