GATED COMPLEX RECURRENT NEURAL NETWORKS

Wolter Moritz, Yao Angela

Institute for Computer Science, Bonn University

UNIVERSITÄT BONN

Abstract

Complex numbers have long been favored for digital signal processing, yet complex representations rarely appear in deep learning architectures. RNNs, widely used to process time series and sequence information, could greatly benefit from complex representations. We present a novel complex gated recurrent cell. When used together with norm-preserving state transition matrices, our complex gated RNN exhibits excellent stability and convergence properties. We demonstrate competitive performance of our complex gated RNN on the synthetic memory and adding task, as well as on the real-world task of human motion prediction.

Introduction

▶ Complex cell states can be more expressive than real ones. In GRU like RNNs the number of parameters is dominated by the weight matrices, which process the cell state. If the state size is h_r, these matrices have h_r^2 parameters. In the complex case these matrices have $2h_c^2$ parameters. For a complex network to have approximately the same amount of parameters the state can be chosen to be up to $h_c = \sqrt{h_r^2/2}$, this is more than $h_r/2$. Counting magnitude and phase we gain some extra space.

▶ Orthogonal and unitary matrices have been shown to be beneficial in machine learning[1, 4, 3]. Unitary matrices are more expressive than orthogonal ones, because their eigenvalues can be spread out over the entire unit circle, while orthogonal matrix eigenvalues are constrained to $-1, 1$.

Contribution

▶ We introduce a novel complex-gated recurrent unit to the best of our knowledge, we are the first to explore such a structure using complex number representations.

▶ We compare experimentally the effects of a bounded versus unbounded non-linearity in recurrent networks, finding evidence countering the commonly held heuristic that only bounded non-linearities should be applied in RNNs. Unbounded non-linearities, in our case, perform better, but must be coupled with the stabilizing measure of using norm-preserving state transition matrices.

▶ Our complex gated network is stable and fast to train; it outperforms state-of-the-art with equal parameters on both synthetic tasks and the real-world application of predicting poses in human motion capture.
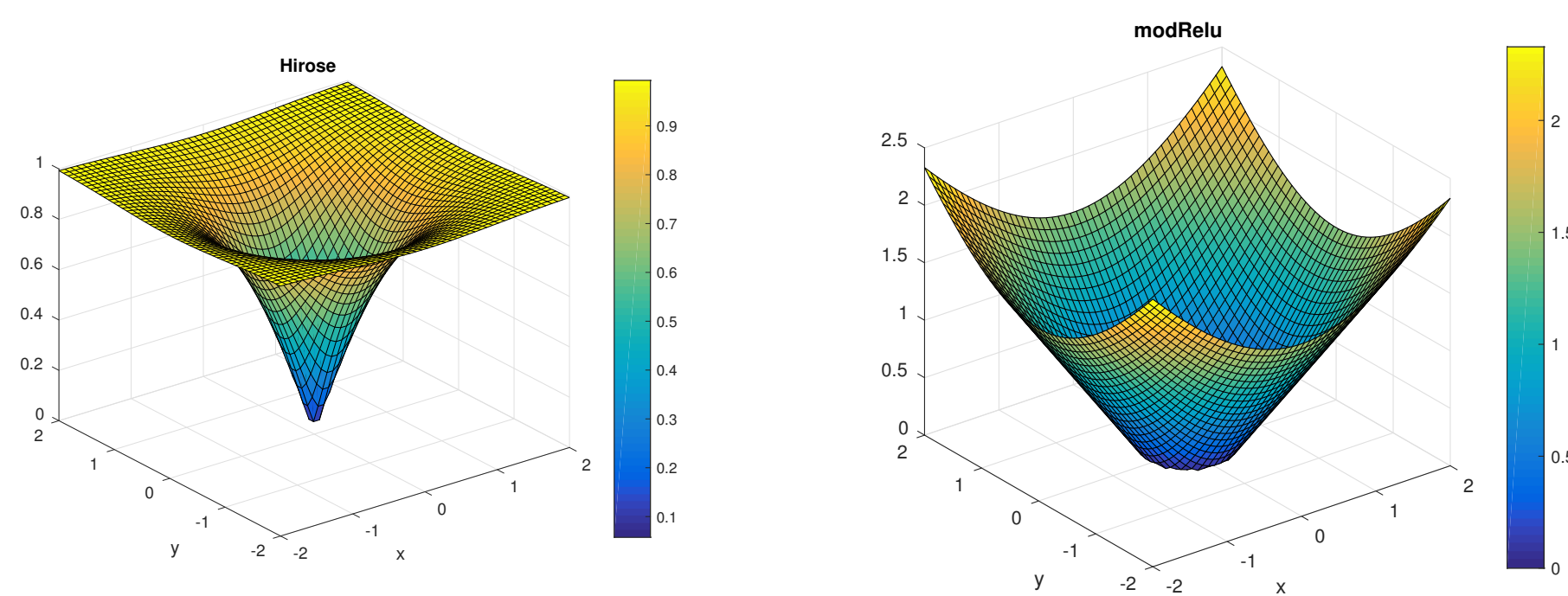
Non-linearities

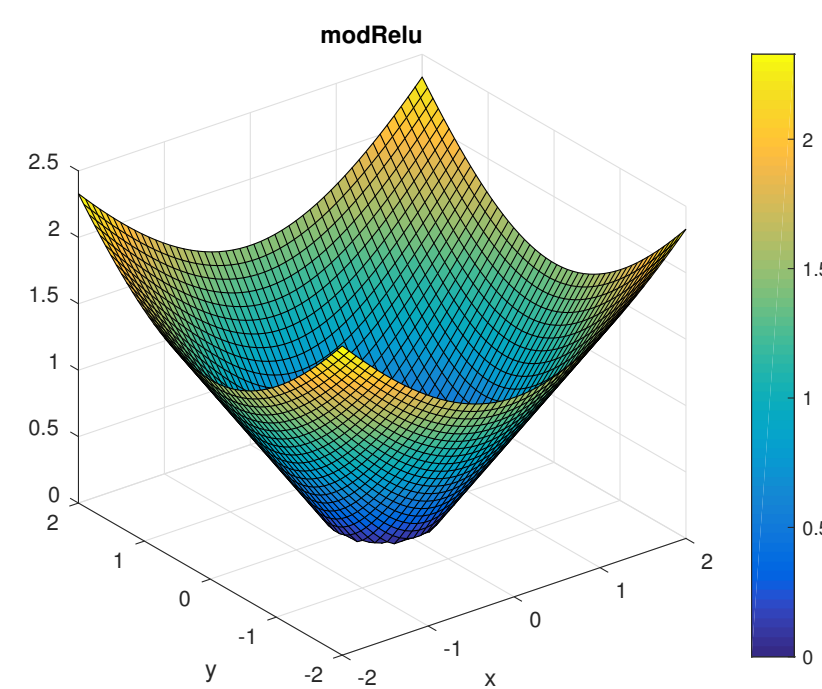
Figure 1: Hirose non-linearity

Figure 2: mod-Relu non-linearity

$$f_{\text{Hirose}}(z) = \tanh\left(\frac{|z|}{m^2}\right) e^{-i \cdot \theta_z} = \tanh\left(\frac{|z|}{m^2}\right) \frac{z}{|z|}. \tag{1}$$

$$f_{\text{modReLU}}(z) = \text{ReLU}(|z| + b) e^{-i \cdot \theta_z} = \text{ReLU}(|z| + b) \frac{z}{|z|}, \tag{2}$$

We compare the bounded non-linearity defined in equation 1 as well as the unbounded one from equation 2.

References

[1] M. Arjovsky, A. Shah, and Y. Bengio.
Unitary evolution recurrent neural networks.
In ICML, 2016.

[2] K. Cho, B. van Merriënboer, Ç. Gülçehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio.
Learning phrase representations using RNN encoder–decoder for statistical machine translation.
In EMNLP, Oct. 2014.

[3] S. L. Hyland and G. Raetsch.
Learning unitary operators with help from u(n).
In AAAI, 2017.

[4] S. Wisdom, T. Powers, J. R. Hershey, J. L. Roux, , and L. Atlas.
Full-capacity unitary recurrent neural networks.
In Advances in Neural Information Processing Systems, 2016.

The complex gated RNN

Following [1] we compute our weight updates using:

$$\mathbf{W}_{k+1} = (\mathbf{I} + \frac{\lambda}{2}\mathbf{A}_k)^{-1}(\mathbf{I} - \frac{\lambda}{2}\mathbf{A}_k)\mathbf{W}_k \qquad \text{where} \qquad \mathbf{A} = \mathbf{W}\nabla_{\overline{\mathbf{w}}}F - \overline{\mathbf{W}}\nabla_{\mathbf{w}}F \tag{3}$$

Inspired by [2] we define our complex memory cell as:

$$\widetilde{\mathbf{z}}_t = \mathbf{W}(\mathbf{g}_r \odot \mathbf{h}_{t-1}) + \mathbf{V}\mathbf{x}_t + \mathbf{b} \tag{4}$$

$$\mathbf{h}_t = \mathbf{g}_z \odot f_a(\widetilde{\mathbf{z}}_t) + (1 - \mathbf{g}_z) \odot \mathbf{h}_{t-1}, \tag{5}$$

With f_a being the Hirose and modRelu non-linearities. We define our gates as:

$$\mathbf{g}_r = f_g(\mathbf{z}_r), \qquad \text{where} \qquad \mathbf{z}_r = \mathbf{W}_r\mathbf{h} + \mathbf{V}_r\mathbf{x}_t + \mathbf{b}_r, \tag{6}$$

$$\mathbf{g}_z = f_g(\mathbf{z}_z), \qquad \text{where} \qquad \mathbf{z}_z = \mathbf{W}_z\mathbf{h} + \mathbf{V}_z\mathbf{x}_t + \mathbf{b}_z. \tag{7}$$

Using the two gate-non-linearities:

$$f_{g\ \text{sum}}(\mathbf{z}) = \sigma(\alpha\Re(\mathbf{z}) + (1 - \alpha)\Im(\mathbf{z})) \tag{8}$$

$$f_{g\ \text{prod}}(\mathbf{z}) = \sigma(\Re(\mathbf{z}))\sigma(\Im(\mathbf{z})) \tag{9}$$

$$\tag{10}$$

With $\alpha \in [0, 1]$. We have observed that the product formulation performs better on the adding problem, while the sum tends to do better on the memory problem. Both activations define mappings from $\mathbb{C} \to \mathbb{R}$.
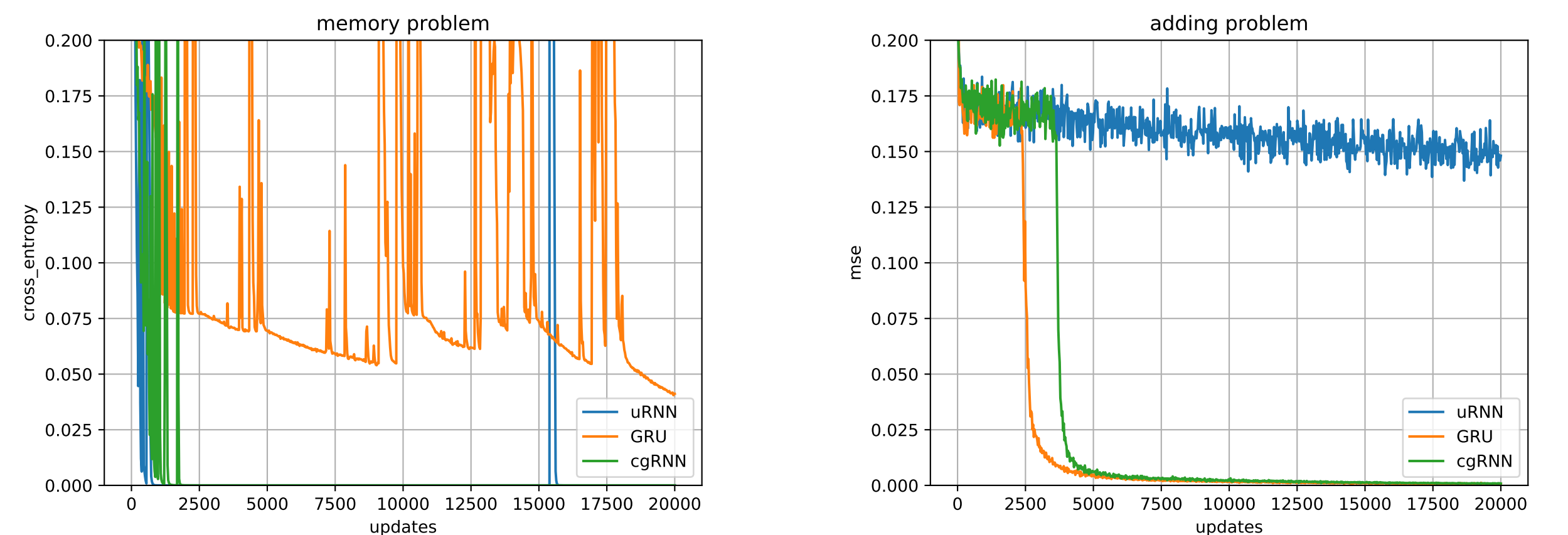
Adding and memory problem

Figure 3: Comparison of the previously proposed uRNN[1](blue, $n_h = 112$), GRU[2](orange, $n_h = 112$) and our own cgRNN(green, $n_h = 80$)
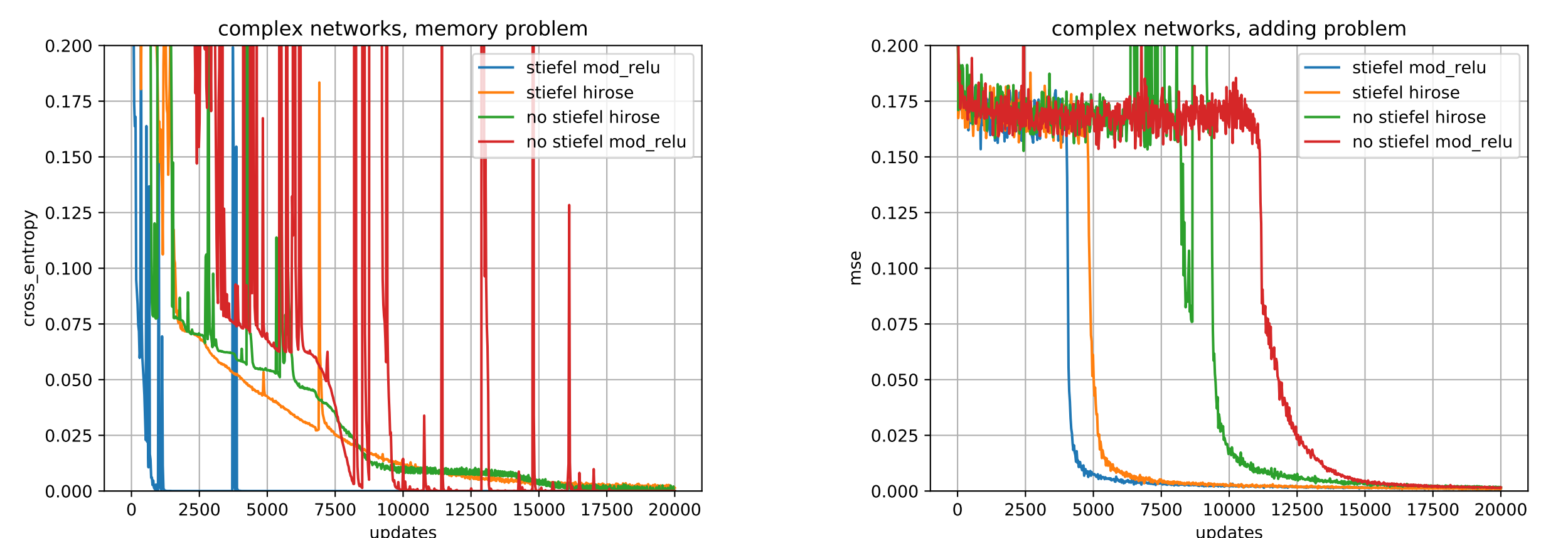

Figure 4: Comparison of non-linearities and norm preserving state transition matrices on the complex gated RNNs for the memory (left) and adding (right) problems for T=250. We find that the unbounded modReLU (see Eq. 2) performs best for both problems, but only if the state transition matrices are kept unitary. Without unitary state-transition matrices, the bounded Hirose non-linearity (see Eq. 1) performs better. We use $n_h = 250$ for the memory and $n_h = 91$ for the adding problem.

Human motion prediction

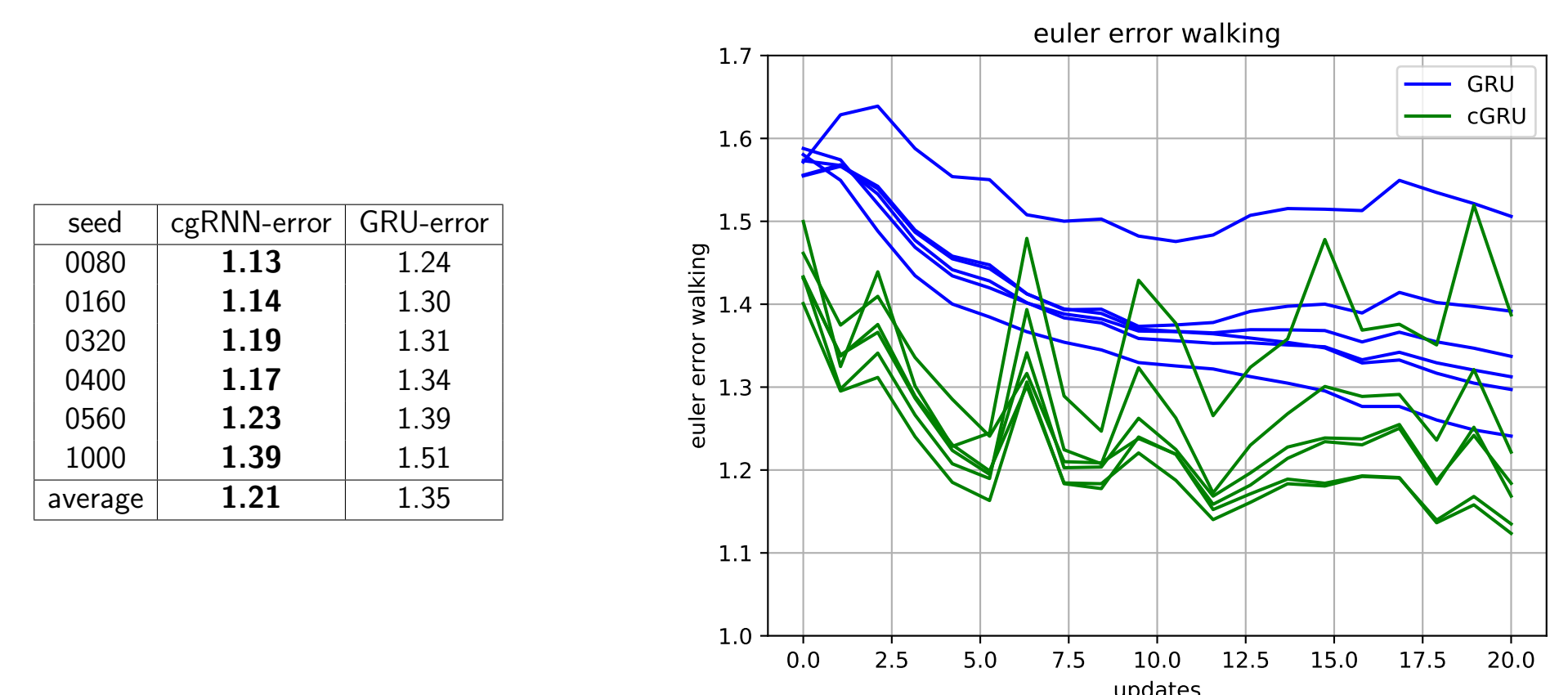| seed	cgRNN-error	GRU-error
0080	**1.13**	1.24
0160	**1.14**	1.30
0320	**1.19**	1.31
0400	**1.17**	1.34
0560	**1.23**	1.39
1000	**1.39**	1.51
average	**1.21**	1.35

Figure 5: Motion prediction Euler angle errors for the complex gated RNN (green) versus GRU (blue), where each line indicates a separate test sequence. The final error after 20000 iterations is shown in the table.