

Unitary and Symmetric Structure in Deep Neural Networks

Dr. Kehelwala Dewage Gayan Maduranga

October 24, 2022

1 Abstract

Recurrent neural networks (RNNs) have been successfully used on a wide range of sequential data problems. A well-known difficulty in using RNNs is the *vanishing or exploding gradient* problem. Recently, there have been several different RNN architectures that try to mitigate this issue by maintaining an orthogonal or unitary recurrent weight matrix. One such architecture is the scaled Cayley orthogonal recurrent neural network (scoRNN) which parameterizes the orthogonal recurrent weight matrix through a scaled Cayley transform. This parametrization contains a diagonal scaling matrix consisting of positive or negative one entries that can not be optimized by gradient descent. Thus the scaling matrix is fixed before training and a hyperparameter is introduced to tune the matrix for each particular task. In the first part of this thesis, we develop a unitary RNN architecture based on a complex scaled Cayley transform. Unlike the real orthogonal case, the transformation uses a diagonal scaling matrix consisting of entries on the complex unit circle which can be optimized using gradient descent and no longer requires the tuning of a hyperparameter. We compare the performance of The scaled Cayley unitary recurrent neural network (scuRNN) with scoRNN and other unitary RNN architectures.

Convolutional Neural Networks (CNNs) is a class of deep neural networks, most commonly applied to analyzing visual imagery. Nowadays deep neural networks also play an important role in understanding biological problems such as modeling RNA sequences and Protein sequences. The second part of the thesis explores deep learning approaches involving recurrent and convolutional networks to directly infer RNA secondary structure or Protein contact map, which has a symmetric feature matrix as output. We develop a CNN architecture with a suitable symmetric parameterization of the convolutional Kernel that naturally produces symmetric feature matrices. We apply this architecture to the inference tasks for the RNA secondary structure or protein contact maps. We compare our symmetrized CNN architecture with the usual convolution network architecture and show that these approaches can improve prediction results while using equal or fewer numbers of machine parameters.