Navigating the Complex Plane: A Curious Exploration of Complex-Valued Activation Functions in Neural Networks

Nicklas Holmberg
nholmber@ucsd.edu
2022

Abstract.
This overview, more of an exploratory glance than a formal technical paper, covers some of the intricacies of Complex Valued Neural Networks (CVNNs). This effort is simply driven out of curiosity and a desire to understand the underlying principles that make CVNNs a promising new direction in neural network research. The purpose of this project is not to present an exhaustive or definitive account, but merely to offer some insights and reflections on a subject that continues to push the boundaries of machine learning within the field of signal processing.

This exploration is an example of the interests I wish to pursue further in graduate school. Many of my side projects, like this one, have been attempts to explore topics at the intersection of mathematics and computer science, particularly within machine learning. Hopefully, these endeavors reflect my curiosity and a growing passion for understanding subjects within the field, and I sincerely hope to continue this journey through more structured and in-depth studies in the future.

Introduction.
In the rapidly evolving field of neural networks, the introduction of Complex-Valued Neural Networks (CVNNs) has opened new possibilities for processing and analyzing complex-valued data. Historically, neural networks have largely centered around real-valued data, with frameworks designed for floating point or integer numbers. However, many domains, from signal processing to bioinformatics, are sometimes better expressed using complex data. Complex numbers, by definition, consist of two parts: a magnitude (or amplitude) and a phase (or angle). In these scenarios, where the data naturally has both magnitude and phase, merely projecting complex data to a real domain can risk losing its inherent richness and nuance. In signal processing for instance, the phase of a signal can dictate crucial aspects like timing or synchronization. This is why developing neural networks capable of efficiently handling complex data is becoming increasingly essential in modern computational programs.

An essential component of developing any neural network (real-valued or otherwise) involves choosing an activation function, whose primary role is to introduce non-linearities into the model. These non-linearities allow the network to capture patterns and relationships in data that would be unattainable with linear transformations alone. While the main goal of an activation function is to introduce non-linearities, there are other considerations that can be crucial for the stable training of neural networks. For instance, bounded activation functions are desirable because they prevent extreme activations, helping in stabilizing the learning process by ensuring values don't explode or vanish as they pass through layers, thereby aiding convergence during training. In the real-valued domain, functions like the sigmoid is often used, defined as

$$\sigma(x) = \frac{1}{1 + e^{-x}}$$

which is analytic and bounded while being non-linear [1]. However, directly applying functions like these to the complex case in CVNNs is not straightforward. Complicating the landscape for CVNNs is Liouville's theorem, which posits that any function that is holomorphic (analytic) throughout the complex plane and remains bounded must be a constant function— which, as mentioned, is a non-starter for neural networks. The appeal of analytic functions in CVNNs is rooted in their smooth, differentiable nature, which can foster predictable behavior and smoother optimization landscapes. Given these constraints, the challenge is to create activation functions for CVNNs that are non-linear, maintain a strong degree of holomorphic behavior, stay bounded, and thereby enable efficient learning dynamics. This underscores one of the many distinct challenges associated with CVNNs when contrasted with their real-valued counterparts.

In this project, I'm focusing on testing the performance of complex activation functions that have been highlighted in past research and their efficacy on Fourier-transformed image data, resembling standard signal processing datasets. The existing literature offers a variety of promising approaches for complex activation functions. Some researchers advocate for functions that have strong holomorphic properties, given their potential benefits. At the same time, there's growing interest in "phasor networks", where the activation functions focus on maintaining magnitude while adjusting phase, with their outputs radiating from the origin to the boundaries of the unit circle [2].

Hopefully this gives a brief overview of the background and reasoning behind this project. As the debate between holomorphic and nonholomorphic complex activation functions is still very much ongoing, my endeavor here is simply to learn about and document their behavior. This project is merely a starting point, and I hope to further explore concepts like these, spanning both machine learning and complex analysis, during my future graduate studies.

Experiment Overview: Exploring Complex-Valued Activation Functions in CVNNs.
In the following experiment, the performance of a few established complex-valued activation functions is tested. The experiment utilizes the MNIST dataset, a standard benchmark data set for image recognition, and applies Fourier Transforms to the images, transforming them into the frequency domain. This transformation aligns with the complex-valued nature of the activation functions being tested, which include ReLU, ModReLU, Complex Leaky ReLU, and a variation of a complex cardioid function. I will define them one by one down below. Also, the exact code I used for the experiment is available in the same GitHub repository as this document, which explains the exact setup of the code executing this experiment.

Without going into too much detail, the code is developed using TensorFlow and Keras, well-known deep learning frameworks. Two distinct neural network architectures are employed: one tailored for complex-valued activations using the cvnn package's specialized layers, and the other for conventional activations like ReLU. For each activation function, a distinct model is trained and subsequently evaluated. For more details, please look at the code directly, available in the same folder as this document.

The activation functions chosen for this experiment represent a diverse set of mathematical properties and behaviors within the context of CVNNs. The Rectified Linear Unit (ReLU) serves as a baseline for comparison, given its widespread use in traditional real-valued neural networks. ReLU introduces non-linearity by allowing positive values to pass through unchanged while setting negative values to zero, defined mathematically as:

$$\mathrm{ReLU}(z) = \max(0, R(z))$$

As we can see, the ReLU activation function is not directly applicable to complex numbers as that part is completely overlooked. Yet, it is a cornerstone in deep learning, having contributed to state-of-the-art performances in various real valued tasks including image classification, natural language processing, and speech recognition [3]. However, as has been stated previously ReLU is conventionally defined for real numbers, and its application to complex-valued neural networks require adaptation or generalization. In this experiment however, we will use it is a measuring stick against the other activation functions being tested.

Given its success in real-valued networks, there have been suggestions to adapt the ReLU activation function for the complex domain. A well-known example of this is the modified ReLU (modReLU), defined as:

$$\sigma_{\text{modReLU}}(z) = \begin{cases} (|z| + b)\frac{z}{|z|} & \text{if } |z| + b \geq 0 \\ 0 & \text{if } |z| + b < 0 \end{cases}$$

where b and c are constants [3]. Just as ReLU allows positive numbers to pass while setting negative values to zero, modReLU allows complex numbers with magnitudes above a certain threshold to pass, and sets others to zero, offering a similar "filtering" effect for complex values.

The Complex Leaky ReLU was developed as another complex-valued generalizations of ReLU [4]. It is an extension of the traditional Leaky ReLU activation function designed specifically for complex-valued inputs. It's defined as:

$$f(z) = \{\text{LeakyReLU}\}(\text{Re}(z)) + i \cdot \{\text{LeakyReLU}\}(\text{Im}(z))$$

where the Leaky ReLU function is given by

$$\{\text{LeakyReLU}\}(x) = \begin{cases} x & \{\text{if }\} x > 0 \\ \alpha x & \{\text{if }\} x \leq 0 \end{cases}$$

and α is a constant that governs the slope of the function for negative values of x. In contrast to the standard ReLU, where negative values are set to zero and thus have no gradient, the Leaky ReLU assigns a non-zero slope to the negative part. The Complex Leaky ReLU operates on both the real and imaginary components separately. If either the real or imaginary part of the complex input is negative, that component is multiplied by α (typically a value close to zero). This ensures that even negative components maintain a non-zero gradient, aiding in gradient propagation during training and preventing the 'dying ReLU' problem. This approach retains the essence of "leakiness" in the complex domain, allowing neural networks to better maintain information and potentially enhancing their learning capabilities.

Lastly, we examine a variant of a complex cardioid functions, a novel activation function tailored for complex-valued data [6]. A salient feature of this function is its magnitude transformation, which aligns closely with the ReLU function when considered on the real axis. The distinguishing factor of the complex cardioid is its focus on input phase over magnitude. The function's output magnitude is modulated by the input phase, yet the input phase remains consistent in the output. Mathematically, this is represented as:

$$f(z) = \frac{(1 + \cos(\angle z)) \cdot z}{2}$$

Under this function, inputs on the positive real axis are scaled by a factor of one, those on the negative real axis by zero, and values with a non-zero imaginary component undergo a scaled transformation from one to zero as their phase transitions from the positive to the negative real axis. Notably, when limited to real inputs, the cardioid function converges to the familiar ReLU function. This inherent link between real-valued and complex-valued activations emphasizes its potential significance in our assessment of complex activation functions in CVNNs.

Metrics for Evaluation

The evaluation of the activation functions is based on these performance metrics, chosen to capture various aspects of each model's performance.

- **Test Loss**: This represents the model's error on the unseen test data. Lower test loss indicates better generalization and predictive accuracy.
- **Test Accuracy**: A direct measure of how many test instances were correctly classified. Higher accuracy indicates better performance in classifying unseen data.
- **ROC AUC Score**: The Receiver Operating Characteristic Area Under the Curve (ROC AUC) measures the model's ability to distinguish between the classes across different thresholds. A value close to 1 indicates a robust model in terms of sensitivity and specificity.
- **Matthews Correlation Coefficient (MCC)**: This is a balanced measure that considers true and false positives and negatives. The MCC is especially informative when classes are imbalanced, providing a value between -1 and 1, where 1 is perfect prediction.
- **Cohen's Kappa**: This statistic measures inter-rater agreement for categorical items, correcting for the chance of agreement. It provides a more robust understanding of how well the model is classifying compared to random classification. Since the data set being tested consists of 10 categories, (digits 0-9) this is a prudent measurement.

Together, these metrics should offer a balanced view of the model's performance. They consider not just accuracy, but also how well the model discriminates between classes and how it performs across different categories and balance scenarios.

Results

Activation Function	Test Loss	Test Accuracy	ROC AUC Score	Matthews Correlation Coefficient	Cohen's Kappa
ReLU	1.58	45.75%	0.80	0.48	0.40
ModReLU	78.68	84.60%	0.92	0.83	0.83
Complex Leaky ReLU	95.93	95.42%	0.98	0.95	0.95
Complex Cardioid	40.72	96.53%	0.98	0.96	0.96

Observations

ReLU (Real-valued):

- Test loss and accuracy suggest moderate performance.
- The ROC AUC Score of 0.8 indicates a good ability of the model to differentiate between positive and negative classes.
- Both the Matthews Correlation Coefficient and Cohen's Kappa suggest that the classifier's performance is better than random guessing but still far from perfect.
- The lower accuracy and challenges in classifying certain digits could be attributed to the "dying ReLU" problem, where neurons can get stuck in a state of inactivity.

ModReLU (Modified Real-valued ReLU for Complex Numbers):

- Significantly improved accuracy compared to ReLU. ModReLU introduces a parameter that allows negative values to be modulated, which can help overcome the "dying ReLU" issue. Its

better performance compared to ReLU suggests that the modulation of negative values helps in capturing more useful features during training.
- High ROC AUC Score, Matthews Correlation Coefficient, and Cohen's Kappa all point towards excellent classifier performance.

Complex Leaky ReLU:
- This activation function gave the best accuracy among all.
- Its metrics, especially ROC AUC, Matthews Correlation Coefficient, and Cohen's Kappa, are very close to 1, indicating near-perfect performance.
- This suggests that introducing a "leaky" component that allows small gradients when the unit is not active can be very effective in complex domains.

Complex Cardioid:
- This function, being a phase-sensitive complex extension of the ReLU, showcases the best performance next to Complex Leaky ReLU.
- The test accuracy, ROC AUC Score, Matthews Correlation Coefficient, and Cohen's Kappa all suggest an excellent ability to classify, leveraging both magnitude and phase information.

Reflection

Upon reviewing our findings, it's evident that the choice of activation function plays a crucial role in model performance. Some functions clearly outperformed others, emphasizing the importance of ongoing research in this domain. The stark contrast in results between the functions suggests that even established methods can be outpaced when we tailor approaches specifically to the nuances of the problem at hand.

Singularities and Activation Performance: The significant test loss observed for ModReLU and Complex Leaky ReLU might hint at the network's activation functions approaching singular points in certain scenarios. When training neural networks, avoiding or circumventing singularities is crucial for stability. If the data or the weights during training push the activations close to these singular regions, it could account for such loss spikes.

Boundary Behavior and Results: The excellent performance metrics of the Complex Cardioid function, especially its high accuracy and Matthews Correlation Coefficient, can be tied back to its inherent boundary behaviors. From a complex analysis perspective, the behavior on boundaries is pivotal, and the cardioid function, with its unique properties on the unit circle, might be leveraging this to achieve superior results on Fourier-transformed data.

In short, the behaviors of different activation functions, whether related to smoothness, boundary actions, or holomorphic properties, are clearly central to their effectiveness. There is a lot more to dig into here, and I deliberately want to limit the scope of this exploration. A more granular breakdown and a series of systematic experiments would be necessary to truly discern the cause-and-effect relationship governing the performance differences across various complex activation functions. My primary intention here was to introduce the problem, take initial steps in testing, and provide some preliminary numerical insights. As I hopefully soon start my graduate studies, I hope to delve deeper into this area, seeking more answers and refining our understanding further.

Footnotes:

1. Pratiwi, Heny, et al. "Sigmoid Activation Function in Selecting the Best Model of Artificial Neural Networks." *J. Phys.: Conf. Ser.*, vol. 1471, 2020, p. 012010.
2. Bassey, Joshua, Xiangfang Li, and Lijun Qian. "A Survey of Complex-Valued Neural Networks." *Texas A&M University System,* 2021.
3. M. Arjovsky et al., "Unitary Evolution Recurrent Neural Networks," 2016.
4. “Empirical Evaluation of Rectified Activations in Convolutional Network," arXiv:1505.00853v2 [cs.LG], 2015.
5. Özdemir, Necati. "Complex valued neural network with Möbius activation function." 2011.
6. Virtue, Patrick, Stella X. Yu, and Michael Lustig. "Better than Real: Complex-valued Neural Nets for MRI Fingerprinting." *University of California, Berkeley / International Computer Science Institute,* 2017.