Framework for Deep and Temporal Complex-Valued Networks

[1]Andy M. Sarroff, [2]Colin Raffel, [1]Michael Casey, [2]Daniel P. W. Ellis

[1]Dept. of Computer Science, Dartmouth College, Hanover, New Hampshire
[2]LabROSA, Dept. of Electrical Engineering, Columbia University, New York
Contact: sarroff@cs.dartmouth.edu

Abstract

We present a framework for building deep and temporal complex-valued networks that contain compositions of holomorphic and non-holomorphic functions. The gradients of a real-valued cost function are back-propagated through the network to adjust the parameters by:

- Using convenience of Wirtinger calculus.
- Leveraging identities of the complex conjugate.

Motivation

Back-propagation for complex valued networks. Easily build deep networks having compositions of holomorphic and non holomorphic functions.

Why fully-complex activations? We may want to have a means of jointly modeling phase-amplitude dependencies for time-frequency representations of audio.

Why split-complex activations? Liouiville's theorem: Every bounded entire function must be constant. We might prioritize boundedness in some components of a model.

Wirtinger Calculus

Let $z \in \mathbb{Z}$, $x, y \in \mathbb{R}$, with

$$f(z) = f(z, \bar{z}) = f(x, y) = u(x, y) + iv(x, y)$$

and $\bar{z}$ denoting complex conjugate of z.

$\mathbb{CR}$ **(Wirtinger) Calculus**:

$\mathbb{R}$-derivative $= \left.\frac{\partial f}{\partial z}\right|_{\bar{z} \text{ constant}}$ $\overline{\mathbb{R}}$-derivative $= \left.\frac{\partial f}{\partial \bar{z}}\right|_{z \text{ constant}}$

Cauchy-Riemann Equations:

$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y}$, $\frac{\partial v}{\partial x} = -\frac{\partial u}{\partial y}$, or equivalently: $\frac{\partial f}{\partial \bar{z}} = 0$.

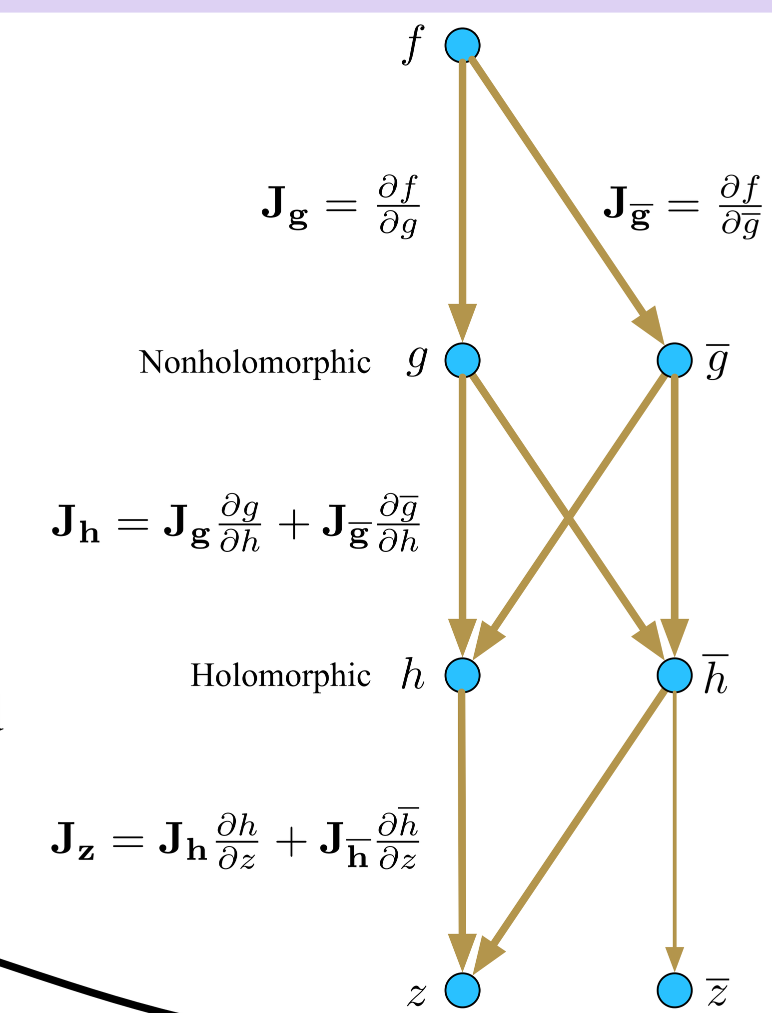
Framework

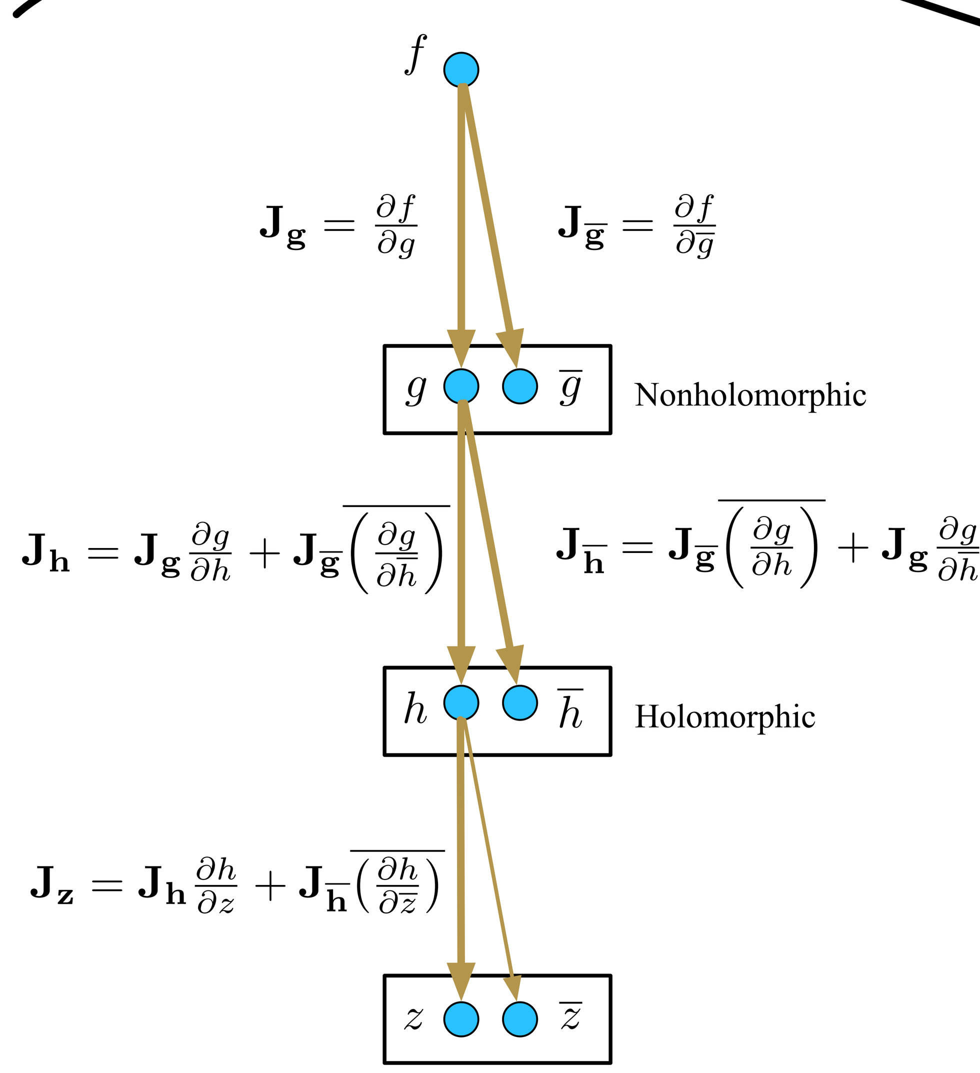
Dependency Graph:

$$y = f(g) = f(g, \bar{g}) = |g|^2 = g\bar{g}$$
$$y = (f \circ g \circ h)(z)$$

We seek to find the Jacobian of f with respect to z, $\mathbf{J_z}$.

If g or h are nonholomorphic then the dependency graph is not straightforward to compute.





Algorithm
Wirtinger calculus provides convenient means to compute partial derivatives of complex-valued functions.

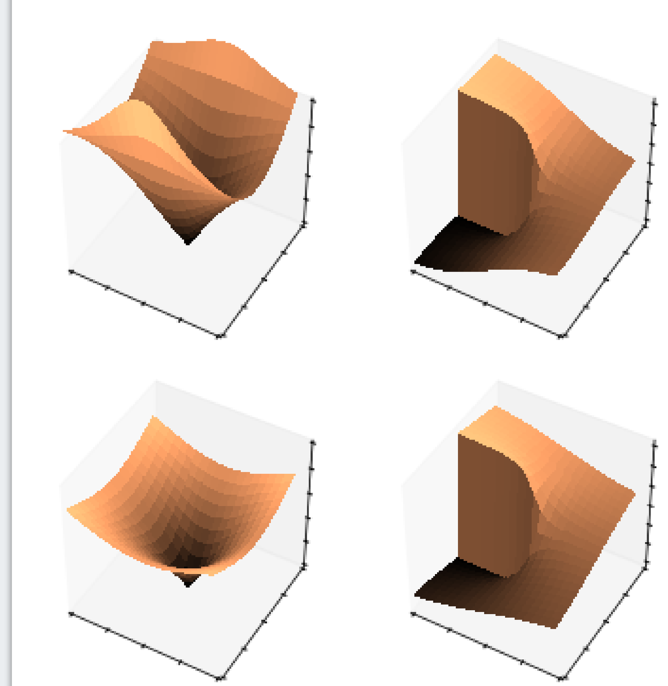
Use the following identities to keep track of two gradients rather than four:

$\frac{\partial \bar{f}}{\partial \bar{z}} = \overline{\left(\frac{\partial f}{\partial z}\right)}$, $\frac{\partial \bar{f}}{\partial z} = \overline{\left(\frac{\partial f}{\partial \bar{z}}\right)}$.

$\frac{\partial f}{\partial \bar{z}} = 0$ when f is holomorphic.

Activation Functions

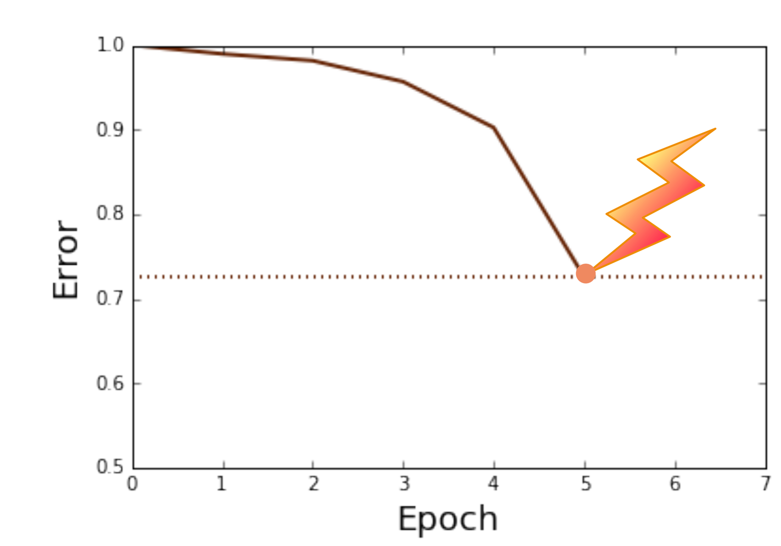
Elementary transcendental functions have desirable characteristics [2] but may be difficult to train.



tanh: holomorphic but unbounded

Squashing and phase-preserving [3]: not holomorphic but bounded

Complex-valued LSTM encounters singularity during training on toy task.



Acknowledgments

Part of this work was performed by the first author while visiting LabROSA, Columbia University. Thanks to Professor John Wright for conversations about back-propagation.

References

[1] Brandwood, D. A complex gradient operator and its application in adaptive array theory. *Communications, Radar and Signal Processing, IEE Proceedings F 130*, 1 (February 1983), 11–16.

[2] Kim, T., and Adalı, T. Approximation by fully complex multilayer perceptrons. *Neural Computation 15*, 7 (July 2003), 1641–1666.

[3] Georgiou, G., and Koutsougeras, C. *Complex domain backpropagation. Circuits and Systems II: Analog and Digital Signal Processing, IEEE Transactions on 39*, 5 (May 1992), 330–334.

[4] Mandic, D. P., and Goh, V. S. L. *Complex Valued Nonlinear Adaptive Filters: Noncircularity, Widely Linear and Neural Models.* Wiley, Chichester, U.K., 2009.