Complex Valued Convolutional Neural Networks for Image Restoration

Mid Thesis to be submitted in the partial fulfillment of the requirements for the degree

MASTER OF SCIENCE

IN

MATHEMATICS AND COMPUTING

by

Amara Datta Dola

(19MA20003)

Under the guidance of

Prof. Mahesh Mohan M R and
Coguide: **Prof. Dipankar Ghosh**

Department of Mathematics
Indian Institute of Technology, Kharagpur

28th Nov, 2023

Department of Mathematics
Indian Institute of Technology,
Kharagpur, India - 721302

CERTIFICATE

This is to certify that we have examined the thesis entitled **Complex Valued Convolutional Neural Networks for Image Restoration**, submitted by **Amara Datta Dola** (**19MA20003**) an undergraduate student of **Department of Mathematics** in partial fulfillment for the award of degree of Master of Science. We hereby accord our approval of it as a study carried out and presented in a manner required for its acceptance in partial fulfillment for the undergraduate Degree for which it has been submitted. The thesis has fulfilled all the requirements as per the regulations of the Institute and has reached the standard needed for submission.

Guide's Signature

Co-Guide's Signature

Date________________

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to my supervisor, Prof. Mahesh Mohan M R, who introduced me to the exciting field of complex valued AI. His profound expertise and insightful guidance have been instrumental in the progress and success of my research.

I am also indebted to the faculty and staff, whose assistance have been invaluable throughout my academic journey. Their willingness to share knowledge and offer advice has greatly enriched my learning experience.

Special thanks are due to my peers and colleagues who have provided me with support and collaboration. Their perspectives and feedback have been a source of encouragement and growth.

Finally, and most importantly, my heartfelt thanks go to my family. Their unwavering support, endless patience, and belief in my abilities have been the cornerstone of my strength and perseverance. This accomplishment is as much theirs as it is mine.

ABSTRACT

Convolutional Neural Networks (CNNs) are a cornerstone in deep learning, especially for computer vision tasks. In the literature hitherto, most research in CNN focuses on real-valued data like images, videos, speech, etc. However, there exist many inherently complex-valued signals like SAR, MRI, wind, etc that are not well studied using CNNs. These complex valued signals are successfully used for diverse practical applications in agriculture, medicine, earth sciences, etc. Similarly, several traditional studies have shown that some computer vision problems are effectively addressed in the complex-valued domain, such as within Fourier transform domains.

In the pre-deep learning era, several studies have indicated that complex signals can be effectively processed using complex-valued systems. This insight has led to the development of a novel variation of the CNN model, which utilizes complex-valued inputs and weights. This approach effectively captures the phase structures of complex valued signals, making it particularly valuable for tasks where such information is crucial, like MRI and remote sensing. However, the techniques developed for real-valued neural networks seldom translates to the case of complex-valued neural networks. One particular challenge is in optimizing complex-valued neural network via back-propogation. The exploration of complex-valued calculus and backpropagation techniques further expands the potential of deep learning, allowing for more nuanced data interpretation and broader application possibilities.

In this report, we study the following aspects of Complex-valued AI:

- Potential of Complex-valued Neural Networks
- Challenges in Optimizing Complex-valued Neural Networks
- A Comprehensive Mathematical Coverage on Effectively Optimizing Complex-valued Neural Networks

Finally, we identify other relevant problems pertaining to complex-valued CNN, which we plan as our future studies.

Contents

Chapter 1

Complex-Valued Neural Networks

Complex-valued neural networks (CVNNs) have emerged as an essential paradigm in the landscape of computational intelligence and machine learning (Hirose 2009). Unlike traditional neural networks, which primarily operate on real-valued data, CVNNs extend this concept to the complex domain, harnessing the potential of complex numbers in representing and processing information. This transition is not merely a straightforward extension of well-understood real-valued neural networks but a fundamental shift in how neural networks understand and interact with data.

Complex numbers, characterized by their real and imaginary components, provide a richer framework for representing data. In the realm of CVNNs, this richness translates into an enhanced ability to capture intricate patterns and relationships in data that would otherwise be elusive or overly simplified in a real-valued context. The implications of this are profound, opening up new avenues for solving complex problems across various scientific and technological domains.

At the heart of CVNNs lies the recognition that many natural and technological phenomena are inherently complex-valued. From electromagnetic signals to quantum mechanical systems, the complex domain offers a more natural and expressive language for describing these phenomena. By leveraging this language, CVNNs offer a more nuanced and effective approach to tasks such as pattern recognition, signal processing, and complex system modeling.

This report delves into the world of complex-valued neural networks, exploring their foundational concepts, motivations, and diverse applications. It aims to shed light on the unique characteristics of CVNNs.

1.1 Introduction

CVNNs mark a significant departure from traditional neural network architectures. They are not merely an adaptation of existing networks to handle complex numbers but represent a fundamentally different approach to neural computation. This section provides a comprehensive overview of the architecture, operation, and unique aspects of CVNNs.

In CVNNs, both the neurons and the synaptic weights are complex-valued, allowing the network to process and learn from data represented as complex numbers. This change necessitates rethinking various aspects of neural network design, from activation functions to learning algorithms. For instance, the complex nature of the data requires the use of activation functions that can operate effectively in the complex domain, such as complex versions of the rectified linear unit (ReLU) or hyperbolic tangent functions (Lee et al. 2022).

The learning algorithms in CVNNs also undergo significant modifications to accommodate the complex nature of the weights and inputs. Gradient descent, a cornerstone of neural network training, is adapted to handle complex derivatives, often employing techniques based on Wirtinger calculus.

The implications of these architectural and algorithmic changes are far-reaching. CVNNs exhibit distinct behaviors and learning dynamics compared to their real-valued counterparts, often resulting in different patterns of convergence and model performance. These differences underscore the importance of developing specialized techniques and methodologies for designing, training, and deploying CVNNs.

As we explore further, the report will highlight the theoretical underpinnings of optimizing CVNNs, its practical implementations, and the unique challenges and opportunities they present.

1.2 Motivation for Complex-Valued Signals and Their Applications

The motivation for exploring and employing complex-valued neural networks (CVNNs) stems from the inherent nature of complex-valued signals in various scientific and technological fields. These signals, characterized by their amplitude and phase, provide a more comprehensive understanding of the phenomena they represent. This section delves into the specific cases where complex-valued signals are crucial and the applications that benefit from the advanced processing capabilities of CVNNs.

1.2.1 Synthetic Aperture Radar for Agriculture

Synthetic Aperture Radar (SAR) technology, widely used in remote sensing, significantly benefits from complex-valued signal processing. SAR (Synthetic Aperture Radar) is a radar imaging technique that uses the movement of the radar antenna to synthesize a large aperture. This allows SAR to produce high-resolution images, regardless of weather conditions or light levels (see Fig. 1.1).

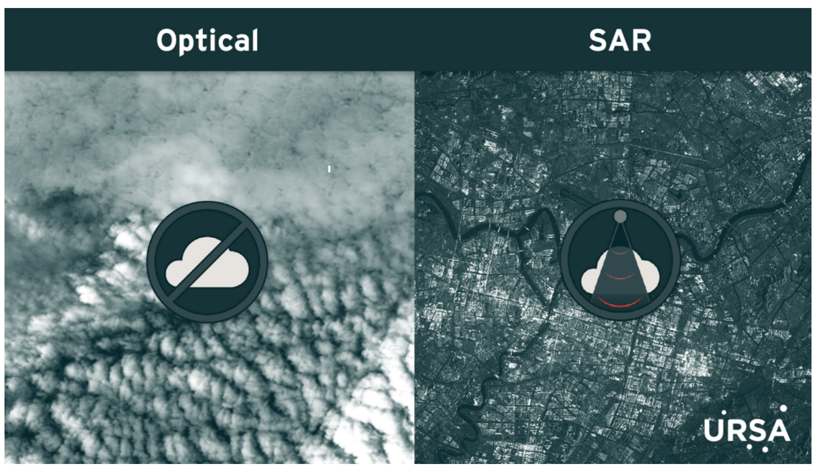

Figure 1.1: Optical vs SAR Imaging for the case of cloudy situation. SAR is able to capture the information whereas optical imaging struggles (Source: NASA).

In agriculture, SAR is instrumental in monitoring and managing agricultural activities. The complex-valued signals in SAR contain vital information about the land surface, including crop growth, soil moisture, and overall agricultural health.

The phase information in SAR images, often disregarded in traditional analysis, can provide unique insights into surface structures and vegetation patterns. CVNNs, with their ability to process and interpret these complex-valued signals, can extract meaningful information that aids in precision farming, crop yield estimation, and sustainable land management practices (LIU et al. 2019).

1.2.2 Magnetic Resonance Imaging for Medicine

Magnetic Resonance Imaging (MRI) is another domain where complex-valued signals play a pivotal role. MRI technology relies on complex-valued data to generate detailed images of internal body structures. The challenge in MRI is the reconstruction of high-quality images from the raw data, which is inherently complex-valued.

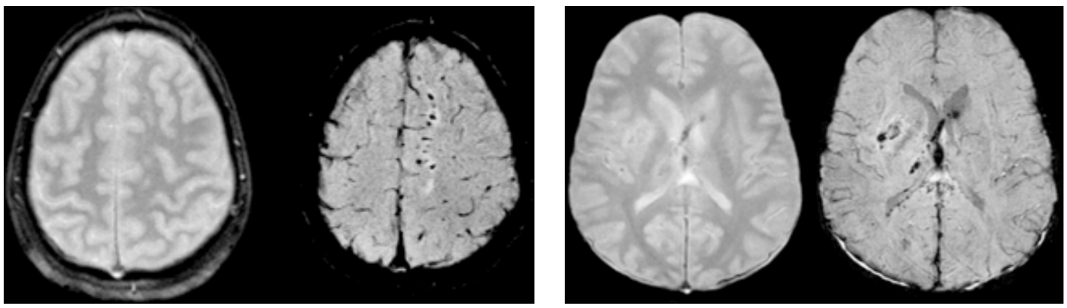

Figure 1.2: Condition: Axonal Injury vs Hemorrhage. For each condition, the left image shows typical MRI (without phase information) and right shows SW-MRI (which incorporates both magnitude and phase information). Note the condition evident in SW-MRI expressed by black patches.

Typical MRIs often discard the phase information. Using Susceptibility Weighted MR phase information is also taken into account. Phase information in such images is influenced by the magnetic properties of the tissues being imaged. This means that different tissues or substances within the body, such as veins, hemorrhage areas, or calcifications, will exhibit distinct signals, sometimes appearing as anti-phase signals compared to their surroundings (see Fig. 1.2). Anti-phase signals occur when the phase of the signal is inverted, indicating variations in susceptibility and often correlating with pathological changes. The MRI images shown compare conditions of axonal injury and hemorrhage, highlighting how phase information can help differentiate between different types of brain injuries and conditions, providing valuable diagnostic information as shown in Fig. 1.2.

CVNNs may offer a sophisticated approach to MRI image reconstruction. By processing the complex-valued MRI data, CVNNs can enhance the clarity and accuracy of the images, leading to better diagnosis and treatment planning in healthcare. The ability of CVNNs to preserve and utilize the phase information in MRI data is particularly beneficial, as it can lead to finer details and contrasts in the reconstructed images.

1.2.3 Image Transforms: Fourier Transform

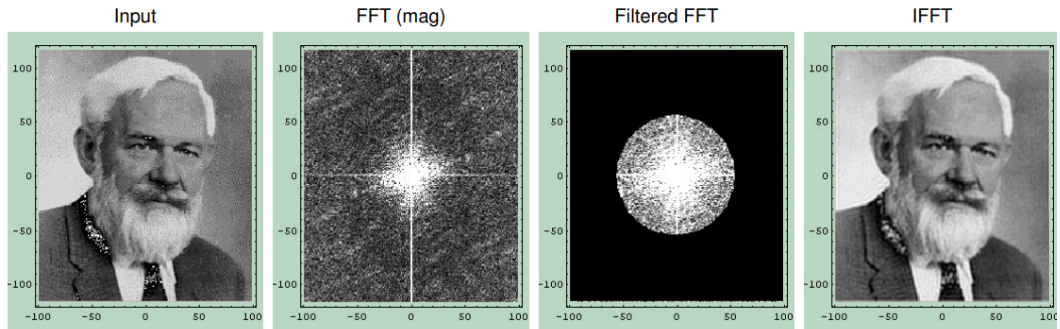

Figure 1.3: Process of Image Restoration using Fourier transform. As natural images have predominantly low frequency (central part of Fourier spectrum) and noise have predominantly high frequency, a low pass filtering in frequency domain removes substantial noise.

Figure 1.3 shows an application where solving in Fourier transformed domain is easier as compared to that in raw or image domain. Starting with the original grayscale image of a noisy potrait image, the process moves to a Fast Fourier Transform (FFT), which translates the image into its frequency domain, revealing the distribution of frequencies as varying intensities in the FFT magnitude representation. A circular low-pass filter is then applied to the FFT to isolate and remove high-frequency noise, which is typically spread outside the central region in the frequency domain. Finally, the Inverse FFT (IFFT) reconstitutes the filtered image back to its spatial domain, resulting in a restored image that ideally has less noise and clearer features than the original.

In applications such as audio processing, telecommunications, and signal filtering, the Fourier Transform's complex-valued output is essential. CVNNs, with their capacity to handle complex-valued data, can effectively process these spectra. This ability is crucial for tasks like noise reduction, signal enhancement, and feature extraction in various signal processing applications.

1.3 Advantages of using complex valued NNs

1.3.1 Orthogonal decision boundaries:

Remark 1: *(a) A single complex-valued neuron with n-inputs is equivalent to two real-valued neurons with 2n-inputs which have a restriction on a set of weight parameters. (Nitta 2003)*
(b) The decision boundary of a single complex-valued neuron consists of two hypersurfaces which intersect orthogonally.(Nitta 2003)

Proof for the case of a single input single neuron (Remark 1(a)).

Proof. Consider a single neuron with a complex weight $W = w_1 + iw_2$ and a complex input $a + ib$. The output is a complex number $c + id$. The matrix form of the transformation is represented as:

$$\begin{pmatrix} w_1 & -w_2 \\ w_2 & w_1 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} c \\ d \end{pmatrix}$$

This can be interpreted as the multiplication of a complex number W with a complex input $(a + ib)$, resulting in a complex output $(c + id)$.

Let M be the matrix representing the transformation:

$$M = \begin{pmatrix} w_1 & -w_2 \\ w_2 & w_1 \end{pmatrix}$$

The magnitude of the complex weight $|W|$ can be represented as $\sqrt{w_1^2 + w_2^2}$. Therefore, we can normalize the matrix M to get:

$$M = |W| \begin{pmatrix} \frac{w_1}{\sqrt{w_1^2+w_2^2}} & \frac{-w_2}{\sqrt{w_1^2+w_2^2}} \\ \frac{w_2}{\sqrt{w_1^2+w_2^2}} & \frac{w_1}{\sqrt{w_1^2+w_2^2}} \end{pmatrix}$$

This normalized matrix can be further expressed in terms of the magnitude $|W|$ and the arguments (or phase) of W, denoted by $\arg(W)$:

$$M = |W| \begin{pmatrix} \cos(\arg(W)) & -\sin(\arg(W)) \\ \sin(\arg(W)) & \cos(\arg(W)) \end{pmatrix}$$

Matrix M is an orthogonal matrix because it satisfies the property $MM^T = I$, where M^T is the transpose of M, and I is the identity matrix. Orthogonal matrices have the property that they preserve the inner product, and thus, the transformation represents a rotation in the complex plane.

Finally, we illustrate this with the transformation of a vector $\begin{pmatrix} x_1 \\ y_1 \end{pmatrix}$ in the complex plane, which is rotated by the matrix M to a new position while preserving its magnitude. This shows that the decision boundary created by a single neuron with complex weights is orthogonal.

The complex number $c + id$, is:

$$c = aw_1 - bw_2,$$
$$d = aw_2 + bw_1.$$

Let activation function (Step function) ϕ applied to the output can be defined as:

$$\phi(x) = \begin{cases} 1 & \text{if } x > c_1, \\ 0 & \text{if } x \leq c_1, \end{cases}$$

where c_1 is a threshold value.

The decision boundary in the input space can then be determined by setting the real part of the neuron's output equal to the threshold c_1:

$$aw_1 - bw_2 = c_1.$$

Solving for b in terms of a gives the equation of the decision boundary:

$$b = \frac{aw_1 - c_1}{w_2}.$$

Given $w_1, w_2 \neq 0$, the slope of the decision boundary is $-\frac{w_1}{w_2}$, which is orthogonal to the direction of the weight vector in the complex plane. □

Proof for the Case of a n-input single neuron (Remark 1(b))

Proof. Consider a single complex-valued neuron with n complex inputs $z_j = x_j + iy_j$ for $j = 1, \ldots, n$, where i is the imaginary unit, and each z_j corresponds to a real part x_j and an imaginary part y_j. The neuron's weight vector is given by complex numbers $w_j = w_j^R + iw_j^I$, where w_j^R and w_j^I are the real and imaginary parts of the weight associated with input z_j. The output of the neuron before activation is a complex number represented by the dot product of the inputs and weights.

The real part of the output is given by:

$$\mathrm{Re}\left(\sum_{j=1}^{n} w_j z_j\right) = \sum_{j=1}^{n}(w_j^R x_j - w_j^I y_j),$$

and the imaginary part of the output is:

$$\mathrm{Im}\left(\sum_{j=1}^{n} w_j z_j\right) = \sum_{j=1}^{n}(w_j^R y_j + w_j^I x_j).$$

The decision boundary is then determined by the set of points for which the real part of the output equals a threshold c_1, and the imaginary part equals another threshold c_2. These conditions can be expressed as two linear equations representing two hypersurfaces in the $2n$-dimensional real space of the inputs $(x_1, \ldots, x_n, y_1, \ldots, y_n)$:

$$\sum_{j=1}^{n}(w_j^R x_j - w_j^I y_j) = c_1, \quad \text{and} \quad \sum_{j=1}^{n}(w_j^R y_j + w_j^I x_j) = c_2.$$

To show that these hypersurfaces intersect orthogonally, consider the gradients of the two functions defining the decision boundaries. The gradients are normal to the hypersurfaces and are given by:

$$\nabla_1 = \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \vdots \\ \frac{\partial f}{\partial x_p} \\ \frac{\partial f}{\partial y_1} \\ \vdots \\ \frac{\partial f}{\partial y_n} \end{pmatrix} = \begin{pmatrix} w_1^R \\ \vdots \\ w_n^R \\ -w_1^I \\ \vdots \\ -w_n^I \end{pmatrix}, \quad \nabla_2 = \begin{pmatrix} w_1^I \\ \vdots \\ w_n^I \\ w_1^R \\ \vdots \\ w_n^R \end{pmatrix}.$$

The orthogonality of the hypersurfaces can be verified by showing that the dot product of the gradients is zero:

$$\langle \nabla_1, \nabla_2 \rangle = \sum_{j=1}^{n} w_j^R w_j^I - w_j^I w_j^R = 0.$$

This dot product is zero, implying that the gradients, and thus the decision boundaries, are orthogonal. Therefore, a single complex-valued neuron with n inputs creates decision boundaries that consist of two orthogonal hypersurfaces in the real 2.

1.3.2 Complex Activation Functions

Real-valued non-linear activations are incapable of maintaining the magnitude and phase information of complex-valued inputs. This inability highlights the need for complex-valued activations in neural networks, particularly to preserve the intricate relationships embedded in complex data's magnitude and phase(Mahesh Mohan M R 2023). Such complex activations are essential for accurately handling and interpreting the characteristics of complex-valued inputs.

1.3.3 Robustness to Noise

Complex-valued neural networks (CV-CNNs) are robust to noise and data distortions, a key advantage in processing real-world data. Their ability to handle complex representations allows them to discern and filter out irrelevant fluctuations in the data more effectively(Chakraborty et al. 2019). This feature is particularly valuable in applications where signal integrity is crucial, such as in medical imaging or telecommunications. By maintaining the integrity of the underlying data structure, CV-CNNs can deliver more accurate and reliable outputs, even in challenging noisy environments.

Chapter 2

Optimizing Complex-valued NNs: Challenges and Solution

2.1 Complex Calculus - Preliminaries

We start by stating some known results from complex functions theory. Throughout this section we use the following notations for complex numbers:

$$z = x + iy \in \mathbb{C} \qquad x, y \in \mathbb{R}$$

And for complex functions:

$$f : \mathbb{C} \to \mathbb{C}$$

$$f(z) = u(z) + iv(z)$$

$$u, v : \mathbb{R} \to \mathbb{R}$$

First, we point out that the complex field $\mathbb{C}$ cannot be ordered in a meaningful way, i.e. there is no total ordering of $\mathbb{C}$ under which the axioms of an ordered field are met. One implication is that the loss function we wish to minimize has to be real valued. To that end, we follow with some needed background about real valued complex functions. We focus on differentiability, as it plays a key role in the optimization process.

Definition 1. *A complex function f is complex differentiable at z, with the derivative $f'(z)$, if the following limit exists*

$$f'(z) = \lim_{h \to 0} \frac{f(z+h) - f(z)}{h}$$

A function that is complex differentiable everywhere is called *entire*. A very useful equivalent definition is given by the Cauchy-Riemann equations.

Definition 2. *A complex function f is complex differentiable at point z if and only if u, v are differentiable (as real functions) there, and the Cauchy-Riemann equations hold at z:*

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y} \qquad \frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x}$$

2.2 Challenges in Optimizing Complex Valued NNs

For gradient descent, we require the gradient to exist differentiability, which means the complex function should be holomorphic (analytic) therefore complex differentiable. But most complex functions are not analytic. Also, if the loss function f is real valued, namely $f(z) = u(z)$, then the CR equations reduce to

$$\frac{\partial u}{\partial x} = \frac{\partial u}{\partial y} = 0$$

If such an f is entire, it is constant. Another result implied by the above is the Liouville Theorem which states that an entire function that is bounded everywhere is constant.

Complex derivative can't be used. Wirtinger derivative provides an alternative approach to complex derivative.

2.3 Wirtinger Calculus

In the following part of this section we present the Wirtinger derivatives, which will be used to adjust gradient based methods to the complex domain. First, we define the differentials with respect to the variables z and its conjugate z^*:

It is evident that if neither the Cauchy Riemann conditions, nor the conjugate Cauchy-Riemann conditions are satisfied for a function f, then the complex derivatives cannot be exploited and the function cannot be expressed neither in terms of h or h^*, as in the case of complex or conjugate-complex differentiable functions. Nevertheless, if f is differentiable in the real sense (i.e., u and v have partial derivatives), we may still find a form of Taylor's series expansion u, v.

Proof. Consider the first order Taylor expansions of u and v at $c = c_1 + ic_2 = (c_1, c_2)$

$$u(c+h) = u(c) + \frac{\partial u}{\partial x}(c)h_1 + \frac{\partial u}{\partial y}(c)h_2 + o(|h|),$$
$$v(c+h) = v(c) + \frac{\partial v}{\partial x}(c)h_1 + \frac{\partial v}{\partial y}(c)h_2 + o(|h|).$$

Multiplying the second relation by i and adding it to the first one, we take:

$$f(c+h) = f(c) + \left(\frac{\partial u}{\partial x}(c) + i\frac{\partial v}{\partial x}(c)\right) h_1 + \left(\frac{\partial u}{\partial y}(c) + i\frac{\partial v}{\partial y}(c)\right) h_2 + o(|h|).$$

To simplify the notation we may define

$$\frac{\partial f}{\partial x}(c) = \frac{\partial u}{\partial x}(c) + i\frac{\partial v}{\partial x}(c) \text{ and } \frac{\partial f}{\partial y}(c) = \frac{\partial u}{\partial y}(c) + i\frac{\partial v}{\partial y}(c)$$

and obtain:

$$f(c+h) = f(c) + \frac{\partial f}{\partial x}(c)h_1 + \frac{\partial f}{\partial y}(c)h_2 + o(|h|).$$

Next, we substitute h_1 and h_2 using the relations $h_1 = \frac{h+h^*}{2}$ and $h_2 = \frac{h-h^*}{2i}$.

$$\begin{aligned} f(c+h) &= f(c) + \frac{1}{2}\left(\frac{\partial f}{\partial x}(c) + \frac{1}{i}\frac{\partial f}{\partial y}(c)\right) h + \frac{1}{2}\left(\frac{\partial f}{\partial x}(c) - \frac{1}{i}\frac{\partial f}{\partial y}(c)\right) h^* + o(|h|) \\ &= f(c) + \frac{1}{2}\left(\frac{\partial f}{\partial x}(c) - i\frac{\partial f}{\partial y}(c)\right) h + \frac{1}{2}\left(\frac{\partial f}{\partial x}(c) + i\frac{\partial f}{\partial y}(c)\right) h^* + o(|h|). \end{aligned}$$

It will be shown this equation is essential for the development of Wirtinger's calculus.

$$f(c+h) = f(c) + \frac{1}{2}\left(\frac{\partial f}{\partial z}(c) - i\frac{\partial f}{\partial z^*}(c)\right) h + \frac{1}{2}\left(\frac{\partial f}{\partial z}(c) + i\frac{\partial f}{\partial z^*}(c)\right) h^* + o(|h|). \quad (2.1)$$

One may notice that in the more general case, where f is real-differentiable, it's Taylor's expansion is casted in terms of both h and h^*. This can be generalized for higher order Taylor's expansion formulas following the same rationale. Observe also that, if f is complex or conjugate-complex differentiable, this relation degenerates (due to the Cauchy Riemann conditions) to the respective Taylor's expansion formula. In this context, the following definitions come naturally.

We define the Wirtinger's derivative (or W-derivative for short) of f at c as follows

$$\frac{\partial f}{\partial z}(c) = \frac{1}{2}\left(\frac{\partial}{\partial x}(c) - i\frac{\partial}{\partial y}(c)\right) = \frac{1}{2}\left(\frac{\partial u}{\partial x}(c) + \frac{\partial v}{\partial y}(c)\right) + \frac{i}{2}\left(\frac{\partial v}{\partial x}(c) - \frac{\partial u}{\partial y}(c)\right).$$

Consequently, the conjugate Wirtinger's derivative (or CW-derivative for short) of f at c is defined by:

$$\frac{\partial f}{\partial z^*}(c) = \frac{1}{2}\left(\frac{\partial f}{\partial x}(c) + i\frac{\partial f}{\partial y}(c)\right) = \frac{1}{2}\left(\frac{\partial u}{\partial x}(c) - \frac{\partial v}{\partial y}(c)\right) + \frac{i}{2}\left(\frac{\partial v}{\partial x}(c) + \frac{\partial u}{\partial y}(c)\right).$$

Note that both the W-derivative and the CW-derivative exist, if f is differentiable in the real sense. In view of these new definitions, equation may now be recasted as follows

$$f(c+h) = f(c) + \frac{\partial f}{\partial z}(c)h + \frac{\partial f}{\partial z^*}(c)h^* + o(|h|).$$

Definition 3. *The Wirtinger derivatives operators are*

$$\frac{\partial}{\partial z} := \frac{1}{2}\left(\frac{\partial}{\partial x} - i\frac{\partial}{\partial y}\right)$$

$$\frac{\partial}{\partial z^*} := \frac{1}{2}\left(\frac{\partial}{\partial x} + i\frac{\partial}{\partial y}\right)$$

The Wirtinger derivatives have some desirable properties. For one, z, z^* are independent variables as

$$\frac{\partial z}{\partial z^*} = \frac{\partial z^*}{\partial z} = 0$$

Also, some dual connections with the conjugate hold for the derivatives as well,

$$\left(\frac{\partial f^*}{\partial z}\right) = \left(\frac{\partial f}{\partial z^*}\right)^*, \quad \left(\frac{\partial f}{\partial z}\right) = \left(\frac{\partial f^*}{\partial z^*}\right)^*$$

Using the Wirtinger derivatives, we can express the total differential of any complex valued function f.

Remark 1. *The differential df of a complex-valued function $f(z) : A \to \mathbb{C}$ with $A \subseteq \mathbb{C}$ can be expressed as*

$$df = \frac{\partial f}{\partial z}dz + \frac{\partial f}{\partial z^*}dz^*$$

Proof. Consider the bivariate functions $F : \mathbb{R}^2 \to \mathbb{C}$ and $u, v : \mathbb{R}^2 \to \mathbb{R}$ associated to $f(z)$ by

$$\forall z = x + iy, \quad F(x, y) = u(x, y) + iv(x, y) = f(z)$$

The total differential of F is given by

$$dF = \frac{\partial F}{\partial x}dx + \frac{\partial F}{\partial y}dy = \frac{\partial u}{\partial x}dx + \frac{\partial v}{\partial x}dy + i\left(\frac{\partial u}{\partial y}dy + \frac{\partial v}{\partial y}dy\right)$$

□

By using the differentials defined above, we can write

$$dx = \frac{1}{2}(dz + dz^*), \quad dy = \frac{1}{2i}(dz - dz^*)$$

Obtaining

$$\begin{aligned} dF &= \frac{1}{2}\left[\left(\frac{\partial u}{\partial x} + i\frac{\partial v}{\partial x}\right) + i\left(\frac{\partial u}{\partial y} - i\frac{\partial v}{\partial y}\right)\right]dz + \frac{1}{2}\left[\left(\frac{\partial u}{\partial x} + i\frac{\partial v}{\partial x}\right) - i\left(\frac{\partial u}{\partial y} - i\frac{\partial v}{\partial y}\right)\right]dz^* \\ &= \frac{\partial f}{\partial z}dz + \frac{\partial f}{\partial z^*}dz^* \end{aligned}$$

2.4 Wirtinger derivative properties

Remark 2. *If u, v is differentiable at a point c, then the following equality holds:*

$$\left(\frac{\partial f}{\partial z}(c)\right)^* = \frac{\partial f^*}{\partial z^*}(c). \tag{2.2}$$

Proof. The conjugate of the Wirtinger derivative of f at c, when f is real-differentiable, can be computed by taking the complex conjugate of the Wirtinger derivative expression, leading to:

$$\begin{aligned} \left(\frac{\partial f}{\partial z}(c)\right)^* &= \frac{1}{2}\left(\frac{\partial u}{\partial x}(c) + \frac{\partial v}{\partial y}(c) - i\left(\frac{\partial v}{\partial x}(c) - \frac{\partial u}{\partial y}(c)\right)\right) \\ &= \frac{1}{2}\left(\frac{\partial u}{\partial x}(c) - \frac{\partial v}{\partial y}(c)\right) + \frac{i}{2}\left(\frac{\partial v}{\partial x}(c) + \frac{\partial u}{\partial y}(c)\right) \\ &= \frac{\partial f^*}{\partial z^*}(c). \end{aligned}$$

This demonstrates the equivalence of the conjugate of the Wirtinger derivative to the Wirtinger derivative of the conjugate function. □

Remark 3. *If u, v is differentiable at a point c, then the following equality holds:*

$$\left(\frac{\partial f}{\partial z^*}(c)\right)^* = \frac{\partial f^*}{\partial z}(c). \tag{2.3}$$

Proof. Applying a similar reasoning as above, the conjugate of the conjugate Wirtinger derivative is obtained by:

$$\begin{aligned}\left(\frac{\partial f}{\partial z^*}(c)\right)^* &= \frac{1}{2}\left(\frac{\partial u}{\partial x}(c) - \frac{\partial v}{\partial y}(c) + i\left(\frac{\partial v}{\partial x}(c) + \frac{\partial u}{\partial y}(c)\right)\right) \\ &= \frac{1}{2}\left(\frac{\partial u}{\partial x}(c) + \frac{\partial v}{\partial y}(c)\right) - \frac{i}{2}\left(\frac{\partial u}{\partial y}(c) - \frac{\partial v}{\partial x}(c)\right) \\ &= \frac{\partial f^*}{\partial z}(c).\end{aligned}$$

This result complements the previous one, showing the symmetry in the behavior of the Wirtinger derivatives with respect to complex conjugation. □

Remark 4. ***Linearity Property of Wirtinger Derivatives:***
Given two complex functions $f(z)$ and $g(z)$, and any two complex constants a and b, the linearity property of Wirtinger derivatives is expressed as:

1. *For the derivative with respect to z:*

$$\frac{\partial}{\partial z}(af(z) + bg(z)) = a\frac{\partial f}{\partial z} + b\frac{\partial g}{\partial z}$$

2. *For the derivative with respect to the conjugate of z (denoted as z^*):*

$$\frac{\partial}{\partial^*}(af(z) + bg(z)) = a\frac{\partial f}{\partial z^*} + b\frac{\partial g}{\partial z^*}$$

Proof. Let $f(z) = f(x, y) = u_f(x, y) + iv_f(x, y), g(z) = g(x, y) = u_g(x, y) + iv_g(x, y)$ be two complex functions and $\alpha, \beta \in \mathbb{C}$, such that $\alpha = \alpha_1 + i\alpha_2, \beta = \beta_1 + i\beta_2$. Then

$$\begin{aligned}r(z) =& \alpha f(z) + \beta g(z) = (\alpha_1 + i\alpha_2)\left(u_f(x, y) + iv_f(x, y)\right) + (\beta_1 + i\beta_2)\left(u_g(x, y) + iv_g(x, y)\right) \\ &= (\alpha_1 u_f(x, y) - \alpha_2 v_f(x, y) + \beta_1 u_g(x, y) - \beta_2 v_g(x, y)) \\ &\quad + i\left(\alpha_1 v_f(x, y) + \alpha_2 u_f(x, y) + \beta_1 v_g(x, y) + \beta_2 u_g(x, y)\right).\end{aligned}$$

Thus, the W-derivative of r will be given by:

$$
\begin{aligned}
\frac{\partial r}{\partial z}(c) =& \frac{1}{2}\left(\frac{\partial u_r}{\partial x}(c) + \frac{\partial v_r}{\partial y}(c)\right) + \frac{i}{2}\left(\frac{\partial v_r}{\partial x}(c) - \frac{\partial u_r}{\partial y}(c)\right) \\
=& \frac{1}{2}\left(\alpha_1 \frac{\partial u_f}{\partial x}(c) - \alpha_2 \frac{\partial v_f}{\partial x}(c) + \beta_1 \frac{\partial u_g}{\partial x}(c) - \beta_2 \frac{\partial v_g}{\partial x}(c) + \alpha_1 \frac{\partial v_f}{\partial y}(c) + \alpha_2 \frac{\partial u_f}{\partial y}(c)\right. \\
&+ \beta_1 \frac{\partial v_g}{\partial y}(c) + \beta_2 \frac{\partial u_g}{\partial y}(c) + \frac{i}{2}\left(\alpha_1 \frac{\partial v_f}{\partial x}(c) + \alpha_2 \frac{\partial u_f}{\partial x}(c) + \beta_1 \frac{\partial v_g}{\partial x}(c) + \beta_2 \frac{\partial u_g}{\partial x}(c) - \alpha_1 \frac{\partial u_f}{\partial y}(c)\right. \\
=& \frac{1}{2}(\alpha_1 + i\alpha_2)\frac{\partial u_f}{\partial x}(c) + \frac{i}{2}(\alpha_1 + i\alpha_2)\frac{\partial v_f}{\partial x}(c) + \frac{1}{2}(\beta_1 + i\beta_2)\frac{\partial u_g}{\partial x}(c) + \frac{i}{2}(\beta_1 + i\beta_2)\frac{\partial v_g}{\partial x}(c) \\
&+ \frac{1}{2}(\alpha_1 + i\alpha_2)\frac{\partial v_f}{\partial y}(c) - \frac{i}{2}(\alpha_1 + i\alpha_2)\frac{\partial u_f}{\partial y}(c) + \frac{1}{2}(\beta_1 + i\beta_2)\frac{\partial v_g}{\partial y}(c) - \frac{i}{2}(\beta_1 + i\beta_2)\frac{\partial u_g}{\partial y}(c \\
=& \alpha\left(\frac{1}{2}\left(\frac{\partial u_f}{\partial x}(c) + \frac{\partial v_f}{\partial y}(c)\right) + \frac{i}{2}\left(\frac{\partial v_f}{\partial x}(c) - \frac{\partial u_f}{\partial y}(c)\right)\right) \\
&+ \beta\left(\frac{1}{2}\left(\frac{\partial u_g}{\partial x}(c) + \frac{\partial v_g}{\partial y}(c)\right) + \frac{i}{2}\left(\frac{\partial v_g}{\partial x}(c) - \frac{\partial u_g}{\partial y}(c)\right)\right) \\
=& \alpha \frac{\partial f}{\partial z}(c) + \beta \frac{\partial g}{\partial z}(c).
\end{aligned}
$$

Similar to the first part, the linearity of the partial derivatives with respect to x and y ensures that the Wirtinger derivative with respect to $\overline{z}$ is also linear.

□

Remark 5. ***Product Rule:***
If f, g are differentiable in the real sense at c, then

$$
\begin{aligned}
\frac{\partial (f \cdot g)}{\partial z}(c) &= \frac{\partial f}{\partial z}(c) g(c) + f(c) \frac{\partial g}{\partial z}(c), \\
\frac{\partial (f \cdot g)}{\partial z^*}(c) &= \frac{\partial f}{\partial z^*}(c) g(c) + f(c) \frac{\partial g}{\partial z^*}(c).
\end{aligned}
$$

Proof. Let $f(z) = f(x, y) = u_f(x, y) + iv_f(x, y), g(z) = g(x, y) = u_g(x, y) + iv_g(x, y)$ be two complex functions differentiable at c. Consider the complex function r defined as $r(z) = f(z)g(z)$. Then

$$
r(z) = (u_f(z) + iv_g(z))(u_g(z) + iv_g(z)) = (u_f u_g - v_f v_g) + i(u_f v_g + v_f u_g).
$$

Hence the W-derivative of r at c is given by:

$$
\begin{aligned}
\frac{\partial r}{\partial z}(c) =& \frac{1}{2}\left(\frac{\partial u_r}{\partial x}(c)+\frac{\partial v_r}{\partial y}(c)\right)+\frac{i}{2}\left(\frac{\partial v_r}{\partial x}(c)-\frac{\partial u_r}{\partial y}(c)\right)\\
=& \frac{1}{2}\left(\frac{\partial\left(u_f u_g - v_f v_g\right)}{\partial x}(c)+\frac{\partial\left(u_f v_g + v_f u_g\right)}{\partial y}(c)\right)+\frac{i}{2}\left(\frac{\partial\left(u_f v_g + v_f u_g\right)}{\partial x}(c)-\frac{\partial\left(u_f u_g - v_f v_g\right)}{\partial y}(c)\right)\\
=& \frac{1}{2}\left(\frac{\partial\left(u_f u_g\right)}{\partial x}(c)-\frac{\partial\left(v_f v_g\right)}{\partial x}(c)+\frac{\partial\left(u_f v_g\right)}{\partial y}(c)+\frac{\partial\left(v_f u_g\right)}{\partial y}(c)\right)\\
&+\frac{i}{2}\left(\frac{\partial\left(u_f v_g\right)}{\partial x}(c)+\frac{\partial\left(v_f u_g\right)}{\partial x}(c)-\frac{\partial\left(u_f u_g\right)}{\partial y}(c)+\frac{\partial\left(v_f v_g\right)}{\partial y}(c)\right)=\\
&+\frac{\partial u_f}{\partial y}(c)v_g(c)+\frac{\partial v_g}{\partial y}(c)u_f(c)+\frac{\partial v_f}{\partial y}(c)u_g(c)+\frac{\partial u_g}{\partial y}(c)v_f(c)\Bigg)\\
&+\frac{i}{2}\Bigg(\frac{\partial u_f}{\partial x}(c)v_g(c)+\frac{\partial v_g}{\partial x}(c)u_f(c)+\frac{\partial v_f}{\partial x}(c)u_g(c)+\frac{\partial u_g}{\partial x}(c)v_f(c)\\
&-\frac{\partial u_f}{\partial y}(c)u_g(c)-\frac{\partial u_g}{\partial y}(c)u_f(c)+\frac{\partial v_f}{\partial y}(c)v_g(c)+\frac{\partial v_g}{\partial y}(c)v_f(c)\Bigg).
\end{aligned}
$$

After factorization we obtain:

$$
\begin{aligned}
\frac{\partial r}{\partial z}(c) =& u_g(c)\left(\frac{1}{2}\left(\frac{\partial u_f}{\partial x}(c)+\frac{\partial v_f}{\partial y}(c)\right)+\frac{i}{2}\left(\frac{\partial v_f}{\partial x}(c)-\frac{\partial u_f}{\partial y}(c)\right)\right)\\
&+v_g(c)\left(\frac{1}{2}\left(-\frac{\partial v_f}{\partial x}(c)+\frac{\partial u_f}{\partial y}(c)\right)+\frac{i}{2}\left(\frac{\partial u_f}{\partial x}(c)+\frac{\partial v_f}{\partial y}(c)\right)\right)\\
&+u_f(c)\left(\frac{1}{2}\left(\frac{\partial u_g}{\partial x}(c)+\frac{\partial v_g}{\partial y}(c)\right)+\frac{i}{2}\left(\frac{\partial v_g}{\partial x}(c)-\frac{\partial u_g}{\partial y}(c)\right)\right)\\
&+v_f(c)\left(\frac{1}{2}\left(-\frac{\partial v_g}{\partial x}(c)+\frac{\partial u_g}{\partial y}(c)\right)+\frac{i}{2}\left(\frac{\partial u_g}{\partial x}(c)+\frac{\partial v_g}{\partial y}(c)\right)\right).
\end{aligned}
$$

Applying the simple rule $1/i=-i$, we take:

$$
\begin{aligned}
\frac{\partial T}{\partial z}(c) =& u_g(c)\left(\frac{1}{2}\left(\frac{\partial u_f}{\partial x}(c)+\frac{\partial v_f}{\partial y}(c)\right)+\frac{i}{2}\left(\frac{\partial v_f}{\partial x}(c)-\frac{\partial u_f}{\partial y}(c)\right)\right)\\
&+iv_g(c)\left(\frac{1}{2}\left(\frac{\partial u_f}{\partial x}(c)+\frac{\partial v_f}{\partial y}(c)\right)+\frac{i}{2}\left(\frac{\partial v_f}{\partial x}(c)-\frac{\partial u_f}{\partial y}(c)\right)\right)\\
&+u_f(c)\left(\frac{1}{2}\left(\frac{\partial u_g}{\partial x}(c)+\frac{\partial v_g}{\partial y}(c)\right)+\frac{i}{2}\left(\frac{\partial v_g}{\partial x}(c)-\frac{\partial u_g}{\partial y}(c)\right)\right)\\
&+iv_f(c)\left(\frac{1}{2}\left(\frac{\partial u_g}{\partial x}(c)+\frac{\partial v_g}{\partial y}(c)\right)+\frac{i}{2}\left(\frac{\partial v_g}{\partial x}(c)-\frac{\partial u_g}{\partial y}(c)\right)\right)\\
=& \left(u_g(c)+iv_g\right)\frac{\partial f}{\partial z}(c)+\left(u_f(c)+iv_f\right)\frac{\partial g}{\partial z}(c),
\end{aligned}
$$

which gives the result. The product rule of the CW-derivative follows from the product rule of the W-derivative :

$$\begin{aligned}\frac{\partial(fg)}{\partial z^*}(c) &= \left(\frac{\partial(fg)^*}{\partial z}(c)\right)^* = \left(\frac{\partial f^*}{\partial z}(c)g^*(c) + \frac{\partial g^*}{\partial z}(c)f^*(c)\right)^* \\ &= \frac{\partial f}{\partial z^*}(c)g(c) + \frac{\partial g}{\partial z^*}(c)f(c)\end{aligned}$$

□

Remark 6. *If f is differentiable in c and $f(c) \neq 0$, then*

$$\begin{aligned}\frac{\partial\left(\frac{1}{f}\right)}{\partial z}(c) &= -\frac{\frac{\partial f}{\partial z}(c)}{f^2(c)}, \\ \frac{\partial\left(\frac{1}{f}\right)}{\partial z^*}(c) &= -\frac{\frac{\partial f}{\partial z^*}(c)}{f^2(c)}.\end{aligned}$$

Proof. Let $f(z) = u_f(x,y) + iv_f(x,y)$ be a complex function, differentiable in the real sense at c, such that $f(c) \neq 0$. Consider the function $r(z) = 1/f(z)$. Then

$$r(z) = \frac{u_f(z)}{u_f^2(z) + v_f^2(z)} - i\frac{v_f(z)}{u_f^2(z) + v_f^2(z)}.$$

For the partial derivatives of u_r, v_r we have:

$$\begin{aligned}\frac{\partial u_r}{\partial x}(c) &= \frac{\frac{\partial u_f}{\partial x}(c)\left(u_f^2(c) + v_f^2(c)\right) - 2u_f^2(c)\frac{\partial u_f}{\partial x}(c) - 2u_f(c)v_f(c)\frac{\partial v_f}{\partial x}(c)}{\left(u_f^2(c) + v_f^2(c)\right)^2} \\ \frac{\partial u_r}{\partial y}(c) &= \frac{\frac{\partial u_f}{\partial y}(c)\left(u_f^2(c) + v_f^2(c)\right) - 2u_f^2(c)\frac{\partial u_f}{\partial y}(c) - 2u_f(c)v_f(c)\frac{\partial v_f}{\partial y}(c)}{\left(u_f^2(c) + v_f^2(c)\right)^2} \\ \frac{\partial v_r}{\partial x}(c) &= -\frac{\frac{\partial v_f}{\partial x}(c)\left(u_f^2(c) + v_f^2(c)\right) - 2v_f^2(c)\frac{\partial v_f}{\partial x}(c) - 2u_f(c)v_f(c)\frac{\partial u_f}{\partial x}(c)}{\left(u_f^2(c) + v_f^2(c)\right)^2} \\ \frac{\partial v_r}{\partial y}(c) &= -\frac{\frac{\partial v_f}{\partial y}(c)\left(u_f^2(c) + v_f^2(c)\right) - 2v_f^2(c)\frac{\partial v_f}{\partial y}(c) - 2u_f(c)v_f(c)\frac{\partial u_f}{\partial y}(c)}{\left(u_f^2(c) + v_f^2(c)\right)^2}\end{aligned}$$

Therefore,

$$\begin{aligned}
\frac{\partial r}{\partial z}(c) =& \frac{1}{2}\left(\frac{\partial u_r}{\partial x}(c) + \frac{\partial v_r}{\partial y}(c)\right) + \frac{i}{2}\left(\frac{\partial v_r}{\partial x}(c) - \frac{\partial u_r}{\partial y}(c)\right) \\
=& \frac{1}{2\left(u_f^2(c)+v_f^2(c)\right)^2}\left(\frac{\partial u_f}{\partial x}(c)\left(-u_f^2(c)+v_f^2(c)+2iu_f(c)v_f(c)\right)\right. \\
&+ \frac{\partial v_f}{\partial x}(c)\left(-2u_f(c)v_f(c) - iu_f^2 + iv_f^2(c)\right) \\
&\left.+\frac{\partial v_f}{\partial y}(c)\left(-u_f^2(c)+v_f^2(c)+2iu_f(c)v_f(c)\right) + \frac{\partial u_f}{\partial y}(c)\left(2u_f(c)v_f(c) + iu_f^2(c) - iv_f^2(c)\right)\right) \\
=& \frac{u_f^2(c) - v_f^2(c) - 2iu_f(c)v_f(c)}{2\left(u_f^2(c)+v_f^2(c)\right)^2}\left(-\left(\frac{\partial u_f}{\partial x}(c) + \frac{\partial v_f}{\partial y}(c)\right) - i\left(\frac{\partial v_f}{\partial x}(c) - \frac{\partial u_f}{\partial y}(c)\right)\right) \\
=& -\frac{\frac{\partial f}{\partial z}(c)}{f^2(c)}
\end{aligned}$$

To prove the corresponding rule of the CW-derivative we apply the reciprocal rule of the W-derivative as well as Propositions:

$$\frac{\partial\left(\frac{1}{f}\right)}{\partial z^*}(c) = \left(\frac{\partial\left(\frac{1}{f^*}\right)}{\partial z}(c)\right)^* = \left(-\frac{\frac{\partial f^*}{\partial z}(c)}{\left(f^*(c)\right)^2}\right)^* = -\frac{\frac{\partial f}{\partial z^*}(c)}{(f(c))^2}$$

□

Remark 7. ***Division Rule****:*
If f, g are differentiable in the real sense i.e u_f, u_g, v_f, v_g are differentiable at c and $g(c) \neq 0$, then

$$\frac{\partial\left(\frac{f}{g}\right)}{\partial z}(c) = \frac{\frac{\partial f}{\partial z}(c)g(c) - f(c)\frac{\partial g}{\partial z}(c)}{g^2(c)},$$
$$\frac{\partial\left(\frac{f}{g}\right)}{\partial z}(c) = \frac{\frac{\partial f}{\partial z^*}(c)g(c) - f(c)\frac{\partial g}{\partial z^*}(c)}{g^2(c)}.$$

Proof. It follows immediately from the multiplication rule and the reciprocal rule $\left(\frac{f(c)}{g(c)} = f(c)\cdot\frac{1}{g(c)}\right)$.

□

Remark 8. ***Chain Rule***

If f, g are differentiable in the real sense i.e u_f, u_g, v_f, v_g are differentiable at c and $g(c) \neq 0$, then

$$\frac{\partial g \circ f}{\partial z}(c) = \frac{\partial g}{\partial z}(f(c))\frac{\partial f}{\partial z}(c) + \frac{\partial g}{\partial z^*}(f(c))\frac{\partial f^*}{\partial z}(c),$$
$$\frac{\partial g \circ f}{\partial z^*}(c) = \frac{\partial g}{\partial z}(f(c))\frac{\partial f}{\partial z^*}(c) + \frac{\partial g}{\partial z^*}(f(c))\frac{\partial f^*}{\partial z^*}(c).$$

Proof. Given, $f : \mathbb{C} \mapsto \mathbb{C}$ and $g : \mathbb{C} \mapsto \mathbb{C}$, we would like to obtain identities for $\frac{\partial(f \circ g)}{dz}$ and $\frac{\partial(f \circ g)}{dz^*}$. Let's write the total differential for $g(z)$:

$$dg = \frac{\partial g}{\partial z}dz + \frac{\partial g}{\partial z^*}dz^*$$

Then the total differential for $g^*(z)$:

$$dg^* = \frac{\partial g^*}{\partial z}dz + \frac{\partial g^*}{\partial z^*}dz^*$$

Let's write the total differential for $f(g)$:

$$d(f \circ g) = \frac{\partial f}{\partial g}dg + \frac{\partial f}{\partial g^*}dg^*$$

Put dg and dg^* in to the equation:

$$d(f \circ g) = \left(\frac{\partial f}{dg}\frac{\partial g}{dz} + \frac{\partial f}{dg^*}\frac{\partial g^*}{dz}\right)dz + \left(\frac{\partial f}{dg}\frac{\partial g}{dz^*} + \frac{\partial f}{dg^*}\frac{\partial g^*}{dz^*}\right)dz^*$$

So, these are the chain rules and they are exactly same with the one we know for real functions! (as thinking $f(g, g^*), g(z, z^*)$ are real multi-variable functions)

$$\frac{\partial(f \circ g)}{dz} = \frac{\partial f}{dg}\frac{\partial g}{dz} + \frac{\partial f}{dg^*}\frac{\partial g^*}{dz}$$
$$\frac{\partial(f \circ g)}{dz^*} = \frac{\partial f}{dg}\frac{\partial g}{dz^*} + \frac{\partial f}{dg^*}\frac{\partial g^*}{dz^*}$$

□

2.5 Optimizing single variable complex valued CNN

Remark 9. *For the aforementioned f, the steepest ascent at point z is obtained by*

$$dz = \frac{\partial f}{\partial z^*} ds$$

Where ds is a real-valued differential. Therefore, the steepest ascent's direction is

$$\frac{\partial f}{\partial z^*}$$

Proof. An important consequence is that if f is a real valued function defined on $\mathbb{C}$, then its first order Taylor's expansion at z is given by:

$$\begin{aligned} f(c+h) &= f(c) + \frac{\partial f}{\partial z}(c)h + \frac{\partial f}{\partial z^*}(c)h^* + o(|h|) \\ &= f(c) + \frac{\partial f}{\partial z}(c)h + \left(\frac{\partial f}{\partial z}(c)h\right)^* + o(|h|) \\ &= f(c) + \Re\left[\frac{\partial f}{\partial z}(c)h\right] + o(|h|). \end{aligned}$$

However, in view of the Cauchy Riemann inequality we have:

$$\begin{aligned} \Re\left[\frac{\partial f}{\partial z}(c)h\right] = \Re\left[\left\langle h, \left(\frac{\partial f}{\partial z}(c)\right)^*\right\rangle_{\mathbb{C}}\right] &\leq \left|\left\langle h, \left(\frac{\partial f}{\partial z}(c)\right)^*\right\rangle_{\mathbb{C}}\right| \\ &\leq |h| \left|\frac{\partial f}{\partial z^*}(c)\right|. \end{aligned}$$

The equality in the above relationship holds, if $h \uparrow \frac{\partial f}{\partial z^*}$ (Both are antiparallel). Hence, the direction of increase of f is $\frac{\partial f}{\partial z^*}$. Direction of steepest descent is -$\frac{\partial f}{\partial z^*}$. Therefore, any gradient descent based algorithm minimizing $f(z)$ is based on the update scheme:

$$z_n = z_{n-1} - \mu \cdot \frac{\partial f}{\partial z^*}(z_{n-1})$$

- □

2.6 Multivariate Wiringer derivative

Definition 4. *Consider a complex function f that takes n complex variables $z_1, z_2, \ldots, z_n$ and maps them to a complex number. The variables z_j can be expressed in terms of their real and imaginary parts: $z_j = x_j + iy_j$. The total derivative of f is:*

$$df = f(z_1 + h_1, z_2 + h_2, \ldots, z_n + h_n) - f(z_1, z_2, \ldots, z_n) \tag{2.4}$$

Let $u(x_1, y_1, \ldots, x_n, y_n)$ and $v(x_1, y_1, \ldots, x_n, y_n)$ represent the real and imaginary parts of f, respectively. Then, we can express the total derivative as:

$$\begin{aligned} du &= u(x_1 + a_1, y_1 + b_1, \ldots) - u(x_1, y_1, \ldots), \\ dv &= v(x_1 + a_1, y_1 + b_1, \ldots) - v(x_1, y_1, \ldots), \end{aligned} \tag{2.5}$$

where a_j and b_j are infinitesimal increments in the real and imaginary parts of z_j, respectively.

The change in f, denoted by df, can thus be written as:

$$df = \sum_{j=1}^{n} \left(\frac{\partial u}{\partial x_j} + i \frac{\partial v}{\partial x_j} \right) a_j + \left(\frac{\partial u}{\partial y_j} + i \frac{\partial v}{\partial y_j} \right) b_j \tag{2.6}$$

We can simplify this further by defining the Multivariate Wirtinger derivatives:

$$\frac{\partial f}{\partial z_j} := \frac{1}{2} \left(\frac{\partial u}{\partial x_j} - i \frac{\partial v}{\partial x_j} \right), \quad \frac{\partial f}{\partial \bar{z}_j} := \frac{1}{2} \left(\frac{\partial u}{\partial y_j} + i \frac{\partial v}{\partial y_j} \right) \tag{2.7}$$

Using these Wirtinger derivatives, we can express the total derivative as:

$$df = \sum_{j=1}^{n} \left(\frac{\partial f}{\partial z_j} h_j + \frac{\partial f}{\partial {z_j}^*} {h_j}^* \right) \tag{2.8}$$

where $h_j = a_j + ib_j$ and ${h_j}^*$ is its complex conjugate.

This form shows that the total derivative of a multivariate complex function can be decomposed into a sum of derivatives with respect to each complex variable and its conjugate.

For the given function f, the total derivative with respect to the complex variables z is then:

$$df = \sum_{j=1}^{n} \left(\frac{\partial f}{\partial z_j} dz_j + \frac{\partial f}{\partial {z_j}^*} d{z_j}^* \right) \tag{2.9}$$

where $dz_j = dx_j + idy_j$ and $d\bar{z}_j = dx_j - idy_j$.

2.7 Optimizing multi-variable complex valued function

Consider a multivariate complex function $f : \mathbb{C}^n \to \mathbb{R}$ and let $Z = (z_1, z_2, \ldots, z_n)$ be a point in the domain of f, where each z_j is a complex variable. The directional derivative of f at Z in the direction of a complex vector $h = (h_1, h_2, \ldots, h_n)$ is given by:

$$df = \sum_{j=1}^{n} \left(\frac{\partial f}{\partial z_j} h_j + \frac{\partial f}{\partial {z_j}^*} {h_j}^* \right). \tag{2.10}$$

Since f is a real valued loss function, it must hold that:

$$df = 2\mathrm{Re} \left(\sum_{j=1}^{n} \frac{\partial f}{\partial z_j} h_j \right). \tag{2.11}$$

The steepest ascent direction is the direction in which this derivative is maximized. Since f is real, the steepest ascent direction is obtained. The maximization of the directional derivative is equivalent to maximizing the real part of the inner product between the gradient vector and the direction vector h.

We define the gradient of f at Z with respect to the complex variables as $\nabla f(Z) = \left(\frac{\partial f}{\partial z_1}, \ldots, \frac{\partial f}{\partial z_n}\right)$, and its conjugate transpose as $\nabla f(Z)^* = \left(\frac{\partial f}{\partial {z_1}^*}, \ldots, \frac{\partial f}{\partial {z_n}^*}\right)^T$.

The steepest ascent direction at Z is then the direction of $\nabla f(Z)$. To move in the direction of steepest ascent, we update Z as follows:

$$Z_{i+1} = Z_i + \lambda \nabla f(Z_i)^* \tag{2.12}$$

where λ is a positive scalar that determines the step size.

In conclusion, the direction of steepest ascent for a real-valued function of complex variables is given by the gradient with respect to those variables λ.

So, for the steepest descent, the direction is $-\nabla f(Z_i)^*$. We do gradient descent as follows:

$$Z_{i+1} = Z_i - \lambda \nabla f(Z_i)^* \tag{2.13}$$

2.8 Conclusion

In this report, we have conducted a systematic study of optimizing complex-valued CNNs. We concluded that normal complex valued differentiation (or analyticity) cannot be used for optimizing complex-valued neural networks; instead, we need to consider a different calculus called Wirtinger calculus. We have derived useful properties of the Wirtinger calculus that is needed for back-propogation. The expressiveness of Wirtinger derivatives are wonderful. Apart from chain rule, every property with Wirtinger derivatives becomes same as that one learned in real calculus.

Chapter 3

Future Works

3.1 Extended Optimizers for CV-CNNs

Complex-valued convolutional neural networks (CV-CNNs) present unique challenges and opportunities for optimization algorithms. Traditional optimizers like Adam and AdaGrad have been primarily designed and tuned for real-valued neural networks. Adapting and extending these optimizers for CV-CNNs could lead to significant improvements in training efficiency and model performance. This involves:

- Developing complex-valued versions of gradient estimators and moment calculations.
- Adjusting learning rate schedules and other hyperparameters to accommodate the properties of complex numbers.
- Investigating the effects of complex-valued batch statistics on the optimization process.

Such advancements would not only enhance the training of CV-CNNs but also deepen our understanding of complex-valued optimization landscapes.

3.2 Batch Normalization for CV-CNNs

Batch normalization has been a pivotal technique in stabilizing and accelerating the training of real-valued neural networks. However, its direct application to complex-valued networks is not straightforward due to the intrinsic properties of complex numbers. Future research in this area could include:

- Exploring novel ways to normalize complex-valued activations, taking into account the interactions between their real and imaginary components.
- Investigating the effects of different normalization schemes on the phase and amplitude aspects of complex-valued feature maps.
- Developing normalization techniques that respect the topological and algebraic structure of the complex number space.

Such developments in batch normalization for CV-CNNs would be crucial for harnessing the full potential of complex-valued representations in deep learning architectures.

3.3 Initialization Strategies for CV-CNNs

Proper initialization of neural network weights is crucial to ensure efficient training and convergence. For CV-CNNs, this aspect becomes even more critical due to the complex nature of the parameters. Future research directions might include:

- Developing initialization methods that account for the phase and magnitude dynamics in complex-valued weights.
- Exploring the effects of different initialization schemes on the learning trajectory and stability of CV-CNNs.
- Examining the interplay between weight initialization and complex-valued activation functions.
- Investigating symmetry-breaking in initialization to avoid stagnation in suboptimal solutions peculiar to complex-valued optimization landscapes.

Advancements in initialization strategies for CV-CNNs will be instrumental in fully leveraging the capabilities of complex numbers in deep learning models.

Bibliography

Chakraborty, R., Wang, J. & Yu, S. X. (2019), ‘Sur-real: Frechet mean and distance transform for complex-valued deep learning’, *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops* pp. 0–0.

Hirose, A. (2009), Complex-valued neural networks: The merits and their origins, *in* ‘2009 International Joint Conference on Neural Networks’, pp. 1237–1244.

Lee, C., Hasegawa, H. & Gao, S. (2022), ‘Complex-valued neural networks: A comprehensive survey’, *IEEE/CAA Journal of Automatica Sinica* **9**(8), 1406–1426.

LIU, C.-A., xin CHEN, Z., SHAO, Y., song CHEN, J., Hasi, T. & zhu PAN, H. (2019), ‘Research advances of sar remote sensing for agriculture applications: A review’, *Journal of Integrative Agriculture* **18**(3), 506–525.
URL: *https://www.sciencedirect.com/science/article/pii/S2095311918620167*

Mahesh Mohan M R, Kartik Srivastava, N. A. (2023), ‘On complex-domain cnn representations for classifying real/complex-valued data’, *ICLR 2023* **50**, 291–303.
URL: *https://openreview.net/forum?id=SmjW4kKLjuU*

Nitta, T. (2003), ‘On the inherent property of the decision boundary in complex-valued neural networks’, *Neurocomputing* **50**, 291–303.
URL: *https://www.sciencedirect.com/science/article/pii/S0925231202005684*