5

Elements of $\mathbb{CR}$ Calculus

The design of adaptive learning algorithms is based on the minimisation of a suitable objective (cost) function, typically a function of the output error of an adaptive filter. This optimisation problem is well understood for real valued adaptive filters where, for instance, the steepest descent approach is based on the iteration

$$\mathbf{w}(k+1) = \mathbf{w}(k) - \mu \nabla_{\mathbf{w}} J(k)$$

where $\mathbf{w}(k)$ is the vector of filter coefficients, μ is a parameter (learning rate) and $J = \frac{1}{2}e^2(k)$ is the cost function, a quadratic function of the output error of the filter.

Although a formalism similar to that used for real valued adaptive filters can also be used for complex valued adaptive filters, notice that in this case the cost function $J(k) = \frac{1}{2}|e(k)|^2 = \frac{1}{2}e(k)e^*(k) = \frac{1}{2}(e_r^2 + e_i^2)$ is a real valued function of complex variable, which gives rise to several important issues:

- Standard complex differentiability is based on the Cauchy–Riemann equations and imposes a stringent structure on complex holomorphic functions;
- Cost functions are real functions of complex variable, that is $J : \mathbb{C} \mapsto \mathbb{R}$, and so they are not differentiable in the complex sense, the Cauchy–Riemann equations do not apply, and we need to develop alternative, more general and relaxed, ways of calculating their gradients;
- It is also desired that these generalised gradients are equivalent to standard complex gradients when applied to holomorphic (analytic) functions.

This chapter provides an overview of complex differentiability for both real valued and complex valued functions of complex variables. The concepts of complex continuity, differentiability and Cauchy–Riemann conditions are first introduced; this is followed by more general $\mathbb{R}$-derivatives. The duality between gradient calculation in $\mathbb{R}$ and $\mathbb{C}$ is then addressed and the so called $\mathbb{CR}$ calculus is introduced for general functions of complex variables. It is shown that the $\mathbb{CR}$ calculus provides a unified framework for computing the Jacobians, Hessians, and gradients of cost functions. This chapter will serve as a basis for the derivation of learning algorithms throughout this book.

Complex Valued Nonlinear Adaptive Filters: Noncircularity, Widely Linear and Neural Models
Danilo Mandic and Vanessa Su Lee Goh

5.1 Continuous Complex Functions

Complex analyticity and singularities of complex functions have been introduced in Section 4.1 and Section 4.1.1. To address complex continuity, consider complex functions $f : D \mapsto \mathbb{C}$, $D \subset \mathbb{C}$

$$f(z) = f(x, y) = u(x, y) + \jmath v(x, y) = \big(u(x, y), v(x, y)\big) \tag{5.1}$$

Limits of complex functions are defined similarly to those in $\mathbb{R}$, $\lim_{z\to z_0} f(z) = \varsigma$ that is, for any $\epsilon > 0$ there exists $\delta > 0$ such that $|f(z) - \varsigma| < \epsilon$ when $0 < |z - z_0| < \delta$.

To define a continuous complex function we need to show that at any $z_0 \in D$

$$\lim_{z\to z_0} f(z) = f(z_0) \tag{5.2}$$

5.2 The Cauchy–Riemann Equations

For a complex function $f(z) = u + \jmath v$, to be differentiable at z, limit (5.3) must converge to a unique complex number no matter how $\Delta z \to 0$. In other words, for $f(z)$ to be analytic, the limit

$$f'(z) = \lim_{\Delta x\to 0, \Delta y\to 0} \frac{\big[u\,(x + \Delta x, y + \Delta y) + \jmath v\,(x + \Delta x, y + \Delta y)\big] - \big[u(x, y) + \jmath v(x, y)\big]}{\Delta x + \jmath\Delta y} \tag{5.3}$$

must exist regardless of how Δz approaches zero.

It is convenient to consider the two following cases [206, 218]

Case 1: $\Delta y = 0$ and $\Delta x \to 0$, which yields

$$\begin{aligned} f'(z) &= \lim_{\Delta x\to 0} \frac{\big[u\,(x + \Delta x, y) + \jmath v\,(x + \Delta x, y)\big] - \big[u(x, y) + \jmath v(x, y)\big]}{\Delta x} \\ &= \lim_{\Delta x\to 0} \frac{u\,(x + \Delta x, y) - u(x, y)}{\Delta x} + \jmath \frac{v\,(x + \Delta x, y) - v(x, y)}{\Delta x} \\ &= \frac{\partial u(x, y)}{\partial x} + \jmath \frac{\partial v(x, y)}{\partial x} \end{aligned} \tag{5.4}$$

Case 2: $\Delta x = 0$ and $\Delta y \to 0$, which yields

$$\begin{aligned} f'(z) &= \lim_{\Delta y\to 0} \frac{\big[u\,(x, y + \Delta y) + \jmath v\,(x, y + \Delta y)\big] - \big[u(x, y) + \jmath v(x, y)\big]}{\jmath\Delta y} \\ &= \lim_{\Delta y\to 0} \frac{u\,(x, y + \Delta y) - u(x, y)}{\jmath\Delta y} + \frac{v\,(x, y + \Delta y) - v(x, y)}{\Delta y} \\ &= \frac{\partial v(x, y)}{\partial y} - \jmath \frac{\partial u(x, y)}{\partial y} \end{aligned} \tag{5.5}$$

For continuity (Section 5.1), the limits from Case 1 and Case 2 must be identical, which yields

$$\frac{\partial u(x, y)}{\partial x} = \frac{\partial v(x, y)}{\partial y}, \qquad \frac{\partial v(x, y)}{\partial x} = -\frac{\partial u(x, y)}{\partial y} \tag{5.6}$$

that is, the expressions for the Cauchy-Riemann equations. Therefore, for a function $f(z) : \mathbb{C} \mapsto \mathbb{C}$ to be holomorphic (analytic in z), the partial derivatives $\partial u(x, y)/\partial x$, $\partial u(x, y)/\partial y$, $\partial v(x, y)/\partial x$, and $\partial v(x, y)/\partial y$, must not only exist – they must also satisfy the Cauchy–Riemann (C–R) conditions. This imposes a great amount of structure on holomorphic functions, which may prove rather stringent in practical applications. To avoid this, a more relaxed definition of a derivative is introduced in Section 5.3, based on the duality between the spaces $\mathbb{C}$ and $\mathbb{R}^2$. A comprehensive account of complex vector and matrix differentiation is given in Appendix A.

The Jacobian matrix of a complex function $f(z) = u + \jmath v$, where $z = x + \jmath y$, is given by

$$\mathbf{J} = \begin{bmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y} \end{bmatrix} \tag{5.7}$$

If z and $f'(z)$ were vectors in $\mathbb{R}^2$, say $z = [x, y]$, $dz = [dz, dy]$, and $df(z) = [du, dv]$, they would have to satisfy

$$df(z) = f'(z)\, dz = dz\, f'(z)$$

As the multiplication in the complex domain is commutative, and a 2×2 dimensional Jacobian matrix $\mathbf{J} = f'(z)$ cannot premultiply a row vector dz, in general function $f'(z)$ cannot lie in the same space as z and $f(z)$, and hence the Jacobian matrix cannot be an arbitrary matrix. We have already shown (see also chapters 12 and 11) that special 2×2 matrices, such as

$$\begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} x & y \\ -y & x \end{bmatrix}$$

are algebraically *isomorphic* with complex variables. These matrices commute only with the matrices of their own kind, and hence the Jacobian matrix in Equation (5.7) has to be of this kind too, thus conforming with the Cauchy–Riemann equations.

5.3 Generalised Derivatives of Functions of Complex Variable

In practical applications, we often need to perform optimisation on complex functions which are not directly analytic in $\mathbb{C}$[1]. Frequently encountered complex functions of complex variable which are not analytic include those which depend on the complex conjugate, and those which use absolute values of complex numbers. For instance, the complex conjugate $f(z) = z^*$ is

[1]For instance, in the Wiener filtering problem, the aim is to find the set of coefficients which minimise the total error power, a real function of complex variable. This function clearly has a minimum, but is not differentiable in $\mathbb{C}$. Similar problems arise in power engineering [100].

not analytic, and the Cauchy–Riemann conditions are not satisfied, as can be seen from its Jacobian

$$\mathbf{J} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \tag{5.8}$$

This is also the case with the class of functions which depend on both $z = x + \jmath y$ and $z^* = x - \jmath y$, for instance

$$J(z, z^*) = zz^* = x^2 + y^2 \quad \Rightarrow \quad \mathbf{J} = \begin{bmatrix} 2x & 2y \\ 0 & 0 \end{bmatrix} \quad \Leftrightarrow \quad \frac{\partial u}{\partial x} \neq \frac{\partial v}{\partial y} \qquad \frac{\partial v}{\partial x} \neq -\frac{\partial u}{\partial y} \tag{5.9}$$

As a consequence, any polynomial in both z and z^*, or any polynomial depending on z^* alone, is not analytic. **Therefore, our usual cost function** $J(k) = \frac{1}{2}e(k)e^*(k) = \frac{1}{2}[e_r^2 + e_i^2]$, **a real function of a complex variable, is not analytic or differentiable in the complex sense, and does not satisfy the Cauchy–Riemann conditions.**

The Cauchy–Riemann conditions therefore impose a very stringent structure on functions of complex variables, and several attempts have been made to introduce more convenient derivatives. As functions of complex variable

$$f(z) \quad \leftrightarrow \quad g(x, y) \quad \text{(real valued bivariate)}$$

can also be viewed as functions of its real and imaginary components, it is natural to ask whether the rules of real gradient calculation may be somehow applied. This way, we may be able to replace the stringent conditions of the standard complex derivative ($\mathbb{C}$-derivative) of a holomorphic function $f : \mathbb{C} \to \mathbb{C}$ with the more relaxed conditions of a real derivative ($\mathbb{R}$-derivative) of bivariate function $g(x, y) : \mathbb{R}^2 \to \mathbb{R}^2$. For convenience, we would like the $\mathbb{R}$-derivative to be equivalent to the $\mathbb{C}$-derivative when applied to holomorphic functions.

Based on our earlier examples of nonanalytic functions $f(z) = z^*$ and $f(z) = |z|^2 = zz^*$, observe that:-

- A function $f(z)$ can be non-holomorphic in the complex variable $z = x + \jmath y$, but still be analytic in real variables x and y, as for instance, $f(z) = z^*$ and $f(z) = zz^* = x^2 + y^2$;
- Both $f(z) = z^*$ and $f(z) = zz^*$ are holomorphic in z for $z^* = const$, and are also holomorphic in z^* when $z = const$.

The main idea behind $\mathbb{CR}$ calculus (also known as Wirtinger calculus[2] [313]) and Brandwood's result [35], is to introduce so-called *conjugate coordinates*, a concept applicable to any complex valued or real valued function of a complex variable, whereby a complex function is formally expressed as a function of both z and z^*, that is[3]

$$f(z) = f(z, z^*) = \Re\{f\} + \jmath\Im\{f\} = u(x, y) + \jmath v(x, y) = g(x, y) \tag{5.10}$$

[2]Wirtinger's result has been used largely by the German speaking DSP community [78].
[3]For an excellent overview we refer to the lecture material 'The Complex Gradient Operator and the $\mathbb{CR}$ Calculus', (ECE275CG-F05v1.3d) by Kenneth Kreutz–Delgado.

Notice that $g(x, y)$ is a real bivariate function associated with the complex univariate function $f(z)$, and that the total differential of the function $g(x, y)$ can be expressed as

$$dg(x, y) = \frac{\partial g(x, y)}{\partial x}dx + \frac{\partial g(x, y)}{\partial y}dy$$

We then have

$$dg(x, y) = \frac{\partial u(x, y)}{\partial x}dx + \jmath\frac{\partial v(x, y)}{\partial x}dx + \frac{\partial u(x, y)}{\partial y}dy + \jmath\frac{\partial v(x, y)}{\partial y}dy$$

and the variable swap

$$dz = dx + \jmath dy \qquad dz^* = dx - \jmath dy$$
$$dx = \frac{1}{2}\big[dz + dz^*\big] \qquad dy = \frac{1}{2\jmath}\big[dz - dz^*\big]$$

yields

$$\begin{aligned} dg(x, y) = &\frac{1}{2}\left[\frac{\partial u(x, y)}{\partial x} + \frac{\partial v(x, y)}{\partial y} + \jmath\Big(\frac{\partial v(x, y)}{\partial x} - \frac{\partial u(x, y)}{\partial y}\Big)\right]dz \\ &+ \frac{1}{2}\left[\frac{\partial u(x, y)}{\partial x} - \frac{\partial v(x, y)}{\partial y} + \jmath\Big(\frac{\partial v(x, y)}{\partial x} + \frac{\partial u(x, y)}{\partial y}\Big)\right]dz^* \end{aligned}$$

The differential $dg(x, y)$ now becomes

$$dg(x, y) = \frac{1}{2}\left[\frac{\partial g(x, y)}{\partial x} - \jmath\frac{\partial g(x, y)}{\partial y}\right]dz + \frac{1}{2}\left[\frac{\partial g(x, y)}{\partial x} + \jmath\frac{\partial g(x, y)}{\partial y}\right]dz^*$$

and hence the differential of the complex function $f(z)$ can be written as

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

5.3.1 $\mathbb{CR}$ Calculus

Using the above formalism [158], for a function $f(z) = f(z, z^*) = g(x, y)$, where f can be either complex valued or real valued, we can formally[4] introduce the $\mathbb{R}$-derivatives:-

- The $\mathbb{R}$-derivative of a real function of a complex variable $f = f(z, z^*)$ is given by

$$\frac{\partial f}{\partial z}\Big|_{z^*=const} = \frac{1}{2}\Big(\frac{\partial f}{\partial x} - \jmath\frac{\partial f}{\partial y}\Big) \tag{5.11}$$

  where the partial derivatives $\partial f/\partial x$ and $\partial f/\partial y$ are **true (non–formal)** partial derivatives of the function $f(z) = f(z, z^*) = g(x, y)$;

[4] As z is not independent of z^* this is only a formalism; a similar formalism has been used to introduce the augmented complex statistics in chapter 12.

- The conjugate $\mathbb{R}$-derivative ($\mathbb{R}^*$-derivative) of a function $f(z) = f(z, z^*)$ is given by

$$\frac{\partial f}{\partial z^*}\Big|_{z=const} = \frac{1}{2}\Big(\frac{\partial f}{\partial x} + \jmath\frac{\partial f}{\partial y}\Big) \tag{5.12}$$

  - ○ For these generalised derivatives, it is assumed that z and z^* are mutually independent, that is

$$\frac{\partial z}{\partial z} = \frac{\partial z^*}{\partial z^*} = 1 \qquad \frac{\partial z}{\partial z^*} = \frac{\partial z^*}{\partial z} = 0$$

  Then, the $\mathbb{R}$-derivatives can be straightforwardly calculated by replacing

$$x = (z + z^*)/2 \qquad y = -\jmath(z - z^*)/2$$

  and using the chain rule, as shown above;

- The $\mathbb{R}$-derivatives can be expressed in terms of $g(x, y) = f(z, z^*)$ as

$$\frac{\partial f(z, z^*)}{\partial z} = \frac{1}{2}\Big(\frac{\partial g(x, y)}{\partial x} - \jmath\frac{\partial g(x, y)}{\partial y}\Big) \qquad \frac{\partial f(z, z^*)}{\partial z^*} = \frac{1}{2}\Big(\frac{\partial g(x, y)}{\partial x} + \jmath\frac{\partial g(x, y)}{\partial y}\Big)$$

  In other words, the analyticity of $f(z) = f(z, z^*)$ with respect to both z and z^* independently is equivalent to the $\mathbb{R}$-differentiability of $g(x, y)$.
- As a consequence, in terms of $\mathbb{R}$-derivatives, function $f(z)$ has two stationary points, at $\partial f(z, z^*)/\partial z = 0$ and $\partial f(z, z^*)/\partial z^* = 0$.

Thus, although real functions of complex variable are not differentiable in $\mathbb{C}$ (the $\mathbb{C}$-derivative does not exist), they are generally differentiable in both x and y and their $\mathbb{R}$-derivatives do exist.

5.3.2 Link between $\mathbb{R}$- and $\mathbb{C}$-derivatives

When considering holomorphic complex functions of complex variables in light of the $\mathbb{CR}$-derivatives [158], we can observe that:-

- If a function $f = f(z, z^*) = g(x, y) = u(x, y) + \jmath v(x, y)$ is holomorphic, then the Cauchy–Riemann conditions are satisfied, that is, $\partial u(x, y)/\partial x = \partial v(x, y)/\partial y$ and $\partial v(x, y)/\partial x = -\partial u(x, y)/\partial y$, and

$$\begin{aligned}
\mathbb{R} - \text{derivative} \quad \frac{1}{2}\left[\frac{\partial f}{\partial x} - \jmath\frac{\partial f}{\partial y}\right] &= \frac{1}{2}\left[\frac{\partial u}{\partial x} + \jmath\frac{\partial v}{\partial x} - \jmath\frac{\partial u}{\partial y} + \frac{\partial v}{\partial y}\right] \\
&= \frac{1}{2}\left[2\frac{\partial u}{\partial x} + 2\jmath\frac{\partial v}{\partial x}\right] = f'(z) \\
\mathbb{R}^* - \text{derivative} \quad \frac{1}{2}\left[\frac{\partial f}{\partial x} + \jmath\frac{\partial f}{\partial y}\right] &= \frac{1}{2}\left[\frac{\partial u}{\partial x} + \jmath\frac{\partial v}{\partial x} + \jmath\frac{\partial u}{\partial y} - \frac{\partial v}{\partial y}\right] = 0
\end{aligned}$$

  that is, for holomorphic functions the $\mathbb{R}^*$-derivative vanishes and the $\mathbb{R}$-derivative is equivalent to the standard complex derivative $f'(z)$;
- In other words, if an $\mathbb{R}$-differentiable function $f(z, z^*)$ is independent of z^*, then the $\mathbb{R}$-derivative of $f(z)$ is equivalent to the standard $\mathbb{C}$-derivative;

- Since for a complex holomorphic function $f(z)$ the $\mathbb{R}^*$-derivative vanishes ($\partial f(z)/\partial z^* = 0$), we can state an *alternative, generalised, form of the Cauchy–Riemann conditions* as

$$\frac{\partial f(z)}{\partial z^*} = 0 \tag{5.13}$$

Thus, holomorphic functions are essentially those which can be written without z^* terms.

Therefore, the $\mathbb{R}$-derivatives are a natural generalisation of the standard complex derivative, and apply to both holomorphic and nonholomorphic functions. We say that complex functions are *real analytic* ($\mathbb{R}$-analytic) over $\mathbb{R}^2$ if they are both $\mathbb{R}$-differentiable and $\mathbb{R}^*$-differentiable [158, 250].

Based on the $\mathbb{CR}$-derivatives, rules of complex differentiation can be obtained by replacing $dz = dx + \jmath dy$ and $dz^* = dx - \jmath dy$. Several of these rules are listed below.

$$\begin{aligned}
\frac{\partial f^*(z)}{\partial z^*} &= \left(\frac{\partial f(z)}{\partial z}\right)^* \\
\frac{\partial f^*(z)}{\partial z} &= \left(\frac{\partial f(z)}{\partial z^*}\right)^* \\
df(z) &= \frac{\partial f(z)}{\partial z}dz + \frac{\partial f(z)}{\partial z^*}dz^* \qquad \text{differential} \\
\frac{\partial f(g(z))}{\partial z} &= \frac{\partial f}{\partial g}\frac{\partial g}{\partial z} + \frac{\partial f}{\partial g^*}\frac{\partial g^*}{\partial z} \qquad \text{chain rule}
\end{aligned} \tag{5.14}$$

In the particular case of cost functions, for instance, the nonholomorphic $f(z, z^*) : \mathbb{C} \times \mathbb{C} \mapsto \mathbb{R}$ given by $f(z) = zz^* = x^2 + y^2 = g(x, y)$, the $\mathbb{R}^*$-derivative

$$\frac{1}{2}\left[\frac{\partial g(x,y)}{\partial x} + \jmath\frac{\partial g(x,y)}{\partial y}\right] = x + \jmath y = z = \frac{\partial f(z,z^*)}{\partial z^*}_{|z=const} = \left(\frac{\partial f^*(z,z^*)}{\partial z}_{|z^*=const}\right)^* \tag{5.15}$$

that is, for a real function of complex variable we have

$$\left(\frac{\partial f}{\partial z}\right)^* = \frac{\partial f}{\partial z^*} \tag{5.16}$$

It is important to highlight again that for general holomorphic complex functions of complex variable $f(z) : \mathbb{C} \mapsto \mathbb{C}$

$$\frac{\partial f(z)}{\partial z} = \underbrace{\frac{\partial f_r(z)}{\partial x} + \jmath\frac{\partial f_i(z)}{\partial x}}_{\mathbb{C}-derivative} \neq \frac{\partial f(z)}{\partial z^*}_{|z=const.} = \frac{1}{2}\underbrace{\left[\frac{\partial f(z)}{\partial x} + \jmath\frac{\partial f(z)}{\partial y}\right]}_{\mathbb{R}^*-derivative} = 0 \tag{5.17}$$

and that in this case the $\mathbb{R}^*$-derivative vanishes.

For instance, for a holomorphic, $\mathbb{C}$-differentiable, function $f(z) = z = x + \jmath y$, we have

$$
\begin{aligned}
&\mathbb{C} \ - \text{derivative} && f'(z) = 1 \\
&\mathbb{R}^* \ - \text{derivative} && \frac{1}{2}\left[\frac{\partial(x+\jmath y)}{\partial x} + \jmath\frac{\partial(x+\jmath y)}{\partial y}\right] = 0 \\
&\mathbb{R} \ - \text{derivative} && \frac{1}{2}\left[\frac{\partial(x+\jmath y)}{\partial x} - \jmath\frac{\partial(x+\jmath y)}{\partial y}\right] = 1
\end{aligned}
$$

5.4 $\mathbb{CR}$-derivatives of Cost Functions

In complex valued adaptive signal processing, it is common to perform optimisation based on scalar functions of complex variables. These are typically quadratic functions of the output error of an adaptive filter, for instance

$$J(k) = \frac{1}{2}|e(k)|^2 = \frac{1}{2}e(k)e^*(k) \tag{5.18}$$

that is, $J(e, e^*) : \mathbb{C} \mapsto \mathbb{R}$, where the error $e(k)$ is a function of the vector of filter parameters $\mathbf{w}(k) = \mathbf{w}^r(k) + \jmath\mathbf{w}^i(k)$, $e(k) = d(k) - \mathbf{x}^T(k)\mathbf{w}(k)$, and the symbols $d(k)$ and $\mathbf{x}(k)$ denote respectively the teaching signal and the tap input vector. We have shown in Equation (5.9) that such functions are not holomorphic, hence the Cauchy–Riemann conditions are not satisfied, and their $\mathbb{C}$-derivative is not defined. However, $J(k)$ is differentiable in both the $\mathbf{w}^r(k)$ and $\mathbf{w}^i(k)$, and to find its stationary points (extrema), both the $\mathbb{R}$- and $\mathbb{R}^*$-derivatives must vanish; this is the essence of Brandwood's result [35].

5.4.1 *The Complex Gradient*

We shall now provide a step by step derivation of the expressions for gradients of cost functions, as their understanding is critical to the derivation of adaptive filtering algorithms in $\mathbb{C}$. Our aim is to show that

$$\nabla_{\mathbf{w}} J(k) = \frac{\partial J(k)}{\partial \mathbf{w}(k)} = \frac{\partial J(k)}{\partial \mathbf{w}^r(k)} + \jmath\frac{\partial J(k)}{\partial \mathbf{w}^i(k)} \tag{5.19}$$

To this end, recall that the relationship between the complex number $z = x + \jmath y$ and a composite real variable $\omega = (a, b) \in \mathbb{R}^2$ is described by[5] (also shown in chapters 11 and 12)

$$\begin{bmatrix} z_k \\ z_k^* \end{bmatrix} = \mathbb{J}\begin{bmatrix} x_k \\ y_k \end{bmatrix} = \begin{bmatrix} 1 & \jmath \\ 1 & -\jmath \end{bmatrix}\begin{bmatrix} x_k \\ y_k \end{bmatrix}$$

[5] We have introduced the relationship between the real and complex gradients and Hessians based on the work by Van Den Bos [30]. As points where the gradient vanishes (stationary points) give us the location of the extrema of the cost function, and the Hessian helps to differentiate between the minimum and maximum, it is important to provide further insight into their calculation.

whereas the relationship between the composite real vector $\boldsymbol{\omega} = [x_1, y_1, \ldots, x_N, y_N]^T \in \mathbb{R}^{2N\times 1}$ and the "augmented" complex vector $\mathbf{v} = \left[z_1, z_1^*, \ldots, z_N, z_N^*\right]^T \in \mathbb{C}^{2N\times 1}$ is given by

$$\mathbf{v} = \mathbf{A}\boldsymbol{\omega} \tag{5.20}$$

where the matrix $\mathbf{A} = diag(\mathbb{J}, \ldots, \mathbb{J}) \in \mathbb{C}^{2N\times 2N}$ is block diagonal. Thus, as $\mathbb{J}$ can be made a unitary matrix (upon scaling by $\frac{1}{\sqrt{2}}$), we have

$$\boldsymbol{\omega} = \mathbf{A}^{-1}\mathbf{v} = \frac{1}{2}\mathbf{A}^H\mathbf{v} = \frac{1}{2}\mathbf{A}^T\mathbf{v}^* \tag{5.21}$$

To establish a relation between the $\mathbb{R}$- and $\mathbb{C}$-derivatives, consider a Taylor series expansion of a real function $f(\boldsymbol{\omega})$ around $\boldsymbol{\omega} = \mathbf{0}$, given by

$$f + \frac{\partial f}{\partial \boldsymbol{\omega}^T}\boldsymbol{\omega} + \frac{1}{2!}\boldsymbol{\omega}^T\frac{\partial^2 f}{\partial \boldsymbol{\omega}\partial \boldsymbol{\omega}^T}\boldsymbol{\omega} + \cdots \tag{5.22}$$

where the vector $\partial f/\partial \boldsymbol{\omega}$ is called the gradient. The connection between the real and complex gradient is then given by

$$\frac{\partial f}{\partial \boldsymbol{\omega}^T}\boldsymbol{\omega} = \frac{1}{2}\frac{\partial f}{\partial \boldsymbol{\omega}^T}\mathbf{A}^H\mathbf{v} \quad \Rightarrow \quad \frac{\partial f}{\partial \mathbf{v}} = \frac{1}{2}\mathbf{A}^*\frac{\partial f}{\partial \boldsymbol{\omega}} \quad \Rightarrow \quad \frac{\partial f}{\partial \boldsymbol{\omega}} = \mathbf{A}^T\frac{\partial f}{\partial \mathbf{v}} \tag{5.23}$$

that is, the real and complex gradient are related by a linear transformation. For illustration, consider the elements of the gradient vector $\partial f/\partial \mathbf{v}$, that is, $\partial f/\partial z$ and $\partial f/\partial z^*$. From Equations (5.11) and (5.12), we then immediately have

$$\begin{aligned} \mathbb{R}-\text{derivative:} \qquad & \frac{\partial f}{\partial z_n} = \frac{1}{2}\left(\frac{\partial f}{\partial x_n} - \jmath\frac{\partial f}{\partial y_n}\right) \\ \mathbb{R}^*-\text{derivative:} \qquad & \frac{\partial f}{\partial z_n^*} = \frac{1}{2}\left(\frac{\partial f}{\partial x_n} + \jmath\frac{\partial f}{\partial y_n}\right) \end{aligned} \tag{5.24}$$

For the particular case of cost functions, real functions of complex variable, we have

$$\left(\frac{\partial f}{\partial z_n}\right)^* = \left(\frac{\partial f}{\partial z_n^*}\right) \qquad\qquad \frac{\partial f}{\partial \mathbf{v}^*} = \left(\frac{\partial f}{\partial \mathbf{v}}\right)^*$$

and also from Equation (5.23)

$$\frac{\partial f}{\partial \boldsymbol{\omega}} = \mathbf{A}^H\frac{\partial f}{\partial \mathbf{v}^*} \tag{5.25}$$

These relationships are very useful in the derivation of the learning algorithms in the steepest descent setting, and especially when addressing their convergence.[6]

[6]Relationships between complex random variables and composite real random variables in terms of their respective probability density functions are given in chapter 12 and Appendix A.

5.4.2 The Complex Hessian

To establish the $\mathbb{CR}$ relationships in terms of second order derivatives, consider the quadratic term in the Taylor series expansion (TSE) in Equation (5.22), given by

$$\frac{1}{2!}\boldsymbol{\omega}^T\mathbf{H}\boldsymbol{\omega} \quad \text{where} \quad \mathbf{H} = \frac{\partial^2 f}{\partial\boldsymbol{\omega}\,\partial\boldsymbol{\omega}^T} \tag{5.26}$$

Using the substitution (5.21), this can be rewritten as

$$\frac{1}{2!}\mathbf{v}^H\mathbf{G}\mathbf{v} \quad \text{where} \quad \mathbf{G} = \frac{\partial^2 f}{\partial\mathbf{v}^*\,\partial\mathbf{v}^T} = \frac{1}{4}\mathbf{A}\mathbf{H}\mathbf{A}^H \tag{5.27}$$

thus, giving the relation between the real Hessian $\mathbf{H}$ and the complex Hessian $\mathbf{G}$.

As $\mathbf{A}^{-1} = \frac{1}{2}\mathbf{A}^H$ and therefore $\mathbf{I} = \mathbf{A}\mathbf{A}^{-1} = \frac{1}{2}\mathbf{A}\mathbf{A}^H$, the characteristic equation for the eigenvalues of the complex Hessian $\mathbf{G}$ is given by

$$\mathbf{G} - \lambda\mathbf{I} = \frac{1}{4}\mathbf{A}\big(\mathbf{H} - 2\lambda\mathbf{I}\big)\mathbf{A}^H \tag{5.28}$$

The roots of $\mathbf{G} - \lambda\mathbf{I}$ are the same as the roots of $\mathbf{H} - 2\lambda\mathbf{I}$, and therefore the eigenvalues of the complex Hessian λ_n^c and the eigenvalues of the real Hessian λ_n^r are related as

$$\lambda_n^c = 2\lambda_n^r, \qquad\qquad n = 1, \ldots, N \tag{5.29}$$

An important consequence is that the conditioning of the matrices $\mathbf{G}$ and $\mathbf{H}$ is the same – this is very useful when studying the relationship between complex valued stochastic gradient algorithms and their counterparts in $\mathbb{R}^2$.

In the particular case of the Newton type of optimisation, the real and complex Newton steps are given by

$$\text{Real:} \quad \mathbf{H}\Delta\boldsymbol{\omega} = -\frac{\partial f}{\partial\boldsymbol{\omega}} \qquad\qquad \text{Complex:} \quad \mathbf{G}\Delta\mathbf{v} = -\frac{\partial f}{\partial\mathbf{v}^*} \tag{5.30}$$

where $\partial f/\partial\boldsymbol{\omega}$ and $\partial f/\partial\mathbf{v}^*$ denote respectively the real gradient and the conjugate of the complex gradient.

5.4.3 Complex Jacobian and Complex Differential

Based on Equation (5.14), the differential of the complex valued vector function F can now be defined as

$$dF(\mathbf{z}, \mathbf{z}^*) = \frac{\partial F(\mathbf{z}, \mathbf{z}^*)}{\partial\mathbf{z}}d\mathbf{z} + \frac{\partial F(\mathbf{z}, \mathbf{z}^*)}{\partial\mathbf{z}^*}d\mathbf{z}^* \tag{5.31}$$

where $\mathbf{z} = [z_1, \ldots, z_N]^T$ and $\mathbf{z}^* = [z_1^*, \ldots, z_N^*]^T$, and

$$F(\mathbf{z}, \mathbf{z}^*) = [f_1(\mathbf{z}, \mathbf{z}^*), \ldots, f_N(\mathbf{z}, \mathbf{z}^*)]^T \tag{5.32}$$

whereas the complex Jacobians $\mathbf{J} = \partial F(\mathbf{z}, \mathbf{z}^*)/\partial \mathbf{z}$ and $\mathbf{J}^c = \partial F(\mathbf{z}, \mathbf{z}^*)/\partial \mathbf{z}^*$ are defined as

$$\mathbf{J} = \begin{bmatrix} \frac{\partial f_1}{\partial z_1} & \cdots & \frac{\partial f_1}{\partial z_N} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_N}{\partial z_1} & \cdots & \frac{\partial f_N}{\partial z_N} \end{bmatrix} \quad \text{and} \quad \mathbf{J}^c = \begin{bmatrix} \frac{\partial f_1}{\partial z_1^*} & \cdots & \frac{\partial f_1}{\partial z_N^*} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_N}{\partial z_1^*} & \cdots & \frac{\partial f_N}{\partial z_N^*} \end{bmatrix} \tag{5.33}$$

Notice that for holomorphic functions $\mathbf{J}^* \neq \mathbf{J}^c$, whereas for real functions of complex variable $\mathbf{J}^* = \mathbf{J}^c$. The complex differential of a real function of complex variable thus becomes

$$dF(\mathbf{z}, \mathbf{z}^*) = \mathbf{J}d\mathbf{z} + \mathbf{J}^* d\mathbf{z}^* = 2\Re\{\mathbf{J}d\mathbf{z}\} = 2\Re\{\mathbf{J}^* d\mathbf{z}^*\} \tag{5.34}$$

The chain rule from Equation (5.14) can also be extended to the vector case as

$$\frac{\partial F(\mathbf{g})}{\partial \mathbf{z}} = \frac{\partial F}{\partial \mathbf{g}}\frac{\partial \mathbf{g}}{\partial \mathbf{z}} + \frac{\partial F}{\partial \mathbf{g}^*}\frac{\partial \mathbf{g}^*}{\partial \mathbf{z}} \quad \text{and} \quad \frac{\partial F(\mathbf{g})}{\partial \mathbf{z}^*} = \frac{\partial F}{\partial \mathbf{g}}\frac{\partial \mathbf{g}}{\partial \mathbf{z}^*} + \frac{\partial F}{\partial \mathbf{g}^*}\frac{\partial \mathbf{g}^*}{\partial \mathbf{z}^*} \tag{5.35}$$

5.4.4 Gradient of a Cost Function

As $\mathbb{C}$-derivatives are not defined for real functions of complex variable, generalised gradient operators can be defined based on the $\mathbb{R}$- and $\mathbb{R}^*$-derivatives in Equations (5.11–5.12) with respect to vectors $\mathbf{z} = [z_1, \ldots, z_N]^T$ and $\mathbf{z}^* = [z_1^*, \ldots, z_N^*]^T$, that is

$$\begin{aligned} \mathbb{R} - \text{derivative:} \qquad & \frac{\partial}{\partial \mathbf{z}} = \frac{1}{2}\left[\frac{\partial}{\partial \mathbf{x}} - \jmath \frac{\partial}{\partial \mathbf{y}}\right] \\ \mathbb{R}^* - \text{derivative:} \qquad & \frac{\partial}{\partial \mathbf{z}^*} = \frac{1}{2}\left[\frac{\partial}{\partial \mathbf{x}} + \jmath \frac{\partial}{\partial \mathbf{y}}\right] \end{aligned} \tag{5.36}$$

where $\mathbf{z} = \mathbf{x} + \jmath\mathbf{y}$ and $\mathbf{z}^* = \mathbf{x} - \jmath\mathbf{y}$ and the elements of the gradient vectors $\partial/\partial\mathbf{z}$ and $\partial/\partial\mathbf{z}^*$ are given by

$$\frac{\partial}{\partial z_n} = \frac{1}{2}\left[\frac{\partial}{\partial x_n} - \jmath \frac{\partial}{\partial y_n}\right] \quad \text{and} \quad \frac{\partial}{\partial z_n^*} = \frac{1}{2}\left[\frac{\partial}{\partial x_n} + \jmath \frac{\partial}{\partial y_n}\right] \tag{5.37}$$

Thus, in expressing the gradient of a scalar function with respect to a complex vector, the derivatives are applied component–wise, for instance, the gradient of $J(e, e^*)$ with respect to the complex weight vector $\mathbf{w} = [w_1, \ldots, w_N]^T$ is given by

$$\nabla_{\mathbf{w}} J(e, e^*) = \frac{\partial J(e, e^*)}{\partial \mathbf{w}} = \left[\frac{\partial J(e, e^*)}{\partial w_1}, \ldots, \frac{\partial J(e, e^*)}{\partial w_N}\right]^T \tag{5.38}$$

Therefore, to optimise a real function J with respect to a complex valued parameter vector $\mathbf{w}$, two conditions must be satisfied

$$\frac{\partial J(e, e^*)}{\partial \mathbf{w}} = \mathbf{0} \quad \text{and} \quad \frac{\partial J(e, e^*)}{\partial \mathbf{w}^*} = \mathbf{0} \tag{5.39}$$

To determine the extrema of real valued cost functions, apply the rules of complex differentiation from Equation (5.31) to the first term of Taylor series expansion in Equation (5.22), to yield

$$\Delta J(e, e^*) = \left[\frac{\partial J}{\partial \mathbf{w}}\right]^T \Delta \mathbf{w} + \left[\frac{\partial J}{\partial \mathbf{w}^*}\right]^T \Delta \mathbf{w}^* = 2\Re\left\{\left[\frac{\partial J}{\partial \mathbf{w}}\right]^H \Delta \mathbf{w}^*\right\} = 2\Re\left\{\left[\frac{\partial J}{\partial \mathbf{w}^*}\right]^T \Delta \mathbf{w}^*\right\}$$

Observe that, although the set of stationary points of the gradient has two solutions $\partial J/\partial \mathbf{w}$ and $\partial J/\partial \mathbf{w}^*$, since $df(z) = df(z, z^*) = dg(x, y)$ and

$$dg(x, y) = \frac{\partial g(x, y)}{\partial x}dx + \frac{\partial g(x, y)}{\partial y}dy$$

the differential df vanishes only if the $\mathbb{R}$-derivative is zero, and *the maximum change of the cost function $J(e, e^*)$ is in the direction of the conjugate gradient*. It is therefore natural to express the gradient of the cost function with respect to the filter parameter vector as

$$\nabla_{\mathbf{w}} J = 2\,\frac{\partial J}{\partial \mathbf{w}^*} = \frac{\partial J}{\partial \mathbf{w}^r} + \jmath \frac{\partial J}{\partial \mathbf{w}^i} \tag{5.40}$$

as this reflects the direction of the maximum change of the gradient.

The stochastic gradient based coefficient update of an adaptive filter can now be expressed as

$$\mathbf{w}(k+1) = \mathbf{w}(k) - \mu \nabla_{\mathbf{w}} J(k)$$

where

$$\mathbf{w}^r(k+1) = \mathbf{w}^r(k) - \mu \frac{\partial J(k)}{\partial \mathbf{w}^r(k)} \quad \text{and} \quad \mathbf{w}^i(k+1) = \mathbf{w}^i(k) - \mu \frac{\partial J(k)}{\partial \mathbf{w}^i(k)} \tag{5.41}$$

For the particular case of a complex linear FIR filter, described in chapter 3, the cost function $J = \frac{1}{2}e(k)e^*(k)$ needs to be minimised with respect to the filter coefficient vector $\mathbf{w}(k)$, where $e(k) = d(k) - \mathbf{x}^T(k)\mathbf{w}(k)$, and $\mathbf{x}(k)$ is the input vector. The stochastic gradient algorithm for the weight update (complex LMS) thus can be expressed as

$$\mathbf{w}(k+1) = \mathbf{w}(k) - \mu \frac{\partial \frac{1}{2}e(k)e^*(k)}{\partial \mathbf{w}^*} = \mathbf{w}(k) + \mu e(k)\mathbf{x}^*(k) \tag{5.42}$$

Derivatives and differentials with respect to complex vectors and matrices are given in Table A.1–A.3 in Appendix A.

Summary: This chapter has introduced elements of $\mathbb{CR}$ calculus for general functions of complex variable. Particular topics include:-

- Complex continuity, differentiability, and the derivation of Cauchy–Riemann equations. It has been shown that the Cauchy–Riemann equations provide a very elegant tool to calculate derivatives of complex holomorphic functions, however, they also impose a great amount of structure on such functions;
- It has been shown that typical cost functions used in adaptive filtering are real functions of complex variables, and as such they do not obey the Cauchy–Riemann conditions. To

deal with this problem, the $\mathbb{R}$-derivatives, that is, generalised derivatives which apply to both complex functions of complex variables and real functions of complex variables have been introduced;

- Relationships between the complex ($\mathbb{C}$-derivatives) and $\mathbb{R}$-derivatives have been established, and a generalised Cauchy–Riemann condition has been introduced;
- An insight into the calculation of complex Jacobians and Hessians has been provided and their relation with the corresponding Jacobians and Hessians in $\mathbb{R}^2$ has been established;
- The $\mathbb{CR}$ calculus has been applied to compute gradients of real functions of complex variable, and the complex steepest descent method has been introduced within this framework.