Quaternion Valued Neural Networks and Nonlinear Adaptive Filters*

B. Che Ujang, C. Jahanchahi, C. Cheong Took and D. P. Mandic

July 16, 2010

Technical Report: TR-ICU-QNN-0024/07-10
Email: {che.che-ujang07, cyrus.jahanchahi05, c.cheong-took, d.mandic}@imperial.ac.uk)
Electrical and Electronic Engineering Department, Imperial College London, SW7 2AZ, U.K.

Abstract

A class of locally analytic transcendental functions suitable for nonlinear adaptive filtering and neural network filtering is proposed. Since the stringent standard analyticity conditions prevent full exploitation of nonlinear quaternionic models, we make use of local analyticity conditions to provide a framework for a generic extension of nonlinear learning algorithms from the real and complex domain. In addition, it is shown that the use of the proposed class of locally analytic transcendental functions in conjunction with widely linear modelling allows to fully exploit the second-order information in the data. Simulations over a range of noncircular synthetic, chaotic and real world three dimensional wind signals support the approach.

1 Introduction

Quaternions are proven to have great potential in three- and four-dimensional data modelling and have found application across the areas of engineering, including computer graphics [1] and robotics [2], due to their convenience over real valued vectorial models. In the statistical signal processing field, quaternions have been employed in adaptive filtering, including Kalman filtering [3] and stochastic gradient type of algorithm, such as the Quaternion Least Mean Square (QLMS) [4]. Despite gaining in popularity, they are still relatively underexplored in nonlinear filtering, neural networks, and blind source separation communities, mainly because of problems due to the lack of analytic nonlinear functions in $\mathbb{H}$. The analyticity in $\mathbb{H}$ is governed by the Cauchy-Riemann-Fueter (CRF) conditions [5] which do not permit a wide range of holomorphic functions in the same way as their wide employment in $\mathbb{R}$ and $\mathbb{C}$; the only globally analytic functions in $\mathbb{H}$ are linear functions and constant values. To partially overcome this issue, a suboptimal solution in the form of "split" nonlinear quaternion function that processes each channel separately in $\mathbb{R}$ instead of $\mathbb{H}$ through a real smooth nonlinearity was employed in [6].

One of the first nonlinear learning algorithms to use the "split" nonlinear quaternion activation function is the Quaternion Multilayer Perceptron (QMLP) [6], which benefitting from quaternion algebra, exhibited enhanced performance over previous vector based algorithms [7] [8]. However, the QMLP neglects the non-commutativity aspect of the quaternion algebra and thus does not exploit the full potential of the processing in the quaternion domain; this issue was addressed with the Split Quaternion Nonlinear Adaptive Filtering Algorithm (SQAFA) [9]. However, the nonlinearities used in SQAFA were still standard real activation functions applied channelwise, thus prohibiting a fully capture of the cross-dynamics across the data channels.

It is important to notice that most practical learning algorithms are gradient descent based [6–9], the operating point moves at every sample interval, and therefore the nonlinearities used only require local analyticity. In analogy to the complex domain, where so called fully complex nonlinearities (elementary transcendental functions) provide means for generic extensions of real neural networks [10] [11], our aim is to show that the class of elementary transcendental functions, such as tanh are also locally analytic in $\mathbb{H}$. This is not possible to achieve using the standard Cauchy-Riemann-Fueter (CRF) conditions [5], which are too restrictive, and to this end we shall explore some recent results on local analyticity [12]. Due to a cumbersome derivation, we prove analytically the possibility of building generic quaternion-valued nonlinear adaptive filters only for the most commonly used tanh function. Since the derivation involves proving a local analyticity of exponential functions, this makes other typical transcendental nonlinear activation functions in $\mathbb{H}$ also suitable for this purpose, as is shown by simulations. This set of results opens possibilities to establish nonlinear learning architectures and learning algorithms in $\mathbb{H}$, in the same way they are established in $\mathbb{R}$ and $\mathbb{C}$ [10] [13–21].

*This material is a part of the submission to IEEE Transactions on Neural Networks.

In this work, we propose a class of fully quaternion locally analytic nonlinear activation functions for nonlinear adaptive filtering. For completeness, we also show that full statistical information in the quaternion domain can be exploited by combining the proposed nonlinear models with so called augmented quaternion statistics and the widely linear model [22]. The benefit of the local analytic conditions used in this work is that they are suitable for polynomials, thus permitting Taylor series expansions, and giving the quaternion based algorithms the same generic form as the corresponding algorithms in $\mathbb{R}$ and $\mathbb{C}$. For simplicity, the analysis and derivation are provided for a single nonlinear perceptron and its widely linear counterpart. Extensions to large-scale architectures can be achieved in the same way as in $\mathbb{R}$ and $\mathbb{C}$, and are out of the scope of this work.

This paper is organized as follows. Section II introduces basic operations of quaternion algebra. Section III reviews the basic concept behind augmented quaternion statistics. This is followed by a review of the analyticity conditions in $\mathbb{H}$ along with the analysis of the quaternion exponential function and quaternion tanh function. Section V derives the proposed QAFA and WLQAFA, whose convergence analysis is provided in Section VI. In Section VII, the performances of QAFA and WLQAFA are compared against the standard models QFIR and AASQAFA through simulations on both benchmark and real-world multidimensional data. The results are analyzed and discussed in Section VIII. The paper concludes in Section IX.

2 Quaternion Algebra

Quaternions are an algebra defined over $\mathbb{R}$, where quaternion variable q is given by

$$q = [q_a, \mathbf{q}] = q_a + q_b\imath + q_c\jmath + q_d\kappa \tag{1}$$

where $q_a, q_b, q_c, q_d \in \mathbb{R}$ and $\imath$, $\jmath$, κ are both orthogonal unit vectors and imaginary units. These orthogonal unit vectors are related by

$$\begin{aligned} \imath\jmath &= \kappa; \quad \jmath\kappa = \imath; \quad \kappa\imath = \jmath; \\ \imath\jmath\kappa &= \imath^2 = \jmath^2 = \kappa^2 = -1 \end{aligned} \tag{2}$$

The addition and subtraction operations in quaternion algebra are similar to those in complex algebra, however, the multiplication and division operate quite differently. The multiplication is given as

$$wx = [w_a, \mathbf{w}][x_a, \mathbf{x}] = [w_a x_a - \mathbf{w}\cdot\mathbf{x}, w_a\mathbf{x} + x_a\mathbf{w} + \mathbf{w}\times\mathbf{x}] \tag{3}$$

where the symbols "$\cdot$" and "$\times$" denote respectively to the dot-product and cross-product. The quaternion multiplication is non-commutative due to the outer product between $\mathbf{w}$ and $\mathbf{x}$. Owing to the non-commutativity aspect of quaternion algebra, the definition of quaternion division is ambiguous; for consistency the quaternion division considered in this work is given by

$$\frac{w}{x} = wx^{-1} \tag{4}$$

Similarly to the complex case, the conjugate of a quaternion q is

$$q^* = [q_a, \mathbf{q}]^* = [q_a, -\mathbf{q}] = q_a - q_b\imath - q_c\jmath - q_d\kappa \tag{5}$$

and the norm square is

$$\| q \|_2^2 = qq^* = q^*q = q_a^2 + q_b^2 + q_c^2 + q_d^2 \tag{6}$$

The three quaternion involutions (self-inverse operators) are defined as

$$\begin{aligned} q^\imath &= -\imath q\imath = q_a + q_b\imath - q_c\jmath - q_d\kappa \\ q^\jmath &= -\jmath q\jmath = q_a - q_b\imath + q_c\jmath - q_d\kappa \\ q^\kappa &= -\kappa q\kappa = q_a - q_b\imath - q_c\jmath + q_d\kappa \end{aligned} \tag{7}$$

In the sequel, all the quantities are treated as quaternion valued, unless stated otherwise.

3 Augmented Quaternion Statistics

The concept of augmented statistics was first introduced to define the notion of second order noncircularity, or improperness, for complex random normal vectors [23], and was subsequently extended to non-normal vectors [24]. In the complex domain $\mathbb{C}$, the second order characteristics of a complex random vector can be fully characterized by its covariance $\mathcal{C}_{\mathbf{zz}}$ and pseudocovariance $\mathcal{P}_{\mathbf{zz}}$, defined as [23]

$$\mathcal{C}_{\mathbf{zz}} = E(\mathbf{z}\mathbf{z}^H) \qquad \mathcal{P}_{\mathbf{zz}} = E(\mathbf{z}\mathbf{z}^T) \tag{8}$$

where $(\cdot)^H$ and $(\cdot)^T$ denote respectively the Hermitian and transpose vector operator, and $\mathbf{z} = \mathbf{x} + \mathbf{y}\imath$ where $\mathbf{x}$ and $\mathbf{y}$ are real-valued. A complex random vector is termed "circular" if its probability distribution is rotation-invariant. This implies that the real and imaginary components have equal variance and are not correlated, that is, the pseudocovariance $\mathcal{P}_{\mathbf{zz}}$ vanishes [10] [17].

3.1 $\mathbb{C}^\eta$-circular Quaternion Random Variables

The concept of augmented statistics was subsequently extended to the quaternion domain, with the restriction that the rotation angle be around $\pi/2$ [25]. A quaternion random variable q that obeys this restriction is said to be $\mathbb{C}^\eta$-circular, and is defined as

$$q \triangleq \mathrm{e}^{\eta\theta} q, \forall\theta \tag{9}$$

for one and only one pure imaginary unit η, where $\eta \in \{\imath, \jmath, \kappa\}$. The symbol $\triangleq$ denotes equality in terms of the probability distribution function (pdf) and the symbol θ represents the angle of rotation. For example, a quaternion random variable q is considered to be $\mathbb{C}^\imath$-circular if it has a circular probability distribution over the real and $\imath$-axis however its $\jmath$ and κ components are not necessarily uncorrelated. Similar definitions can be given for $\mathbb{C}^\jmath$-circular and $\mathbb{C}^\kappa$-circular quaternion random variables.

3.2 $\mathbb{H}$-circular Quaternion Random Variables

The restriction over the angle of rotation for $\mathbb{C}^\eta$-circular random variable proves too rigid in practical scenarios and an improvement was proposed allowing for a pdf along any two arbitrary axis of rotation to be circular [26]. A quaternion random variable q that satisfies this condition is said to be $\mathbb{H}$-circular, or $\mathbb{Q}$-proper, and is defined as

$$q \triangleq \mathrm{e}^{\eta\theta} q, \forall\theta \tag{10}$$

for all the pure imaginary units $\eta \in \{\imath, \jmath, \kappa\}$. A $\mathbb{H}$-circular quaternion random variable is circular in all its dimensions, meaning that the real, $\imath$, $\jmath$ and κ components are all circular with respect to each other.

3.3 Augmented Second-Order Statistics of Quaternion Random Vectors

To build a generic framework for second-order statistical analysis in the quaternion domain, similarly to the complex case, it was shown that the covariance alone is not sufficient to fully describe the complete second-order information within the quaternion random vector. In order to deal with $\mathbb{H}$-improper signals, it is shown that we also need to employ complementary covariance matrices [27]. These complementary covariance matrices are termed the $\imath$-covariance $\mathcal{C}_{\mathbf{q}^\imath}$, $\jmath$-covariance $\mathcal{C}_{\mathbf{q}^\jmath}$ and κ-covariance $\mathcal{C}_{\mathbf{q}^\kappa}$, and are given by [22]

$$\mathcal{C}_{\mathbf{q}^\imath} = E\{\mathbf{q}\mathbf{q}^{\imath H}\}; \quad \mathcal{C}_{\mathbf{q}^\jmath} = E\{\mathbf{q}\mathbf{q}^{\jmath H}\}; \quad \mathcal{C}_{\mathbf{q}^\kappa} = E\{\mathbf{q}\mathbf{q}^{\kappa H}\} \tag{11}$$

Thus, the complete second-order characteristics of the quaternion random vector can be captured by the augmented covariance matrix $\mathcal{C}^a_{\mathbf{q}}$ of an augmented vector $\mathbf{q}^a = [\mathbf{q}^T \mathbf{q}^{\imath T} \mathbf{q}^{\jmath T} \mathbf{q}^{\kappa T}]^T$, given by

$$\mathcal{C}^a_{\mathbf{q}} = E\{\mathbf{q}^{\mathrm{a}}\mathbf{q}^{\mathrm{aH}}\} = \begin{bmatrix} \mathcal{C}_{\mathbf{qq}} & \mathcal{C}_{\mathbf{q}^\imath} & \mathcal{C}_{\mathbf{q}^\jmath} & \mathcal{C}_{\mathbf{q}^\kappa} \\ \mathcal{C}^H_{\mathbf{q}^\imath} & \mathcal{C}_{\mathbf{q}^\imath\mathbf{q}^\imath} & \mathcal{C}_{\mathbf{q}^\imath\mathbf{q}^\jmath} & \mathcal{C}_{\mathbf{q}^\imath\mathbf{q}^\kappa} \\ \mathcal{C}^H_{\mathbf{q}^\jmath} & \mathcal{C}_{\mathbf{q}^\jmath\mathbf{q}^\imath} & \mathcal{C}_{\mathbf{q}^\jmath\mathbf{q}^\jmath} & \mathcal{C}_{\mathbf{q}^\jmath\mathbf{q}^\kappa} \\ \mathcal{C}^H_{\mathbf{q}^\kappa} & \mathcal{C}_{\mathbf{q}^\kappa\mathbf{q}^\imath} & \mathcal{C}_{\mathbf{q}^\kappa\mathbf{q}^\jmath} & \mathcal{C}_{\mathbf{q}^\kappa\mathbf{q}^\kappa} \end{bmatrix} \tag{12}$$

where the submatrices in (12) are calculated according to

$$\begin{aligned} \mathcal{C}_{\boldsymbol{\delta}} &= E\{\mathbf{q}\boldsymbol{\delta}^H\} & \mathcal{C}_{\boldsymbol{\alpha\beta}} &= E\{\boldsymbol{\alpha}\boldsymbol{\beta}^{\mathbf{H}}\} \\ \boldsymbol{\delta} &\in \{\mathbf{q}^\imath, \mathbf{q}^\jmath, \mathbf{q}^\kappa\} & \boldsymbol{\alpha}, \boldsymbol{\beta} &\in \{\mathbf{q}, \mathbf{q}^\imath, \mathbf{q}^\jmath, \mathbf{q}^\kappa\} \end{aligned} \tag{13}$$

An $\mathbb{H}$-circular quaternion random variable has the property that $\mathbf{q}$ is not correlated with its quaternion involutions $\mathbf{q}^\imath$, $\mathbf{q}^\jmath$ and $\mathbf{q}^\kappa$, that is

$$E\{\mathbf{q}\mathbf{q}^{\imath H}\} = \mathbf{0}; \quad \mathbf{E}\{\mathbf{q}\mathbf{q}^{\jmath\mathbf{H}}\} = \mathbf{0}; \quad \mathbf{E}\{\mathbf{q}\mathbf{q}^{\kappa\mathbf{H}}\} = \mathbf{0} \tag{14}$$

giving the covariance matrix $\mathcal{C}^a_{\mathbf{q}}$ in (12) of a $\mathbb{H}$-circular random vector in the form[1]

$$\mathcal{C}^a_{\mathbf{q}} = E\{\mathbf{q}^{\mathrm{a}}\mathbf{q}^{\mathrm{aH}}\} = \begin{bmatrix} \mathcal{C}_{\mathbf{qq}} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathcal{C}_{\mathbf{q}^\imath\mathbf{q}^\imath} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathcal{C}_{\mathbf{q}^\jmath\mathbf{q}^\jmath} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathcal{C}_{\mathbf{q}^\kappa\mathbf{q}^\kappa} \end{bmatrix} \tag{15}$$

To exploit the complete second-order statistics of quaternion valued signals, a filtering model similar to the widely linear model in $\mathbb{C}$ needs to be considered [10] [28]. The quaternion widely linear model is based on the augmented basis that builds the matrix $\mathcal{C}^a_{\mathbf{q}}$ (12), and can described by [22] [29]

$$y = \mathbf{w}^{\mathrm{aT}}\mathbf{x}^{\mathrm{a}} = \mathbf{g}^T\mathbf{x} + \mathbf{h}^T\mathbf{x}^\imath + \mathbf{u}^T\mathbf{x}^\jmath + \mathbf{v}^T\mathbf{x}^\kappa \tag{16}$$

[1] Any other basis comprising four combinations out of $\{\mathbf{q}, \mathbf{q}^*, \mathbf{q}^\imath, \mathbf{q}^\jmath, \mathbf{q}^\kappa\}$ and their conjugates is equally valid. The basis proposed in [22] and used here, $\mathbf{q}^a = [\mathbf{q}^T \mathbf{q}^{\imath T} \mathbf{q}^{\jmath T} \mathbf{q}^{\kappa T}]^T$ provides most convenient representation, as shown in the augmented covariance structure for $\mathbb{H}$-circular signals in (12).

where $\mathbf{g}(n)$, $\mathbf{h}(n)$, $\mathbf{u}(n)$ and $\mathbf{v}(n)$ are the weight vectors, $\mathbf{x}(n)$ is the input signal, $\mathbf{x}^{\imath}(n)$, $\mathbf{x}^{\jmath}(n)$ and $\mathbf{x}^{\kappa}(n)$ are respectively its $\imath$, $\jmath$ and κ involutions, $\mathbf{w^a} = [\mathbf{g}^T \quad \mathbf{h}^T \quad \mathbf{u}^T \quad \mathbf{v}^T]^T$ is the augmented weight vector, and $\mathbf{x^a} = [\mathbf{x}^T \quad \mathbf{x}^{\imath T} \quad \mathbf{x}^{\jmath T} \quad \mathbf{v}^{\kappa T}]^T$ is the augmented random input vector.

4 Nonlinear Functions in $\mathbb{H}$

In $\mathbb{C}$, the analyticity of a complex function $f(z) = u(x,y) + v(x,y)\imath$ is governed by the Cauchy-Riemann (CR) equations, given by

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y}; \qquad \frac{\partial v}{\partial x} = -\frac{\partial u}{\partial y} \tag{17}$$

that is, for a complex function $f(z)$ to be analytic in $\mathbb{C}$, the derivatives along the real and imaginary axis have to be equal, that is

$$\frac{\partial f}{\partial x} + \frac{\partial f}{\partial y}\imath = 0 \Leftrightarrow \frac{\partial f}{\partial z^*} = 0 \tag{18}$$

where $z = x + y\imath$. By continuity, the analyticity in the quaternion domain can be defined by the Generalized Cauchy-Riemann (GCR) conditions, given by [30]

$$\frac{\partial f}{\partial q_a} = -\frac{\partial f}{\partial q_b}\imath; \qquad \frac{\partial f}{\partial q_a} = -\frac{\partial f}{\partial q_c}\jmath; \qquad \frac{\partial f}{\partial q_a} = -\frac{\partial f}{\partial q_d}\kappa \tag{19}$$

where $q = q_a + q_b\imath + q_c\jmath + q_d\kappa$. Fueter further modified these conditions to propose the Cauchy-Riemann-Fueter (CRF) condition given by [5]

$$\frac{\partial f}{\partial q_a} + \frac{\partial f}{\partial q_b}\imath + \frac{\partial f}{\partial q_c}\jmath + \frac{\partial f}{\partial q_d}\kappa = 0 \Leftrightarrow \frac{\partial f}{\partial q^*} = 0 \tag{20}$$

It can be shown that only linear quaternion functions and constant quaternion values satisfy the CRF condition [5]; limiting the scope for nonlinear adaptive filtering in $\mathbb{H}$ which requires holomorphic nonlinear functions. To further relax the quaternion analyticity condition in $\mathbb{H}$, a "local" analyticity condition was proposed in [12], by using a complex representation of a quaternion to give

$$\frac{\partial f}{\partial q_a} = -\frac{\partial f}{\partial \alpha}\hat{\zeta} \tag{21}$$

where $\hat{\zeta}$ and α are given by

$$\hat{\zeta} = \frac{q_b\imath + q_c\jmath + q_d\kappa}{\alpha} \tag{22}$$

$$\alpha = \sqrt{q_b^2 + q_c^2 + q_d^2} \tag{23}$$

The term "local" here refers to the fact that this representation uses "imaginary" unit $\hat{\zeta}$ which depends on the values of q_b, q_c and q_d [12].

4.1 Fully Quaternion Functions

Before introducing fully- (as opposed to split-) quaternion nonlinearities, recall that due to adaptive filters producing time varying outputs and therefore nonlinearities applied to them require only local analyticity. It is desirable that such functions share some properties of fully complex nonlinearities, suitable for nonlinear filtering applications, given by [31]:

- $f(z) = u(x,y) + v(x,y)\imath$ is nonlinear in x and y;
- $f(z)$ has no singularities and is always bounded for all values of z;
- The partial derivatives $\frac{\partial u}{\partial x}$, $\frac{\partial v}{\partial y}$, $\frac{\partial v}{\partial x}$ and $\frac{\partial u}{\partial y}$ are continuous and bounded;
- $\frac{\partial u}{\partial x}\frac{\partial v}{\partial y} \neq \frac{\partial v}{\partial x}\frac{\partial u}{\partial y}$ to ensure continuous learning.

The so called fully complex activation functions satisfy locally all the characteristics above [11]. Notice that fulfilling the third and fourth characteristics is equivalent with fulfilling the CR conditions (18). To address the possibility of finding fully quaternion nonlinearities in $\mathbb{H}$ in order to provide a rigorous basis for nonlinear quaternion-valued adaptive filtering, the class of functions that satisfies the local analyticity condition in (21) is termed a fully quaternion nonlinearity in the sense of local analyticity. We shall now employ the local analyticity condition in (21) to evaluate the analyticity of a function at a given point. We begin by analysing the local derivative within the $\hat{\zeta}$-plane (with $\hat{\zeta}$ fixed) to obtain the relationship [12]

$$\begin{aligned}\frac{\partial f}{\partial \alpha} &= \frac{\partial q_b}{\partial \alpha}\frac{\partial f}{\partial q_b} + \frac{\partial q_c}{\partial \alpha}\frac{\partial f}{\partial q_c} + \frac{\partial q_d}{\partial \alpha}\frac{\partial f}{\partial q_d} \\ \alpha\frac{\partial f}{\partial \alpha} &= q_b\frac{\partial f}{\partial q_b} + q_c\frac{\partial f}{\partial q_c} + q_d\frac{\partial f}{\partial q_d}\end{aligned} \tag{24}$$

Using the relationship defined in (24) along with $\hat{\zeta}$ in (22) and α in (23), the right hand side of the analyticity condition in (21) is expanded along the orthogonal-axis vectors $\imath$, $\jmath$ and κ as

$$-\left(\frac{\partial f}{\partial \alpha}\right)\left(\hat{\zeta}\right) = -\left(\frac{q_b}{\alpha}\frac{\partial f}{\partial q_b} + \frac{q_c}{\alpha}\frac{\partial f}{\partial q_c} + \frac{q_d}{\alpha}\frac{\partial f}{\partial q_d}\right)\left(\frac{q_b\imath + q_c\jmath + q_d\kappa}{\alpha}\right) \tag{25}$$

Therefore, by continuity, the characteristics of a fully quaternion locally analytic nonlinearity suitable for gradient based learning are given by

- $f(q) = u(q_a, \alpha) + v(q_a, \alpha)\hat{\zeta}$ is nonlinear in q_a and α;
- $f(q)$ has no singularities and is always bounded for all values of q;
- The partial derivatives $\frac{\partial u}{\partial q_a}$, $\frac{\partial v}{\partial \alpha}$, $\frac{\partial v}{\partial q_a}$ and $\frac{\partial u}{\partial \alpha}$ are continuous and bounded;
- $\frac{\partial u}{\partial q_a}\frac{\partial v}{\partial \alpha} \neq \frac{\partial v}{\partial q_a}\frac{\partial u}{\partial \alpha}$ to ensure continuous learning.

We shall now examine the analyticity of the quaternion exponential function e^q, as it serves as a building block to construct other transcendental nonlinear quaternion functions.

4.2 Quaternion Exponential Function

To examine the analyticity of e^q, we first need to extend the notion of exponential function into $\mathbb{H}$; this is not straightforward, due to the non-commutativity of the quaternion product. There exist several definitions of the quaternion exponential function [32]; for convenience, we consider the following exponential function (p.9 [33])

$$\mathrm{e}^q = \mathrm{e}^{q_a + q_b\imath + q_c\jmath + q_d\kappa} = \mathrm{e}^{q_a}\mathrm{e}^{q_b\imath + q_c\jmath + q_d\kappa} \tag{26}$$

Expanding the term e^q term using the Euler formula leads to

$$\begin{aligned}\mathrm{e}^q &= \mathrm{e}^{q_a}\Big(\cos(\alpha) + \sin(\alpha)\hat{\zeta}\Big) && (27)\\ &= \mathrm{e}^{q_a}\left(\cos(\alpha) + \frac{q_b\sin(\alpha)\imath}{\alpha} + \frac{q_c\sin(\alpha)\jmath}{\alpha} + \frac{q_d\sin(\alpha)\kappa}{\alpha}\right) && (28)\end{aligned}$$

where α and $\hat{\zeta}$ are defined in (23) and (22). To examine whether such quaternion exponential function satisfies the analyticity condition in (21), we first differentiate (28) with respect to q_a to give the left hand side of (21), that is

$$\frac{\partial \mathrm{e}^q}{\partial q_a} = \mathrm{e}^{q_a}\Big(\cos(\alpha) + \sin(\alpha)\hat{\zeta}\Big) \tag{29}$$

Next, (28) is differentiated with respect to α to obtain the right hand side of (21) as

$$-\frac{\partial \mathrm{e}^q}{\partial \alpha}\hat{\zeta} = -\left(\frac{q_b}{\alpha}\frac{\partial \mathrm{e}^q}{\partial q_b} + \frac{q_c}{\alpha}\frac{\partial \mathrm{e}^q}{\partial q_c} + \frac{q_d}{\alpha}\frac{\partial \mathrm{e}^q}{\partial q_d}\right)\left(\frac{q_b\imath + q_c\jmath + q_d\kappa}{\alpha}\right) \tag{30}$$

The result of such differentiation is given by (see Appendix A for a full derivation)

$$-\frac{\partial \mathrm{e}^q}{\partial \alpha}\hat{\zeta} = \mathrm{e}^{q_a}\Big(\cos(\alpha) + \sin(\alpha)\hat{\zeta}\Big) \tag{31}$$

A comparison of (29) and (31) shows that the exponential function defined in (26) satisfies the local analyticity condition (21). It is now straightforward to introduce the local derivative of the exponential function, given as

$$\frac{\partial \mathrm{e}^q}{\partial q} = \frac{\partial \mathrm{e}^q}{\partial q_a} = \mathrm{e}^q \tag{32}$$

Observe that, as desired, this result is a generic extension of the real and complex exponential function derivatives. In addition, as gradient based learning algorithms are local, this result provides a basis for introducing other common nonlinearities, such as the elementary transcendental functions, and a vehicle for a class of fully quaternion nonlinear adaptive filters.

4.3 Local Analyticity of the Quaternion tanh Function

Similarly to the complex domain, $\tanh(q)$ in $\mathbb{H}$ can be defined as

$$\tanh(q) = \frac{\mathrm{e}^{2q} - 1}{\mathrm{e}^{2q} + 1} \tag{33}$$

Proceeding in a similar manner as when determining the analyticity of e^q, we shall first expand $\tanh(q)$ in (33) using the Euler formula, leading to (full derivation is given in Appendix B)

$$\tanh(q) = \frac{\mathrm{e}^{4q_a} - 1 + 2\mathrm{e}^{2q_a}\sin(2\alpha)}{\mathrm{e}^{4q_a} + 2\mathrm{e}^{2q_a}\cos(2\alpha) + 1} \tag{34}$$

To prove the local analyticity, the left hand side of (21) is obtained by differentiating (34) with respect to q_a and, the right hand side of (21) is obtained by differentiating (34) with respect to α, resulting in (detailed derivations are given in Appendix C)

$$\begin{aligned} \frac{\partial \tanh(q)}{\partial q_a} &= \frac{4\mathrm{e}^{6q_a}\cos(2\alpha) + 8\mathrm{e}^{4q_a} + 4\mathrm{e}^{2q_a}\cos(2\alpha)}{\left(\mathrm{e}^{4q_a} + 2\mathrm{e}^{2q_a}\cos(2\alpha) + 1\right)^2} + \frac{\left(4\mathrm{e}^{2q_a}\sin(2\alpha) - 4\mathrm{e}^{6q_a}\sin(2\alpha)\right)}{\left(\mathrm{e}^{4q_a} + 2\mathrm{e}^{2q_a}\cos(2\alpha) + 1\right)^2}\hat{\zeta} & (35) \\ &= -\frac{\partial \tanh(q)}{\partial \alpha}\hat{\zeta} & (36) \end{aligned}$$

and illustrating that $\tanh(q)$ is a locally analytic quaternion function. The expression for a local derivative of $\tanh(q)$, is obtained analogously to the complex case; to this end we shall first define $\mathrm{sech}(q)$ as

$$\mathrm{sech}(q) = \frac{2}{\mathrm{e}^{q} + \mathrm{e}^{-q}} \tag{37}$$

By expanding (37) into its Euler form and then squaring (full derivation can be found in Appendix D), we have

$$\mathrm{sech}^2(q) = \frac{4\mathrm{e}^{6q_a}\cos(2\alpha) + 8\mathrm{e}^{4q_a} + 4\mathrm{e}^{2q_a}\cos(2\alpha)}{\left(\mathrm{e}^{4q_a} + 2\mathrm{e}^{2q_a}\cos(2\alpha) + 1\right)^2} + \frac{-4\mathrm{e}^{6q_a}\sin(2\alpha) + 4\mathrm{e}^{2q_a}\sin(2\alpha)}{\left(\mathrm{e}^{4q_a} + 2\mathrm{e}^{2q_a}\cos(2\alpha) + 1\right)^2}\hat{\zeta} \tag{38}$$

Comparing the definition for $\mathrm{sech}^2(q)$ in (38) with $\frac{\partial \tanh(q)}{\partial q_a} = -\frac{\partial \tanh(q)}{\partial \alpha}\hat{\zeta}$ in (36) shows that they are equivalent; therefore, we have shown that, as desired,

$$\frac{\partial \tanh(q)}{\partial q} = \mathrm{sech}^2(q) \tag{39}$$

5 Nonlinear Adaptive Filtering in $\mathbb{H}$

The cost function in quaternion-valued adaptive filtering is given by

$$\begin{aligned} E(n) &= e_a^2(n) + e_b^2(n) + e_c^2(n) + e_d^2(n) \\ &= e(n)e^*(n) & (40) \end{aligned}$$

where the error $e(n) = d(n) - y(n)$ with $d(n)$ and $y(n)$ corresponding respectively to the desired signal and output signal. The terms $e_a(n)$, $e_b(n)$, $e_c(n)$ and $e_d(n)$ denote the error component in the real part, $\imath$ part, $\jmath$ part and κ part.

5.1 Review of Nonlinear Quaternion Adaptive Filtering Algorithms

All current nonlinear quaternion-based adaptive filtering algorithms employ a "split" quaternion nonlinear function, that is a real function such as tanh applied componentwise. The output signal $y(n)$ is defined as [6] [9]

$$y(n) = \Phi_s(\mathbf{w}^T(n)\mathbf{x}(n)) = \Phi_a\big(\mathbf{w}^T(n)\mathbf{x}(n)\big) + \Phi_b\big(\mathbf{w}^T(n)\mathbf{x}(n)\big)\imath + \Phi_c\big(\mathbf{w}^T(n)\mathbf{x}(n)\big)\jmath + \Phi_d\big(\mathbf{w}^T(n)\mathbf{x}(n)\big)\kappa \tag{41}$$

where $\Phi_s(\cdot)$ denotes the "split" quaternion nonlinearity, $\mathbf{w}(n)$ is the adaptive filter weight vector and $\mathbf{x}(n)$ is the filter input. Function Φ_a is a real-valued nonlinear activation function applied to the real part of $\mathbf{w}^T(n)\mathbf{x}(n)$, Φ_b to the $\imath$ part, Φ_c to the $\jmath$ part and Φ_d to the κ part. This "split" quaternion function is analytic only componentwise, meaning that we are not fully exploiting the couplings between the $\{1,\imath,\jmath,\kappa\}$ axes (channels). Notice that the odd-symmetry property still applies to the split quaternion function, that is $\Phi_s^{'*}\big(\mathbf{w}^T(n)\mathbf{x}(n)\big) = \Phi_s^{'}\big(\mathbf{x}^H(n)\mathbf{w}(n)\big)$. Existing nonlinear quaternion based algorithms such as the QMLP learning algorithm derived for the feedforward nonlinear neural network architecture [6], minimize the cost function (40) through a gradient descent weight update specified by $\mathbf{w}(n+1) = \mathbf{w}(n) - \mu\nabla_{\mathbf{w}}E(n)$, where the gradient $\nabla_{\mathbf{w}}E(n)$ is given by [34]

$$\nabla_{\mathbf{w}}E(n) = \frac{\partial E}{\partial \mathbf{w}_a} + \frac{\partial E}{\partial \mathbf{w}_b}\imath + \frac{\partial E}{\partial \mathbf{w}_c}\jmath + \frac{\partial E}{\partial \mathbf{w}_d}\kappa \tag{42}$$

For fair comparison with our proposed algorithms, we shall consider the simplified QMLP with only one neuron, that is, Split-Quaternion Finite Impulse Response (QFIR) weight update given by [9]

$$\mathbf{w}(n+1) = \mathbf{w}(n) - \mu\Big(e(n)\Phi_s^{'}\big(\mathbf{w}^T(n)\mathbf{x}(n)\big)\mathbf{x}^*(n)\Big) \tag{43}$$

where $\Phi_s^{'}(\cdot)$ denotes the derivative of the split quaternion function $\Phi_s(\cdot)$ and μ is a real-valued learning rate. One of the main drawbacks of the QFIR learning algorithm is that it ignores the non-commutativity of the quaternion product, leading to a suboptimal performance. An improvement over the QFIR algorithm was achieved based on the SQAFA in [9] which takes into account the non-commutativity aspect of quaternion algebra resulting in superior performance. The weight update of SQAFA can be expressed as [9]

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \mu\Big(2e(n)\Phi_s^{'}\big(\mathbf{x}^H(n)\mathbf{w}^*(n)\big)\mathbf{x}^*(n) - \Phi_s^{'}\big(\mathbf{w}^T(n)\mathbf{x}(n)\big)\mathbf{x}^*(n)e^*(n)\Big) \tag{44}$$

The SQAFA was further improved to tackle the large dynamic range of the signal by adding an adaptive amplitude of the activation function to give the AASQAFA [9]. The output $y(n)$ of AASQAFA is given by

$$\begin{aligned}\Phi_s\big(\mathbf{w}^T(n)\mathbf{x}(n)\big) &= \lambda_a(n)\bar{\Phi}_a\big(\mathbf{w}^T(n)\mathbf{x}(n)\big) + \lambda_b(n)\bar{\Phi}_b\big(\mathbf{w}^T(n)\mathbf{x}(n)\big)\imath \\ &+ \lambda_c(n)\bar{\Phi}_c\big(\mathbf{w}^T(n)\mathbf{x}(n)\big)\jmath + \lambda_d(n)\bar{\Phi}_d\big(\mathbf{w}^T(n)\mathbf{x}(n)\big)\kappa \end{aligned} \tag{45}$$

where $\lambda_a(n)$ is the real-valued amplitude of the nonlinearity for the real part of the quaternion, $\lambda_b(n)$ for the $\imath$ part, $\lambda_c(n)$ for the $\jmath$ part and $\lambda_d(n)$ for the κ part. The term $\bar{\Phi}_a$ refers to the unit amplitude of the nonlinear function in the real part, $\bar{\Phi}_b$ in the $\imath$ part, $\bar{\Phi}_c$ in the $\jmath$ part and $\bar{\Phi}_d$ in the κ part. The parameter λ is made adaptive according to [35] [36]

$$\lambda(n+1) = \lambda(n) - \rho\nabla_{\lambda}E(n) \tag{46}$$

where ρ is a real-valued learning rate. Due to the limitation of the "split" quaternion nonlinearity, the update for the amplitudes are performed componentwise, given by [9]

$$\lambda_i(n+1) = \lambda_i(n) + \rho e_i(n)\bar{\Phi}_i\big(\mathbf{w}^T(n)\mathbf{x}(n)\big), \quad i \in \{a, b, c, d\} \tag{47}$$

We shall next employ the proposed fully quaternion activation functions leading to structurally simpler yet more powerful algorithms.

5.2 Derivation of the Quaternion Adaptive Filtering Algorithm (QAFA)

The Quaternion Adaptive Filtering Algorithm (QAFA) employs a fully quaternion function instead of the "split" quaternion function with the output $y(n)$ given by

$$y(n) = \Phi\big(\mathbf{w}^T(n)\mathbf{x}(n)\big) \tag{48}$$

where $\Phi(\cdot)$ is the fully quaternion nonlinearity such as the $\tanh(q)$ introduced in Section IV. In order to derive the QAFA, we shall express the cost function (40) as

$$\begin{aligned} E(n) &= \Big(d(n) - y(n)\Big)\Big(d^*(n) - y^*(n)\Big) \\ &= d(n)d^*(n) - d(n)y^*(n) - y(n)d^*(n) + y(n)y^*(n) \end{aligned} \tag{49}$$

The error gradient $\nabla_{\mathbf{w}}E(n)$ of QAFA is then calculated as

$$\nabla_{\mathbf{w}}E(n) = -d(n)\nabla_{\mathbf{w}}y^*(n) - \nabla_{\mathbf{w}}y(n)d^*(n) + y(n)\nabla_{\mathbf{w}}y^*(n) + \nabla_{\mathbf{w}}y(n)y^*(n) \tag{50}$$

and the expressions for $\nabla_{\mathbf{w}}y(n)$ and $\nabla_{\mathbf{w}}y^*(n)$ are given by (the full derivation is given in Appendix E)

$$\begin{aligned} \nabla_{\mathbf{w}}y(n) &= -\Phi^{'}\big(\mathbf{w}^T(n)\mathbf{x}(n)\big)2\mathbf{x}^*(n) \\ \nabla_{\mathbf{w}}y^*(n) &= \Phi^{'*}\big(\mathbf{w}^T(n)\mathbf{x}(n)\big)4\mathbf{x}^*(n) \end{aligned} \tag{51}$$

Substitute the terms $\nabla_{\mathbf{w}}y^*(n)$ and $\nabla_{\mathbf{w}}y(n)$ into (50) to obtain the QAFA weight update in the form

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \mu\Big(2e(n)\Phi^{'}\big(\mathbf{x}^H(n)\mathbf{w}^*(n)\big)\mathbf{x}^*(n) - \Phi^{'}\big(\mathbf{w}^T(n)\mathbf{x}(n)\big)\mathbf{x}^*(n)e^*(n)\Big) \tag{52}$$

where $\Phi^{'}(\cdot)$ is the local derivative of the fully quaternion function.

5.3 Derivation of Widely Linear Quaternion Adaptive Filtering Algorithm (WLQAFA)

We shall now extend the QAFA to fully capture the second-order statistics of the signal by incorporating the quaternion widely linear model [22] [29] into its derivation, resulting in the Widely Linear Quaternion Adaptive Filtering Algorithm (WLQAFA)[2]. Based on the widely linear model introduced in Section III, the output $y(n)$ of WLQAFA is defined as

$$y(n) = \Phi\big(\mathbf{g}^T(n)\mathbf{x}(n) + \mathbf{h}^T(n)\mathbf{x}^{\imath}(n) + \mathbf{u}^T(n)\mathbf{x}^{\jmath}(n) + \mathbf{v}^T(n)\mathbf{x}^{\kappa}(n)\big) = \Phi\big(\mathrm{net}(n)\big) \tag{53}$$

where $\mathrm{net}(n) = \mathbf{g}^T(n)\mathbf{x}(n) + \mathbf{h}^T(n)\mathbf{x}^{\imath}(n) + \mathbf{u}^T(n)\mathbf{x}^{\jmath}(n) + \mathbf{v}^T(n)\mathbf{x}^{\kappa}(n)$ is the widely linear part, to which the function $\Phi(\cdot)$ is applied.
The weight updates of the WLQAFA are made gradient adaptive according to

$$\begin{aligned} \mathbf{g}(n+1) &= \mathbf{g}(n) - \mu\nabla_{\mathbf{g}}E(n); \qquad & \mathbf{h}(n+1) &= \mathbf{h}(n) - \mu\nabla_{\mathbf{h}}E(n) \\ \mathbf{u}(n+1) &= \mathbf{u}(n) - \mu\nabla_{\mathbf{u}}E(n); \qquad & \mathbf{v}(n+1) &= \mathbf{v}(n) - \mu\nabla_{\mathbf{v}}E(n) \end{aligned} \tag{54}$$

The error gradient $\nabla_{\mathbf{w}}E(n)$ in (52) is equivalent to $\nabla_{\mathbf{g}}E(n)$, hence

$$\mathbf{g}(n+1) = \mathbf{g}(n) + \mu\Big(2e(n)\Phi^{'*}\big(\mathrm{net}(n)\big)\mathbf{x}^*(n) - \Phi^{'}\big(\mathrm{net}(n)\big)\mathbf{x}^*(n)e^*(n)\Big) \tag{55}$$

where $\Phi^{'}\big(\mathrm{net}(n)\big)$ is the fully quaternion derivatives. The error gradient $\nabla_{\mathbf{h}}E(n)$ is given by

$$\nabla_{\mathbf{h}}E(n) = -d(n)\nabla_{\mathbf{h}}y^*(n) - \nabla_{\mathbf{h}}y(n)d^*(n) + y(n)\nabla_{\mathbf{h}}y^*(n) + \nabla_{\mathbf{h}}y(n)y^*(n) \tag{56}$$

Following on in the same manner, the terms $\nabla_{\mathbf{h}}y(n)$ and $\nabla_{\mathbf{h}}y^*(n)$ are both calculated as

$$\begin{aligned} \nabla_{\mathbf{h}}y(n) &= -\Phi^{'}\big(\mathrm{net}(n)\big)2\mathbf{x}^{\imath *}(n) \\ \nabla_{\mathbf{h}}y^*(n) &= \Phi^{'*}\big(\mathrm{net}(n)\big)4\mathbf{x}^{\imath *}(n) \end{aligned} \tag{57}$$

Substituting $\nabla_{\mathbf{h}}y(n)$ and $\nabla_{\mathbf{h}}y^*(n)$ into the error gradient $\nabla_{\mathbf{h}}E(n)$ in (56) yields

$$\mathbf{h}(n+1) = \mathbf{h}(n) + \mu\Big(2e(n)\Phi^{'*}\big(\mathrm{net}(n)\big)\mathbf{x}^{\imath *}(n) - \Phi^{'}\big(\mathrm{net}(n)\big)\mathbf{x}^{\imath *}(n)e^*(n)\Big) \tag{58}$$

Proceeding in a similar manner, the weight updates for $\mathbf{u}(n)$ and $\mathbf{v}(n)$ are found to be

$$\begin{aligned} \mathbf{u}(n+1) &= \mathbf{u}(n) + \mu\Big(2e(n)\Phi^{'*}\big(\mathrm{net}(n)\big)\mathbf{x}^{\jmath *}(n) - \Phi^{'}\big(\mathrm{net}(n)\big)\mathbf{x}^{\jmath *}(n)e^*(n)\Big) \\ \mathbf{v}(n+1) &= \mathbf{v}(n) + \mu\Big(2e(n)\Phi^{'*}\big(\mathrm{net}(n)\big)\mathbf{x}^{\kappa *}(n) - \Phi^{'}\big(\mathrm{net}(n)\big)\mathbf{x}^{\kappa *}(n)e^*(n)\Big) \end{aligned} \tag{59}$$

For convenience of representation, the final weight update of the WLQAFA can be written in an augmented form as

$$\mathbf{w}^{\mathbf{a}}(n+1) = \mathbf{w}^{\mathbf{a}}(n) + \mu\Big(2e(n)\Phi^{'*}\big(\mathrm{net}(n)\big)\mathbf{x}^{a*}(n) - \Phi^{'}\big(\mathrm{net}(n)\big)\mathbf{x}^{a*}(n)e^*(n)\Big) \tag{60}$$

6 Convergence Analysis of QAFA and WLQAFA

The convergence criterion employed in this work is given by

$$\|\bar{e}(n)\|_2^2 < \|\tilde{e}(n)\|_2^2 \tag{61}$$

where $\bar{e}$ and $\tilde{e}$ are respectively the a posteriori output error and the a priori output error. To proceed with the analysis, we will make two widely used general assumptions [37]

- the learning rate μ is small;
- at convergence, $\tilde{e}(n)$ is statistically independent of $\mathbf{x}(n)$.

[2] A full account of widely linear modelling in $\mathbb{C}$ is given in [10]

6.1 Convergence of QAFA

The a posteriori output error $\bar{e}$ and the a priori output error $\tilde{e}$ are given by

$$\begin{aligned} \bar{e}(n) &= d(n) - \Phi\big(\mathbf{w}^T(n+1)\mathbf{x}(n)\big) \\ \tilde{e}(n) &= d(n) - \Phi\big(\mathbf{w}^T(n)\mathbf{x}(n)\big) \end{aligned} \tag{62}$$

and can be related by the first order Taylor series expansion [38]

$$\|\bar{e}(n)\|_2^2 = \|\tilde{e}(n)\|_2^2 + \Delta\mathbf{w}^H(n)\frac{\partial\|\tilde{e}(n)\|_2^2}{\partial\mathbf{w}(n)} \tag{63}$$

where $\frac{\partial\|\tilde{e}(n)\|_2^2}{\partial\mathbf{w}(n)}$ is the QAFA error gradient, and the term $\Delta\mathbf{w}^H(n)$ in (63) is obtained from (52) as

$$\Delta\mathbf{w}^H = \mu\big[2\mathbf{x}^T(n)\Phi^{'*}\big(\mathbf{x}^H(n)\mathbf{w}^*(n)\big)\tilde{e}^*(n) - \tilde{e}(n)\mathbf{x}^T(n)\Phi'^*\big(\mathbf{w}^T(n)\mathbf{x}(n)\big)\big] \tag{64}$$

The term $\frac{\partial\|\tilde{e}(n)\|_2^2}{\partial\mathbf{w}(n)}$ is given by

$$\frac{\partial\|\tilde{e}(n)\|_2^2}{\partial\mathbf{w}(n)} = -\big[4\tilde{e}(n)\Phi'\big(\mathbf{x}^H(n)\mathbf{w}^*(n)\big)\mathbf{x}^*(n) - 2\Phi'\big(\mathbf{w}^T(n)\mathbf{x}(n)\big)\mathbf{x}^*(n)\tilde{e}^*(n)\big] \tag{65}$$

Substitute (64) - (65) into the Taylor series expansion (63) to yield

$$\begin{aligned} \|\bar{e}(n)\|_2^2 &= \|\tilde{e}(n)\|_2^2 - \mu\Big(\big[2\mathbf{x}^T(n)\Phi'^*\big(\mathbf{x}^H(n)\mathbf{w}^*(n)\big)\tilde{e}^*(n) \\ &- \tilde{e}(n)\mathbf{x}^T(n)\Phi'^*\big(\mathbf{w}^T(n)\mathbf{x}(n)\big)\big]\big[4\tilde{e}(n)\Phi'\big(\mathbf{x}^H(n)\mathbf{w}^*(n)\big)\mathbf{x}^*(n) - 2\Phi'\big(\mathbf{w}^T(n)\mathbf{x}(n)\big)\mathbf{x}^*(n)\tilde{e}^*(n)\big]\Big) \end{aligned} \tag{66}$$

Applying the independence assumptions and factorizing the term $\|\tilde{e}(n)\|_2^2$ gives

$$\|\bar{e}(n)\|_2^2 = \|\tilde{e}(n)\|_2^2\Big[1 - 10\mu\mathbf{x}^T(n)\mathbf{x}^*(n)\|\Phi'\big(\mathbf{w}^T(n)\mathbf{x}(n)\big)\|_2^2\Big] \tag{67}$$

The condition for convergence in (61) is satisfied for

$$0 < 10\mu\mathbf{x}^T(n)\mathbf{x}^*(n)\|\Phi'\big(\mathbf{w}^T(n)\mathbf{x}(n)\big)\|_2^2 < 1 \tag{68}$$

Solving for μ we obtain the range of the stepsize for QAFA to converge

$$0 < \mu < \frac{1}{10\mathbf{x}^T(n)\mathbf{x}^*(n)\|\Phi'\big(\mathbf{w}^T(n)\mathbf{x}(n)\big)\|_2^2} \tag{69}$$

6.2 Convergence of WLQAFA

The a posteriori output error $\bar{e}$ and the a priori output error $\tilde{e}$ of the WLQAFA are rewritten as

$$\begin{aligned} \bar{e}(n) &= d(n) - \Phi\big(\mathbf{w}^{aT}(n+1)\mathbf{x}^a(n)\big) \\ \tilde{e}(n) &= d(n) - \Phi\big(\mathbf{w}^{aT}(n)\mathbf{x}^a(n)\big) \end{aligned} \tag{70}$$

To accommodate the widely linear model, the Taylor Series in (63) is modified to

$$\|\bar{e}(n)\|_2^2 = \|\tilde{e}(n)\|_2^2 + \Delta\mathbf{w}^{aH}(n)\frac{\partial\|\tilde{e}(n)\|_2^2}{\partial\mathbf{w}^a(n)} \tag{71}$$

where $\Delta\mathbf{w}^{aH}(n)$ and $\frac{\partial\|\tilde{e}(n)\|_2^2}{\partial\mathbf{w}^a(n)}$ are respectively the Hermitian of the WLQAFA weight update and the error gradient. Following in the same manner as for QAFA,

$$\Delta\mathbf{w}^{aH}(n) = \mu\big[2\mathbf{x}^{aT}(n)\Phi^{'}\big(\mathrm{net}(n)\big)\tilde{e}^*(n) - \tilde{e}(n)\mathbf{x}^{aT}(n)\Phi^{'*}\big(\mathrm{net}(n)\big)\big] \tag{72}$$

$$\frac{\partial\|\tilde{e}(n)\|_2^2}{\partial\mathbf{w}^a(n)} = = -\big[4\tilde{e}(n)\Phi^{'*}\big(\mathrm{net}(n)\big)\mathbf{x}^{a*}(n) - 2\Phi^{'}\big(\mathrm{net}(n)\big)\mathbf{x}^{a*}(n)\tilde{e}^*(n)\big] \tag{73}$$

and the final bounds on μ so that the WLQAFA converges are given by

$$0 < \mu < \frac{1}{10\mathbf{x}^{aT}(n)\mathbf{x}^{a*}(n)\|\Phi'\big(\mathrm{net}(n)\big)\|_2^2} \tag{74}$$

Note that the upper bound of μ for the WLQAFA in (74) is smaller than that of QAFA in (69), due to the larger size of the augmented input vector $\mathbf{x}^a(n)$.

7 Simulations

A comprehensive comparison of the performances is provided for nonlinear FIR filters trained with QFIR [9], AASQAFA [9] and the proposed algorithms based on fully quaternion nonlinear functions, QAFA and WLQAFA. The filter length is denoted by L and the $\tanh(q)$ nonlinear activation function was used. The stepsize for the adaptive amplitude of AASQAFA was set $\rho = 0.4$. The optimal AASQAFA initial amplitude was set to be $\lambda(0) = 1$. The performance was measured in terms of prediction gain R_p defined as [37]

$$R_p = 10 \log_{10} \frac{\sigma_x^2}{\sigma_e^2} \tag{75}$$

where σ_x^2 and σ_e^2 denote respectively the estimated variance of the input and error. The three quaternion valued processes considered were the synthetic linear AR(4) process [10] with a varying degree of circularity, the noncircular chaotic four-dimensional Saito signal [39], and the real-world three-dimensional wind field.

7.1 Linear AR (4)

For the purpose of this experiment, the input tap length was chosen to be $L = 3$, prediction horizon $M = 1$ and the learning rate $\mu = 10^{-2}$. In the first set of experiments, the performances of WLQAFA, QAFA, AASQAFA and QFIR were analyzed for a linear AR (4) process with a varying degree of circularity of the driving quaternion quadruply white Gaussian noise (QWGN) $\epsilon(n)$. A total of 100 independent simulation trials were conducted and averaged for the linear AR (4) process given by

$$r(n) = 1.79r(n-1) - 1.85r(n-2) + 1.27r(n-3) - 0.41r(n-4) + \epsilon(n) \tag{76}$$

Figure 1 shows the learning curves for a $\mathbb{H}$-circular quaternion white Gaussian noise as the driving noise of the linear AR(4) process. Observe that the proposed WLQAFA and QAFA had the fastest convergence, followed by the AASQAFA and QFIR. Figure 2 depicts the learning curves for the input $\mathbb{C}^i$-circular white Gaussian

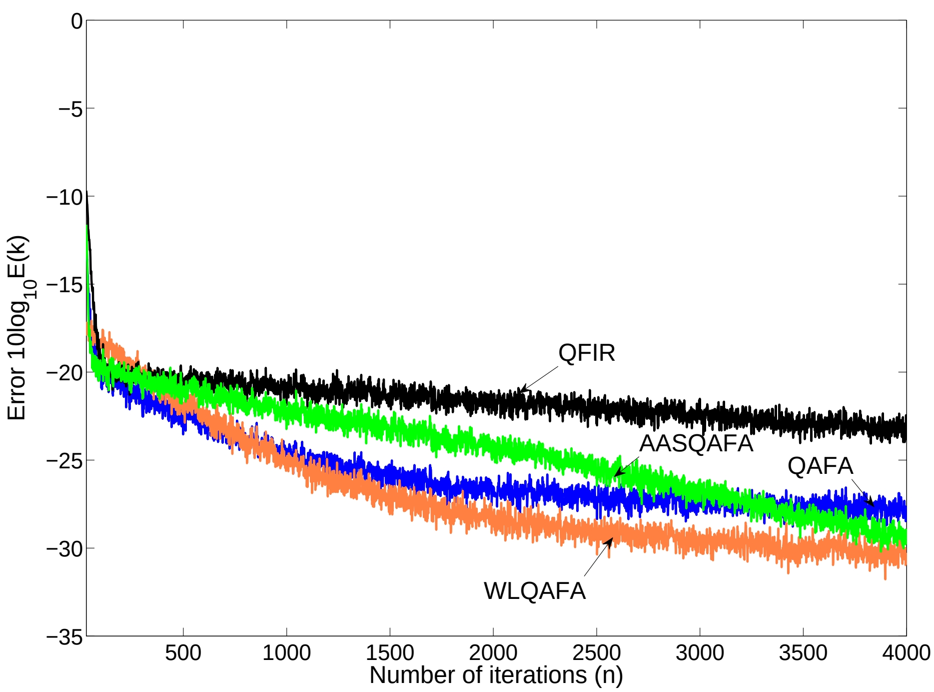


Figure 1: Learning curves for QFIR [9], AASQAFA [9], QAFA and WLQAFA on the prediction of linear AR (4) signal (76) driven by $\mathbb{H}$-circular white Gaussian noise.

noise[3] for all of the algorithms considered. Similar to the previous case, the WLQAFA and QAFA had the fastest convergence, however, in this case, as desired, the steady-state results for WLQAFA and QAFA were equivalent. In the case of $\mathbb{C}^j$ and $\mathbb{C}^k$ white Gaussian noises, similar performances were obtained and are omitted in this work for conciseness. Figure 3 shows learning curves for all the algorithms considered using a noncircular

[3]The notion of $\mathbb{C}^\eta$ circularity refers to only having a pair of axis exhibiting complex circularity.

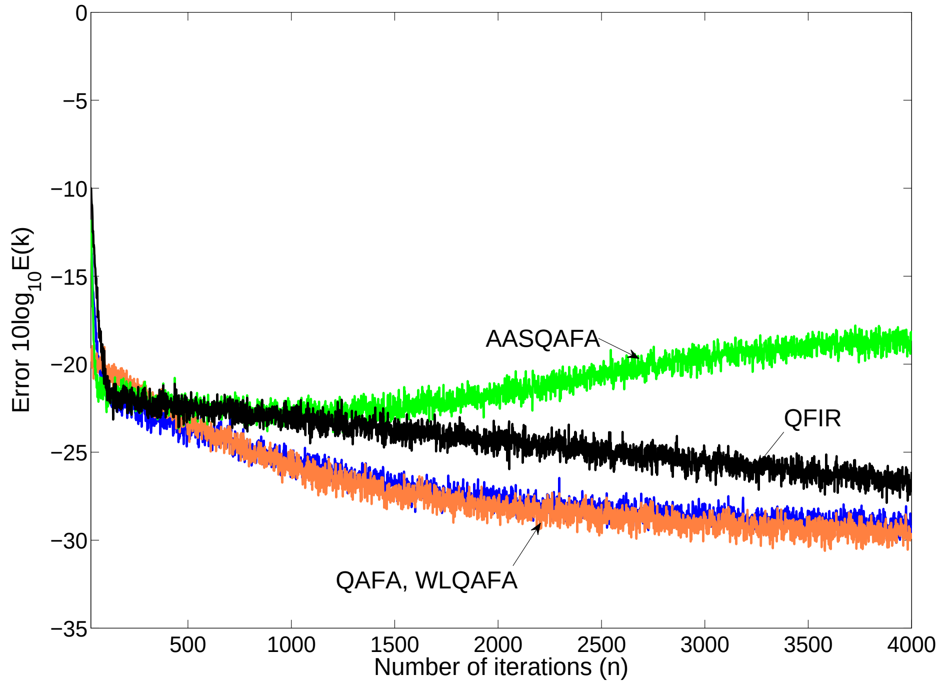


Figure 2: Learning curves for QFIR [9], AASQAFA [9], QAFA and WLQAFA on the prediction of linear AR (4) signal (76) driven by $\mathbb{C}^i$-circular white Gaussian noise.

white Gaussian noise as the input; the WLQAFA and QAFA had superior performances over the AASQAFA and QFIR. It can also be seen that the steady-state of WLQAFA was similar than that of QAFA as they both designed to cater for any AR type of processes. Table 1 reviews a comparison of prediction gains R_p between

Table 1: Prediction Gain R_p for a Linear AR (4) Process With Varying Degree of Noncircularity

Algorithms	$\mathbb{H}$-circular	$\mathbb{C}^i$-circular	$\mathbb{C}^j$-circular	$\mathbb{C}^k$-circular	Noncircular
WLQAFA	20.22dB	20.93dB	20.91dB	20.88dB	21.58dB
QAFA	19.46dB	20.04dB	19.99dB	20.01dB	20.45dB
AASQAFA	18.09dB	15.75dB	15.35dB	15.66dB	17.01dB
QFIR	16.58dB	18.11dB	18.11dB	18.05dB	18.04dB

the WLQAFA, QAFA, AASQAFA and QFIR for the prediction of linear AR (4) process with varying classes of input circularity. In all the cases, the proposed algorithms, WLQAFA and QAFA, had superior performance over the AASQAFA and QFIR, illustrating the power of the fully quaternion function over the "split" quaternion function. Also from Table 1, the use of the quaternion widely linear model for noncircular data is fully justified, as indicated by a higher prediction gain of WLQAFA over the QAFA.

7.2 Four-dimensional Saito's Chaotic Circuit

The four state variables and five parameters that govern Saito's chaotic circuit are given by [39]

$$\begin{aligned} \begin{bmatrix} \frac{\partial x_1}{\partial \tau} \\ \frac{\partial y_1}{\partial \tau} \end{bmatrix} &= \begin{bmatrix} -1 & 1 \\ -\alpha_1 & -\alpha_1\beta_1 \end{bmatrix} \begin{bmatrix} x_1 - \eta\rho_1 h(z) \\ y_1 - \eta\frac{\rho_1}{\beta_1} h(z) \end{bmatrix} \\ \begin{bmatrix} \frac{\partial x_2}{\partial \tau} \\ \frac{\partial y_2}{\partial \tau} \end{bmatrix} &= \begin{bmatrix} -1 & 1 \\ -\alpha_2 & -\alpha_2\beta_2 \end{bmatrix} \begin{bmatrix} x_2 - \eta\rho_2 h(z) \\ y_2 - \eta\frac{\rho_2}{\beta_2} h(z) \end{bmatrix} \end{aligned} \qquad (77)$$

where τ is the time constant of the chaotic circuit and $h(z)$ is the normalized hysteresis value given by [39]

$$h(z) = \begin{cases} 1, & z \geq -1 \\ -1, & z \leq 1 \end{cases} \qquad (78)$$

The parameters z, ρ_1 and ρ_2 are given as $z = x_1 + x_2$, $\rho_1 = \frac{\beta_1}{1-\beta_1}$ and $\rho_2 = \frac{\beta_2}{1-\beta_2}$. The Saito chaotic signal was initialized with the following parameters: η=1.3, α_1=7.5, α_2=15, β_1=0.16 and β_2=0.097, and is shown

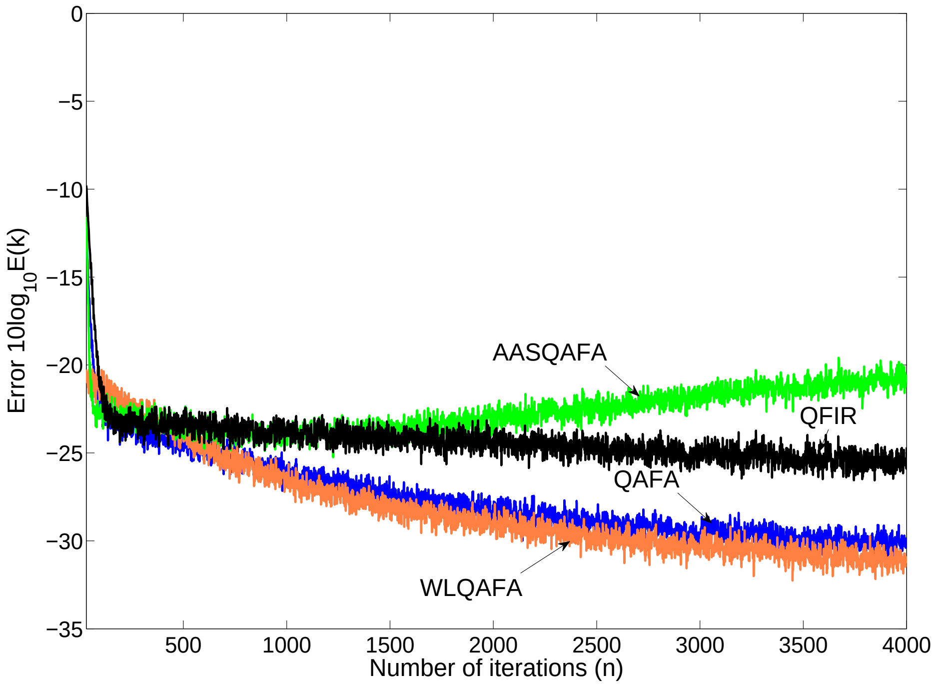


Figure 3: Learning curves for QFIR [9], AASQAFA [9], QAFA and WLQAFA on the prediction of linear AR (4) signal (76) driven by noncircular quaternion Gaussian noise.

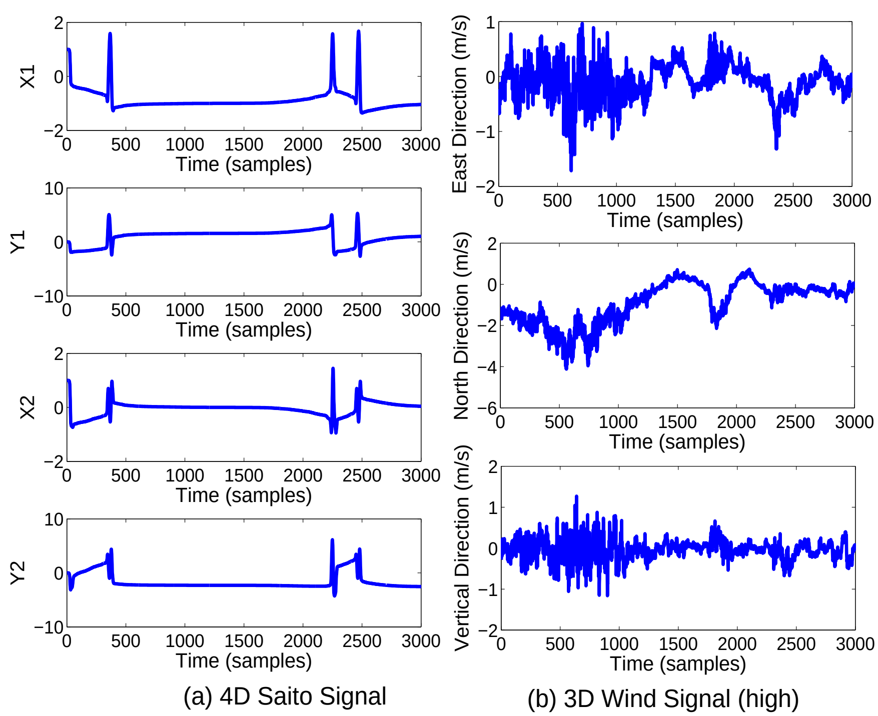


Figure 4: Noncircular signals used in simulations. Left: The 4D Saito Signal. Right: The 3D wind signal (high region).

dimension-wise in Figure 4(a). Figure 5 depicts the performances of the algorithms considered in terms of prediction horizon M (with fixed stepsize $\mu = 10^{-2}$) and stepsize μ (with fixed prediction horizon M=1). Observe that the WLQAFA outperformed all the other algorithms by a margin greater than 2dB.

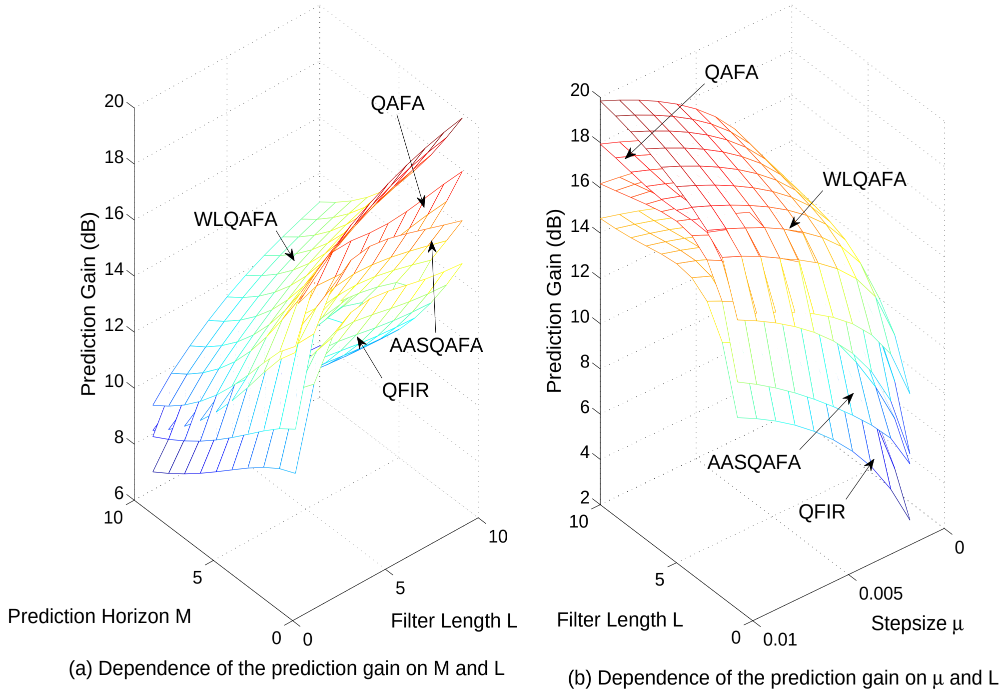


Figure 5: The performance of WLQAFA, QAFA, AASQAFA and QFIR on the prediction of the noncircular 4D Saito signal.

7.3 Wind Forecasting

In this set of simulations, a single realization of three-dimensional wind field was used as the input[4]. Figure 4(b) shows the wind field in the high region used dimension-wise, and Figure 6 illustrates the performances of WLQAFA, QAFA, AASQAFA and QFIR as a function of prediction horizon M and stepsize μ. The performance of WLQAFA was better than that of QAFA; this was closely followed by AASQAFA, whereas the performance of the QFIR was the poorest. Figure 7(a) and Figure 7(b) show respectively the wind field in the medium region and low region and Figure 8 and Figure 9 depict its corresponding prediction gain for the WLQAFA, QAFA, AASQAFA and QFIR. The results obtained are similar to those for the wind field in the high region.

7.4 Three-dimensional Rossler Signal

The Rossler attractor is governed by coupled partial differential equations [40]

$$\begin{aligned} \frac{\partial x}{\partial t} &= -y - z; \\ \frac{\partial y}{\partial t} &= x + \alpha y; \\ \frac{\partial z}{\partial t} &= \beta + z(x - \rho) \end{aligned} \tag{79}$$

where α, ρ and β >0. The Rossler attractor implemented was initialized with the following parameters: $\alpha = 0.1$, $\rho = 14$ and $\beta = 0.1$. The resulting Rossler signal is shown component-wise in Fig 10(a). Figure 11 depicts the performance of the algorithms in terms of prediction horizon M (with fixed stepsize $\mu = 10^{-2}$) and stepsize μ (with fixed prediction horizon M=1). It is clearly seen that the WLQAFA has superior performance over the QAFA followed by the QFIR and AASQAFA.

[4]The wind data were sampled at 32 Hz and recorded by the 3D WindMaster anemometer provided by Gill Instruments.

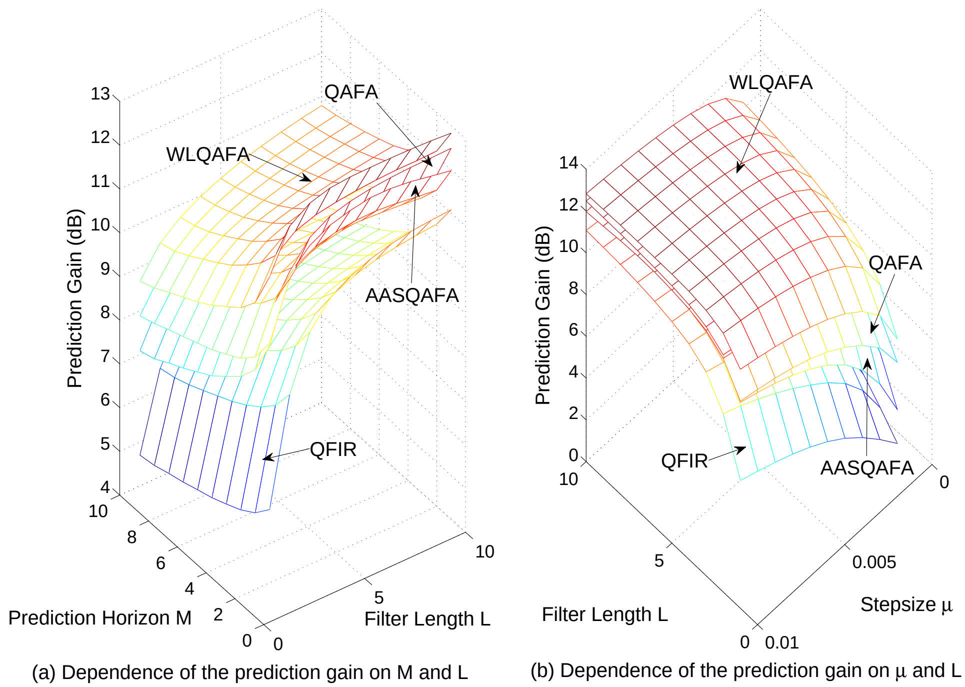


Figure 6: The performance of WLQAFA, QAFA, AASQAFA and QFIR on the prediction of a 3D wind signal in the high region.

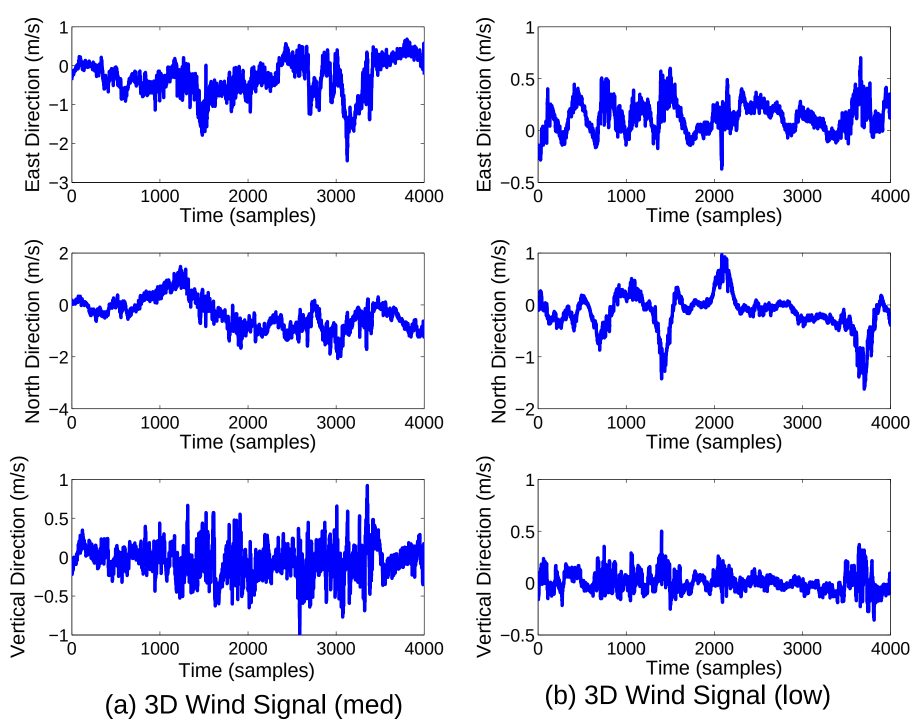
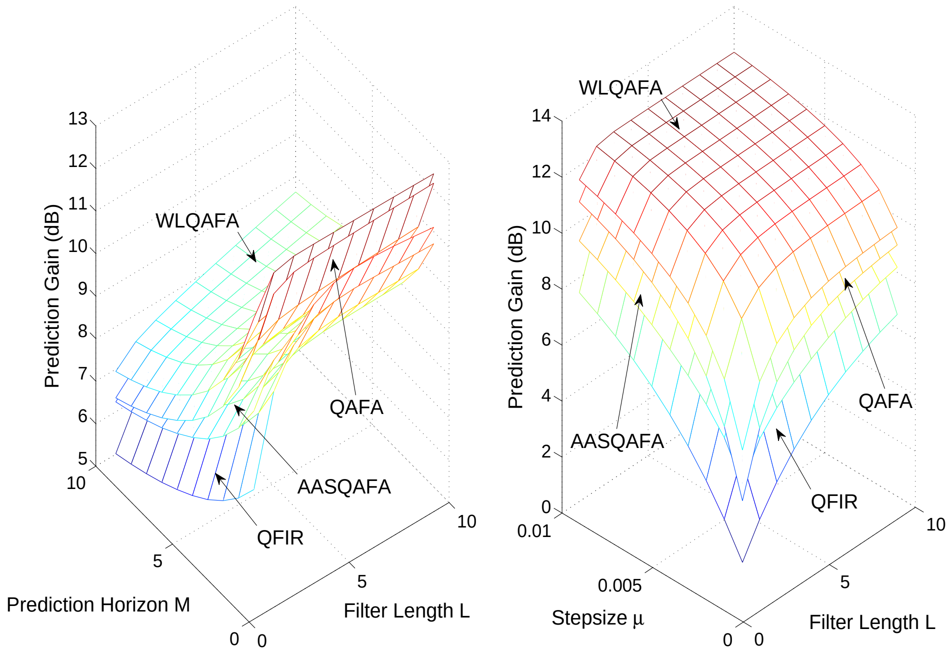
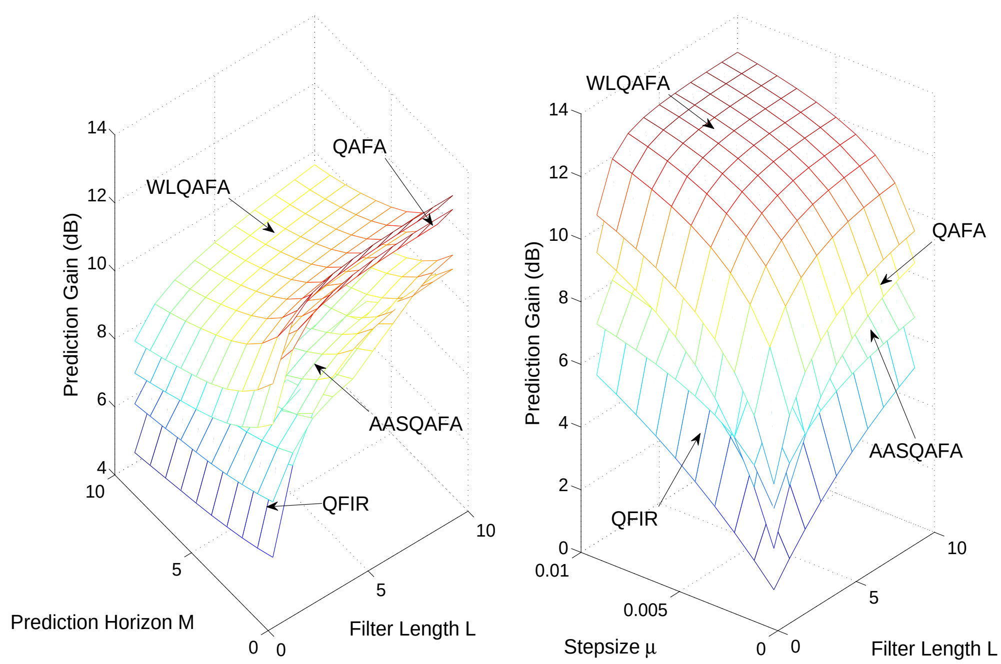


Figure 7: Noncircular signals used in simulations. Left: The 3D wind signal (medium region). Right: The 3D wind signal (low region).



(a) Dependence of the prediction gain on M and L
(b) Dependence of the prediction gain on μ and L

Figure 8: The performance of WLQAFA, QAFA, AASQAFA and QFIR on the prediction of a 3D wind signal in the medium region.



(a) Dependence of the prediction gain on M and L
(b) Dependence of the prediction gain on μ and L

Figure 9: The performance of WLQAFA, QAFA, AASQAFA and QFIR on the prediction of a 3D wind signal in the low region.

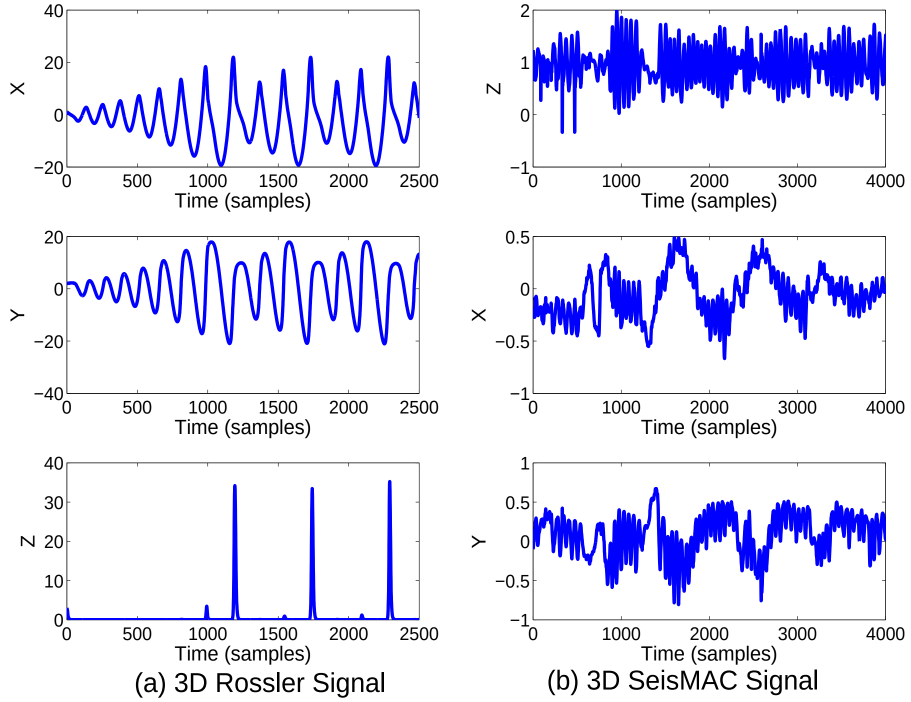


Figure 10: The time waveforms on the Rossler signal (left) and Seismac signal (right).

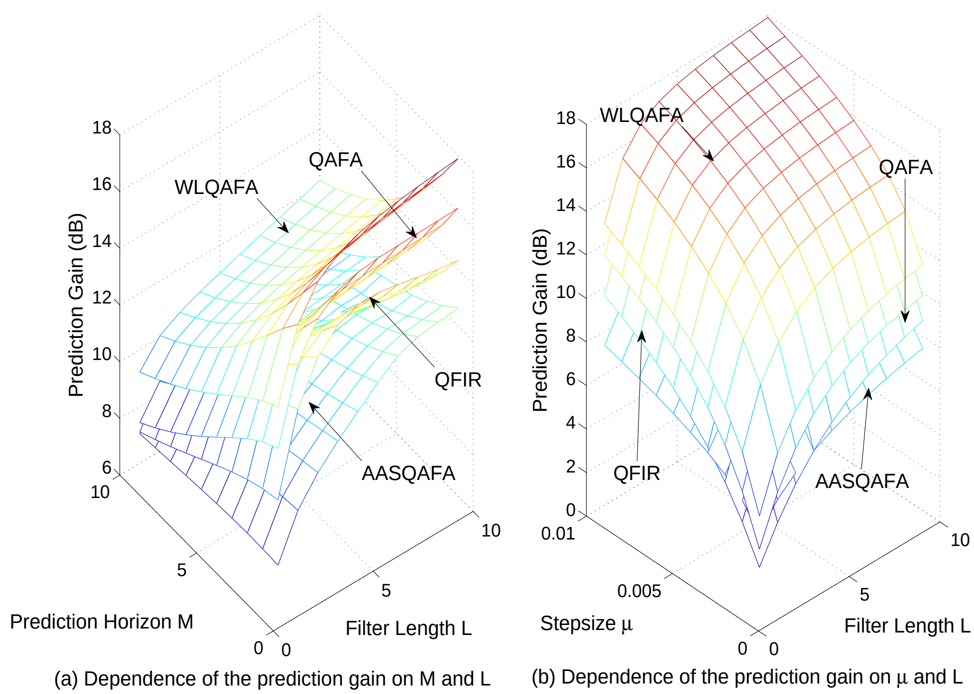


Figure 11: The performance of WLQAFA, QAFA, AASQAFA and QFIR on the prediction of a 3D Rossler signal.

7.5 Forecasting Seismic Field

In this simulation, a three-dimensional seismic field was used as the input[5]. Figure 10(b) shows the seismic field dimension-wise, and Figure 12 illustrates the performances of WLQAFA, QAFA, AASQAFA and QFIR as a function of prediction horizon M and stepsize μ. The performance of WLQAFA was significantly better than that of QAFA followed by AASQAFA and QFIR.

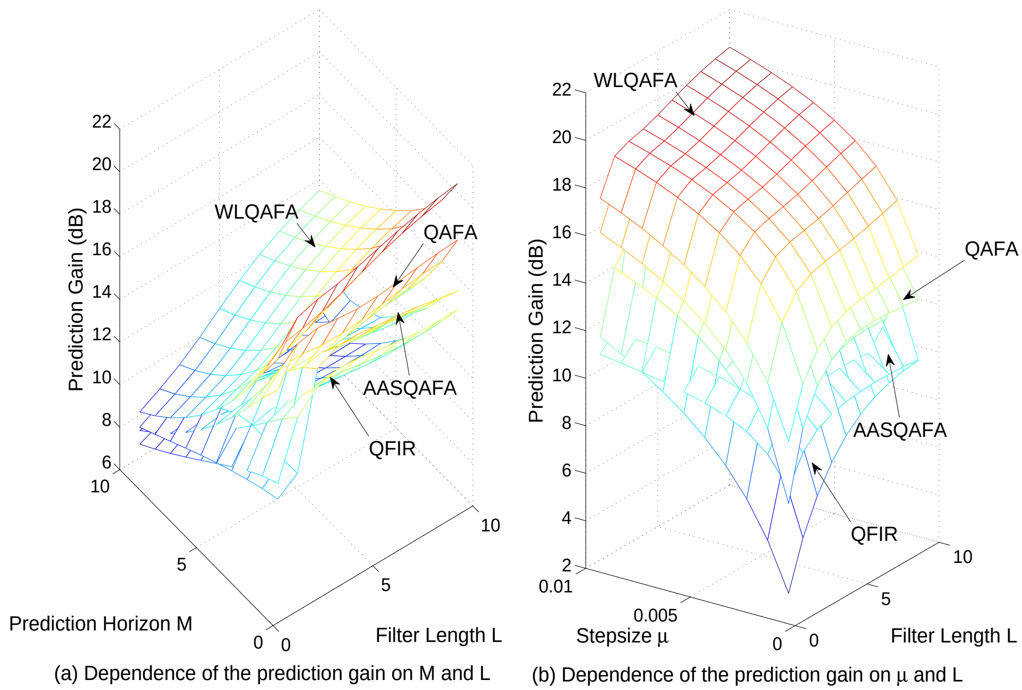


Figure 12: The performance of WLQAFA, QAFA, AASQAFA and QFIR on the prediction of a 3D Seismic Field.

8 Discussion

The performances of the filters with the proposed locally analytic fully quaternion activation functions were generally better than those of the existing AASQAFA and QFIR. The widely linear version outperformed the QAFA, due to the implementation of the quaternion widely linear model that fully captures the second-order statistics of quaternion signals. In order to create a class of fully quaternion function that is suitable for quaternion-valued adaptive filtering, it is essential to examine the possibility of employing other fully complex transcendental functions [11] as locally analytic fully quaternion functions. In Section IV, we have established that the exponential function e^q is locally analytic and, given that summations and products of analytic functions are analytic as well as quotients (provided the denominator does not vanish), the $\tanh(q)$ function is also locally analytic because it can be expressed in terms of e^q as

$$\tanh(q) = \frac{\sinh(q)}{\cosh(q)} = \frac{e^q - e^{-q}}{e^q + e^{-q}} = \frac{e^{2q} - 1}{e^{2q} + 1} \tag{80}$$

This was verified by a rigorous derivation given in Appendix D. By continuity, the other quaternion transcendental functions are also locally analytic. In the complex domain, Duch et al. have shown that if a set of functions are fully analytic, then their performances should be similar [41]. In the same spirit, Figure 7 confirms by simulations that the other elementary transcendental functions give similar performance as the locally analytic function $\tanh(q)$. Thus, the fully complex functions from $\mathbb{C}$ can be extended to fully quaternion functions in $\mathbb{H}$ and are consistent with the observations in [41]. For convenience, the class of locally analytic fully quaternion

[5]The seismic wave data was sampled at 500 Hz in each axis and is recorded by the SeisMAC software provided by Suitable Systems.

functions and their derivatives are given below

$$\tanh(q) \quad : \quad \frac{\partial \tanh(q)}{\partial q} = \mathrm{sech}^2(q) \tag{81}$$

$$\tan(q) \quad : \quad \frac{\partial \tan(q)}{\partial q} = \sec^2(q) \tag{82}$$

$$\sin(q) \quad : \quad \frac{\partial \sin(q)}{\partial q} = \cos(q) \tag{83}$$

$$\arctan(q) \quad : \quad \frac{\partial \arctan(q)}{\partial q} = (1+q^2)^{-1} \tag{84}$$

$$\arcsin(q) \quad : \quad \frac{\partial \arcsin(q)}{\partial q} = (1-q^2)^{-1/2} \tag{85}$$

$$\sinh(q) \quad : \quad \frac{\partial \sinh(q)}{\partial q} = \cosh(q) \tag{86}$$

$$\mathrm{arctanh}(q) \quad : \quad \frac{\partial \mathrm{arctanh}(q)}{\partial q} = (1-q^2)^{-1} \tag{87}$$

$$\mathrm{arcsinh}(q) \quad : \quad \frac{\partial \mathrm{arcsinh}(q)}{\partial q} = (1+q^2)^{-1} \tag{88}$$

Another factor to consider is the computational complexity of the algorithms which is summarised in Table 2.

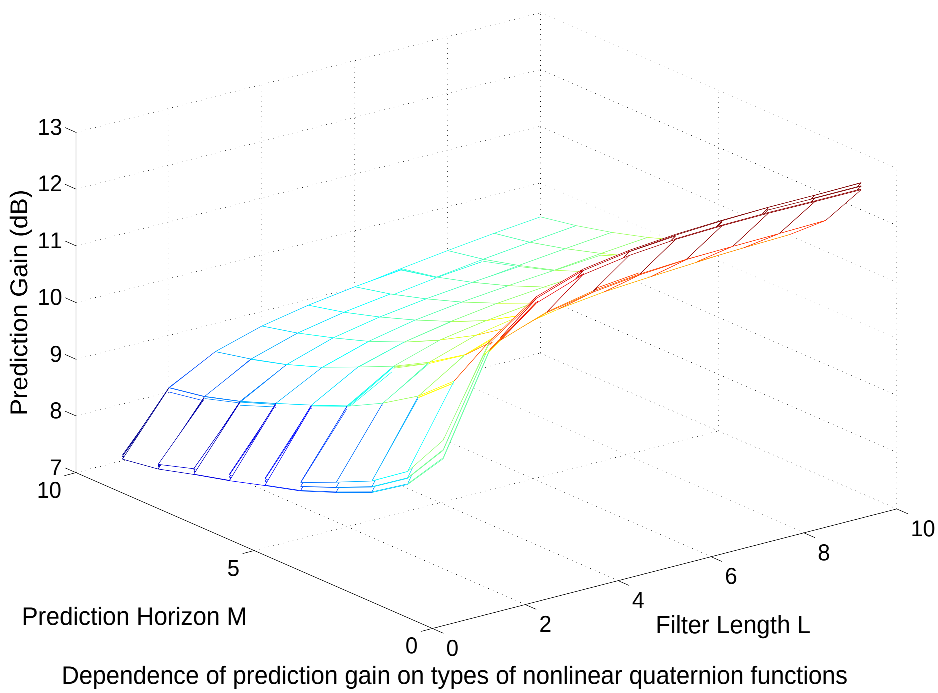


Figure 13: Prediction gains of QAFA for $\tan(q)$, $\sin(q)$, $\arctan(q)$, $\arcsin(q)$, $\sinh(q)$, $\mathrm{arctanh}(q)$ and $\mathrm{arcsinh}(q)$ for the prediction of 3D wind signal.

Table 2: Computational complexities of the algorithms considered

Algorithms	Multiplications	Additions
QFIR	36L+20	28L+15
AASQAFA	68L+36	54L+19
QAFA	68L+36	54L+24
WLQAFA	272L+144	208L+38

The computational complexity of the AASQAFA, QAFA is $\mathcal{O}(68L)$; the QFIR has the lowest computational complexity of $\mathcal{O}(36L)$ and the WLQAFA has the highest computational complexity of $\mathcal{O}(272L)$. The QAFA

algorithm thus represents an improvement from our previous proposed algorithm AASQAFA [9] in terms of performance and simplicity, while maintaining similar computational complexity.
In summary, the advantages of proposed class of QAFA and WLQAFA algorithms based on fully quaternion locally analytic nonlinearities, are

- The performances of algorithms based on fully quaternion locally analytic functions, QAFA and WLQAFA, were superior compared to those based on the split quaternion functions, AASQAFA and QFIR, as the fully quaternion nonlinearities (81) - (88) provide a direct manipulation of the quaternion signal, instead of the channelwise processing in $\mathbb{R}$;
- The widely linear model (16) enables the WLQAFA to fully capture the quaternion second order statistics, and hence offers a further performance enhancement over the standard linear model employed in QAFA, AASQAFA and QFIR;
- The fully quaternion based QAFA is a reasonable choice as it allows a trade off between performance and computational complexity.

Future works will include fusing the atmospheric parameters, such as wind temperature [42] [43].

9 Conclusion

A class of quaternion-valued nonlinear functions suitable for stochastic gradient based training of quaternion valued nonlinear adaptive filters has been proposed. The existing learning algorithms either neglect the non-commutativity aspect of quaternion, thus proving inadequate for the modelling of three and four-dimensional processes, or are unable to provide an accurate estimate due to the use of the suboptimal split-quaternion function. A class of fully quaternion activation functions has been derived according to the local analyticity condition which enables the extension of fully complex nonlinear activation functions to the quaternion domain. The proposed fully quaternion algorithms (QAFA and WLQAFA) have been shown to exhibit excellent performance on the prediction of four-dimensional synthetic and three-dimensional real-world vector signals. The WLQAFA has been shown to achieve enhanced performance due to the utilization of the quaternion widely linear model and the associated augmented quaternion statistics, which fully captures the second-order information within quaternion-valued signals and enable the processing of both second-order circular (proper) and improper processes. Simulations over a range of noncircular synthetic signals and real world three-dimensional wind recordings illustrate the benefit of the proposed approach.

.1 Analyticity of the exponential function e^q

To calculate the term $-\frac{\partial \mathrm{e}^q}{\partial \alpha}\hat{\zeta}$, we first need to evaluate the terms $\frac{\partial \mathrm{e}^q}{\partial q_b}$, $\frac{\partial \mathrm{e}^q}{\partial q_c}$ and $\frac{\partial \mathrm{e}^q}{\partial q_d}$. The term $\frac{\partial \mathrm{e}^q}{\partial q_b}$ is derived by differentiating (28) with respect to q_b to yield

$$\begin{aligned}
\frac{\partial \mathrm{e}^q}{\partial q_b} &= \mathrm{e}^{q_a}\frac{\partial}{\partial q_b}\Big(\cos(\alpha)+\frac{q_b\sin(\alpha)\imath}{\alpha}+\frac{q_c\sin(\alpha)\jmath}{\alpha}+\frac{q_d\sin(\alpha)\kappa}{\alpha}\Big)\\
&= \mathrm{e}^{q_a}\Big(-\sin(\alpha)\frac{\partial\alpha}{\partial q_b}+\frac{q_b}{\alpha}\cos(\alpha)\frac{\partial\alpha}{\partial q_b}\imath+(\frac{\partial}{\partial q_b}\frac{q_b}{\alpha})\sin(\alpha)\imath+\frac{q_c}{\alpha}\cos(\alpha)\frac{\partial\alpha}{\partial q_b}\jmath\\
&\quad+(\frac{\partial}{\partial q_b}\frac{q_c}{\alpha})\sin(\alpha)\jmath+\frac{q_d}{\alpha}\cos(\alpha)\frac{\partial\alpha}{\partial q_b}\kappa+(\frac{\partial}{\partial q_b}\frac{q_d}{\alpha})\sin(\alpha)\kappa\Big)\\
&= \mathrm{e}^{q_a}\Big(\frac{-q_b\sin(\alpha)}{\alpha}+\frac{q_b^2\cos(\alpha)\imath}{\alpha^2}+\frac{\big(q_c^2+q_d^2\big)\sin(\alpha)\imath}{\alpha^3}+\frac{q_bq_c\cos(\alpha)\jmath}{\alpha^2}\\
&\quad-\frac{q_bq_c\sin(\alpha)\jmath}{\alpha^3}+\frac{q_bq_d\cos(\alpha)\kappa}{\alpha^2}-\frac{q_bq_d\sin(\alpha)\kappa}{\alpha^3}\Big)
\end{aligned}\tag{89}$$

Proceeding in the same manner, the terms $\frac{\partial \mathrm{e}^q}{\partial q_c}$ and $\frac{\partial \mathrm{e}^q}{\partial q_d}$ are calculated as

$$\begin{aligned}
\frac{\partial \mathrm{e}^q}{\partial q_c} &= \mathrm{e}^{q_a}\Big(\frac{-q_c\sin(\alpha)}{\alpha}+\frac{q_bq_c\cos(\alpha)\imath}{\alpha^2}-\frac{q_bq_c\sin(\alpha)\imath}{\alpha^3}+\frac{q_c^2\cos(\alpha)\jmath}{\alpha^2}\\
&\quad+\frac{\big(q_b^2+q_d^2\big)\sin(\alpha)\jmath}{\alpha^3}+\frac{q_cq_d\cos(\alpha)\kappa}{\alpha^2}-\frac{q_cq_d\sin(\alpha)\kappa}{\alpha^3}\Big)
\end{aligned}\tag{90}$$

$$\begin{aligned}
\frac{\partial \mathrm{e}^q}{\partial q_d} &= \mathrm{e}^{q_a}\Big(\frac{-q_d\sin(\alpha)}{\alpha}+\frac{q_bq_d\cos(\alpha)\imath}{\alpha^2}-\frac{q_bq_d\sin(\alpha)\imath}{\alpha^3}+\frac{q_cq_d\cos(\alpha)\jmath}{\alpha^2}\\
&\quad-\frac{\big(q_cq_d\big)\sin(\alpha)\jmath}{\alpha^3}+\frac{q_d^2\cos(\alpha)\kappa}{\alpha^2}+\frac{\big(q_b^2+q_c^2\big)\sin(\alpha)\kappa}{\alpha^3}\Big)
\end{aligned}\tag{91}$$

Substituting the terms defined in (89), (90) and (91) into the analyticity condition specified in (30) results in

$$-\frac{\partial e^q}{\partial \alpha}\hat{\zeta} = e^{q_a}\Big(\frac{-\sin(\alpha)}{\alpha^2}\big(q_b^2+q_c^2+q_d^2\big)+\frac{q_b\cos(\alpha)}{\alpha^3}\imath+\frac{q_c\cos(\alpha)}{\alpha^3}\jmath+\frac{q_d\cos(\alpha)}{\alpha^3}\kappa\Big)\hat{\zeta} \tag{92}$$

We can simplify (92) by substituting the definition of $\hat{\zeta}$ in (22) and α in (23) to give

$$\begin{aligned}-\frac{\partial e^q}{\partial \alpha}\hat{\zeta} &= e^{q_a}\Big(-\sin(\alpha)+\frac{q_b\cos(\alpha)\imath}{\alpha}+\frac{q_c\cos(\alpha)\jmath}{\alpha}+\frac{q_d\cos(\alpha)\kappa}{\alpha}\Big)\Big(-\hat{\zeta}\Big)\\ &= e^{q_a}\Big(-\sin(\alpha)+\cos(\alpha)\hat{\zeta}\Big)\Big(-\hat{\zeta}\Big)\\ &= e^{q_a}\Big(\cos(\alpha)+\sin(\alpha)\hat{\zeta}\Big)\end{aligned} \tag{93}$$

.2 Euler form of $\tanh(q)$

The function $\tanh(q)$ in terms of the Euler formula is given by

$$\begin{aligned}\tanh(q) &= \frac{e^{2q_a}\cos(2\alpha)-1+e^{2q_a}\sin(2\alpha)\hat{\zeta}}{e^{2q_a}\cos(2\alpha)+1+e^{2q_a}\sin(2\alpha)\hat{\zeta}}\\ &= \frac{e^{4q_a}\big(\cos^2(2\alpha)+\sin^2(2\alpha)\big)+2e^{2q_a}\cos(2\alpha)\hat{\zeta}+1}{e^{4q_a}\big(\cos^2(2\alpha)+\sin^2(2\alpha)\big)+2e^{2q_a}\cos(2\alpha)-1}\\ &= \frac{e^{4q_a}-1+2e^{2q_a}\sin(2\alpha)}{e^{4q_a}+2e^{2q_a}\cos(2\alpha)+1}\end{aligned} \tag{94}$$

.3 Local Analyticity of $\tanh(q)$

To examine the local analyticity of $\tanh(q)$, we first apply the quaternion local analyticity condition in (21) to (94) to show that

$$\frac{\partial\tanh(q)}{\partial q_a} = -\Big(\frac{q_b}{\alpha}\frac{\partial\tanh(q)}{\partial q_b}+\frac{q_c}{\alpha}\frac{\partial\tanh(q)}{\partial q_c}+\frac{q_d}{\alpha}\frac{\partial\tanh(q)}{\partial q_d}\Big)\Big(\frac{q_b\imath+q_c\jmath+q_d\kappa}{\alpha}\Big) \tag{95}$$

Similarly to the case of quaternion exponential functions, we obtain the term $\frac{\partial\tanh(q)}{\partial q_a}$ by differentiating (94) with respect to q_a, to give

$$\begin{aligned}\frac{\partial\tanh(q)}{\partial q_a} &= \frac{\partial}{\partial q_a}\Big(\frac{e^{4q_a}-1}{e^{4q_a}+2e^{2q_a}\cos(2\alpha)+1}+\frac{2e^{2q_a}\sin(2\alpha)\hat{\zeta}}{e^{4q_a}+2e^{2q_a}\cos(2\alpha)+1}\Big)\\ &= \frac{4e^{6q_a}\cos(2\alpha)+8e^{4q_a}+4e^{2q_a}\cos(2\alpha)}{\big(e^{4q_a}+2e^{2q_a}\cos(2\alpha)+1\big)^2}+\frac{\big(4e^{2q_a}\sin(2\alpha)-4e^{6q_a}sin(2\alpha)\big)}{\big(e^{4q_a}+2e^{2q_a}\cos(2\alpha)+1\big)^2}\hat{\zeta}\end{aligned} \tag{96}$$

In order to determine the remaining terms in (95), define

$$u = 2e^{2q_a}\sin(2\alpha) \tag{97}$$

$$v = e^{4q_a}+2e^{2q_a}\cos(2\alpha)+1 \tag{98}$$

We can then substitute u and v into (94) and expand $\hat{\zeta}$ according to (22) to yield

$$\begin{aligned}\tanh(q) &= \frac{e^{4q_a}-1+u\hat{\zeta}}{v}\\ &= \frac{e^{4q_a}-1}{v}+\frac{uq_b\imath}{v\alpha}+\frac{uq_c\jmath}{v\alpha}+\frac{uq_d\kappa}{v\alpha}\end{aligned} \tag{99}$$

Proceeding in a manner similar to when determining the analyticity of e^q, the term $\frac{\partial \tanh(q)}{\partial q_b}$ is obtained by differentiating (95) with respect to q_b, resulting in

$$\begin{aligned}
\frac{\partial \tanh(q)}{\partial q_b} &= \frac{\partial}{\partial q_b}\left(\frac{\mathrm{e}^{4q_a}-1}{v} + \frac{uq_b\imath}{v\alpha} + \frac{uq_c\jmath}{v\alpha} + \frac{uq_d\kappa}{v\alpha}\right) \\
&= \frac{\left(\mathrm{e}^{4q_a}-1\right)\left(4\mathrm{e}^{2q_a}q_b\sin(2\alpha)\right)}{v^2} + \frac{(v\alpha)\left(\frac{\partial uq_b}{\partial q_b}\right) - (uq_b)\left(\frac{\partial v\alpha}{\partial q_b}\right)}{(v\alpha)^2}\imath + \frac{(v\alpha)\left(\frac{\partial uq_c}{\partial q_b}\right) - (uq_c)\left(\frac{\partial v\alpha}{\partial q_b}\right)}{(v\alpha)^2}\jmath \\
&\quad + \frac{(v\alpha)\left(\frac{\partial uq_d}{\partial q_b}\right) - (uq_d)\left(\frac{\partial v\alpha}{\partial q_b}\right)}{(v\alpha)^2}\kappa \\
&= \frac{\left(\mathrm{e}^{4q_a}-1\right)\left(4\mathrm{e}^{2q_a}q_b\sin(2\alpha)\right)}{v^2} + \left(\frac{v\alpha u + v4\mathrm{e}^{2q_a}q_b^2\cos(2\alpha) - \frac{uvq_b^2}{\alpha} + uq_b^2 4\mathrm{e}^{2q_a}\sin(2\alpha)}{(v\alpha)^2}\right)\imath \\
&\quad + \left(\frac{v4\mathrm{e}^{2q_a}q_bq_c\cos(2\alpha) - \frac{uvq_bq_c}{\alpha} + uq_bq_c4\mathrm{e}^{2q_a}\sin(2\alpha)}{(v\alpha)^2}\right)\jmath \\
&\quad + \left(\frac{v4\mathrm{e}^{2q_a}q_bq_d\cos(2\alpha) - \frac{uvq_bq_d}{\alpha} + uq_bq_d4\mathrm{e}^{2q_a}\sin(2\alpha)}{(v\alpha)^2}\right)\kappa
\end{aligned} \tag{100}$$

Noticing that u, v and α are functions of the variables q_b, q_c and q_d, the terms $\frac{\partial \tanh(q)}{\partial q_c}$ and $\frac{\partial \tanh(q)}{\partial q_d}$ become

$$\begin{aligned}
\frac{\partial \tanh(q)}{\partial q_c} &= \frac{\left(\mathrm{e}^{4q_a}-1\right)\left(4\mathrm{e}^{2q_a}q_c\sin(2\alpha)\right)}{v^2} + \left(\frac{v4\mathrm{e}^{2q_a}q_bq_c\cos(2\alpha) - \frac{uvq_bq_c}{\alpha} + uq_bq_c4\mathrm{e}^{2q_a}\sin(2\alpha)}{(v\alpha)^2}\right)\imath \\
&+ \left(\frac{v\alpha u + v4\mathrm{e}^{2q_a}q_c^2\cos(2\alpha) - \frac{uvq_c^2}{\alpha} + uq_c^2 4\mathrm{e}^{2q_a}\sin(2\alpha)}{(v\alpha)^2}\right)\jmath \\
&+ \left(\frac{v4\mathrm{e}^{2q_a}q_cq_d\cos(2\alpha) - \frac{uvq_cq_d}{\alpha} + uq_cq_d4\mathrm{e}^{2q_a}\sin(2\alpha)}{(v\alpha)^2}\right)\kappa
\end{aligned} \tag{101}$$

$$\begin{aligned}
\frac{\partial \tanh(q)}{\partial q_d} &= \frac{\left(\mathrm{e}^{4q_a}-1\right)\left(4\mathrm{e}^{2q_a}q_d\sin(2\alpha)\right)}{v^2} + \left(\frac{v4\mathrm{e}^{2q_a}q_bq_d\cos(2\alpha) - \frac{uvq_bq_d}{\alpha} + uq_bq_d4\mathrm{e}^{2q_a}\sin(2\alpha)}{(v\alpha)^2}\right)\imath \\
&+ \left(\frac{v4\mathrm{e}^{2q_a}q_cq_d\cos(2\alpha) - \frac{uvq_cq_d}{\alpha} + uq_cq_d4\mathrm{e}^{2q_a}\sin(2\alpha)}{(v\alpha)^2}\right)\jmath \\
&+ \left(\frac{v\alpha u + v4\mathrm{e}^{2q_a}q_d^2\cos(2\alpha) - \frac{uvq_d^2}{\alpha} + uq_d^2 4\mathrm{e}^{2q_a}\sin(2\alpha)}{(v\alpha)^2}\right)\kappa
\end{aligned} \tag{102}$$

Replacing (100), (101) and (102) to the right hand of side of (95) yields

$$\begin{aligned}
-\frac{\partial \tanh(q)}{\partial \alpha}\hat{\zeta} &= \left(\frac{\left(\mathrm{e}^{4q_a}-1\right)\left(4\mathrm{e}^{2q_a}\sin(2\alpha)\left(q_b^2+q_c^2+q_d^2\right)\right)}{(v\alpha)^2} + \frac{v4q_b\mathrm{e}^{2q_a}\cos(2\alpha) + u4q_b\mathrm{e}^{2q_a}\sin(2\alpha)}{v^2\alpha}\imath\right. \\
&\quad \left. + \frac{v4q_c\mathrm{e}^{2q_a}\cos(2\alpha) + u4q_c\mathrm{e}^{2q_a}\sin(2\alpha)}{v^2\alpha}\jmath + \frac{v4q_d\mathrm{e}^{2q_a}\cos(2\alpha) + u4q_d\mathrm{e}^{2q_a}\sin(2\alpha)}{v^2\alpha}\kappa\right)\left(-\hat{\zeta}\right)
\end{aligned} \tag{103}$$

Next, the terms u (97) and v (98) in (103) are expanded to give

$$\begin{aligned}
-\frac{\partial \tanh(q)}{\partial \alpha}\hat{\zeta} &= \left(\frac{\left(\mathrm{e}^{4q_a}-1\right)\left(4\mathrm{e}^{2q_a}\sin(2\alpha)\left(q_b^2+q_c^2+q_d^2\right)\right)}{\left(\left(\mathrm{e}^{4q_a}+2\mathrm{e}^{2q_a}\cos(2\alpha)+1\right)\alpha\right)^2}\right. \\
&\quad + \frac{4q_b\mathrm{e}^{6q_a}\cos(2\alpha) + 4q_b\mathrm{e}^{2q_a}\cos(2\alpha) + 8q_b\mathrm{e}^{4q_a}\left(\cos^2(2\alpha)+\sin^2(2\alpha)\right)}{\left(\mathrm{e}^{4q_a}+2\mathrm{e}^{2q_a}\cos(2\alpha)+1\right)^2\alpha}\imath \\
&\quad + \frac{4q_c\mathrm{e}^{6q_a}\cos(2\alpha) + 4q_c\mathrm{e}^{2q_a}\cos(2\alpha) + 8q_c\mathrm{e}^{4q_a}\left(\cos^2(2\alpha)+\sin^2(2\alpha)\right)}{\left(\mathrm{e}^{4q_a}+2\mathrm{e}^{2q_a}\cos(2\alpha)+1\right)^2\alpha}\jmath \\
&\quad \left. + \frac{4q_d\mathrm{e}^{6q_a}\cos(2\alpha) + 4q_d\mathrm{e}^{2q_a}\cos(2\alpha) + 8q_d\mathrm{e}^{4q_a}\left(\cos^2(2\alpha)+\sin^2(2\alpha)\right)}{\left(\mathrm{e}^{4q_a}+2\mathrm{e}^{2q_a}\cos(2\alpha)+1\right)^2\alpha}\kappa\right)\left(-\hat{\zeta}\right)
\end{aligned} \tag{104}$$

To simplify (104) further, we employ $\sin^2(\alpha)+\cos^2(\alpha)=1$ to give

$$-\frac{\partial\tanh(q)}{\partial\alpha}\hat{\zeta} = \Bigg(\frac{\big(4\mathrm{e}^{6q_a}\sin(2\alpha)-4\mathrm{e}^{2q_a}\sin(2\alpha)\big)}{\big(\mathrm{e}^{4q_a}+2\mathrm{e}^{2q_a}\cos(2\alpha)+1\big)^2} + \frac{4\mathrm{e}^{6q_a}\cos(2\alpha)+8\mathrm{e}^{4q_a}+4\mathrm{e}^{2q_a}\cos(2\alpha)}{\big(\mathrm{e}^{4q_a}+2\mathrm{e}^{2q_a}\cos(2\alpha)+1\big)^2}\big(\frac{q_b\imath+q_c\jmath+q_d\kappa}{\alpha}\big)\Bigg)\Big(-\hat{\zeta}\Big) \tag{105}$$

Further substituting $\hat{\zeta}$ in (22) and α in (23) into (105) gives

$$\begin{aligned}-\frac{\partial\tanh(q)}{\partial\alpha}\hat{\zeta} &= \Bigg(\frac{\big(4\mathrm{e}^{6q_a}\sin(2\alpha)-4\mathrm{e}^{2q_a}\sin(2\alpha)\big)}{\big(\mathrm{e}^{4q_a}+2\mathrm{e}^{2q_a}\cos(2\alpha)+1\big)^2}+\frac{4\mathrm{e}^{6q_a}\cos(2\alpha)+8\mathrm{e}^{4q_a}+4\mathrm{e}^{2q_a}\cos(2\alpha)}{\big(\mathrm{e}^{4q_a}+2\mathrm{e}^{2q_a}\cos(2\alpha)+1\big)^2}\hat{\zeta}\Bigg)\Big(-\hat{\zeta}\Big)\\ &= \frac{4\mathrm{e}^{6q_a}\cos(2\alpha)+8\mathrm{e}^{4q_a}+4\mathrm{e}^{2q_a}\cos(2\alpha)}{\big(\mathrm{e}^{4q_a}+2\mathrm{e}^{2q_a}\cos(2\alpha)+1\big)^2}+\frac{4\mathrm{e}^{2q_a}\sin(2\alpha)-4\mathrm{e}^{6q_a}\sin(2\alpha)}{\big(\mathrm{e}^{4q_a}+2\mathrm{e}^{2q_a}\cos(2\alpha)+1\big)^2}\hat{\zeta}\end{aligned} \tag{106}$$

.4 A Local Derivative of $\tanh(q)$

We shall first expand (37) into its Euler formula to give

$$\begin{aligned}\mathrm{sech}(q) &= \frac{2}{\mathrm{e}^{q_a}\big(\cos(\alpha)+\sin(\alpha)\hat{\zeta}\big)+\mathrm{e}^{-q_a}\big(\cos(\alpha)-\sin(\alpha)\hat{\zeta}\big)}\\ &= \frac{2\mathrm{e}^{q_a}\big(\cos(\alpha)-\sin(\alpha)\hat{\zeta}\big)+2\mathrm{e}^{-q_a}\big(\cos(\alpha)+\sin(\alpha)\hat{\zeta}\big)}{\mathrm{e}^{2q_a}+2\big(\cos^2(\alpha)-\sin^2(\alpha)\big)+\mathrm{e}^{-2q_a}}\\ &= \frac{2\mathrm{e}^{3q_a}\big(\cos(\alpha)-\sin(\alpha)\hat{\zeta}\big)+2\mathrm{e}^{q_a}\big(\cos(\alpha)+\sin(\alpha)\hat{\zeta}\big)}{\mathrm{e}^{4q_a}+2\mathrm{e}^{2q_a}\big(\cos^2(\alpha)-\sin^2(\alpha)\big)+1}\end{aligned} \tag{107}$$

and apply the identity $\cos^2(\alpha)-\sin^2(\alpha)=\cos(2\alpha)$ to give

$$\mathrm{sech}(q)=\frac{2\mathrm{e}^{3q_a}\big(\cos(\alpha)-\sin(\alpha)\hat{\zeta}\big)+2\mathrm{e}^{q_a}\big(\cos(\alpha)+\sin(\alpha)\hat{\zeta}\big)}{\mathrm{e}^{4q_a}+2\mathrm{e}^{2q_a}\cos(2\alpha)+1} \tag{108}$$

Upon squaring (108) results in

$$\begin{aligned}\mathrm{sech}^2(q) &= \frac{4\mathrm{e}^{6q_a}\big(\cos^2(\alpha)-\sin^2(\alpha)\big)+4\mathrm{e}^{4q_a}\big(2\cos^2(\alpha)+2\sin^2(\alpha)\big)+4\mathrm{e}^{2q_a}\big(\cos^2(\alpha)-\sin^2(\alpha)\big)}{\big(\mathrm{e}^{4q_a}+2\mathrm{e}^{2q_a}\cos(2\alpha)+1\big)^2}\\ &\quad+\frac{-8\mathrm{e}^{6q_a}\sin(\alpha)\cos(\alpha)+8\mathrm{e}^{2q_a}\sin(\alpha)\cos(\alpha)}{\big(\mathrm{e}^{4q_a}+2\mathrm{e}^{2q_a}\cos(2\alpha)+1\big)^2}\hat{\zeta}\end{aligned} \tag{109}$$

and substituting $2\sin(\alpha)\cos(\alpha)=\sin(2\alpha)$ yields

$$\mathrm{sech}^2(q) = \frac{4\mathrm{e}^{6q_a}\cos(2\alpha)+8\mathrm{e}^{4q_a}+4\mathrm{e}^{2q_a}\cos(2\alpha)}{\big(\mathrm{e}^{4q_a}+2\mathrm{e}^{2q_a}\cos(2\alpha)+1\big)^2}+\frac{-4\mathrm{e}^{6q_a}\sin(2\alpha)+4\mathrm{e}^{2q_a}\sin(2\alpha)}{\big(\mathrm{e}^{4q_a}+2\mathrm{e}^{2q_a}\cos(2\alpha)+1\big)^2}\hat{\zeta} \tag{110}$$

.5 Derivation of $\nabla_{\mathbf{w}}y^*(n)$ and $\nabla_{\mathbf{w}}y(n)$

The terms $\mathbf{w}^T(n)\mathbf{x}(n)$ and $\mathbf{x}^H(n)\mathbf{w}^*(n)$ are first expanded as (due to space limitation, the time index "n" has been dropped) :

$$\mathbf{w}^T(n)\mathbf{x}(n) = \begin{bmatrix}\mathbf{w}_a^T\mathbf{x}_a-\mathbf{w}_b^T\mathbf{x}_b-\mathbf{w}_c^T\mathbf{x}_c-\mathbf{w}_d^T\mathbf{x}_d\\ \mathbf{w}_a^T\mathbf{x}_b+\mathbf{w}_b^T\mathbf{x}_a+\mathbf{w}_c^T\mathbf{x}_d-\mathbf{w}_d^T\mathbf{x}_c\\ \mathbf{w}_a^T\mathbf{x}_c+\mathbf{w}_c^T\mathbf{x}_a+\mathbf{w}_d^T\mathbf{x}_b-\mathbf{w}_b^T\mathbf{x}_d\\ \mathbf{w}_a^T\mathbf{x}_d+\mathbf{w}_d^T\mathbf{x}_a+\mathbf{w}_b^T\mathbf{x}_c-\mathbf{w}_c^T\mathbf{x}_b\end{bmatrix} \tag{111}$$

$$\mathbf{x}^H(n)\mathbf{w}^*(n) = \begin{bmatrix}\mathbf{w}_a^T\mathbf{x}_a-\mathbf{w}_b^T\mathbf{x}_b-\mathbf{w}_c^T\mathbf{x}_c-\mathbf{w}_d^T\mathbf{x}_d\\ -\mathbf{w}_a^T\mathbf{x}_b-\mathbf{w}_b^T\mathbf{x}_a-\mathbf{w}_c^T\mathbf{x}_d+\mathbf{w}_d^T\mathbf{x}_c\\ -\mathbf{w}_a^T\mathbf{x}_c-\mathbf{w}_c^T\mathbf{x}_a-\mathbf{w}_d^T\mathbf{x}_b+\mathbf{w}_b^T\mathbf{x}_d\\ -\mathbf{w}_a^T\mathbf{x}_d-\mathbf{w}_d^T\mathbf{x}_a-\mathbf{w}_b^T\mathbf{x}_c+\mathbf{w}_c^T\mathbf{x}_b\end{bmatrix} \tag{112}$$

The gradients $\nabla_{\mathbf{w}}y^*(n)$ and $\nabla_{\mathbf{w}}y(n)$ are defined as

$$\nabla_{\mathbf{w}}y(n) = \nabla_{\mathbf{w}_a}y(n)+\nabla_{\mathbf{w}_b}y(n)\imath+\nabla_{\mathbf{w}_c}y(n)\jmath+\nabla_{\mathbf{w}_d}y(n)\kappa \tag{113}$$

$$\nabla_{\mathbf{w}}y^*(n) = \nabla_{\mathbf{w}_a}y^*(n)+\nabla_{\mathbf{w}_b}y^*(n)\imath+\nabla_{\mathbf{w}_c}y^*(n)\jmath+\nabla_{\mathbf{w}_d}y^*(n)\kappa \tag{114}$$

The odd-symmetry property also applies to the fully quaternion function and is given by

$$\Phi^{'*}\big(\mathbf{w}^T(n)\mathbf{x}(n)\big) = \Phi^{'}\big(\mathbf{x}^H(n)\mathbf{w}(n)\big) \tag{115}$$

The derivatives in (113) can be calculated from the expansions (111) while using (115), resulting in

$$\begin{aligned}
\nabla_{\mathbf{w}_a} y(n) &= \Phi^{'}\big(\mathbf{w}^T(n)\mathbf{x}(n)\big)(\mathbf{x}_a + \mathbf{x}_b\imath + \mathbf{x}_c\jmath + \mathbf{x}_d\kappa) \\
\nabla_{\mathbf{w}_b} y(n)\imath &= \Phi^{'}\big(\mathbf{w}^T(n)\mathbf{x}(n)\big)(-\mathbf{x}_b + \mathbf{x}_a\imath - \mathbf{x}_d\jmath + \mathbf{x}_c\kappa)\imath \\
&= \Phi^{'}\big(\mathbf{w}^T(n)\mathbf{x}(n)\big)(-\mathbf{x}_a - \mathbf{x}_b\imath + \mathbf{x}_c\jmath + \mathbf{x}_d\kappa) \\
\nabla_{\mathbf{w}_c} y(n)\jmath &= \Phi^{'}\big(\mathbf{w}^T(n)\mathbf{x}(n)\big)(-\mathbf{x}_c + \mathbf{x}_d\imath + \mathbf{x}_a\jmath - \mathbf{x}_b\kappa)\jmath \\
&= \Phi^{'}\big(\mathbf{w}^T(n)\mathbf{x}(n)\big)(-\mathbf{x}_a + \mathbf{x}_b\imath - \mathbf{x}_c\jmath + \mathbf{x}_d\kappa) \\
\nabla_{\mathbf{w}_d} y(n)\kappa &= \Phi^{'}\big(\mathbf{w}^T(n)\mathbf{x}(n)\big)(-\mathbf{x}_d - \mathbf{x}_c\imath + \mathbf{x}_b\jmath + \mathbf{x}_a\kappa)\kappa \\
&= \Phi^{'}\big(\mathbf{w}^T(n)\mathbf{x}(n)\big)(-\mathbf{x}_a + \mathbf{x}_b\imath + \mathbf{x}_c\jmath - \mathbf{x}_d\kappa)
\end{aligned} \tag{116}$$

where the symbols $\Phi^{'}(\cdot)$ denotes the derivative of the fully quaternion function. Similarly, the remaining derivatives in (114) are calculated from (112) to give

$$\begin{aligned}
\nabla_{\mathbf{w}_a} y^*(n) &= \Phi^{'*}\big(\mathbf{w}^T(n)\mathbf{x}(n)\big)(\mathbf{x}_a - \mathbf{x}_b\imath - \mathbf{x}_c\jmath - \mathbf{x}_d\kappa) \\
\nabla_{\mathbf{w}_b} y^*(n)\imath &= \Phi^{'*}\big(\mathbf{w}^T(n)\mathbf{x}(n)\big)(-\mathbf{x}_b - \mathbf{x}_a\imath + \mathbf{x}_d\jmath - \mathbf{x}_c\kappa)\imath \\
&= \Phi^{'*}\big(\mathbf{w}^T(n)\mathbf{x}(n)\big)(\mathbf{x}_a - \mathbf{x}_b\imath - \mathbf{x}_c\jmath - \mathbf{x}_d\kappa) \\
\nabla_{\mathbf{w}_c} y^*(n)\jmath &= \Phi^{'*}\big(\mathbf{w}^T(n)\mathbf{x}(n)\big)(-\mathbf{x}_c - \mathbf{x}_d\imath - \mathbf{x}_a\jmath + \mathbf{x}_b\kappa)\jmath \\
&= \Phi^{'*}\big(\mathbf{w}^T(n)\mathbf{x}(n)\big)(\mathbf{x}_a - \mathbf{x}_b\imath - \mathbf{x}_c\jmath - \mathbf{x}_d\kappa) \\
\nabla_{\mathbf{w}_d} y^*(n)\kappa &= \Phi^{'*}\big(\mathbf{w}^T(n)\mathbf{x}(n)\big)(-\mathbf{x}_d + \mathbf{x}_c\imath - \mathbf{x}_b\jmath - \mathbf{x}_a\kappa)\kappa \\
&= \Phi^{'*}\big(\mathbf{w}^T(n)\mathbf{x}(n)\big)(\mathbf{x}_a - \mathbf{x}_b\imath - \mathbf{x}_c\jmath - \mathbf{x}_d\kappa)
\end{aligned} \tag{117}$$

Finally, substituting (116) into (113) yields

$$\nabla_{\mathbf{w}} y(n) = -\Phi^{'}\big(\mathbf{w}^T(n)\mathbf{x}(n)\big)2\mathbf{x}^*(n) \tag{118}$$

and substituting (117) into (114) gives

$$\nabla_{\mathbf{w}} y^*(n) = \Phi^{'*}\big(\mathbf{w}^T(n)\mathbf{x}(n)\big)4\mathbf{x}^*(n) \tag{119}$$

which is applied in the derivation of QAFA and WLQAFA. Similar derivations hold for the weight vectors **h**, **u** and **v**.

References

[1] S. B. Choe and J. J. Faraway, "Modeling head and hand orientation during motion using quaternions," *Journal of Aerospace*, vol. 113, pp. 186–192, 2004.

[2] J. C. K. Chou, "Quaternions kinematic and dynamic differential equations," *IEEE Transactions on Robotics and Automation*, vol. 8, no. 1, pp. 53–64, 1992.

[3] D. Choukkroun, I. Y. Bar Itzhack, and Y. Ohsman, "Novel quaternion Kalman filter," *IEEE Transactions on Aerospace and Electronics Systems*, vol. 42, no. 1, pp. 174–190, 2006.

[4] C. Cheong Took and D. P. Mandic, "The quaternion LMS algorithm for adaptive filtering of hypercomplex processes," *IEEE Transactions on Signal Processing*, vol. 57, no. 4, pp. 1316–1327, 2009.

[5] A. Sudbery, "Quaternionic analysis," *Mathematical Proceedings of the Cambridge Philosophical Society*, vol. 85, no. 2, pp. 199–225, 1979.

[6] P. Arena, L. Fortuna, G. Muscato, and M. G. Xibilia, *Neural Networks in Multidimensional Domains*, Lecture Notes in Control and Information Sciences (Springer Verlag), Vol. 234, 1998.

[7] T. Nitta and H. de Garis, "A 3D vector version of the back-propagation algorithm," *In Proceedings of International Joint Conference on Neural Networks (IJCNN)*, vol. 2, pp. 511–516, 1992.

[8] T. Nitta, "A back-propagation algorithm for neural networks based on 3D vector product," *In Proceedings of International Joint Conference on Neural Networks (IJCNN)*, vol. 1, pp. 589–592, 1993.

[9] B. Che Ujang, C. Cheong Took, and D. P. Mandic, "Split quaternion nonlinear adaptive filtering," *Neural Networks*, vol. 23, no. 3, pp. 426–434, 2010.

[10] D. P. Mandic and V. S. L. Goh, *Complex valued nonlinear adaptive filters: noncircularity, widely linear and neural models*, Wiley, 2009.

[11] T. Kim and T. Adali, "Approximation by fully complex multilayer perceptrons," *Neural Computation*, vol. 15, no. 7, pp. 1641–1666, 2003.

[12] S. De Leo and P. Rotelli, "Quaternion analyticity," *Applied Mathematics Letters*, vol. 16, no. 7, pp. 1077–1081, 2003.

[13] A. Hirose, *Complex-valued neural networks:theories and applications*, World Scientific Publishing, 2003.

[14] A. Hirose and H. Onishi, "Proposal of relative-minimization learning for behavior stabilization of complex-valued recurrent neural networks," *Neurocomputing*, vol. 24, no. 1-3, pp. 163–171, 1999.

[15] I. Aizenberg and C. Moraka, "Multilayer feedforward neural network based on multi-valued neurons (MLMVN) and a backpropagation learning algorithm," *Soft Computing*, vol. 11, no. 2, pp. 169–183, 2007.

[16] I. Aizenberg, N. N. Aizenberg, and J. P. L. Vandewalle, *Multi-valued and universal binary neurons*, Springer-Verlag New York, 2000.

[17] D. P. Mandic, S. Javidi, S. L. Goh, A. Kuh, and K. Aihara, "Complex valued prediction of wind profile using augmented complex statistics," *Renewable Energy*, vol. 34, no. 1, pp. 196–210, 2009.

[18] S. L.Goh, M. Chen, D. H. Popovic, K. Aihara, D. Obradovic, and D. P. Mandic, "Complex valued forecasting of wind profile," *Renewable Energy*, vol. 31, no. 11, pp. 1733–1750, 2006.

[19] S. Javidi, M. Pedzisz, S. L. Goh, and D. P. Mandic, "The augmented complex least mean square algorithm with application to adaptive prediction problems," *IAPR Workshop on Cognitive Information Processing 2008*, pp. 54–57, 2008.

[20] S. L.Goh and D. P. Mandic, "An augmented extended Kalman filter algorithm for complex-valued recurrent neural networks," *Neural Computation*, vol. 19, no. 4, pp. 1039–1055, 2007.

[21] S. L.Goh and D. P. Mandic, "An augmented CRTRL for complex-valued recurrent neural networks," *Neural Computation*, vol. 20, no. 10, pp. 1061–1066, 2007.

[22] C. Cheong Took and D. P. Mandic, "Augmented second-order statistics of quaternion random process," *Signal Processing (2010)*, doi:10.1016/j.sigpro.2010.06.024.

[23] F. D. Neeser and J. L. Massey, "Proper complex random processes with applications to information theory," *IEEE Transactions on Information Theory*, vol. 39, no. 4, pp. 1293–1302, 1993.

[24] B. Picinbono, "On circularity," *IEEE Signal Processing Letters*, vol. 42, no. 12, pp. 3473–3482, 1994.

[25] N. N. Vakhania, "Random vectors with values in quaternion Hilbert spaces," *Theories of Probability and its Applications*, vol. 43, no. 1, pp. 99–115, 1999.

[26] P. O. Amblard and N. Le Bihan, "On properness of quaternion valued random variables," *In Proceedings of International Conference on Mathemathics (IMA) in Signal Processing*, pp. 23–26, 2004.

[27] J. Via, D. Ramirez, and I. Santamaria, "Properness and widely linear processing of quaternion random vectors," *IEEE Transactions on Information Theory*, vol. 56, no. 7, pp. 3502–3515, 2010.

[28] B. Picinbono and P. Chevalier, "Widely linear estimation with complex data," *IEEE Transactions on Signal Processing*, vol. 43, no. 8, pp. 2030–2033, 1995.

[29] C. Cheong Took and D. P. Mandic, "A quaternion widely linear adaptive filter," *IEEE Transactions on Signal Processing*, vol. 58, no. 8, pp. 4427–4431, 2010.

[30] E. Stiefel, "On Cauchy-Riemann equations in higher dimensions," *Journal of Research of the National Bureau of Standards*, vol. 48, no. 5, pp. 395–398, 1952.

[31] C. You and D. Hong, "Nonlinear blind equalization shemes using complex-valued multilayer feedforward neural networks," *IEEE Transactions on Neural Networks*, vol. 9, no. 6, pp. 1442–1455, 1998.

[32] F. F. Brackx, "The exponential function of a quaternion variable," *Applicable Analysis*, vol. 8, pp. 265–276, 1979.

[33] L. Shi, "Exploration in quaternion colour," M.S. thesis, Computer Science, Simon Fraser University, 2005.

[34] C. A. Deavours, "The quaternion calculus," *The American Mathematical Monthly*, vol. 80, no. 9, pp. 995–1008, 1973.

[35] E. Trentin, "Networks with trainable amplitude of activation functions," *Neural Networks*, vol. 14, no. 4-5, pp. 471–493, 2001.

[36] S. L. Goh and D. P. Mandic, "Recurrent neural networks with trainable amplitude of activation functions," *Neural Networks*, vol. 16, no. 8, pp. 1095–1100, 2003.

[37] D. P. Mandic and J. A. Chambers, *Recurrent Neural Networks for Prediction: Learning Algorithms, Architectures and Stability*, Wiley, 2001.

[38] E. Soria Olivas, J. Maravilla, J. F. Guerrero Martinez, M. Martinez Sober, and J. Espi Lopez, "An easy demonstration of the optimum value of the adaptation constant in the LMS algorithm," *IEEE Transactions on Education*, vol. 41, no. 1, pp. 81, 1998.

[39] K. Mitsubori and T. Saito, "Torus doubling and hyperchaos in a five dimensional hysteresis circuit," *In Proceedings of 1994 IEEE International Symposium on Circuit and Systems*, vol. 6, pp. 113–116, 1994.

[40] S. H. Strogartz, *Nonlinear dynamics and chaos: with applications to physics, biology, chemistry and engineering (studies in nonlinearity) 1st edition*, Boulder , CO: Westview Press, 2001.

[41] W. Duch and N. Jankowski, "Survey of neural transfer functions," *Neural Computing Survey*, vol. 2, pp. 163–212, 1999.

[42] D. P. Mandic, M. Golz, A. Kuh, D. Obradovic, and T. Tanaka, *Signal Processing Techniques for Knowledge Extraction and Information Fusion*, Springer, 2008.

[43] D. P. Mandic, D. Obradovic, A. Kuh, T. Adali, U. Trutschell, M. Golz, P. De Wilde, J. Barria, A. Constantinides, and J. Chamber, "Data fusion for modern engineering applications: an overview," *In Proceedings of International Conference on Artificial Neural Networks (ICANN)*, pp. 715–721, 2005.