2014 International Joint Conference on Neural Networks (IJCNN)
July 6-11, 2014, Beijing, China

The $\mathbb{HC}$ Calculus, Quaternion Derivatives and Caylay-Hamilton Form of Quaternion Adaptive Filters and Learning Systems

Yili Xia, Cyrus Jahanchahi, Dongpo Xu and Danilo P. Mandic

***Abstract*— We introduce a novel and unifying framework for the calculation of gradients of both quaternion holomorphic functions and nonholomorphic real functions of quaternion variables. This is achieved by considering the isomorphism between the quaternion domain $\mathbb{H}$ and the bivariate complex domain $\mathbb{C}\times\mathbb{C}$, and by exploiting complex calculus to simplify the quaternion gradient calculation. The validation of the proposed $\mathbb{HC}$ calculus is performed against the existing $\mathbb{HR}$ calculus, and its convenience is illustrated in the context of gradient-based quaternion optimisation as well as in adaptive learning systems. Quaternion adaptive filtering algorithms and a dynamical perceptron update are next derived based on the bivariate complex representation of quaternions and the $\mathbb{HC}$ calculus. Simulations on both synthetic and real-world multidimensional signals support the analysis.**

I. INTRODUCTION

QUATERNION signal processing is a rapidly growing area, as it is convenient to cast many 3D and 4D problems into the quaternion domain $\mathbb{H}$ to exploit the 'coupled' nature of information across the data channels. This includes modelling of rotations in computer graphics, tracking in aeronautics, and 3D color imaging [1], [2]. Recent mathematical tools to support these developments include the quaternion singular value decomposition, quaternion Fourier transform, quaternion independent component analysis, augmented statistics and Taylor series expansion [3]–[9].

However, gradient based optimisation in $\mathbb{H}$ is quite restrictive, as the standard Cauchy-Riemann-Fueter conditions [10] do not admit derivatives of nonholomorphic real-valued cost functions. It is only recently that the $\mathbb{HR}$ calculus has made possible the differentiation of both holomorphic and nonholomorphic functions in $\mathbb{H}$, through exploiting the duality with their isomorphic quadrivariate real functions [11]. Other existing quaternion gradients are also useful, but typically do not consider a quaternion as an entity, for instance, by treating the real part and the vector part of a quaternion separately [1], [12], [13].

Our aim here is to build upon the $\mathbb{HR}$ calculus, in order to introduce an intuitive and rigorous framework for calculating derivatives of both holomorphic and nonholomorphic functions of quaternion variables. To this end, we consider the Cayley-Dickson construction of a quaternion (as a bivariate complex vector [3], [5]), in order to benefit from a direct use of the well established complex domain algebra. In this way, we obtain a hypercomplex extension of the $\mathbb{CR}$ calculus [14], [15], thus inheriting its ability to provide gradients of both holomorphic and nonholomorphic functions, such as typical cost functions (real-valued error power) in signal processing. The proposed $\mathbb{HC}$ calculus is derived by considering the duality between the quaternion derivatives and the corresponding bivariate complex $\mathbb{CR}$ derivatives. Unlike the existing quaternion gradient calculations [1], [11]–[13], the $\mathbb{HC}$ calculus is able to decompose quaternion derivatives of both holomorphic and nonholomorphic functions of quaternion variables into a pair of complex-valued derivatives instead of four real-valued derivatives. It thus provides more compact and convenient expressions for quaternion derivatives, at a reduced computational complexity. The principle and the usefulness of the proposed $\mathbb{HC}$ calculus is illustrated in the context of quaternion gradient optimisation, such as in the identification of stationary points, direction of maximum change problems, as well as in adaptive filtering applications.

Y. Xia is with the School of Information Science and Engineering, Southeast University, 2 Sipailou, Nanjing 210096, PR China. (e-mail:yili.xia06@gmail.com).

C. Jahanchahi, Dongpo Xu and D. P. Mandic are with the Department of Electrical and Electronic Engineering, Imperial College London, SW7 2AZ London, UK. (e-mail: cyrus.jahanchahi05, dongpo.xu, d.mandic@imperial.ac.uk).

II. THE CAYLEY-DICKSON CONSTRUCTION

Quaternions are an associative but not commutative algebra over $\mathbb{R}$, defined as

$$\mathbb{H}: \ \{q = q_a + \imath q_b + \jmath q_c + \kappa q_d \mid q_a, q_b, q_c, q_d \in \mathbb{R}\} \qquad (1)$$

where $\imath$, $\jmath$, and κ are the imaginary units, for which $\imath\jmath = \kappa$, $\jmath\kappa = \imath$, $\kappa\imath = \jmath$, and $\imath^2 = \jmath^2 = \kappa^2 = \imath\jmath\kappa = -1$. It is convenient to view quaternions as a pair of complex numbers via the Cayley-Dickson construction. Letting $\mathbb{C}^2$ be a bivariate vector space over the complex numbers, any quaternion $q \in \mathbb{H}$ can be considered as a point $(x, y) \in \mathbb{C}^2$ [3], [16], whereby

$$q = (q_a + \imath q_b) + \jmath(q_c - \imath q_d) = x + \jmath y \qquad (2)$$

In this way, the quaternion conjugate q^* is given by

$$q^* = (q_a - \imath q_b) - \jmath(q_c - \imath q_d) = x^* - \jmath y \qquad (3)$$

while the Cayley-Dickson form of quaternion involutions [17] becomes

$$\begin{aligned} q^\imath &= -\imath q \imath = q_a + \imath q_b - \jmath q_c - \kappa q_d = x - \jmath y \\ q^\jmath &= -\jmath q \jmath = q_a - \imath q_b + \jmath q_c - \kappa q_d = x^* + \jmath y^* \\ q^\kappa &= -\kappa q \kappa = q_a - \imath q_b - \jmath q_c + \kappa q_d = x^* - \jmath y^* \end{aligned} \qquad (4)$$

Similarly, for the involution conjugates we have

$$q^{\imath *} = x^* + \jmath y, \ \ q^{\jmath *} = x - \jmath y^*, \ \ q^{\kappa *} = x + \jmath y^* \qquad (5)$$

The complex-valued quaternion components x, x^*, y, y^* in (2) and (3) can now be expressed based on quaternion involutions in (4) or conjugate involutions in (5) as

$$\begin{aligned} x &= \frac{q+q^{\imath}}{2}, \quad x^* = \frac{q^{\jmath}+q^{\kappa}}{2} \\ y &= \frac{-\jmath(q-q^{\imath})}{2}, \quad y^* = \frac{-\jmath(q^{\jmath}-q^{\kappa})}{2} \end{aligned} \tag{6}$$

$$\begin{aligned} x &= \frac{q^{\jmath*}+q^{\kappa*}}{2}, \quad x^*=\frac{q^*+q^{\imath*}}{2} \\ y &= \frac{\jmath(q^*-q^{\imath*})}{2}, \quad y^*=\frac{\jmath(q^{\jmath*}-q^{\kappa*})}{2} \end{aligned} \tag{7}$$

III. The $\mathbb{HC}$ Calculus

To establish the duality between the derivatives of a quaternion-valued function $f(q, q^{\imath}, q^{\jmath}, q^{\kappa}) \in \mathbb{H}$ and the corresponding 'composite' bivariate complex function $g = g(x, x^*, y, y^*) \in \mathbb{C}^2$, we employ (4) and (6) to write

$$\begin{aligned} f(q, q^{\imath}, q^{\jmath}, q^{\kappa}) &= u(x, x^*, y, y^*) + \jmath v(x, x^*, y, y^*) \\ &= g(x, x^*, y, y^*) \end{aligned} \tag{8}$$

where $u(x, x^*, y, y^*)$ and $v(x, x^*, y, y^*)$ are functions of complex argument, the bases of which are 1 and $\imath$. For the bivariate complex function $g(x, x^*, y, y^*) \in \mathbb{C}^2$ that is differentiable with respect to each of[1] $\{x, x^*, y, y^*\}$, the total differential

$$\begin{aligned} dg(x, x^*, y, y^*) &= \frac{\partial g}{\partial x}dx+\frac{\partial g}{\partial x^*}dx^*+\frac{\partial g}{\partial y}dy+\frac{\partial g}{\partial y^*}dy^* \\ &= \frac{\partial u}{\partial x}dx+\jmath\frac{\partial v}{\partial x}dx+\frac{\partial u}{\partial x^*}dx^*+\jmath\frac{\partial v}{\partial x^*}dx^* \\ &+ \frac{\partial u}{\partial y}dy+\jmath\frac{\partial v}{\partial y}dy+\frac{\partial u}{\partial y^*}dy^*+\jmath\frac{\partial v}{\partial y^*}dy^* \end{aligned} \tag{9}$$

Using (6) to link the complex and quaternion differentials

$$\begin{aligned} dx &= \frac{dq+dq^{\imath}}{2}, \quad dx^*=\frac{dq^{\jmath}+dq^{\kappa}}{2} \\ dy &= \frac{dq-dq^{\imath}}{2\jmath}, \quad dy^*=\frac{dq^{\jmath}-dq^{\kappa}}{2\jmath} \end{aligned} \tag{10}$$

allows for (9) to be written in a compact form

$$\begin{aligned} dg &= \frac{1}{2}\left(\frac{\partial g}{\partial x}-\frac{\partial g}{\partial y}\jmath\right)dq+\frac{1}{2}\left(\frac{\partial g}{\partial x}+\frac{\partial g}{\partial y}\jmath\right)dq^{\imath} \\ &+ \frac{1}{2}\left(\frac{\partial g}{\partial x^*}-\frac{\partial g}{\partial y^*}\jmath\right)dq^{\jmath}+\frac{1}{2}\left(\frac{\partial g}{\partial x^*}+\frac{\partial g}{\partial y^*}\jmath\right)dq^{\kappa} \end{aligned} \tag{11}$$

Note that, unlike in the complex domain, due to the non-commutative nature of the quaternion product, that is, $q_1 q_2 \neq q_2 q_1$, the position of the imaginary unit $\jmath$ in (11) cannot be swapped with those of the partial derivatives $\frac{\partial g}{\partial y}$ and $\frac{\partial g}{\partial y^*}$. With this in mind, the total differential of the quaternion function $f(q, q^{\imath}, q^{\jmath}, q^{\kappa})$ can be written as

$$df(q, q^{\imath}, q^{\jmath}, q^{\kappa}) = \frac{\partial f}{\partial q}dq+\frac{\partial f}{\partial q^{\imath}}dq^{\imath}+\frac{\partial f}{\partial q^{\jmath}}dq^{\jmath}+\frac{\partial f}{\partial q^{\kappa}}dq^{\kappa} \tag{12}$$

[1]This means that for a $g(x, x^*, y, y^*)$, when performing the partial derivative of g with respect to a specific variable from $\{x, x^*, y, y^*\}$, other variables are treated as constants. For instance, $\frac{\partial x^*}{\partial x} = 0$, a result already established in $\mathbb{C}$ [14], also it is obvious that from (2), we have $\frac{\partial y}{\partial x} = \frac{\partial y^*}{\partial x} = 0$.

A. The $\mathbb{HC}$-derivatives

Since the quaternion function f and its dual bivariate complex function g have the same derivative, upon comparing (12) and (11), we obtain the set of identities which we refer to as the $\mathbb{HC}$-derivatives, given by

$$\begin{bmatrix} \frac{\partial f(q,q^{\imath},q^{\jmath},q^{\kappa})}{\partial q} \\ \frac{\partial f(q,q^{\imath},q^{\jmath},q^{\kappa})}{\partial q^{\imath}} \\ \frac{\partial f(q,q^{\imath},q^{\jmath},q^{\kappa})}{\partial q^{\jmath}} \\ \frac{\partial f(q,q^{\imath},q^{\jmath},q^{\kappa})}{\partial q^{\kappa}} \end{bmatrix}^T = \frac{1}{2} \begin{bmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial x^*} \\ \frac{\partial f}{\partial y} \\ \frac{\partial f}{\partial y^*} \end{bmatrix}^T \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ -\jmath & \jmath & 0 & 0 \\ 0 & 0 & -\jmath & \jmath \end{bmatrix} \tag{13}$$

where $[\cdot]^T$ is the transpose operator and a shorthand notation $f = f(x,\ x^*,\ y,\ y^*)$ was used on the right hand side (RHS).

B. The $\mathbb{HC}^*$-derivatives

In analogy to the complex $\mathbb{CR}^*$-derivative, to arrive at the $\mathbb{HC}^*$-derivatives, we consider the representation of the complex components $\{x, x^*, y, y^*\}$ in the conjugate basis $\{q^*, q^{\imath*}, q^{\jmath*}, q^{\kappa*}\}$ in (7), to give

$$\begin{aligned} dx &= \frac{1}{2}(dq^{\jmath*} + dq^{\kappa*}),\ dx^* = \frac{1}{2}(dq^* + dq^{\imath*}) \\ dy &= \frac{\jmath}{2}(dq^* - dq^{\imath*}),\ dy^* = \frac{\jmath}{2}(dq^{\jmath*} - dq^{\kappa*}) \end{aligned} \tag{14}$$

Substituting (14) into (9), we arrive at

$$\begin{aligned} dg &= \frac{1}{2}\left(\frac{\partial g}{\partial x^*} + \frac{\partial g}{\partial y}\jmath\right)dq^*+\frac{1}{2}\left(\frac{\partial g}{\partial x^*} - \frac{\partial g}{\partial y}\jmath\right)dq^{\imath*} \\ &+ \frac{1}{2}\left(\frac{\partial g}{\partial x} + \frac{\partial g}{\partial y^*}\jmath\right)dq^{\jmath*}+\frac{1}{2}\left(\frac{\partial g}{\partial x} - \frac{\partial g}{\partial y^*}\jmath\right)dq^{\kappa*} \end{aligned} \tag{15}$$

Using the conjugate bases $\{q^*,\ q^{\imath*},\ q^{\jmath*},\ q^{\kappa*}\}$ to express function $f = f(q^*,\ q^{\imath*},\ q^{\jmath*},\ q^{\kappa*})$, gives the differential

$$df = \frac{\partial f}{\partial q^*}dq^* + \frac{\partial f}{\partial q^{\imath*}}dq^{\imath*} + \frac{\partial f}{\partial q^{\jmath*}}dq^{\jmath*} + \frac{\partial f}{\partial q^{\kappa*}}dq^{\kappa*} \tag{16}$$

By comparing (15) and (16), we obtain the set of $\mathbb{HC}^*$-derivatives in a vector-matrix form as

$$\begin{bmatrix} \frac{\partial f(q,q^{\imath},q^{\jmath},q^{\kappa})}{\partial q^*} \\ \frac{\partial f(q,q^{\imath},q^{\jmath},q^{\kappa})}{\partial q^{\imath*}} \\ \frac{\partial f(q,q^{\imath},q^{\jmath},q^{\kappa})}{\partial q^{\jmath*}} \\ \frac{\partial f(q,q^{\imath},q^{\jmath},q^{\kappa})}{\partial q^{\kappa*}} \end{bmatrix}^T = \frac{1}{2} \begin{bmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial x^*} \\ \frac{\partial f}{\partial y} \\ \frac{\partial f}{\partial y^*} \end{bmatrix}^T \begin{bmatrix} 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ \jmath & -\jmath & 0 & 0 \\ 0 & 0 & \jmath & -\jmath \end{bmatrix} \tag{17}$$

C. Analogy with the $\mathbb{CR}$ calculus

Out of the eight $\mathbb{HC}$ and $\mathbb{HC}^*$ derivatives in (13) and (17), of particular interest are $\partial f/\partial q$ and $\partial f/\partial q^*$, given by

$$\frac{\partial f}{\partial q} = \frac{1}{2}\left(\frac{\partial f}{\partial x} - \frac{\partial f}{\partial y}\jmath\right) \quad \mathbb{HC}-\text{derivative} \tag{18}$$

$$\frac{\partial f}{\partial q^*} = \frac{1}{2}\left(\frac{\partial f}{\partial x^*} + \frac{\partial f}{\partial y}\jmath\right) \quad \mathbb{HC}^*-\text{derivative} \tag{19}$$

These can be regarded as a generic extension of the complex $\mathbb{CR}$ calculus [14], [15] into $\mathbb{H}$ via the Cayley-Dickson construction.

For a complex function $f = f(z, z^*)$ which is differentiable with respect to z and z^* independently, where $z = z_x + \imath z_y$, the $\mathbb{CR}$ calculus gives

$$\frac{\partial f}{\partial z} = \frac{1}{2}\left(\frac{\partial f}{\partial z_x} - \frac{\partial f}{\partial z_y}\imath\right) \quad \mathbb{CR} - \text{derivative} \tag{20}$$

$$\frac{\partial f}{\partial z^*} = \frac{1}{2}\left(\frac{\partial f}{\partial z_x} + \frac{\partial f}{\partial z_y}\imath\right) \quad \mathbb{CR}^* - \text{derivative} \tag{21}$$

where on the RHS, we used a shorthand notation $f = f(z_x, z_y)$, while for the real component z_x we have $z_x = z_x^*$. The main difference lies in the placement of imaginary unit vectors; in the complex domain both $\frac{\partial f}{\partial z_y}\imath$ and $\imath\frac{\partial f}{\partial z_y}$ are valid, whereas in $\mathbb{H}$, due to the noncommutative nature of the quaternion product, $\frac{\partial f}{\partial y}\jmath \neq \jmath\frac{\partial f}{\partial y}$.

D. The duality between the $\mathbb{HC}$ and $\mathbb{HR}$ calculus

The partial derivatives with respect to $\{x, x^*, y, y^*\}$ on RHS of (13) and (17) can also be expanded using the $\mathbb{CR}$ calculus as,

$$\begin{aligned}
\frac{\partial f(x,x^*)}{\partial x} &= \frac{1}{2}\left(\frac{\partial f}{\partial q_a} - \frac{\partial f}{\partial q_b}\imath\right)\\
\frac{\partial f(x,x^*)}{\partial x^*} &= \frac{1}{2}\left(\frac{\partial f}{\partial q_a} + \frac{\partial f}{\partial q_b}\imath\right)\\
\frac{\partial f(y,y^*)}{\partial y} &= \frac{1}{2}\left(\frac{\partial f}{\partial q_c} + \frac{\partial f}{\partial q_d}\imath\right)\\
\frac{\partial f(y,y^*)}{\partial y^*} &= \frac{1}{2}\left(\frac{\partial f}{\partial q_c} - \frac{\partial f}{\partial q_d}\imath\right)
\end{aligned}$$

to give the set of derivatives

$$\begin{bmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial x^*} \\ \frac{\partial f}{\partial y} \\ \frac{\partial f}{\partial y^*} \end{bmatrix}^T = \frac{1}{2} \begin{bmatrix} \frac{\partial f}{\partial q_a} \\ \frac{\partial f}{\partial q_b} \\ \frac{\partial f}{\partial q_c} \\ \frac{\partial f}{\partial q_d} \end{bmatrix}^T \begin{bmatrix} 1 & 1 & 0 & 0 \\ -\imath & \imath & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & \imath & -\imath \end{bmatrix} \tag{22}$$

Substituting (22) into (13) and (17), we arrive at the $\mathbb{HR}$ and $\mathbb{HR}^*$ derivatives, given by [11]

$$\begin{bmatrix} \frac{\partial f}{\partial q} \\ \frac{\partial f}{\partial q^\imath} \\ \frac{\partial f}{\partial q^\jmath} \\ \frac{\partial f}{\partial q^\kappa} \end{bmatrix}^T = \frac{1}{4} \begin{bmatrix} \frac{\partial f}{\partial q_a} \\ \frac{\partial f}{\partial q_b} \\ \frac{\partial f}{\partial q_c} \\ \frac{\partial f}{\partial q_d} \end{bmatrix}^T \begin{bmatrix} 1 & 1 & 1 & 1 \\ -\imath & -\imath & \imath & \imath \\ -\jmath & \jmath & -\jmath & \jmath \\ -\kappa & \kappa & \kappa & -\kappa \end{bmatrix} \tag{23}$$

$$\begin{bmatrix} \frac{\partial f}{\partial q^*} \\ \frac{\partial f}{\partial q^{\imath *}} \\ \frac{\partial f}{\partial q^{\jmath *}} \\ \frac{\partial f}{\partial q^{\kappa *}} \end{bmatrix}^T = \frac{1}{4} \begin{bmatrix} \frac{\partial f}{\partial q_a} \\ \frac{\partial f}{\partial q_b} \\ \frac{\partial f}{\partial q_c} \\ \frac{\partial f}{\partial q_d} \end{bmatrix}^T \begin{bmatrix} 1 & 1 & 1 & 1 \\ \imath & \imath & -\imath & -\imath \\ \jmath & -\jmath & \jmath & -\jmath \\ \kappa & -\kappa & -\kappa & \kappa \end{bmatrix} \tag{24}$$

Remark 1: Both the $\mathbb{HC}$ and $\mathbb{HR}$ calculus give identical results for a quaternion derivative, however, via the Cayley-Dickson construction of quaternion variables, the $\mathbb{HC}$ calculus given in (13) and (17) provides a more compact and convenient expression, via two complex-valued derivatives, as compared with four real-valued derivatives within the $\mathbb{HR}$ calculus.

Remark 2: The computational complexity of the proposed $\mathbb{HC}$ calculus is lower than that of the $\mathbb{HR}$ calculus, due to the compact representation of quaternion variables via the Cayley-Dickson construction. For instance, to calculate $\frac{\partial f}{\partial q}$, the $\mathbb{HC}$ derivative requires 12 real multiplications and 10 real additions, whereas the $\mathbb{HR}$ derivative requires 28 real multiplications and 28 real additions.

E. Some useful derivatives using the $\mathbb{HC}$ calculus

1) *Derivative of the holomorphic function* $f(q) = q$. Using (2) and the $\mathbb{HC}$-derivative in (18), we have $\frac{\partial f}{\partial q} = \frac{1}{2}(\frac{\partial q}{\partial x} - \frac{\partial q}{\partial y}\jmath) = \frac{1}{2}(1 - \jmath\jmath) = 1$. This is equivalent to the standard Cauchy-Riemann-Fueter (CRF) derivative [10], which gives $f'(q) = 1$.
2) *Derivative of the nonholomorphic function* $f(q, q^*) = |q|^2 = qq^*$. Keeping in mind that in the complex domain $\mathbb{C}$ $\frac{\partial x^*}{\partial x} = \frac{\partial y}{\partial x} = \frac{\partial x^*}{\partial y} = 0$ and using the product rule, the $\mathbb{HC}$-derivative in (18) and the $\mathbb{HC}^*$-derivative in (19), we have

$$\begin{aligned}
\frac{\partial f}{\partial q} &= \frac{\partial q}{\partial q}q^* + q\frac{\partial q^*}{\partial q}\\
&= q^* + \frac{q}{2}\left(\frac{\partial q^*}{\partial x} - \frac{\partial q^*}{\partial y}\jmath\right)\\
&= q^* + \frac{q}{2}(0 - (-\jmath)\jmath)\\
&= q^* - \frac{q}{2}
\end{aligned} \tag{25}$$

Note that the main difference between the derivatives of a nonholomorphic function in $\mathbb{H}$ and $\mathbb{C}$ lies in the second term on the RHS of (25). Unlike the derivative in $\mathbb{C}$, where $\partial z^*/\partial z = 0$, its quaternion counterpart $\partial q^*/\partial q$ does not vanish, but is equal to $-\frac{1}{2}$. This is because within the Cayley-Dickson construction of q and q^*, their 'real' parts are no longer identical, as shown in (2) and (4), however, their 'imaginary' parts are complex conjugates one another.
3) *Derivatives with respect to quaternion involutions*. Consider the derivative $\frac{\partial q}{\partial q^\imath}$. Since $q = x + \jmath y$ and $q^\imath = x - \jmath y$, using the $\mathbb{HC}$-derivatives in (13), we have $\frac{\partial q}{\partial q^\imath} = \frac{1}{2}(\frac{\partial q}{\partial x} + \frac{\partial q^\imath}{\partial y}\jmath) = 1 + \jmath\jmath = 0$. Next, consider the derivative $\frac{\partial q}{\partial q^{\imath *}}$. Using the $\mathbb{HC}^*$-derivatives in (17), we have $\frac{\partial q}{\partial q^{\imath *}} = \frac{1}{2}(\frac{\partial q}{\partial x^*} - \frac{\partial q}{\partial y}\jmath) = \frac{1}{2}(0 - \jmath\jmath) = \frac{1}{2}$. In summary, the $\mathbb{HC}$ calculus gives

$$\frac{\partial q}{\partial q^\eta} = 0, \ \frac{\partial q^*}{\partial q^\eta} = \frac{1}{2}, \ \forall \eta \in \{\imath, \jmath, \kappa\} \tag{26}$$

IV. Applications of Quaternion Gradient

We next illustrate the usefulness of the $\mathbb{HC}$ calculus in gradient type optimisation of scalar functions of quaternion vectors, widely employed as cost functions in MSE estimation in learning systems.

A. Stationary points of the quaternion gradient

Let $f = f(\mathbf{q}, \mathbf{q}^\imath, \mathbf{q}^\jmath, \mathbf{q}^\kappa) = f(\mathbf{q}^*, \mathbf{q}^{\imath *}, \mathbf{q}^{\jmath *}, \mathbf{q}^{\kappa *})$ be real-valued and $\mathbf{q} = [q_1, \ldots, q_N]^T$. Applying the $\mathbb{HC}$-gradient in (19) component-wise, we have

$$\nabla_{\mathbf{q}} f = \mathbf{0} \Leftrightarrow \nabla_{\mathbf{x}} f = \nabla_{\mathbf{y}} f = \mathbf{0} \tag{27}$$

where $\nabla_{\mathbf{q}} = [\frac{\partial}{\partial q_1}, \ldots, \frac{\partial}{\partial q_N}]^T$ is the quaternion gradient operator with respect to $\mathbf{q}$ and the result on the RHS follows from component-wise equating the 'real' and 'imaginary' parts of $\frac{\partial f}{\partial \mathbf{q}}$ to zero via the $\mathbb{HC}$-derivative in (18). Since f is real, we also have $\nabla_{\mathbf{x}^*} f = (\nabla_{\mathbf{x}} f)^* = \mathbf{0}$ and $\nabla_{\mathbf{y}^*} f = (\nabla_{\mathbf{y}} f)^* = \mathbf{0}$, and

$$\nabla_{\mathbf{x}^*} f = \nabla_{\mathbf{y}^*} f = \nabla_{\mathbf{x}} f = \nabla_{\mathbf{y}} f = \mathbf{0} \Leftrightarrow \nabla_{\alpha} f = \mathbf{0} \tag{28}$$

where $\alpha \in \{q^{\imath}, q^{\jmath}, q^{\kappa}, q^*, q^{\imath *}, q^{\jmath *}, q^{\kappa *}\}$. The result on the RHS can be obtained by applying either the $\mathbb{HC}$-derivatives in (13) or the $\mathbb{HC}^*$-derivatives in (17) component-wise. Therefore, unlike the complex case where the stationary points of a real function of complex variables $f(\mathbf{z}, \mathbf{z}^*)$ are defined by both $\nabla_{\mathbf{z}} f = \mathbf{0}$ and $\nabla_{\mathbf{z}^*} f = \mathbf{0}$ [14], [15], in $\mathbb{H}$ this requirement is $\nabla_{\beta} f = \mathbf{0}$, where $\beta \in \{q, q^{\imath}, q^{\jmath}, q^{\kappa}, q^*, q^{\imath *}, q^{\jmath *}, q^{\kappa *}\}$. Comparing with the $\mathbb{HR}$ calculus in [11], this gives a more comprehensive solution by considering both the quaternion involutions and their conjugates.

B. Direction of maximum change of the quaternion gradient

For real-valued scalar function $f = f(\mathbf{q}, \mathbf{q}^{\imath}, \mathbf{q}^{\jmath}, \mathbf{q}^{\kappa}) = f(\mathbf{x}, \mathbf{x}^*, \mathbf{y}, \mathbf{y}^*)$, the first order Taylor series expansion

$$df = (\nabla_{\mathbf{x}} f)^T d\mathbf{x} + (\nabla_{\mathbf{x}^*} f)^T d\mathbf{x}^* + (\nabla_{\mathbf{y}} f)^T d\mathbf{y} + (\nabla_{\mathbf{y}^*} f)^T d\mathbf{y}^*$$

Since f is real, we have $(\nabla_{\mathbf{x}} f)^T d\mathbf{x} = ((\nabla_{\mathbf{x}^*} f)^T d\mathbf{x}^*)^*$ and $(\nabla_{\mathbf{y}} f)^T d\mathbf{y} = ((\nabla_{\mathbf{y}^*} f)^T d\mathbf{y}^*)^*$ [15], the above equation can be simplified as

$$df = 2\Re\left((\nabla_{\mathbf{x}} f)^T d\mathbf{x} + (\nabla_{\mathbf{y}} f)^T d\mathbf{y}\right) \tag{29}$$

where $\Re(\cdot)$ is the real part operator. Upon applying the $\mathbb{HC}$-derivative in (18) and (1) component-wise, we can expand $(\nabla_{\mathbf{q}} f)^T d\mathbf{q}$ as

$$\begin{aligned}(\nabla_{\mathbf{q}} f)^T d\mathbf{q} &= \frac{1}{2}\left((\nabla_{\mathbf{x}} f)^T - (\nabla_{\mathbf{y}} f)^T \jmath\right)(d\mathbf{x} + \jmath d\mathbf{y}) \\ &= \frac{1}{2}\left((\nabla_{\mathbf{x}} f)^T d\mathbf{x} + (\nabla_{\mathbf{y}} f)^T d\mathbf{y}\right. \\ &\quad \left.+ (\nabla_{\mathbf{x}} f)^T \jmath d\mathbf{y} - (\nabla_{\mathbf{x}} f)^T \jmath d\mathbf{y}\right)\end{aligned} \tag{30}$$

Keeping in mind that f is real-valued, this results in complex-valued functions $\{\nabla_{\mathbf{x}} f,\ \nabla_{\mathbf{y}} f,\ d\mathbf{x},\ d\mathbf{y}\}$ for which the imaginary unit is $\imath$. However, the existence of $\jmath$ in the third and fourth terms on the RHS of (30) means that $\Re((\nabla_{\mathbf{x}} f)^T \jmath d\mathbf{y}) = \Re((\nabla_{\mathbf{y}} f)^T \jmath d\mathbf{x}) = 0$, and hence

$$\Re\left((\nabla_{\mathbf{q}} f)^T d\mathbf{q}\right) = \frac{1}{2}\Re\left((\nabla_{\mathbf{x}} f)^T d\mathbf{x} + (\nabla_{\mathbf{y}} f)^T d\mathbf{y}\right) \tag{31}$$

Comparing with (29) and using the fact that for real functions of quaternion variables $\frac{\partial f}{\partial q^*} = (\frac{\partial f}{\partial q})^*$, we have

$$df = 4\Re\left((\nabla_{\mathbf{q}} f)^T d\mathbf{q}\right) = 4\Re\left((\nabla_{\mathbf{q}^*} f)^H d\mathbf{q}\right) = 4\left\langle \nabla_{\mathbf{q}^*} f, d\mathbf{q}\right\rangle$$

The inner product satisfies $\langle \nabla_{\mathbf{q}^*} f, d\mathbf{q}\rangle \leq \|\nabla_{\mathbf{q}^*} f\| \|d\mathbf{q}\|$ (Schwarz inequality), whereby the equality stands only when $\nabla_{\mathbf{q}^*} f$ is collinear with $\|d\mathbf{q}\|$, indicating that the maximum change of the derivative df occurs when $d\mathbf{q}$ is in the direction of the conjugate gradient $\nabla_{\mathbf{q}^*} f$. In practice, this conjugate gradient can be obtained by using the $\mathbb{HC}^*$-derivatives in (17) as $\nabla_{\mathbf{q}^*} f = \nabla_{\mathbf{x}^*} f + \nabla_{\mathbf{y}} f \jmath$.

V. The use of $\mathbb{HC}$ calculus in adaptive filtering applications

In gradient based optimization problems in adaptive filtering, the task is to minimise the real cost function $J(e, e^*) = E[|e|^2] = E[ee^*]$, where $E[\cdot]$ is the statistical expectation operator, and $e = d - y$ is the error between the desired signal d and the filter output $y = \mathbf{w}^T \mathbf{q}$, where $\mathbf{w} = [w_1, \ldots, w_N]^T$ forms filter coefficients, $\mathbf{q}$ defines the input vector and N is the filter length.

Quaternion Wiener filter. To find the closed form of the optimum weights $\mathbf{w}_{\text{o}}$ in an MSE minimisation problem, we expand the cost function as $J(e, e^*) = E[ee^*] = E[dd^*] + E[yy^*] - E[yd^*] - E[dy^*]$. From Section IV-B the conjugate gradient $\nabla_{\mathbf{w}^*} J$ defines the direction of the maximum change, while since $\frac{\partial \mathbf{w}}{\partial \mathbf{w}^*}$ does not vanish but equals to $-\frac{1}{2}\mathbf{I}$ (see Example 2 in Section III-E for more detail), we have

$$\begin{aligned}\nabla_{\mathbf{w}^*} J &= \nabla_{\mathbf{w}^*}(\mathbf{w}^T \mathbf{R}_{qq} \mathbf{w}^*) - \nabla_{\mathbf{w}^*}(\mathbf{w}^T \mathbf{r}_{qd}) - \nabla_{\mathbf{w}^*}(\mathbf{r}_{dq}^T \mathbf{w}^*) \\ &= \mathbf{w}^T \mathbf{R}_{qq} - \frac{1}{2}(\mathbf{R}_{qq}\mathbf{w}^*)^T + \frac{1}{2}\mathbf{r}_{qd}^T - \mathbf{r}_{dq}^T\end{aligned} \tag{32}$$

where $\mathbf{R}_{qq} = E[\mathbf{q}\mathbf{q}^H]$, $\mathbf{r}_{qd} = E[\mathbf{q}d^*]$ and $\mathbf{r}_{dq} = E[d\mathbf{q}^*]$. From $(\mathbf{w}^T \mathbf{R}_{qq})^* = (\mathbf{R}_{qq}^H \mathbf{w}^*)^T = (\mathbf{R}_{qq}\mathbf{w}^*)^T$ and $\mathbf{r}_{qd}^T = (\mathbf{r}_{dq}^T)^*$ and setting (32) to zero yields the Wiener filtering solution

$$\mathbf{w}_{\text{o}}^T \mathbf{R}_{qq} = \mathbf{r}_{dq}^T \Rightarrow \mathbf{w}_{\text{o}}^T = \mathbf{r}_{dq}^T \mathbf{R}_{qq}^{-1} \tag{33}$$

To achieve an online implementation of the quaternion Wiener filtering solution, at time instant k, the expectation operation $E[\cdot]$ is practically replaced by the weighted summation operation, resulting in

$$\begin{aligned}J(k) &= \sum_{n=0}^{k} \lambda^{k-n} e(n) e^*(n) \\ \mathbf{r}_{qd}(k) &= \sum_{n=0}^{k} \lambda^{k-n} \mathbf{q}(n) d^*(n) \\ \mathbf{r}_{dq}(k) &= \sum_{n=0}^{k} \lambda^{k-n} d(n) \mathbf{q}^*(n) \\ &= \lambda \mathbf{r}_{dq}(k-1) + d(k)\mathbf{q}^*(k)\end{aligned} \tag{34}$$

$$\begin{aligned}\mathbf{R}_{qq}(k) &= \sum_{n=0}^{k} \lambda^{k-n} \mathbf{q}(n) \mathbf{q}^H(n) \\ &= \lambda \mathbf{R}_{qq}(k-1) + \mathbf{q}(k)\mathbf{q}^H(k)\end{aligned} \tag{35}$$

where the forgetting factor $\lambda \in$ (0, 1] and the covariance matrix inverse $\mathbf{R}_{qq}^{-1}$ is recursively estimated by considering (35) and using Woodbury's identity as,

$$\begin{aligned}\mathbf{R}_{qq}^{-1}(k) &= \lambda^{-1}\mathbf{R}_{qq}^{-1}(k-1) \\ &\quad - \frac{\lambda^{-2}\mathbf{R}_{qq}^{-1}(k-1)\mathbf{q}(k)\mathbf{q}^H(k)\mathbf{R}_{qq}^{-1}(k-1)}{1 + \lambda^{-1}\mathbf{q}^H(k)\mathbf{R}_{qq}^{-1}(k-1)\mathbf{q}(k)}\end{aligned} \tag{36}$$

hence,

$$\mathbf{w}_{\text{o}}^T(k) = \mathbf{r}_{dq}^T(k)\mathbf{R}_{qq}^{-1}(k) \tag{37}$$

Quaternion recursive least squares (QRLS). The QRLS aims to recursively solve the Wiener filtering problem in (37).

To comply with the standard literature [18] , from (36), we redefine

$$\mathbf{P}(k) = \mathbf{R}_{qq}^{-1}(k) = \lambda^{-1}\mathbf{P}(k-1) - \lambda^{-2}\frac{\mathbf{p}\mathbf{p}^H}{c}$$
$$\mathbf{p} = \mathbf{P}(k-1)\mathbf{q}(k)$$
$$c = 1 + \lambda^{-1}\mathbf{q}^H(k)\mathbf{P}(k-1)\mathbf{q}(k) \tag{38}$$

By substituting (34) and (38) into (37), we obtain a recursive update for the weight vector $\mathbf{w}_o(k)$ as [21]

$$\begin{aligned}
\mathbf{w}_o^T(k) &= \mathbf{r}_{dq}^T(k)\mathbf{P}(k)\\
&= (\lambda\mathbf{r}_{dq}^T(k-1) + d(k)\mathbf{q}^H(k))(\lambda^{-1}\mathbf{P}(k-1) - \lambda^{-2}\frac{\mathbf{p}\mathbf{p}^H}{c})\\
&= \mathbf{r}_{dq}^T(k-1)\mathbf{P}(k-1) + \lambda^{-1}d(k)\mathbf{q}^H(k)\mathbf{P}(k-1)\\
&\quad - \lambda^{-1}\mathbf{r}_{dq}^T(k-1)\frac{\mathbf{p}\mathbf{p}^H}{c} - \lambda^{-2}d(k)\mathbf{q}^H(k)\frac{\mathbf{p}\mathbf{p}^H}{c}\\
&= \mathbf{w}_o^T(k-1) + \lambda^{-1}(d(k)c - \mathbf{w}_o^T(k-1)\mathbf{q}(k)\\
&\quad - \lambda^{-1}d(k)\mathbf{q}^H(k)\mathbf{P}(k-1)\mathbf{q}(k))\frac{\mathbf{p}^H}{c}\\
&= \mathbf{w}_o^T(k-1) + \lambda^{-1}\frac{e(k)\mathbf{p}^H}{c}
\end{aligned} \tag{39}$$

where $e(k) = d(k) - \mathbf{w}_o^T(k-1)\mathbf{q}(k)$ is the *a priori* error.

Quaternion least mean square (QLMS) and its variants. Based on the instantaneous cost function $J(k) = e(k)e^*(k)$, where $e(k) = d(k) - \mathbf{w}^T(k)\mathbf{q}(k)$, and using the $\mathbb{HC}$-derivative and the $\mathbb{HC}^*$-derivative in (18) and (19), the weight update of the QLMS becomes [7]

$$\begin{aligned}
\mathbf{w}(k+1) &= \mathbf{w}(k) - \mu\nabla_{\mathbf{w}^*}J(k)\\
&= \mathbf{w}(k) - \mu\big(e(k)\big(\nabla_{\mathbf{w}^*}e^*(k)\big) + \big(\nabla_{\mathbf{w}^*}e(k)\big)e^*(k)\big)\\
&= \mathbf{w}(k) + \mu\big(e(k)\mathbf{q}^*(k) - \frac{1}{2}\mathbf{q}(k)e^*(k)\big)
\end{aligned} \tag{40}$$

where μ is the step-size and $\nabla_{\mathbf{w}^*}J(k)$ is analysed in Section IV-B which shows that the direction of the maximum change of the quaternion gradient is in the direction of the conjugate gradient. Using the $\mathbb{HC}$ calculus in (18) and (19) to arrive at $\nabla_{\mathbf{w}^*}e^*(k) = -\mathbf{q}^*(k)$ and $\nabla_{\mathbf{w}^*}e(k) = \frac{1}{2}\mathbf{q}(k)$, gives the QLMS update in its original form [7] with a scalar constant absorbed in μ.

Another QLMS form in [19] considers the involution gradient (I-gradient) given by

$$\nabla_{\mathbf{w}^\eta}J(k) = \frac{\partial J(k)}{\partial \mathbf{w}^\imath} + \frac{\partial J(k)}{\partial \mathbf{w}^\jmath} + \frac{\partial J(k)}{\partial \mathbf{w}^\kappa},\ \forall \eta \in \{\imath, \jmath, \kappa\} \tag{41}$$

so that we arrive at

$$\begin{aligned}
\mathbf{w}(k+1) &= \mathbf{w}(k) - \mu\nabla_{\mathbf{w}^\eta}J(k)\\
&= \mathbf{w}(k) - \mu\big(\left(\nabla_{\mathbf{w}^\eta}e(k)\right)e^*(k) + e(k)\left(\nabla_{\mathbf{w}^\eta}e^*(k)\right)\big)
\end{aligned}$$

Using the $\mathbb{HC}^*$-derivative and the $\mathbb{HC}$-derivative in (19) and (18) to yield $\nabla_{\mathbf{w}^\eta}e(k) = \mathbf{0}$ and $\nabla_{\mathbf{w}^\eta}e^*(k) = -\frac{3}{2}e(k)\mathbf{q}^*(k)$ respectively (see Example 3 in Section III-E for more detail), gives

$$\mathbf{w}(k+1) = \mathbf{w}(k) + \frac{3}{2}\mu e(k)\mathbf{q}^*(k) \tag{42}$$

which is in the same generic form as that of the complex LMS [14], apart from the factor of $\frac{3}{2}$ which simply scales μ.

Quaternion affine projection algorithm (QAPA). The quaternion least mean square (QLMS) and its variants also highlighted the need for faster converging practical algorithms. The stochastic gradient based normalised QLMS (NQLMS) can solve this issue only partially whereas the fast converging quaternion recursive least squares (QRLS) is computationally demanding. To that end, the quaternion affine projection algorithm (QAPA), based on the affine subspace projections, has been introduced for quaternion-valued adaptive filtering [20]. Structurally, the QAPA spans the range between the NQLMS and QRLS both in terms of performance and computational requirements. In practical terms, whereas the NQLMS updates the weight vector based only on the current input vector $\mathbf{q}(k)$, the QAPA employs K past input vectors to form the $N \times K$ data matrix

$$\mathbf{Q}(k) = [\mathbf{q}(k-K+1), \ldots, \mathbf{q}(k)] \tag{43}$$

The aim of QAPA is to minimise adaptively the squared Euclidean norm of the change in the weight vector $\mathbf{w}(k)$, that is

$$\begin{aligned}
&\text{minimise} && \|\triangle\mathbf{w}(k+1)\|^2 = \|\mathbf{w}(k+1) - \mathbf{w}(k)\|^2\\
&\text{subject to} && \mathbf{d}^T(k) = \mathbf{w}^T(k+1)\mathbf{Q}(k)
\end{aligned} \tag{44}$$

where $\mathbf{d}(k) = [d(k-K+1), \ldots, d(k)]^T$ denotes the $K \times 1$ desired signal vector. Using the Lagrange multipliers, the above constrained optimisation problem can be solved by considering the following cost function

$$\begin{aligned}
J(k) &= \|\mathbf{w}(k+1) - \mathbf{w}(k)\|^2 + \Re[(\mathbf{d}^T(k)\\
&\quad - \mathbf{w}^T(k+1)\mathbf{Q}(k))\boldsymbol{\lambda}^*]
\end{aligned} \tag{45}$$

where the symbol $\Re[\cdot]$ denotes the real part of a quaternion variable and the $K \times 1$ dimensional vector $\boldsymbol{\lambda}$ comprises Lagrange multipliers. Using the proposed $\mathbb{HC}^*$-derivative in (19), we have

$$\begin{aligned}
\frac{\partial J(k)}{\partial \mathbf{w}^*(k+1)} &= \mathbf{w}(k+1) - \mathbf{w}(k) - \frac{1}{2}(\mathbf{w}^*(k+1) - \mathbf{w}^*(k))\\
&\quad - \frac{1}{2}(\mathbf{Q}^*(k)\boldsymbol{\Lambda} - \frac{1}{2}\mathbf{Q}(k)\boldsymbol{\Lambda}^*)
\end{aligned} \tag{46}$$

Setting (46) to zero, the weight update of QAPA can be obtained as

$$\mathbf{w}(k+1) - \mathbf{w}(k) = \frac{1}{2}\mathbf{Q}^*(k)\boldsymbol{\Lambda} \tag{47}$$

Using the fact that $\mathbf{e}^T(k) = \mathbf{d}^T(k) - \mathbf{y}^T(k) = (\mathbf{w}^T(k+1) - \mathbf{w}^T(k))\mathbf{Q}(k)$, and based on (47), $\boldsymbol{\Lambda}$ can be solved as

$$\boldsymbol{\Lambda} = 2(\mathbf{Q}^T(k)\mathbf{Q}^*(k))^{-1}\mathbf{e}(k) \tag{48}$$

which gives

$$\mathbf{w}(k+1) = \mathbf{w}(k) + \mathbf{Q}^*(k)(\mathbf{Q}^T(k)\mathbf{Q}^*(k))^{-1}\mathbf{e}(k) \tag{49}$$

Note that to prevent the normalisation matrix $\mathbf{Q}^T(k)\mathbf{Q}^*(k)$ within (49) from becoming singular, a small regularisation term $\varepsilon\mathbf{I} \in \mathbb{H}^{K\times K}$ is typically added to the identity matrix, whereas a step size μ is also incorporated to control the convergence and the steady state performance, giving the final weight update of QAPA in the form

$$\mathbf{w}(k+1)=\mathbf{w}(k)+\mu\mathbf{Q}^*(k)\left(\mathbf{Q}^T(k)\mathbf{Q}^*(k)+\varepsilon\mathbf{I}\right)^{-1}\mathbf{e}(k) \quad (50)$$

Observe that when the observation length of the data matrix $\mathbf{Q}(k)$ is $K = 1$, QAPA degenerates into the normalised QLMS (NQLMS).

Widely linear QLMS (WLQLMS). The WLQLMS is based on the widely linear estimation model $y(k) = \mathbf{w}^{aT}(k)\mathbf{q}^a(k)$ to deal with the generality of quaternion signals (both proper and improper) [7], [22], where $\mathbf{q}^a(k) = [\mathbf{q}(k), \mathbf{q}^{\imath}(k), \mathbf{q}^{\jmath}(k), \mathbf{q}^{\kappa}(k)]^T$ is the augmented input vector and $\mathbf{w}^a(k) = [\mathbf{u}(k), \mathbf{v}(k), \mathbf{g}(k), \mathbf{h}(k)]^T$ is the associated weight vector. Using the gradient defined as $\nabla_{\mathbf{w}^{a*}} J(k)$ and the $\mathbb{HC}^*$-derivative in (19), we arrive at

$$\mathbf{w}^a(k+1) = \mathbf{w}^a(k) + \mu\big(e^a(k)\mathbf{q}^{a*}(k) - \frac{1}{2}\mathbf{q}^a(k)e^{a*}(k)\big) \quad (51)$$

where $e^a(k) = d(k) - \mathbf{w}^{aT}(k)\mathbf{q}^a(k)$, and (51) is the exact WLQLMS update given in [7].

Quaternion nonlinear gradient descent (QNGD). We now consider the gradient based optimisation problem in the context of nonlinear adaptive filtering and neural networks, where a fully quaternion nonlinear activation function is employed to give the system output $y(k)$ as [24]

$$y(k) = \Phi(\mathbf{w}^T(k)\mathbf{q}(k)) \quad (52)$$

where $\Phi(\cdot)$ is a fully quaternion nonlinearity such as the $\tanh(\cdot)$. Based on the instantaneous cost function $J(k) = e(k)e^*(k)$, where $e(k) = d(k) - y(k)$, the weight update of the QNGD becomes

$$\begin{aligned}\mathbf{w}(k+1) &= \mathbf{w}(k) - \mu\nabla_{\mathbf{w}} J(k)\\ &= \mathbf{w}(k) - \mu\big(\left(\nabla_{\mathbf{w}} e(k)\right)e^*(k) + e(k)\left(\nabla_{\mathbf{w}} e^*(k)\right)\big)\end{aligned}$$

Using the $\mathbb{HC}$-gradient in (18), we have

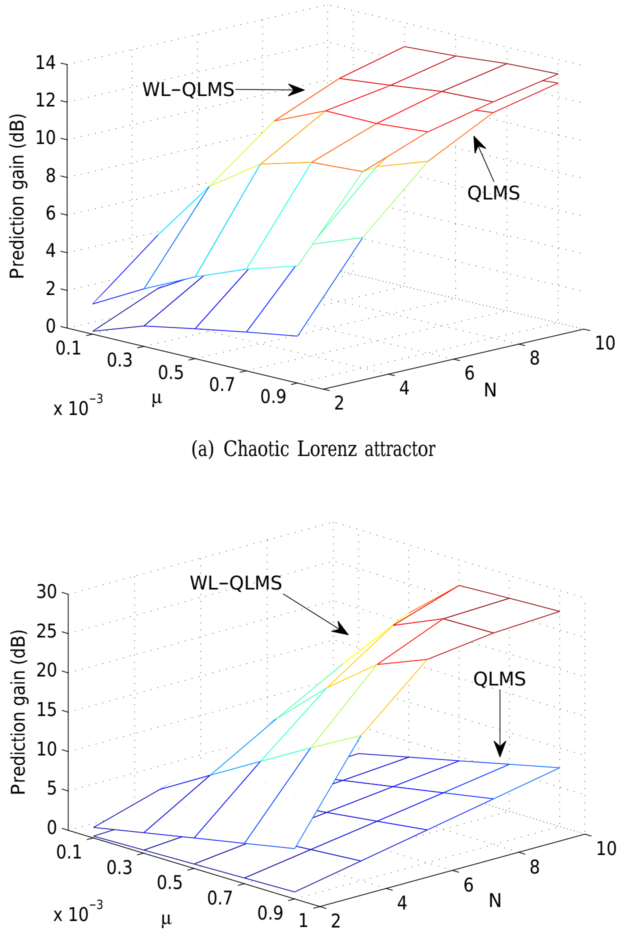
$$\begin{aligned}\nabla_{\mathbf{w}} e^*(k) &= -\frac{\partial\Phi(\mathbf{q}^H(k)\mathbf{w}^*(k))}{\partial\mathbf{w}^*(k)}\\ &= -\Phi'(\mathbf{q}^H(k)\mathbf{w}^*(k))\mathbf{q}^*(k) \end{aligned} \quad (53)$$

while the use of $\mathbb{HC}^*$-gradient in (19) gives

$$\begin{aligned}\nabla_{\mathbf{w}} e(k) &= -\frac{\partial\Phi(\mathbf{w}^T(k)\mathbf{q}(k))}{\partial\mathbf{w}^*(k)}\\ &= \frac{1}{2}\Phi'(\mathbf{w}^T(k)\mathbf{q}(k))\mathbf{q}(k) \end{aligned} \quad (54)$$

so that the weight update of QNGD has the form [24]

$$\begin{aligned}\mathbf{w}(k+1) = \;&\mathbf{w}(k) + \mu e(k)\Phi'(\mathbf{q}^H(k)\mathbf{w}^*(k))\mathbf{q}^*(k)\\ &- \frac{\mu}{2}\Phi'(\mathbf{w}^T(k)\mathbf{q}(k))\mathbf{q}(k)e^*(k) \end{aligned} \quad (55)$$



(a) Chaotic Lorenz attractor

(b) 3D Tai Chi body motion

Fig. 1. Performance of WL-QLMS and QLMS derived based on both the proposed $\mathbb{HC}$ calculus and the $\mathbb{HR}$ calculus on the one-step ahead prediction of the signals considered. (a) Chaotic Lorenz attractor, and (b) 3D Tai Chi body motion.

VI. Simulations

To validate the quaternion adaptive filtering algorithms derived based on the proposed $\mathbb{HC}$ calculus, we compared the performances of the QLMS and WL-QLMS algorithms in a one-step ahead prediction setting. The performance was assessed using the prediction gain $\mathrm{R_p}$, defined as

$$\mathrm{R_p} = 10\log_{10}\frac{\hat{\sigma}_q^2}{\hat{\sigma}_e^2} \quad (56)$$

where $\hat{\sigma}_q^2$ and $\hat{\sigma}_e^2$ denote the estimated variances of the input and the prediction error respectively. The employed test signals which were constructed as full quaternions via the Cayley-Dickson construction as described in Section II include:

- The chaotic Lorenz signal, governed by coupled partial differential equations

$$\frac{\partial x}{\partial t} = \alpha(y-x)$$
$$\frac{\partial y}{\partial t} = x(\rho - z) - y$$
$$\frac{\partial z}{\partial t} = xy - \beta z$$

where $\alpha = 10$, $\rho = 28$ and $\beta = 8/3$ [23].

- A real-world 3D noncircular and nonstationary body motion signal [24]. Two 3D gyroscopes were placed on the left arm and the right arm of an athlete performing Tai Chi movements, and 3D motion data were recorded using the XSense MTx 3DOF orientation tracker. The movement of the left arm was used as a pure quaternion input with real part set to be zero for simulations.

Fig. 1 illustrates the performances of WL-QLMS and standard QLMS derived based on the proposed $\mathbb{HC}$ calculus and the $\mathbb{HR}$ calculus for the prediction of the chaotic Lorenz attractor and 3D Tai Chi body motion over a range of the filter parameters. The advantage of WL-QLMS over standard QLMS resulted from the use of widely linear estimation model in order to incorporate the full second order statistics within the considered signals [7]. As expected, the performances of the $\mathbb{HC}$ calculus based quaternion adaptive filtering algorithms were identical to those based on the $\mathbb{HR}$ calculus, however, the advantage of the proposed $\mathbb{HC}$ calculus over the $\mathbb{HR}$ calculus lies in the simplified quaternion gradient calculation, and the reduced computational cost, as discussed in **Remark 2** in Section III.

VII. Conclusion

A unifying framework, referred to as the $\mathbb{HC}$ calculus, has been proposed for the calculation of gradients of both quaternion holomorphic functions and nonholomorphic real functions of quaternion variables. This has been achieved by making use of the isomorphism between quaternion involutions and bivariate complex vectors via the Cayley-Dickson construction of quaternion variables. Unlike the $\mathbb{HR}$ calculus [11], which considers quaternion derivatives as four individual real components, the proposed $\mathbb{HC}$ calculus directly exploits complex algebra, thus providing more compact and convenient expressions for quaternion derivatives. The usefulness of the $\mathbb{HC}$ calculus has been illustrated in quaternion gradient optimisation, for constrained and unconstrained gradient descent problems in adaptive signal processing and in learning systems. Simulations on synthetic and real-world multidimensional signals support the analysis.

Acknowledgment

This work was supported by the Natural Science Foundation of China under Grant Nos. 61201173, 61271058, the Doctoral Fund of Ministry of Education of China under Grant No. 20110092110008, and the Natural Science Foundation of Jiangsu Province under Grant No. BK2011060.

References

[1] A. M. Sabatini, "Quaternion-based extended Kalman filter for determining orientation and magnetic sensing," *IEEE Trans. Biomed. Eng.*, vol. 53, no. 7, pp. 1346-1356, 2006.

[2] S.-C. Pei and C.-M. Cheng, "Color image processing by using binary quaternion-moment-preserving thresholding technique," *IEEE Trans. Image Process.*, vol. 8, no. 5, pp. 614-628, 1999.

[3] N. Le Bihan and J. Mars, "Singular value decomposition of quaternion matrices: A new tool for vector-sensor signal processing," *Signal Process.*, vol. 84, no. 7, pp. 1177-1199, 2004.

[4] T. A. Ell and S. J. Sangwine, "Hypercomplex Fourier transforms of color images," *IEEE Trans. Image Process.*, vol. 16, no. 1, pp. 22-35, 2007.

[5] S. Said and N. Le Bihan and S. J. Sangwine, "Fast complexified quaternion Fourier transform," *IEEE Trans. Signal Process.*, vol. 56, no. 4, pp. 1522-1531, 2008.

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

[7] C. Cheong-Took and D. P. Mandic, "A quaternion widely linear adaptive filter," *IEEE Trans. Signal Process.*, vol. 58, no. 8, pp. 4427-4431, 2010.

[8] J. Via, D. P. Palomar, L. Vielva and I. Santamaria, "Quaternion ICA from second-order statistics," *IEEE Trans. Signal Process.*, vol. 59, no. 4, pp. 1586-1600, 2011.

[9] J. Navarro-Moreno, R. M. Fernandez-Alcala and J. C. Ruiz-Molina, "A quaternion widely linear series expansion and its applications," *IEEE Signal Process. Lett.*, vol. 19, no. 12, pp. 868-871, 2012.

[10] I. Frenkel and M. Libine, "Quaternionic analysis, representation theory, and physics," *Adv. Math.*, vol. 218, no. 6, pp. 1806-1877, 2008.

[11] D. P. Mandic, C. Jahanchahi and C. Cheong-Took, "A quaternion gradient operator and its applications," *IEEE Signal Process. Lett.*, vol. 18, no. 1, pp. 47-50, 2011.

[12] C. A. Deavours, "The quaternion calculus," *Amer. Math. Month.*, vol. 80, pp. 995-1008, 1973.

[13] S. De Leo and P. Rotelli, "Quaternion analyticity," *Appl. Math. Lett.*, vol. 16, no. 7, pp. 1077-1081, 2003.

[14] D. Brandwood, "A complex gradient operator and its application in adaptive array theory," *Proc. Inst. Elect. Eng. F, Commun., Radar, Signal Process.*, vol. 130, no. 1, pp. 11-16, 1983.

[15] K. Kreutz-Delgado, "The complex gradient operator and the $\mathbb{CR}$-calculus," *Lecture Supplement ECE275A, University of California, San Diego*, 2006.

[16] S. Miron, N. Le Bihan and J. I. Mars, "Quaternion-MUSIC for vector-sensor array," *IEEE Trans. Signal Process.*, vol. 54, no. 4, pp. 1218-1229, 2006.

[17] T. A. Ell and S. J. Sangwine, "Quaternion involutions and anti-involutions," *Comput. Math. Applicat.*, vol. 53, no. 1, pp. 137-143, 2007.

[18] A. H. Sayed, *Fundamentals of Adaptive Filtering.* New York: Wiley, 2003.

[19] C. Jahanchahi, D. P. Mandic and C. Cheong-Took, "On gradient calculation in quaternion adaptive filtering," *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, pp. 3773-3776, 2012.

[20] C. Jahanchahi, C. Cheong-Took and D. P. Mandic, "A class of quaternion valued affine projection algorithms," *Signal Process.*, vol. 93, no. 7, pp. 1712-1723, 2013.

[21] C. Jahanchahi, C. Cheong-Took and D. P. Mandic, "The widely linear quaternion recursive least squares filter," *2nd International Workshop on Cognitive Information Processing*, pp. 87-92, 2010.

[22] J. Via, D. Ramirez and I. Santamaria, "Properness and widely linear processing of quaternion random vectors," *IEEE Trans. Inf. Theory*, vol. 56, no. 7, pp. 3502-3515, 2010.

[23] E. N. Lorenz, "Deterministic nonperiodic flow," *J. Atmos. Sci*, vol. 20, no. 2, pp. 130-141, 1963.

[24] B. C. Ujang, C. Cheong-Took and D. P. Mandic, "Quaternion-valued nonlinear adaptive filtering," *IEEE Trans. Neural Netw.*, vol. 22, no. 8, pp. 1193-1206, 2011.