The Theory of Quaternion Matrix Derivatives

Dongpo Xu and Danilo P. Mandic, *Fellow, IEEE*

***Abstract*—A systematic framework for the calculation of the derivatives of quaternion matrix functions with respect to quaternion matrix variables is introduced. The proposed approach is equipped with the matrix product and chain rules and applies to both analytic and nonanalytic functions of quaternion variables. This rectifies a mathematical shortcut in the existing methods, which incorrectly use the traditional product rule. We also show that within the proposed framework, the derivatives of quaternion matrix functions can be calculated directly, without using quaternion differentials or resorting to the isomorphism with real vectors. Illustrative examples show how the proposed quaternion matrix derivatives can be used as an important tool for solving optimization problems in signal processing applications.**

***Index Terms*—GHR calculus, Jacobian, non-analytic functions, quaternion differentials, quaternion matrix derivatives.**

I. INTRODUCTION

Quaternion signal processing has recently attracted considerable research interest in areas including image processing [1]–[3], computer graphics [4], aerospace and satellite tracking [5], [6], modeling of wind profile [7]–[9], processing of polarized waves [10]–[12], and design of space-time block codes [13]–[16]. Recent mathematical tools to support these developments include the quaternion singular value decomposition [10], quaternion Fourier transform [17], [18], augmented quaternion statistics [19]–[21] and Taylor series expansion [22]. However, gradient based optimisation techniques in quaternion algebra have experienced slow progress, as the quaternion analyticity conditions are rather stringent. For example, the generalised Cauchy-Riemann condition [23] restricts the class of quaternion analytic functions to linear functions and constants. One attempt to relax this constraint is the so-called Cauchy-Riemann-Fueter (CRF) condition [24], however, even the polynomial functions do not satisfy the CRF condition. The slice regular condition was proposed in [25], [26], to enable the derivatives of polynomials and power series with one-sided quaternion coefficients, however, the product and composition of two slice regular functions are generally not slice regular.

In quaternion statistical signal processing, a common optimization objective is to minimize a real cost function of quaternion variables, typically in the form of error power, $f(q) = |e(q)|^2$, however, such a function is obviously non-analytic according to quaternion analysis [24], [27], [28] and therefore a direct use of quaternion derivatives is not possible. To circumvent this problem, the so called pseudo-derivatives are often employed, which treat f as a real analytic function of the four real components of quaternion variable, and then take the real derivatives with respect to these independent real parts, separately. However, this approach makes the computations cumbersome and tedious, even for very simple algorithms. An alternative and more elegant approach that can deal with non-analytic functions directly in the quaternion domain is the HR calculus [29], which takes the derivatives of f with respect to a quaternion variable and its involutions. The HR calculus has been utilized in quaternion independent component analysis [30], nonlinear adaptive filtering [31], affine projection algorithms [32], and Kalman filtering [33]. However, the traditional product rule does not apply within the HR calculus because of the non-commutativity of quaternion product. The recently proposed generalized HR (GHR) calculus [34] rectifies this issue by making use of the quaternion rotation. It also comprises a novel product rule and chain rule and is a natural extension of the complex CR (or Wirtinger) calculus [35]–[37], which has been instrumental for the developments in complex-valued signal processing [38]–[40] and optimization [41]. In [34], the authors provide a systematic treatment of the derivatives of quaternion scalar functions which depend on quaternion argument, however, the more general matrix case was not considered. Problems where the unknown parameter is a quaternion matrix are wide ranging, from array signal processing [10], [11] to space-time coding [13]–[15], and quaternion orthogonal designs [16].

The derivatives of real matrix functions are well understood and have been studied in [42]–[44]. For the complex-valued vector case, the mathematical foundations for derivatives have been considered in [36], where the major contribution is the notion of complex gradient and the condition of stationary point in the context of optimization. This work was further extended to second order derivatives together with a duality relationship between the complex gradient and Hessian and their real bivariate counterparts [45]. A systematic treatment of all the related concepts is available in [37]. More general complex matrix derivatives have been thoroughly addressed in [46], [47].

Our aim here is to establish a systematic theory for calculating the derivatives of matrix functions with respect to quaternion matrix variables. To this end, the GHR calculus for scalars is used to develop a new calculus for functions of quaternion matrices. The 'vectorise' (vec) operator and the Jacobian matrix play an important role in the proposed calculus, allowing, for the first time, for general matrix product and chain rules. In

Manuscript received November 10, 2014; revised January 26, 2015; accepted January 27, 2015. Date of publication February 03, 2015; date of current version February 18, 2015. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Subhrakanti Dey. This work was supported by the National Natural Science Foundation of China (No. 61301202), and by the Research Fund for the Doctoral Program of Higher Education of China (No. 20122304120028).

D. Xu is with the School of Mathematics and Statistics, Northeast Normal University, Changchun 130024, China, with the College of Science, Harbin Engineering University, Harbin 150001, China, and with the Department of Electrical and Electronic Engineering, Imperial College London, London SW7 2AZ, U.K. (e-mail: dongpoxu@gmail.com).

D. P. Mandic is with the Department of Electrical and Electronic Engineering, Imperial College London, London SW7 2AZ, U.K. (e-mail: d.mandic@imperial.ac.uk).

Digital Object Identifier 10.1109/TSP.2015.2399865

TABLE I
NOTATION FOR FUNCTIONS AND VARIABLES

Function type	Scalar variable $q \in \mathbb{H}$	Vector variable $\boldsymbol{q} \in \mathbb{H}^{N\times 1}$	Matrix variable $\boldsymbol{Q} \in \mathbb{H}^{N\times S}$
Scalar function $f \in \mathbb{H}$	$f(q)$	$f(\boldsymbol{q})$	$f(\boldsymbol{Q})$
Vector function $\boldsymbol{f} \in \mathbb{H}^{M\times 1}$	$\boldsymbol{f}(q)$	$\boldsymbol{f}(\boldsymbol{q})$	$\boldsymbol{f}(\boldsymbol{Q})$
Matrix function $\boldsymbol{F} \in \mathbb{H}^{M\times P}$	$\boldsymbol{F}(q)$	$\boldsymbol{F}(\boldsymbol{q})$	$\boldsymbol{F}(\boldsymbol{Q})$

addition, the proposed rules are generic and reduce to scalar calculus rules when the matrices involved are of order one. For a real scalar function of quaternion matrix variable, the necessary conditions for the optimality can be found by either setting the derivative of the function with respect to the quaternion matrix variable or its quaternion involutions to zero. Meanwhile, the direction of maximum rate of change of the function is given by the Hermitian of derivative of the function with respect to the quaternion matrix variable. Our results therefore offer a generalization of the results for scalar functions of vector variables and make possible a direct calculation of the quaternion matrix derivatives without use of the quaternion differentials. The proposed theory is useful for numerous optimization problems which involve quaternion matrix parameters.

The rest of this paper is organized as follows. In Section II, some basic concepts of quaternion algebra and the quaternion differential are introduced and the GHR derivatives are defined and compared with the complex CR derivatives. The definition and rules of the quaternion matrix derivatives are given in Section III. Section IV contains important results, related to conditions for finding stationary points, and the steepest descent method. In Section V, several key results are comprised into tables and some more practical results are derived based on the proposed theory. Finally, Section VI concludes the paper. Some of the detailed proofs are given in the appendices.

II. PRELIMINARIES

A. Notations

We use bold-face upper case letters to denote matrices, bold-face lower case letters for column vectors, and standard lower case letters for scalar quantities. Notation for functions and variables is shown in Table I. Superscripts $(\cdot)^*$, $(\cdot)^T$ and $(\cdot)^H$ denote respectively the quaternion conjugate, transpose and Hermitian (i.e., transpose and quaternion conjugate), while the operators $\mathfrak{R}(\boldsymbol{A})$, $\mathrm{Tr}(\boldsymbol{A})$ and $\|\boldsymbol{A}\|$ denote the real part, trace and norm of $\boldsymbol{A}$, and $\otimes$ and $\odot$ denote the Kronecker and Hadamard product. The operator $\mathrm{vec}(\cdot)$ vectorizes a matrix by stacking its columns, $\boldsymbol{I}_N$ is the identity matrix of dimension N, and $\boldsymbol{0}_{N\times S}$ denotes the $N \times S$ zero matrix. By $\mathrm{reshape}(\cdot)$ we refer to any linear reshaping operator of the matrix, examples of such operators are the transpose $(\cdot)^T$ and $\mathrm{vec}(\cdot)$.

B. Quaternion Algebra

Quaternions are an associative but not commutative algebra over $\mathbb{R}$, defined as[1]

$$\mathbb{H} = \{q_a + iq_b + jq_c + kq_d \mid q_a, q_b, q_c, q_d \in \mathbb{R}\} \tag{1}$$

[1]For advanced reading on quaternions, we refer to [48], and to [49] for results on matrices of quaternions.

where $\{1, i, j, k\}$ is a basis of $\mathbb{H}$, and the imaginary units i, j and k satisfy $i^2 = j^2 = k^2 = ijk = -1$, which implies $ij = k = -ji$, $jk = i = -kj$, $ki = j = -ik$. For any quaternion

$$q = q_a + iq_b + jq_c + kq_d = Sq + Vq \tag{2}$$

the scalar (real) part is denoted by $q_a = Sq = \mathfrak{R}(q)$, while the vector part $Vq = \mathfrak{I}(q) = iq_b + jq_c + kq_d$ comprises the three imaginary parts. Quaternions form a noncommutative algebra, i.e., in general for $p, q \in \mathbb{H}$, $pq \neq qp$. The conjugate of a quaternion q is defined as $q^* = Sq - Vq$, while the conjugate of the product satisfies $(pq)^* = q^*p^*$. The modulus of a quaternion is defined as $|q| = \sqrt{qq^*}$, and obeys $|pq| = |p||q|$. The inner product of p and q is defined as $< p, q >= \mathfrak{R}(p^*q)$ The inverse of a quaternion $q \neq 0$ is $q^{-1} = q^*/|q|^2$, an important property of the inverse is $(pq)^{-1} = q^{-1}p^{-1}$. If $|q| = 1$, we call q a *unit* quaternion. A quaternion q is said to be *pure* if $\mathfrak{R}(q) = 0$, then $q^* = -q$ and $q^2 = -|q|^2$. Thus, a *pure unit* quaternion is a square root of -1, such as the imaginary units i, j and k.

Quaternions can also be written in the *polar* form $q = |q|(\cos\theta + \hat{q}\sin\theta)$, where $\hat{q} = Vq/|Vq|$ is a pure unit quaternion and $\theta = \arccos(S_q/|q|)$ is the angle (or argument). We shall next introduce the quaternion rotation, key concepts for the material in this work.

Definition 2.1 (Quaternion Rotation [48, p. 81]): For any quaternion q, the transformation

$$q^{\mu} \triangleq \mu q \mu^{-1} \tag{3}$$

geometrically describes a 3-dimensional rotation of the vector part of q by an angle 2θ about the vector part of μ, where $\mu = |\mu|(\cos\theta + \hat{\mu}\sin\theta)$ is any non-zero quaternion.

In particular, if μ in (3) is a pure unit quaternion, then the quaternion rotation (3) becomes the quaternion involution [50]. Important properties of the quaternion rotation (the proof of (4) and (5) is given in Appendix I) are

$$q^{\mu} = q^{\left(\frac{\mu}{|\mu|}\right)}, (pq)^{\mu} = p^{\mu}q^{\mu}, pq = q^{p}p = qp^{(q^*)}, \forall p, q \in \mathbb{H} \tag{4}$$

$$q^{\mu\nu} = (q^{\nu})^{\mu},\ q^{\mu *} \triangleq (q^*)^{\mu} = (q^{\mu})^* \triangleq q^{*\mu},\ \forall \nu, \mu \in \mathbb{H} \tag{5}$$

where $\frac{\mu}{|\mu|}$ is an unit quaternion, that is, $|\frac{\mu}{|\mu|}| = 1$. Hence, the quaternion μ in (3) does not need to be an unit quaternion due to $q^{\mu} = q^{\left(\frac{\mu}{|\mu|}\right)}$. Note that the real representation in (1) can be easily generalized to a general orthogonal basis $\{1, i^{\mu}, j^{\mu}, k^{\mu}\}$ given in [48], where the following properties hold

$$i^{\mu}i^{\mu} = j^{\mu}j^{\mu} = k^{\mu}k^{\mu} = i^{\mu}j^{\mu}k^{\mu} = -1 \tag{6}$$

TABLE II
SUMMARY OF RESULTS FOR QUATERNION MATRIX DIFFERENTIALS

Function	$\boldsymbol{A}$	$\alpha\boldsymbol{Q}\beta$	$\boldsymbol{P}+\boldsymbol{Q}$	$\text{Tr}(\boldsymbol{Q})$	$\boldsymbol{PQ}$	$\boldsymbol{P}\otimes\boldsymbol{Q}$
Differential	$\boldsymbol{0}$	$\alpha(d\boldsymbol{Q})\beta$	$d\boldsymbol{P}+d\boldsymbol{Q}$	$\text{Tr}(d\boldsymbol{Q})$	$(d\boldsymbol{P})\boldsymbol{Q}+\boldsymbol{P}(d\boldsymbol{Q})$	$(d\boldsymbol{P})\otimes\boldsymbol{Q}+\boldsymbol{P}\otimes(d\boldsymbol{Q})$

Function	$\boldsymbol{Q}^{\mu}$	$\boldsymbol{Q}^{\mu*}$	$\text{vec}(\boldsymbol{Q})$	$\text{reshape}(\boldsymbol{Q})$	$\boldsymbol{Q}^{-1}$	$\boldsymbol{P}\odot\boldsymbol{Q}$
Differential	$(d\boldsymbol{Q})^{\mu}$	$(d\boldsymbol{Q})^{\mu*}$	$\text{vec}(d\boldsymbol{Q})$	$\text{reshape}(d\boldsymbol{Q})$	$-\boldsymbol{Q}^{-1}(d\boldsymbol{Q})\boldsymbol{Q}^{-1}$	$(d\boldsymbol{P})\odot\boldsymbol{Q}+\boldsymbol{P}\odot(d\boldsymbol{Q})$

Moreover, quaternion rotations around the quaternions $\{\mu, \mu i, \mu j, \mu k\}$ are given by

$$q^{\mu}=q_a+i^{\mu}q_b+j^{\mu}q_c+k^{\mu}q_d,\ q^{\mu i}=q_a+i^{\mu}q_b-j^{\mu}q_c-k^{\mu}q_d$$
$$q^{\mu j}=q_a-i^{\mu}q_b+j^{\mu}q_c-k^{\mu}q_d,\ q^{\mu k}=q_a-i^{\mu}q_b-j^{\mu}q_c+k^{\mu}q_d \tag{7}$$

which allows us to express the four real-valued components of a quaternion q as

$$q_a=\frac{1}{4}(q+q^{\mu i}+q^{\mu j}+q^{\mu k}),\ q_b=\frac{-i^{\mu}}{4}(q+q^{\mu i}-q^{\mu j}-q^{\mu k})$$
$$q_c=\frac{-j^{\mu}}{4}(q-q^{\mu i}+q^{\mu j}-q^{\mu k}),\ q_d=\frac{-k^{\mu}}{4}(q-q^{\mu i}-q^{\mu j}+q^{\mu k}) \tag{8}$$

C. Quaternion Differential

The differential has the same size as the matrix it is applied to, and can be found component-wise, that is, $(d\boldsymbol{Q})_{m,p}=d(\boldsymbol{Q})_{m,p}$. A convenient way to find the differentials of $\boldsymbol{F}(\boldsymbol{Q})$ is to calculate the difference

$$\boldsymbol{F}(\boldsymbol{Q}+d\boldsymbol{Q})-\boldsymbol{F}(\boldsymbol{Q})=\text{First-order}(d\boldsymbol{Q})+\text{Higher-order}(d\boldsymbol{Q}) \tag{9}$$

Then $d\boldsymbol{F}(\boldsymbol{Q})=\text{First}-\text{order}(d\boldsymbol{Q})$, comprising the first order terms of the expression for $d\boldsymbol{Q}$ above. This definition complies with the multiplicative and associative rules

$$d(\alpha\boldsymbol{Q}\beta)=\alpha(d\boldsymbol{Q})\beta,\quad d(\boldsymbol{P}+\boldsymbol{Q})=d\boldsymbol{P}+d\boldsymbol{Q} \tag{10}$$

where $\alpha,\beta\in\mathbb{H}$. If $\boldsymbol{P}$ and $\boldsymbol{Q}$ are product-conforming matrices, it can be verified that the differential of their product is

$$d(\boldsymbol{PQ})=(d\boldsymbol{P})\boldsymbol{Q}+\boldsymbol{P}(d\boldsymbol{Q}) \tag{11}$$

Some of the most important results on quaternion matrix differentials are summarized in Table II, assuming $\boldsymbol{A}$, $\boldsymbol{B}$, and α,β to be quaternion constants, and $\boldsymbol{P}$, $\boldsymbol{Q}$ to be quaternion matrix variables.

The following lemma is useful to distinguish the GHR derivatives from the differential of a quaternion function.

Lemma 2.1: Let $\boldsymbol{Q}\in\mathbb{H}^{N\times S}$, $\boldsymbol{A}_n\in\mathbb{H}^{M\times NS}$ and $\mu\in\mathbb{H}$, $\mu\neq 0$. If $\boldsymbol{A}_n$ satisfies any of the following equations

$$\boldsymbol{A}_1 d\text{vec}(\boldsymbol{Q}^{\mu})+\boldsymbol{A}_2 d\text{vec}(\boldsymbol{Q}^{\mu i})+\boldsymbol{A}_3 d\text{vec}(\boldsymbol{Q}^{\mu j})+\boldsymbol{A}_4 d\text{vec}(\boldsymbol{Q}^{\mu k})=\boldsymbol{0} \tag{12}$$
$$\boldsymbol{A}_1 d\text{vec}(\boldsymbol{Q}^{\mu *})+\boldsymbol{A}_2 d\text{vec}(\boldsymbol{Q}^{\mu i*})+\boldsymbol{A}_3 d\text{vec}(\boldsymbol{Q}^{\mu j*})+\boldsymbol{A}_4 d\text{vec}(\boldsymbol{Q}^{\mu k*})=\boldsymbol{0} \tag{13}$$

for all $d\boldsymbol{Q}^{\mu}\in\mathbb{H}^{N\times S}$, then $\boldsymbol{A}_n=\boldsymbol{0}_{M\times NS}$ for $n\in\{1,2,3,4\}$.

Proof: The proof of Lemma 2.1 is given in Appendix II. ■

D. The GHR Calculus

The quaternion derivative in quaternion analysis is defined only for analytic functions of quaternion variables. However, in engineering problems, often the goal is to minimize a measure of error power, typically a real scalar function of quaternion variables, that is

$$f(q)=|e(q)|^2=e(q)e^*(q) \tag{14}$$

Notice that according to the definition of analytic (regular) function given in [24]–[28], the function f is not analytic. In order to take the derivative of such functions, the HR calculus extends the classical idea of the complex CR calculus [35]–[37] to the quaternion field, whereby the HR derivatives are given by [29]

$$\begin{bmatrix}\frac{\partial f}{\partial q}\\ \frac{\partial f}{\partial q^{i}}\\ \frac{\partial f}{\partial q^{j}}\\ \frac{\partial f}{\partial q^{k}}\end{bmatrix}=\frac{1}{4}\begin{bmatrix}1 & -i & -j & -k\\ 1 & -i & j & k\\ 1 & i & -j & k\\ 1 & i & j & -k\end{bmatrix}\begin{bmatrix}\frac{\partial f}{\partial q_a}\\ \frac{\partial f}{\partial q_b}\\ \frac{\partial f}{\partial q_c}\\ \frac{\partial f}{\partial q_d}\end{bmatrix} \tag{15}$$

Recently, a complementary version of HR derivative is proposed in [34], [52]–[54], that is

$$\begin{bmatrix}\frac{\partial f}{\partial q}\\ \frac{\partial f}{\partial q^{i}}\\ \frac{\partial f}{\partial q^{j}}\\ \frac{\partial f}{\partial q^{k}}\end{bmatrix}^T=\frac{1}{4}\begin{bmatrix}\frac{\partial f}{\partial q_a}\\ \frac{\partial f}{\partial q_b}\\ \frac{\partial f}{\partial q_c}\\ \frac{\partial f}{\partial q_d}\end{bmatrix}^T\begin{bmatrix}1 & 1 & 1 & 1\\ -i & -i & i & i\\ -j & j & -j & j\\ -k & k & k & -k\end{bmatrix} \tag{16}$$

It is important to note that the traditional product rule is not valid for the HR derivatives (15) and (16), see Example 2.1.

Example 2.1: Find the HR derivative of the real scalar function $f:\mathbb{H}\to\mathbb{R}$ given by

$$f(q)=|q|^2=q^*q=q_a^2+q_b^2+q_c^2+q_d^2 \tag{17}$$

where $q=q_a+iq_b+jq_c+kq_d$, $q_a,q_b,q_c,q_d\in\mathbb{R}$.

Solution: By (15) or (16), the HR derivative of $f(q)$ is

$$\frac{\partial(|q|^2)}{\partial q}=\frac{1}{4}\left(\frac{\partial|q|^2}{\partial q_a}-\frac{\partial|q|^2}{\partial q_b}i-\frac{\partial|q|^2}{\partial q_c}j-\frac{\partial|q|^2}{\partial q_d}k\right) \tag{18}$$
$$=\frac{1}{4}(2q_a-2q_b i-2q_c j-2q_d k)=\frac{1}{2}q^* \tag{19}$$

When we misuse the product rule of the HR derivatives, we have $\frac{\partial|q|^2}{\partial q}=q^*\frac{\partial q}{\partial q}+\frac{\partial q^*}{\partial q}q=q^*-\frac{1}{2}q$. In contrast, using the novel product rule (44) and (30), we have $\frac{\partial|q|^2}{\partial q}=q^*\frac{\partial q}{\partial q}+\frac{\partial q^*}{\partial q^{q}}q=q^*-\frac{1}{2}q^*=\frac{1}{2}q^*$. This shows that the HR derivatives (15) and (16) do not admit the product rule, however, the GHR derivatives solve this problem. ■

We shall now introduce the GHR derivatives (the derivation of GHR calculus is given in Appendix III).

Definition 2.2 (The GHR Derivatives [34]): Let $q = q_a + iq_b + jq_c + kq_d$, where $q_a, q_b, q_c, q_d \in \mathbb{R}$. Then, the left GHR derivatives of the function f with respect to q^μ and $q^{\mu*}$ ($\mu \neq 0, \mu \in \mathbb{H}$) are defined as

$$\frac{\partial_l f}{\partial q^\mu} = \frac{1}{4}\left(\frac{\partial f}{\partial q_a} - \frac{\partial f}{\partial q_b} i^\mu - \frac{\partial f}{\partial q_c} j^\mu - \frac{\partial f}{\partial q_d} k^\mu\right) \quad (20)$$

$$\frac{\partial_l f}{\partial q^{\mu*}} = \frac{1}{4}\left(\frac{\partial f}{\partial q_a} + \frac{\partial f}{\partial q_b} i^\mu + \frac{\partial f}{\partial q_c} j^\mu + \frac{\partial f}{\partial q_d} k^\mu\right) \quad (21)$$

while the right GHR derivatives are defined as

$$\frac{\partial_r f}{\partial q^\mu} = \frac{1}{4}\left(\frac{\partial f}{\partial q_a} - i^\mu \frac{\partial f}{\partial q_b} - j^\mu \frac{\partial f}{\partial q_c} - k^\mu \frac{\partial f}{\partial q_d}\right) \quad (22)$$

$$\frac{\partial_r f}{\partial q^{\mu*}} = \frac{1}{4}\left(\frac{\partial f}{\partial q_a} + i^\mu \frac{\partial f}{\partial q_b} + j^\mu \frac{\partial f}{\partial q_c} + k^\mu \frac{\partial f}{\partial q_d}\right) \quad (23)$$

where $\frac{\partial f}{\partial q_a}$, $\frac{\partial f}{\partial q_b}$, $\frac{\partial f}{\partial q_c}$ and $\frac{\partial f}{\partial q_d}$ are the partial derivatives of f with respect to q_a, q_b, q_c and q_d, respectively, and the set $\{1, i^\mu, j^\mu, k^\mu\}$ is a general orthogonal basis of $\mathbb{H}$.

For space consideration, without loss in generality, in the sequel we only consider the left GHR derivatives and write $\frac{\partial_l f}{\partial q^\mu} = \frac{\partial f}{\partial q^\mu}$ and $\frac{\partial_l f}{\partial q^{\mu*}} = \frac{\partial f}{\partial q^{\mu*}}$. The GHR derivatives can be regarded as a generalization of the complex CR derivatives in [36], [37], [46], however, there are significant differences:

- *Placement of imaginary units* i^μ, j^μ, k^μ. In (20) and (22), the terms $\frac{\partial f}{\partial q_a}$, $\frac{\partial f}{\partial q_b}$, $\frac{\partial f}{\partial q_c}$ and $\frac{\partial f}{\partial q_d}$ cannot be swapped with i^μ, j^μ, k^μ due to the non-commutativity of quaternion product. However, the multiplication operator is commutative in the complex domain.
- *The differential of* $f(q)$. From (98) and (106), we have $df = \frac{\partial f}{\partial q} dq + \frac{\partial f}{\partial q^i} dq^i + \frac{\partial f}{\partial q^j} dq^j + \frac{\partial f}{\partial q^k} dq^k$ and $df = \frac{\partial f}{\partial q^*} dq^* + \frac{\partial f}{\partial q^{i*}} dq^{i*} + \frac{\partial f}{\partial q^{j*}} dq^{j*} + \frac{\partial f}{\partial q^{k*}} dq^{k*}$. In comparison in [36], the complex differential is $df = \frac{\partial f}{\partial z} dz + \frac{\partial f}{\partial z^*} dz^*$.
- *The derivatives of* $f(q)$. From (20) and (21), we have $\frac{\partial q}{\partial q} = \frac{\partial q^*}{\partial q^*} = 1$ and $\frac{\partial q}{\partial q^*} = \frac{\partial q^*}{\partial q} = -\frac{1}{2}$. In contrast, the complex CR derivatives $\frac{\partial z}{\partial z^*} = \frac{\partial z^*}{\partial z} = 0$.
- *Product rule*. By Corollary 3.1, we have $\frac{\partial (fg)}{\partial q} = f \frac{\partial g}{\partial q} + \frac{\partial f}{\partial q^g} g$. In contrast, the product rule of the complex CR derivatives is $\frac{\partial (fg)}{\partial z} = f \frac{\partial g}{\partial z} + \frac{\partial f}{\partial z} g$, as given in [36], [37].
- *Chain rule*. When the matrices in (45) are of order one, we have $\frac{\partial f(g(q))}{\partial q} = \frac{\partial f}{\partial g}\frac{\partial g}{\partial q} + \frac{\partial f}{\partial g^i}\frac{\partial g^i}{\partial q} + \frac{\partial f}{\partial g^j}\frac{\partial g^j}{\partial q} + \frac{\partial f}{\partial g^k}\frac{\partial g^k}{\partial q}$. In contrast, the chain rule of complex CR derivatives is $\frac{\partial f(g(z))}{\partial q} = \frac{\partial f}{\partial g}\frac{\partial g}{\partial z} + \frac{\partial f}{\partial g^*}\frac{\partial g^*}{\partial z}$ in [36], [37], [40].

Observe that for $\mu \in \{1, i, j, k\}$, the HR derivatives (15) and (16) are a special case of the right GHR derivatives (22) and (20), which are more concise and easier to understand. In particular, we show that for real functions of quaternion variables, such as the standard real mean square error (MSE) objective function, the left and right GHR derivatives are identical, as shown in (24). This indicates that the choice of the left/right GHR derivative is irrelevant for practical applications of quaternion optimization, a major current source of confusion in the quaternion community.

Lemma 2.2: Let $f(q) : \mathbb{H} \to \mathbb{R}$. Then the following holds

$$\frac{\partial f}{\partial q^\mu} = \frac{\partial_r f}{\partial q^\mu}, \quad \frac{\partial f}{\partial q^{\mu*}} = \frac{\partial_r f}{\partial q^{\mu*}} \quad (24)$$

$$\left(\frac{\partial f}{\partial q^\mu}\right)^\nu = \frac{\partial f}{\partial q^{\nu\mu}}, \quad \left(\frac{\partial f}{\partial q^{\mu*}}\right)^\nu = \frac{\partial f}{\partial q^{\nu\mu*}} \quad (25)$$

$$\left(\frac{\partial f}{\partial q^\mu}\right)^* = \frac{\partial f}{\partial q^{\mu*}}, \quad \left(\frac{\partial f}{\partial q^{\mu*}}\right)^* = \frac{\partial f}{\partial q^\mu} \quad (26)$$

Proof: The proof can be found in Appendix IV. ■

Lemma 2.3: Let $f(q) : \mathbb{H} \to \mathbb{H}$. Then the following holds

$$\frac{\partial(\alpha f \beta)}{\partial q^\mu} = \alpha \frac{\partial f}{\partial q^{\beta\mu}} \beta, \quad \frac{\partial(\alpha f \beta)}{\partial q^{\mu*}} = \alpha \frac{\partial f}{\partial q^{\beta\mu*}} \beta \quad (27)$$

$$\frac{\partial_r(\alpha f \beta)}{\partial q^\mu} = \alpha \frac{\partial_r f}{\partial q^{(\alpha^{-1}\mu)}} \beta, \quad \frac{\partial_r(\alpha f \beta)}{\partial q^{\mu*}} = \alpha \frac{\partial_r f}{\partial q^{(\alpha^{-1}\mu*)}} \beta \quad (28)$$

where α and β are quaternion constants.

Proof: The proof can be found in Appendix IV. ■

One example is when $f(q) = q$ and $\mu = 1$, then we have

$$\begin{aligned} \frac{\partial(\alpha q \beta)}{\partial q} &= \alpha \frac{\partial q}{\partial q^\beta} \beta = \alpha \Re(\beta), \\ \frac{\partial(\alpha q \beta)}{\partial q^*} &= \alpha \frac{\partial q}{\partial q^{\beta*}} \beta = -\frac{1}{2} \alpha \beta^* \end{aligned} \quad (29)$$

Another example is when $f(q) = q^*$ and $\mu = 1$, we have

$$\begin{aligned} \frac{\partial(\alpha q^* \beta)}{\partial q} &= \alpha \frac{\partial q^*}{\partial q^\beta} \beta = -\frac{1}{2} \alpha \beta^*, \\ \frac{\partial(\alpha q^* \beta)}{\partial q^*} &= \alpha \frac{\partial q^*}{\partial q^{\beta*}} \beta = \alpha \Re(\beta) \end{aligned} \quad (30)$$

Example 2.2: Find the HR derivative of the real scalar function $J : \mathbb{H} \to \mathbb{R}$ given by

$$J(w) = |e(w)|^2 = |d - wx|^2 \quad (31)$$

where x and d are quaternion constants.

Solution: When we apply the product rule in conjunction with the HR derivatives, we obtain

$$\begin{aligned} \frac{\partial J}{\partial w^*} &= e^* \frac{\partial e}{\partial w^*} + \frac{\partial e^*}{\partial w^*} e = -e^* \frac{\partial (wx)}{\partial w^*} - \frac{\partial (x^* w^*)}{\partial w^*} e \\ &= -e^* \frac{\partial w}{\partial w^*} x - x^* \frac{\partial w^*}{\partial w^*} e = \frac{1}{2} e^* x - x^* e \end{aligned} \quad (32)$$

In contrast to the incorrect result in (32), using the novel product rule (44) and (30), we have

$$\begin{aligned} \frac{\partial J}{\partial w^*} &= e^* \frac{\partial e}{\partial w^*} + \frac{\partial e^*}{\partial w^{e*}} e = -e^* \frac{\partial (wx)}{\partial w^*} - \frac{\partial (x^* w^*)}{\partial w^{e*}} e \\ &= \frac{1}{2} e^* x^* - x^* \frac{\partial w^*}{\partial w^{e*}} e = \frac{1}{2} e^* x^* - x^* \Re(e) = -\frac{1}{2} e x^* \end{aligned} \quad (33)$$

This indicates the generic and general nature of the GHR derivatives with respect to the HR derivatives, which do not admit the product and chain rules. We should, however, mention that the HR calculus is perfectly correct when these rules are not used. ■

III. Definition of Quaternion Matrix Derivatives

In the differentiation of matrix functions with respect to a matrix variable $\boldsymbol{Q}$, it is always assumed that all the elements of $\boldsymbol{Q}$ are linearly independent.

TABLE III
NOTATION FOR QUATERNION DERIVATIVES

Function type	Differential	Derivatives wrt $q, \boldsymbol{q}, \boldsymbol{Q}$	Derivatives wrt $q^*, \boldsymbol{q}^*, \boldsymbol{Q}^*$	Order of derivatives
$f(q)$	$df = \sum_{\mu\in\{1,i,j,k\}} a_\mu dq^\mu$ $df = \sum_{\mu\in\{1,i,j,k\}} b_\mu dq^{\mu*}$	$\mathcal{D}_q f(q) = a_1$	$\mathcal{D}_{q^*} f(q) = b_1$	1×1
$f(\boldsymbol{q})$	$df = \sum_{\mu\in\{1,i,j,k\}} \boldsymbol{a}_\mu^T d\boldsymbol{q}^\mu$ $df = \sum_{\mu\in\{1,i,j,k\}} \boldsymbol{b}_\mu^T d\boldsymbol{q}^{\mu*}$	$\mathcal{D}_{\boldsymbol{q}} f(\boldsymbol{q}) = \boldsymbol{a}_1^T$	$\mathcal{D}_{\boldsymbol{q}^*} f(\boldsymbol{q}) = \boldsymbol{b}_1^T$	$1\times N$
$f(\boldsymbol{Q})$	$df = \sum_{\mu\in\{1,i,j,k\}} \mathrm{vec}^T(\boldsymbol{A}_\mu)d\mathrm{vec}(\boldsymbol{Q}^\mu)$ $df = \sum_{\mu\in\{1,i,j,k\}} \mathrm{vec}^T(\boldsymbol{B}_\mu)d\mathrm{vec}(\boldsymbol{Q}^{\mu*})$	$\mathcal{D}_{\boldsymbol{Q}} f(\boldsymbol{Q}) = \mathrm{vec}^T(\boldsymbol{A}_1)$	$\mathcal{D}_{\boldsymbol{Q}^*} f(\boldsymbol{Q}) = \mathrm{vec}^T(\boldsymbol{B}_1)$	$1\times NS$
$f(\boldsymbol{Q})$	$df = \sum_{\mu\in\{1,i,j,k\}} \mathrm{Tr}\{\boldsymbol{A}_\mu^T d\boldsymbol{Q}^\mu\}$ $df = \sum_{\mu\in\{1,i,j,k\}} \mathrm{Tr}\{\boldsymbol{B}_\mu^T d\boldsymbol{Q}^*\}$	$\frac{\partial f}{\partial \boldsymbol{Q}} = \boldsymbol{A}_1$	$\frac{\partial f}{\partial \boldsymbol{Q}^*} = \boldsymbol{B}_1$	$N\times S$
$\boldsymbol{f}(q)$	$d\boldsymbol{f} = \sum_{\mu\in\{1,i,j,k\}} \boldsymbol{c}_\mu dq^\mu$ $d\boldsymbol{f} = \sum_{\mu\in\{1,i,j,k\}} \boldsymbol{d}_\mu dq^{\mu*}$	$\mathcal{D}_q \boldsymbol{f}(q) = \boldsymbol{c}_1$	$\mathcal{D}_{q^*} \boldsymbol{f}(q) = \boldsymbol{d}_1$	$M\times 1$
$\boldsymbol{f}(\boldsymbol{q})$	$d\boldsymbol{f} = \sum_{\mu\in\{1,i,j,k\}} \boldsymbol{C}_\mu d\boldsymbol{q}^\mu$ $d\boldsymbol{f} = \sum_{\mu\in\{1,i,j,k\}} \boldsymbol{D}_\mu d\boldsymbol{q}^{\mu*}$	$\mathcal{D}_{\boldsymbol{q}} \boldsymbol{f}(\boldsymbol{q}) = \boldsymbol{C}_1$	$\mathcal{D}_{\boldsymbol{q}^*} \boldsymbol{f}(\boldsymbol{q}) = \boldsymbol{D}_1$	$M\times N$
$\boldsymbol{f}(\boldsymbol{Q})$	$d\boldsymbol{f} = \sum_{\mu\in\{1,i,j,k\}} \boldsymbol{\alpha}_\mu d\mathrm{vec}(\boldsymbol{Q}^\mu)$ $d\boldsymbol{f} = \sum_{\mu\in\{1,i,j,k\}} \boldsymbol{\beta}_\mu d\mathrm{vec}(\boldsymbol{Q}^{\mu*})$	$\mathcal{D}_{\boldsymbol{Q}} \boldsymbol{f}(\boldsymbol{q}) = \boldsymbol{\alpha}_1$	$\mathcal{D}_{\boldsymbol{Q}^*} \boldsymbol{f}(\boldsymbol{q}) = \boldsymbol{\beta}_1$	$M\times NS$
$\boldsymbol{F}(q)$	$d\mathrm{vec}(\boldsymbol{F}) = \sum_{\mu\in\{1,i,j,k\}} \boldsymbol{g}_\mu dq^\mu$ $d\mathrm{vec}(\boldsymbol{F}) = \sum_{\mu\in\{1,i,j,k\}} \boldsymbol{h}_\mu dq^{\mu*}$	$\mathcal{D}_q \boldsymbol{F}(q) = \boldsymbol{g}_1$	$\mathcal{D}_{q^*} \boldsymbol{F}(q) = \boldsymbol{h}_1$	$MP\times 1$
$\boldsymbol{F}(\boldsymbol{q})$	$d\mathrm{vec}(\boldsymbol{F}) = \sum_{\mu\in\{1,i,j,k\}} \boldsymbol{G}_\mu d\boldsymbol{q}^\mu$ $d\mathrm{vec}(\boldsymbol{F}) = \sum_{\mu\in\{1,i,j,k\}} \boldsymbol{H}_\mu d\boldsymbol{q}^{\mu*}$	$\mathcal{D}_{\boldsymbol{q}} \boldsymbol{F}(\boldsymbol{q}) = \boldsymbol{G}_1$	$\mathcal{D}_{\boldsymbol{q}^*} \boldsymbol{F}(\boldsymbol{q}) = \boldsymbol{H}_1$	$MP\times N$
$\boldsymbol{F}(\boldsymbol{Q})$	$d\mathrm{vec}(\boldsymbol{F}) = \sum_{\mu\in\{1,i,j,k\}} \boldsymbol{\zeta}_\mu d\mathrm{vec}(\boldsymbol{Q}^\mu)$ $d\mathrm{vec}(\boldsymbol{F}) = \sum_{\mu\in\{1,i,j,k\}} \boldsymbol{\xi}_\mu d\mathrm{vec}(\boldsymbol{Q}^{\mu*})$	$\mathcal{D}_{\boldsymbol{Q}} \boldsymbol{F}(\boldsymbol{Q}) = \boldsymbol{\zeta}_1$	$\mathcal{D}_{\boldsymbol{Q}^*} \boldsymbol{F}(\boldsymbol{Q}) = \boldsymbol{\xi}_1$	$MP\times NS$

For a scalar function $f(\boldsymbol{q})$ of an $N\times 1$ vector $\boldsymbol{q}$, the GHR derivatives are comprised in the $1\times N$ vector

$$\frac{\partial f}{\partial \boldsymbol{q}^\mu} = \left[\frac{\partial f}{\partial q_1^\mu}, \dots, \frac{\partial f}{\partial q_N^\mu}\right], \quad \frac{\partial f}{\partial \boldsymbol{q}^{\mu*}} = \left[\frac{\partial f}{\partial q_1^{\mu*}}, \dots, \frac{\partial f}{\partial q_N^{\mu*}}\right] \tag{34}$$

The gradient of $f(\boldsymbol{q}) \in \mathbb{H}$ is then the vector

$$\nabla_{\boldsymbol{q}^\mu} f \triangleq \left(\frac{\partial f}{\partial \boldsymbol{q}^\mu}\right)^T, \quad \nabla_{\boldsymbol{q}^{\mu*}} f \triangleq \left(\frac{\partial f}{\partial \boldsymbol{q}^{\mu*}}\right)^T \tag{35}$$

The existing quaternion gradients are summarized in [55], illustrating that the CRF-gradient [27] is not defined for real functions of quaternion variables, and the calculation of the pseudo-gradient (also known as component-wise gradients) is cumbersome and tedious [7], [51], making the derivation of quaternion optimization algorithms prone to errors. The HR-gradient [29] and the I-gradient [56] are a step forward, but do not admit the product and chain rules, complicating the calculation of gradient of nonlinear quaternion functions. On the contrary, the GHR-gradients defined in (35) comprise the product rule (40)-(41) and chain rule (45)–(48), while the gradient $\nabla_{\boldsymbol{q}^*} f$ denotes the direction of the maximum rate of change of real function $f(\boldsymbol{q})$, see Corollary 4.1.Further, if $\boldsymbol{f}$ is an $M\times 1$ vector function of a vector variable $\boldsymbol{q}$, then the $M\times N$ matrices

$$\frac{\partial \boldsymbol{f}}{\partial \boldsymbol{q}^\mu} = \begin{bmatrix} \frac{\partial f_1}{\partial \boldsymbol{q}^\mu} \\ \vdots \\ \frac{\partial f_M}{\partial \boldsymbol{q}^\mu} \end{bmatrix}, \quad \frac{\partial \boldsymbol{f}}{\partial \boldsymbol{q}^{\mu*}} = \begin{bmatrix} \frac{\partial f_1}{\partial \boldsymbol{q}^{\mu*}} \\ \vdots \\ \frac{\partial f_M}{\partial \boldsymbol{q}^{\mu*}} \end{bmatrix} \tag{36}$$

are called the derivatives or Jacobian matrices of $\boldsymbol{f}$. Generalizing these concepts to matrix functions of matrices, we arrive at the following definition.

Definition 3.1: Let $\boldsymbol{F} : \mathbb{H}^{N\times S} \to \mathbb{H}^{M\times P}$. Then the GHR derivatives (or Jacobian matrices) of $\boldsymbol{F}$ with respect to $\boldsymbol{Q}^\mu, \boldsymbol{Q}^{\mu*}$ ($\mu \in \mathbb{H}, \mu \neq 0$) are the $MP\times NS$ matrices

$$\mathcal{D}_{\boldsymbol{Q}^\mu}\boldsymbol{F} = \frac{\partial \mathrm{vec}\boldsymbol{F}(\boldsymbol{Q})}{\partial \mathrm{vec}(\boldsymbol{Q}^\mu)}, \quad \mathcal{D}_{\boldsymbol{Q}^{\mu*}}\boldsymbol{F} = \frac{\partial \mathrm{vec}\boldsymbol{F}(\boldsymbol{Q})}{\partial \mathrm{vec}(\boldsymbol{Q}^{\mu*})} \tag{37}$$

The transposes of the Jacobian matrices $\mathcal{D}_{\boldsymbol{Q}^\mu}\boldsymbol{F}$ and $\mathcal{D}_{\boldsymbol{Q}^{\mu*}}\boldsymbol{F}$ are called the gradients.

Using the matrix derivative notations in Definition 3.1, the differentials of the scalar function f in (98) and (106) can be extended to the following matrix case

$$d\mathrm{vec}(\boldsymbol{F}) = \sum_{\mu\in\{1,i,j,k\}} (\mathcal{D}_{\boldsymbol{Q}^\mu}\boldsymbol{F}) d\mathrm{vec}(\boldsymbol{Q}^\mu) \tag{38}$$

$$d\mathrm{vec}(\boldsymbol{F}) = \sum_{\mu\in\{1,i,j,k\}} (\mathcal{D}_{\boldsymbol{Q}^{\mu*}}\boldsymbol{F}) d\mathrm{vec}(\boldsymbol{Q}^{\mu*}) \tag{39}$$

This is different from the complex-valued matrix variable case in [46], where $d\mathrm{vec}(\boldsymbol{F}) = (\mathcal{D}_{\boldsymbol{Z}}\boldsymbol{F})d\mathrm{vec}(\boldsymbol{Z}) + (\mathcal{D}_{\boldsymbol{Z}^*}\boldsymbol{F})d\mathrm{vec}(\boldsymbol{Z}^*)$. It can be shown by using Lemma 2.1 that the matrix derivatives in (38) and (39) are unique.

Table III shows the connection between the differentials and derivatives of the function types in Table I. In Table III, $q \in \mathbb{H}$, $\boldsymbol{q} \in \mathbb{H}^{N\times 1}$, $\boldsymbol{Q} \in \mathbb{H}^{N\times S}$, $f \in \mathbb{H}$, $\boldsymbol{f} \in \mathbb{H}^{M\times 1}$ and $\boldsymbol{F} \in \mathbb{H}^{M\times P}$. Furthermore, $a_\mu, b_\mu \in \mathbb{H}$, $\boldsymbol{a}_\mu, \boldsymbol{b}_\mu \in \mathbb{H}^{N\times 1}$, $\boldsymbol{A}_\mu, \boldsymbol{B}_\mu \in \mathbb{H}^{N\times S}$, $\boldsymbol{c}_\mu, \boldsymbol{d}_\mu \in \mathbb{H}^{M\times 1}$, $\boldsymbol{C}_\mu, \boldsymbol{D}_\mu \in \mathbb{H}^{M\times N}$, $\boldsymbol{\alpha}_\mu, \boldsymbol{\beta}_\mu \in \mathbb{H}^{M\times NS}$, $\boldsymbol{g}_\mu, \boldsymbol{h}_\mu \in \mathbb{H}^{MP\times 1}$, $\boldsymbol{G}_\mu, \boldsymbol{H}_\mu \in \mathbb{H}^{MP\times N}$, $\boldsymbol{\zeta}_\mu, \boldsymbol{\xi}_\mu \in \mathbb{H}^{MP\times NS}$, and each of these may be a function of q, $\boldsymbol{q}$ or $\boldsymbol{Q}$.

A. Product Rule

In [34]s, we have given an example to show that the traditional product rule is not valid for the HR calculus of quaternion scalar variable. Now, we shall generalize the product rules in [34] to the case of quaternion matrix variable.

Theorem 3.1: Let $\boldsymbol{H}:\mathbb{H}^{N\times S}\rightarrow\mathbb{H}^{M\times P}$ be given by $\boldsymbol{H}=\boldsymbol{F}\boldsymbol{G}$, where $\boldsymbol{F}:\mathbb{H}^{N\times S}\rightarrow\mathbb{H}^{M\times R}$ and $\boldsymbol{G}:\mathbb{H}^{N\times S}\rightarrow\mathbb{H}^{R\times P}$. Then, the following relations hold

$$\mathcal{D}_{\boldsymbol{Q}^{\mu}}\boldsymbol{H}=(\boldsymbol{I}_P\otimes\boldsymbol{F})\mathcal{D}_{\boldsymbol{Q}^{\mu}}\boldsymbol{G}+\mathcal{D}_{\boldsymbol{Q}^{\mu}}(\boldsymbol{F}\boldsymbol{G})|_{\boldsymbol{G}=const} \quad (40)$$
$$\mathcal{D}_{\boldsymbol{Q}^{\mu*}}\boldsymbol{H}=(\boldsymbol{I}_P\otimes\boldsymbol{F})\mathcal{D}_{\boldsymbol{Q}^{\mu*}}\boldsymbol{G}+\mathcal{D}_{\boldsymbol{Q}^{\mu*}}(\boldsymbol{F}\boldsymbol{G})|_{\boldsymbol{G}=const} \quad (41)$$

Proof: The differential of $\boldsymbol{H}$ can be expressed as

$$d\boldsymbol{H}=\boldsymbol{F}(d\boldsymbol{G})\boldsymbol{I}_P+d(\boldsymbol{F}\boldsymbol{G})|_{\boldsymbol{G}=const} \quad (42)$$

By using the differentials of $\boldsymbol{F}$ and $\boldsymbol{G}$ after applying the $\mathrm{vec}(\cdot)$ operator, we have

$$\begin{aligned}&d\mathrm{vec}(\boldsymbol{H})\\&=(\boldsymbol{I}_P\otimes\boldsymbol{F})d\mathrm{vec}(\boldsymbol{G})+d\mathrm{vec}(\boldsymbol{F}\boldsymbol{G})|_{\boldsymbol{G}=const}\\&=\sum_{\mu\in\{1,i,j,k\}}(\boldsymbol{I}_P\otimes\boldsymbol{F})\,(\mathcal{D}_{\boldsymbol{Q}^{\mu}}\boldsymbol{G})d\mathrm{vec}(\boldsymbol{Q}^{\mu})\\&\quad+\sum_{\mu\in\{1,i,j,k\}}\mathcal{D}_{\boldsymbol{Q}^{\mu}}(\boldsymbol{F}\boldsymbol{G})|_{\boldsymbol{G}=const}\,d\mathrm{vec}(\boldsymbol{Q}^{\mu}) \quad (43)\\&=\sum_{\mu\in\{1,i,j,k\}}[(\boldsymbol{I}_P\otimes\boldsymbol{F})\mathcal{D}_{\boldsymbol{Q}^{\mu}}\boldsymbol{G}+\mathcal{D}_{\boldsymbol{Q}^{\mu}}(\boldsymbol{F}\boldsymbol{G})|_{\boldsymbol{G}=const}]d\mathrm{vec}(\boldsymbol{Q}^{\mu})\end{aligned}$$

where (4) and (10) are used in the last equality. Hence, the derivatives of $\boldsymbol{F}\boldsymbol{G}$ with respect to $\boldsymbol{Q}^{\mu}$ can be identified as in (40). The second equality can be proved in similar manner. ■

Corollary 3.1: Let $f,g:\mathbb{H}\rightarrow\mathbb{H}$. Then, the following novel product rules hold

$$\frac{\partial(fg)}{\partial q}=f\frac{\partial g}{\partial q}+\frac{\partial f}{\partial q^{g}}g,\quad \frac{\partial(fg)}{\partial q^{*}}=f\frac{\partial g}{\partial q^{*}}+\frac{\partial f}{\partial q^{g*}}g \quad (44)$$

Proof: If $\mu=1$ and the matrices involved in (40) are of order one, then $\frac{\partial(fg)}{\partial q}=f\frac{\partial g}{\partial q}+\frac{\partial(fg)}{\partial q}|_{g=const}$. Upon using (27), we have $\frac{\partial(fg)}{\partial q}=f\frac{\partial g}{\partial q}+\frac{\partial(fg)}{\partial q}|_{g=const}=f\frac{\partial g}{\partial q}+\frac{\partial f}{\partial q^{g}}g$. Hence, the first part of (44) follows, and the second part can be proved in a similar way. ■

B. Chain Rule

A major advantage of the matrix derivatives defined in Definition 3.1 is that the chain rule can be obtained in a very simple form, as stated in the following theorem.

Theorem 3.2: Let $U\subseteq\mathbb{H}^{N\times S}$ and suppose $\boldsymbol{G}:U\rightarrow\mathbb{H}^{M\times P}$ has the GHR derivatives at an interior point $\boldsymbol{Q}$ of the set U. Let $V\subseteq\mathbb{H}^{M\times P}$ be such that $\boldsymbol{G}(\boldsymbol{Q})\in V$ for all $\boldsymbol{Q}\in U$. Assume $\boldsymbol{F}:V\rightarrow\mathbb{H}^{R\times T}$ has GHR derivatives at an inner point $\boldsymbol{G}(\boldsymbol{Q})\in V$, then the GHR derivatives of the composite function $\boldsymbol{H}(\boldsymbol{Q})\triangleq\boldsymbol{F}(\boldsymbol{G}(\boldsymbol{Q}))$ are as follows:

$$\mathcal{D}_{\boldsymbol{Q}^{\mu}}\boldsymbol{H}=\sum_{\nu\in\{1,i,j,k\}}(\mathcal{D}_{\boldsymbol{G}^{\nu}}\boldsymbol{F})(\mathcal{D}_{\boldsymbol{Q}^{\mu}}\boldsymbol{G}^{\nu}) \quad (45)$$
$$\mathcal{D}_{\boldsymbol{Q}^{\mu*}}\boldsymbol{H}=\sum_{\nu\in\{1,i,j,k\}}(\mathcal{D}_{\boldsymbol{G}^{\nu}}\boldsymbol{F})(\mathcal{D}_{\boldsymbol{Q}^{\mu*}}\boldsymbol{G}^{\nu}) \quad (46)$$
$$\mathcal{D}_{\boldsymbol{Q}^{\mu}}\boldsymbol{H}=\sum_{\nu\in\{1,i,j,k\}}(\mathcal{D}_{\boldsymbol{G}^{\nu*}}\boldsymbol{F})(\mathcal{D}_{\boldsymbol{Q}^{\mu}}\boldsymbol{G}^{\nu*}) \quad (47)$$
$$\mathcal{D}_{\boldsymbol{Q}^{\mu*}}\boldsymbol{H}=\sum_{\nu\in\{1,i,j,k\}}(\mathcal{D}_{\boldsymbol{G}^{\nu*}}\boldsymbol{F})(\mathcal{D}_{\boldsymbol{Q}^{\mu*}}\boldsymbol{G}^{\nu*}) \quad (48)$$

Proof: From (38), we have

$$d\mathrm{vec}(\boldsymbol{H})=d\mathrm{vec}(\boldsymbol{F})=\sum_{\nu\in\{1,i,j,k\}}(\mathcal{D}_{\boldsymbol{G}^{\nu}}\boldsymbol{F})d\mathrm{vec}(\boldsymbol{G}^{\nu}) \quad (49)$$

The differential of $d\mathrm{vec}(\boldsymbol{G}^{\nu})$ is given by

$$d\mathrm{vec}(\boldsymbol{G}^{\nu})=\sum_{\mu\in\{1,i,j,k\}}(\mathcal{D}_{\boldsymbol{Q}^{\mu}}\boldsymbol{G}^{\nu})d\mathrm{vec}(\boldsymbol{Q}^{\mu}) \quad (50)$$

By substituting (50) into (49), we have

$$d\mathrm{vec}(\boldsymbol{H})=\sum_{\mu,\nu\in\{1,i,j,k\}}(\mathcal{D}_{\boldsymbol{G}^{\nu}}\boldsymbol{F})(\mathcal{D}_{\boldsymbol{Q}^{\mu}}\boldsymbol{G}^{\nu})d\mathrm{vec}(\boldsymbol{Q}^{\mu}) \quad (51)$$

According to (38), the derivatives of $\boldsymbol{H}$ with respect to $\boldsymbol{Q}^{\mu}$ can now be identified as in (45). The other equalities can be proved in a similar manner. ■

IV. Quaternion Optimization Using GHR Calculus

The objective functions in engineering applications are often real-valued and thus non-analytic. We next show how to use matrix GHR derivatives to find stationary points for scalar real-valued functions dependent on quaternion matrices, together with the directions where such functions have maximum rates of change.

A. Stationary Points

We shall now identify five equivalent ways which can be used to find stationary points of $f(\boldsymbol{Q})\in\mathbb{R}$, a necessary condition for optimality.

Lemma 4.1: Let $f:\mathbb{H}^{N\times S}\rightarrow\mathbb{R}$. Then the following holds

$$(\mathcal{D}_{\boldsymbol{Q}}f)^{\nu}=\mathcal{D}_{\boldsymbol{Q}^{\nu}}f,\quad \mathcal{D}_{\boldsymbol{Q}^{\nu*}}f=(\mathcal{D}_{\boldsymbol{Q}}f)^{\nu*} \quad (52)$$

Proof: Using (25) and (26), the lemma follows. ■

Theorem 4.1: Let $f:\mathbb{H}^{N\times S}\rightarrow\mathbb{R}$, and let $\boldsymbol{Q}=\boldsymbol{Q}_a+i\boldsymbol{Q}_b+j\boldsymbol{Q}_c+k\boldsymbol{Q}_d$, where $\boldsymbol{Q}_a,\boldsymbol{Q}_b,\boldsymbol{Q}_c,\boldsymbol{Q}_d\in\mathbb{R}^{N\times S}$. A stationary point of the function $f(\boldsymbol{Q})=g(\boldsymbol{\xi})$ can be founded by one of the following five equivalent conditions

$$\mathcal{D}_{\boldsymbol{\xi}}g(\boldsymbol{\xi})=\boldsymbol{0}\ \Leftrightarrow\ \mathcal{D}_{\boldsymbol{\zeta}}f(\boldsymbol{Q})=\boldsymbol{0}\ \Leftrightarrow\ \mathcal{D}_{\boldsymbol{Q}}f(\boldsymbol{Q})=\boldsymbol{0} \quad (53)$$
$$\mathcal{D}_{\boldsymbol{\xi}}g(\boldsymbol{\xi})=\boldsymbol{0}\ \Leftrightarrow\ \mathcal{D}_{\boldsymbol{\zeta}^{*}}f(\boldsymbol{Q})=\boldsymbol{0}\ \Leftrightarrow\ \mathcal{D}_{\boldsymbol{Q}^{*}}f(\boldsymbol{Q})=\boldsymbol{0} \quad (54)$$

where $\boldsymbol{\xi}=[\boldsymbol{Q}_a,\boldsymbol{Q}_b,\boldsymbol{Q}_c,\boldsymbol{Q}_d]$ and $\boldsymbol{\zeta}=[\boldsymbol{Q},\boldsymbol{Q}^{i},\boldsymbol{Q}^{j},\boldsymbol{Q}^{k}]$.

Proof: In [42], a stationary point is defined as point where the derivatives of the function vanish. Thus, $\mathcal{D}_{\boldsymbol{\xi}}g(\boldsymbol{\xi})=\boldsymbol{0}$ gives a stationary point by definition. Applying the chain rule (45) on both sides of $f(\boldsymbol{Q})=g(\boldsymbol{\xi})$, gives

$$\sum_{\nu\in\{1,i,j,k\}}(\mathcal{D}_{\boldsymbol{Q}^{\nu}}f)(\mathcal{D}_{\boldsymbol{\xi}}\boldsymbol{Q}^{\nu})=\mathcal{D}_{\boldsymbol{\xi}}g \quad (55)$$

From $\boldsymbol{Q}^{\nu}=\boldsymbol{Q}_a+i^{\nu}\boldsymbol{Q}_b+j^{\nu}\boldsymbol{Q}_c+k^{\nu}\boldsymbol{Q}_d$, it then follows that $\mathcal{D}_{\boldsymbol{\xi}}\boldsymbol{Q}^{\nu}=[\boldsymbol{I},i^{\nu}\boldsymbol{I},j^{\nu}\boldsymbol{I},k^{\nu}\boldsymbol{I}]$, where $\boldsymbol{I}\in\mathbb{R}^{NS\times NS}$ is the identity matrix. A substitution of these results into (55), gives

$$(\mathcal{D}_{\boldsymbol{\zeta}}f)\boldsymbol{J}=(\mathcal{D}_{\boldsymbol{\zeta}}f)\begin{bmatrix}\boldsymbol{I}&i\boldsymbol{I}&j\boldsymbol{I}&k\boldsymbol{I}\\\boldsymbol{I}&i\boldsymbol{I}&-j\boldsymbol{I}&-k\boldsymbol{I}\\\boldsymbol{I}&-i\boldsymbol{I}&j\boldsymbol{I}&-k\boldsymbol{I}\\\boldsymbol{I}&-i\boldsymbol{I}&-j\boldsymbol{I}&k\boldsymbol{I}\end{bmatrix}=\mathcal{D}_{\boldsymbol{\xi}}g \quad (56)$$

TABLE IV
MATRIX DERIVATIVES OF FUNCTIONS OF THE TYPE $f(\boldsymbol{q})$ and $\boldsymbol{f}(\boldsymbol{q})$

$f(\boldsymbol{q})$	$\mathcal{D}_{\boldsymbol{q}} f$	$\mathcal{D}_{\boldsymbol{q}^*} f$	Note
$\boldsymbol{a}^T\boldsymbol{q}\beta$	$\boldsymbol{a}^T\Re(\beta)$	$-\frac{1}{2}\boldsymbol{a}^T\beta^*$	$\forall\boldsymbol{a}\in\mathbb{H}^{N\times 1},\beta\in\mathbb{H}$
$\boldsymbol{a}^T\boldsymbol{q}^*\beta$	$-\frac{1}{2}\boldsymbol{a}^T\beta^*$	$\boldsymbol{a}^T\Re(\beta)$	$\forall\boldsymbol{a}\in\mathbb{H}^{N\times 1},\beta\in\mathbb{H}$
$\alpha\boldsymbol{q}^T\boldsymbol{b}$	$\alpha\Re(\boldsymbol{b}^T)$	$-\frac{1}{2}\alpha\boldsymbol{b}^H$	$\forall\boldsymbol{b}\in\mathbb{H}^{N\times 1},\alpha\in\mathbb{H}$
$\alpha\boldsymbol{q}^H\boldsymbol{b}$	$-\frac{1}{2}\alpha\boldsymbol{b}^H$	$\alpha\Re(\boldsymbol{b}^T)$	$\forall\boldsymbol{b}\in\mathbb{H}^{N\times 1},\alpha\in\mathbb{H}$
$\boldsymbol{a}^T\boldsymbol{q}\alpha\boldsymbol{q}^T\boldsymbol{b}$	$\boldsymbol{a}^T\Re(\alpha\boldsymbol{q}^T\boldsymbol{b})+\boldsymbol{a}^T\boldsymbol{q}\alpha\Re(\boldsymbol{b}^T)$	$-\frac{1}{2}\boldsymbol{a}^T(\alpha\boldsymbol{q}^T\boldsymbol{b})^*-\frac{1}{2}\boldsymbol{a}^T\boldsymbol{q}\alpha\boldsymbol{b}^H$	$\forall\boldsymbol{a},\boldsymbol{b}\in\mathbb{H}^{N\times 1},\alpha\in\mathbb{H}$
$\boldsymbol{a}^T\boldsymbol{q}\alpha\boldsymbol{q}^H\boldsymbol{b}$	$\boldsymbol{a}^T\Re(\alpha\boldsymbol{q}^H\boldsymbol{b})-\frac{1}{2}\boldsymbol{a}^T\boldsymbol{q}\alpha\boldsymbol{b}^H$	$-\frac{1}{2}\boldsymbol{a}^T(\alpha\boldsymbol{q}^H\boldsymbol{b})^*+\boldsymbol{a}^T\boldsymbol{q}\alpha\Re(\boldsymbol{b}^T)$	$\forall\boldsymbol{a},\boldsymbol{b}\in\mathbb{H}^{N\times 1},\alpha\in\mathbb{H}$
$\boldsymbol{a}^T\boldsymbol{q}^*\alpha\boldsymbol{q}^T\boldsymbol{b}$	$-\frac{1}{2}\boldsymbol{a}^T(\alpha\boldsymbol{q}^T\boldsymbol{b})^*+\boldsymbol{a}^T\boldsymbol{q}^*\alpha\Re(\boldsymbol{b}^T)$	$\boldsymbol{a}^T\Re(\alpha\boldsymbol{q}^T\boldsymbol{b})-\frac{1}{2}\boldsymbol{a}^T\boldsymbol{q}^*\alpha\boldsymbol{b}^H$	$\forall\boldsymbol{a},\boldsymbol{b}\in\mathbb{H}^{N\times 1},\alpha\in\mathbb{H}$
$\boldsymbol{a}^T\boldsymbol{q}^*\alpha\boldsymbol{q}^H\boldsymbol{b}$	$-\frac{1}{2}\boldsymbol{a}^T(\alpha\boldsymbol{q}^H\boldsymbol{b})^*-\frac{1}{2}\boldsymbol{a}^T\boldsymbol{q}^*\alpha\boldsymbol{b}^H$	$\boldsymbol{a}^T\Re(\alpha\boldsymbol{q}^H\boldsymbol{b})+\boldsymbol{a}^T\boldsymbol{q}^*\alpha\Re(\boldsymbol{b}^T)$	$\forall\boldsymbol{a},\boldsymbol{b}\in\mathbb{H}^{N\times 1},\alpha\in\mathbb{H}$
$\boldsymbol{q}^T\boldsymbol{A}\boldsymbol{q}$	$\boldsymbol{q}^T\boldsymbol{A}+\Re((\boldsymbol{A}\boldsymbol{q})^T)$	$-\frac{1}{2}\boldsymbol{q}^T\boldsymbol{A}-\frac{1}{2}(\boldsymbol{A}\boldsymbol{q})^H$	$\forall\boldsymbol{A}\in\mathbb{H}^{N\times N}$
$\boldsymbol{q}^H\boldsymbol{A}\boldsymbol{q}^*$	$-\frac{1}{2}\boldsymbol{q}^H\boldsymbol{A}-\frac{1}{2}(\boldsymbol{A}\boldsymbol{q}^*)^H$	$\boldsymbol{q}^H\boldsymbol{A}+\Re((\boldsymbol{A}\boldsymbol{q}^*)^T)$	$\forall\boldsymbol{A}\in\mathbb{H}^{N\times N}$
$\boldsymbol{q}^T\boldsymbol{A}\boldsymbol{q}^*$	$-\frac{1}{2}\boldsymbol{q}^T\boldsymbol{A}+\Re((\boldsymbol{A}\boldsymbol{q}^*)^T)$	$\boldsymbol{q}^T\boldsymbol{A}-\frac{1}{2}(\boldsymbol{A}\boldsymbol{q}^*)^H$	$\forall\boldsymbol{A}\in\mathbb{H}^{N\times N}$
$\boldsymbol{q}^H\boldsymbol{A}\boldsymbol{q}$	$\boldsymbol{q}^H\boldsymbol{A}-\frac{1}{2}(\boldsymbol{A}\boldsymbol{q})^H$	$-\frac{1}{2}\boldsymbol{q}^H\boldsymbol{A}+\Re((\boldsymbol{A}\boldsymbol{q})^T)$	$\forall\boldsymbol{A}\in\mathbb{H}^{N\times N}$
$\boldsymbol{A}\boldsymbol{q}\beta$	$\boldsymbol{A}\Re(\beta)$	$-\frac{1}{2}\boldsymbol{A}\beta^*$	$\forall\boldsymbol{A}\in\mathbb{H}^{N\times N},\beta\in\mathbb{H}$
$\boldsymbol{A}\boldsymbol{q}^*\beta$	$-\frac{1}{2}\boldsymbol{A}\beta^*$	$\boldsymbol{A}\Re(\beta)$	$\forall\boldsymbol{A}\in\mathbb{H}^{N\times N},\beta\in\mathbb{H}$
$\alpha\boldsymbol{q}^T\boldsymbol{A}$	$\alpha\Re(\boldsymbol{A}^T)$	$-\frac{1}{2}\alpha\boldsymbol{A}^H$	$\forall\boldsymbol{A}\in\mathbb{H}^{N\times N},\alpha\in\mathbb{H}$
$\alpha\boldsymbol{q}^H\boldsymbol{A}$	$-\frac{1}{2}\alpha\boldsymbol{A}^H$	$\alpha\Re(\boldsymbol{A}^T)$	$\forall\boldsymbol{A}\in\mathbb{H}^{N\times N},\alpha\in\mathbb{H}$

where $\mathcal{D}_{\boldsymbol{\zeta}} f=[\mathcal{D}_{\boldsymbol{Q}} f,\mathcal{D}_{\boldsymbol{Q}^i} f,\mathcal{D}_{\boldsymbol{Q}^j} f,\mathcal{D}_{\boldsymbol{Q}^k} f]$ and $\boldsymbol{J}$ is the $4NS\times 4NS$ matrix in (56), which satisfies

$$\boldsymbol{J}\boldsymbol{J}^H=\begin{bmatrix}\boldsymbol{I} & i\boldsymbol{I} & j\boldsymbol{I} & k\boldsymbol{I}\\ \boldsymbol{I} & i\boldsymbol{I} & -j\boldsymbol{I} & -k\boldsymbol{I}\\ \boldsymbol{I} & -i\boldsymbol{I} & j\boldsymbol{I} & -k\boldsymbol{I}\\ \boldsymbol{I} & -i\boldsymbol{I} & -j\boldsymbol{I} & k\boldsymbol{I}\end{bmatrix}\begin{bmatrix}\boldsymbol{I} & \boldsymbol{I} & \boldsymbol{I} & \boldsymbol{I}\\ -i\boldsymbol{I} & -i\boldsymbol{I} & i\boldsymbol{I} & i\boldsymbol{I}\\ -j\boldsymbol{I} & j\boldsymbol{I} & -j\boldsymbol{I} & j\boldsymbol{I}\\ -k\boldsymbol{I} & k\boldsymbol{I} & k\boldsymbol{I} & -k\boldsymbol{I}\end{bmatrix}=4\boldsymbol{I}_{4NS} \tag{57}$$

From (56) and (57), the equalities in (53) are equivalent. The other equivalent relations can be proved by Lemma 4.1. ■

B. Direction of Maximum Rate of Change

We next investigate how to find the maximum rate of change of $f(\boldsymbol{Q})\in\mathbb{R}$, a key condition in steepest descent methods, such as quaternion adaptive filters.

Theorem 4.2: Let $f:\mathbb{H}^{N\times S}\rightarrow\mathbb{R}$. Then, the gradient $[\mathcal{D}_{\boldsymbol{Q}^*} f(\boldsymbol{Q})]^T=[\mathcal{D}_{\boldsymbol{Q}} f(\boldsymbol{Q})]^H$ defines the direction of the maximum rate of change of f with respect to $\mathrm{vec}(\boldsymbol{Q}^*)$.

Proof: From (38), (52), we have

$$df=\sum_{\mu\in\{1,i,j,k\}}(\mathcal{D}_{\boldsymbol{Q}^\mu} f)d\mathrm{vec}(\boldsymbol{Q}^\mu)=\sum_{\mu\in\{1,i,j,k\}}(\mathcal{D}_{\boldsymbol{Q}} f)^\mu\,(d\mathrm{vec}(\boldsymbol{Q}))^\mu \tag{58}$$

Using (4) and (52), we further obtain

$$\begin{aligned}df&=\textstyle\sum_{\mu\in\{1,i,j,k\}}[(\mathcal{D}_{\boldsymbol{Q}} f)d\mathrm{vec}(\boldsymbol{Q})]^\mu\\&=4\Re\left[(\mathcal{D}_{\boldsymbol{Q}} f)d\mathrm{vec}(\boldsymbol{Q})\right]=4\Re\left[(\mathcal{D}_{\boldsymbol{Q}^*} f)^*d\mathrm{vec}(\boldsymbol{Q})\right]\\&=4\left\langle(\mathcal{D}_{\boldsymbol{Q}^*} f)^T,d\mathrm{vec}(\boldsymbol{Q})\right\rangle=4\left\langle(\mathcal{D}_{\boldsymbol{Q}} f)^H,d\mathrm{vec}(\boldsymbol{Q})\right\rangle\end{aligned} \tag{59}$$

where $<\cdot,\cdot>$ is the Euclidean inner product between real vectors in $\mathbb{R}^{4NS\times 1}$. By applying the Cauchy-Schwartz inequality to (59), we obtain

$$|df|=4\left|\left\langle(\mathcal{D}_{\boldsymbol{Q}^*} f)^T,d\mathrm{vec}(\boldsymbol{Q})\right\rangle\right|\leq 4\left\|\mathcal{D}_{\boldsymbol{Q}} f\right\|\left\|d\mathrm{vec}(\boldsymbol{Q})\right\| \tag{60}$$

which indicates that the maximum change of f occurs when $d\mathrm{vec}(\boldsymbol{Q})$ is in the direction of $(\mathcal{D}_{\boldsymbol{Q}^*} f)^T=(\mathcal{D}_{\boldsymbol{Q}} f)^H$. Thus, the steepest descent method can be expressed as

$$\mathrm{vec}(\boldsymbol{Q}_{n+1})=\mathrm{vec}(\boldsymbol{Q}_n)-\eta(\mathcal{D}_{\boldsymbol{Q}} f(\boldsymbol{Q}_n))^H,\quad f\in\mathbb{R} \tag{61}$$

where $\eta>0$ is the step size, and $\boldsymbol{Q}_{n+1}$ is the value of the unknown matrix after n iterations. ■

Corollary 4.1: Let $f:\mathbb{H}^{N\times 1}\rightarrow\mathbb{R}$. Then the gradient $\nabla_{\boldsymbol{q}^*} f=\left(\frac{\partial f}{\partial\boldsymbol{q}^*}\right)^T=\left(\frac{\partial f}{\partial\boldsymbol{q}}\right)^H$ defines the direction of the maximum rate of change of f with respect to $\boldsymbol{q}^*$.

Proof: The proof follows directly from Theorem 4.2. ■

V. ENABLING OF QUATERNION DERIVATIVES IN SIGNAL PROCESSING APPLICATIONS

A. Derivatives of $f(\boldsymbol{q})$

Let $f:\mathbb{H}^{N\times 1}\rightarrow\mathbb{H}$ be $f(\boldsymbol{q})=\boldsymbol{q}^H\boldsymbol{A}\boldsymbol{q}$. This kind of function frequently appears in quaternion filter optimization [7]–[10] and array signal processing [10]. For example, the optimization of the output power $\boldsymbol{q}^H\boldsymbol{A}\boldsymbol{q}$, where $\boldsymbol{q}$ is the vector of filter coefficients and $\boldsymbol{A}$ is the input covariance matrix. i.e., $\boldsymbol{A}^H=\boldsymbol{A}$. Using the product rule in Theorem 3.1, we have $\mathcal{D}_{\boldsymbol{q}^*} f=\boldsymbol{q}^H\mathcal{D}_{\boldsymbol{q}^*}(\boldsymbol{A}\boldsymbol{q})+\mathcal{D}_{\boldsymbol{q}^*}(\boldsymbol{q}^H\boldsymbol{A}\boldsymbol{q})|_{\boldsymbol{A}\boldsymbol{q}=const}=-\frac{1}{2}\boldsymbol{q}^H\boldsymbol{A}+\Re(\boldsymbol{A}\boldsymbol{q})^T=\frac{1}{2}(\boldsymbol{q}^H\boldsymbol{A})^*$. Some results for such functions are shown in Table IV, assuming $\boldsymbol{a}\in\mathbb{H}^{N\times 1}$, $\boldsymbol{A}\in\mathbb{H}^{N\times N}$ to be constant, and $\boldsymbol{q}\in\mathbb{H}^{N\times 1}$ to be vector variable.

1) Quaternion Least Mean Square: This section derives the quaternion least mean square (QLMS) adaptive filtering algorithm [7], [29], [56] using the left/right GHR derivatives. The cost function to be minimized is a real-valued function

$$J(n) = |e(n)|^2 = e^*(n)e(n) \tag{62}$$

where

$$e(n) = d(n) - y(n), \quad y(n) = \boldsymbol{w}^T(n)\boldsymbol{x}(n) \tag{63}$$

and $d(n), y(n) \in \mathbb{H}, \boldsymbol{w}(n), \boldsymbol{x}(n) \in \mathbb{H}^{N\times 1}$. To illustrate the versatility of GHR calculus, the QLMS algorithm shall be derived using the left/right GHR derivatives. The weight update of QLMS based on the left GHR derivatives is given by

$$\boldsymbol{w}(n+1) = \boldsymbol{w}(n) - \eta\, \left(\mathcal{D}_{\boldsymbol{w}} J(n)\right)^H \tag{64}$$

where $\eta > 0$ is the step size and $(\mathcal{D}_{\boldsymbol{w}} J(n))^H$ denotes the left gradient of $J(n)$ with respect to $\boldsymbol{w}^*$. Using the results in Table IV, the left gradient in (64) can be calculated as

$$\begin{aligned}
\mathcal{D}_{\boldsymbol{w}} J &= \mathcal{D}_{\boldsymbol{w}} \left((d - \boldsymbol{w}^T\boldsymbol{x})^* (d - \boldsymbol{w}^T\boldsymbol{x}) \right) \\
&= \mathcal{D}_{\boldsymbol{w}}(d^*d) - \mathcal{D}_{\boldsymbol{w}}(d^*\boldsymbol{w}^T\boldsymbol{x}) - \mathcal{D}_{\boldsymbol{w}}(\boldsymbol{x}^H\boldsymbol{w}^*d) \\
&\quad + \mathcal{D}_{\boldsymbol{w}}(\boldsymbol{x}^H\boldsymbol{w}^*\boldsymbol{w}^T\boldsymbol{x}) \\
&= -d^*\Re(\boldsymbol{x}^T) + \frac{1}{2}\boldsymbol{x}^H d^* - \frac{1}{2}\boldsymbol{x}^H(\boldsymbol{w}^T\boldsymbol{x})^* + \boldsymbol{x}^H\boldsymbol{w}^*\Re(\boldsymbol{x}^T) \\
&= -\frac{1}{2}\boldsymbol{x}^T d^* + \frac{1}{2}\boldsymbol{x}^T(\boldsymbol{w}^T\boldsymbol{x})^* \\
&= -\frac{1}{2}\boldsymbol{x}^T(d - \boldsymbol{w}^T\boldsymbol{x})^* = -\frac{1}{2}\boldsymbol{x}^T e^*
\end{aligned} \tag{65}$$

where time index 'n' is omitted for convenience. Then, the update rule for QLMS becomes

$$\boldsymbol{w}(n+1) = \boldsymbol{w}(n) + \eta\, e(n)\boldsymbol{x}^*(n) \tag{66}$$

where the constant $\frac{1}{2}$ in (65) is absorbed into η.

Based upon the right GHR derivatives, the QLMS update becomes

$$\boldsymbol{w}(n+1) = \boldsymbol{w}(n) - \eta\, \left(\mathcal{D}^R_{\boldsymbol{w}} J(n)\right)^H \tag{67}$$

where $\eta > 0$ is the step size, and the right gradient of $J(n)$ with respect to $\boldsymbol{w}^*$, $(\mathcal{D}^R_{\boldsymbol{w}} J(n))^H$, can be calculated by

$$\begin{aligned}
\mathcal{D}^R_{\boldsymbol{w}} J &= \mathcal{D}^R_{\boldsymbol{w}} \left((d - \boldsymbol{w}^T\boldsymbol{x})^* (d - \boldsymbol{w}^T\boldsymbol{x}) \right) \\
&= \mathcal{D}^R_{\boldsymbol{w}}(d^*d) - \mathcal{D}^R_{\boldsymbol{w}}(d^*\boldsymbol{w}^T\boldsymbol{x}) - \mathcal{D}^R_{\boldsymbol{w}}(\boldsymbol{x}^H\boldsymbol{w}^*d) \\
&\quad + \mathcal{D}^R_{\boldsymbol{w}}\left(\boldsymbol{x}^H\boldsymbol{w}^*\boldsymbol{w}^T\boldsymbol{x}\right) \\
&= -\Re(d^*)\boldsymbol{x}^T + \frac{1}{2}\boldsymbol{x}^T d - \frac{1}{2}\boldsymbol{x}^T(\boldsymbol{w}^T\boldsymbol{x}) + \Re(\boldsymbol{x}^H\boldsymbol{w}^*)\boldsymbol{x}^T \\
&= -\frac{1}{2}\boldsymbol{x}^T d^* + \frac{1}{2}\boldsymbol{x}^T(\boldsymbol{x}^H\boldsymbol{w}^*) \\
&= -\frac{1}{2}\boldsymbol{x}^T(d - \boldsymbol{w}^T\boldsymbol{x})^* = -\frac{1}{2}\boldsymbol{x}^T e^*
\end{aligned} \tag{68}$$

From (67) and (68), the update rule of QLMS becomes

$$\boldsymbol{w}(n+1) = \boldsymbol{w}(n) + \eta\, e(n)\boldsymbol{x}^*(n) \tag{69}$$

Remark 5.1: From (66) and (69), notice that the QLMS derived using the left GHR derivatives is exactly the same as that using the right GHR derivatives, a consequence of the cost function $J(n)$ in (62) being real-valued and Lemma 2.2. In addition, if we start from $e(n) = d(n) - \mathbf{w}^H(n)\mathbf{x}(n)$ in (63), the final update rule of QLMS would become $\mathbf{w}(n+1) = \mathbf{w}(n) + \eta\, \mathbf{x}(n)e^*(n)$. The QLMS algorithm in (66) is therefore a generic extension of complex LMS [57] to the case of quaternion vector.

Remark 5.2: Comparing (66) with the QLMS given in [51], we can see that the GHR-QLMS is essentially the same as the QLMS derived using the pseudo-gradient in [51], however, the use of pseudo-gradient is cumbersome and tedious [7], [51]. Notice that the GHR-QLMS is different and more compact than the original QLMS [7] based on pseudo-gradient, the HR-QLMS [29] based on the HR-gradient, and the I-QLMS [56] based on the I-gradient. The difference from the original QLMS arises due to the rigorous use of the non-commutativity of quaternion product within the GHR calculus in (65). The difference from the HR-QLMS and I-QLMS is also due to the rigorous use of the novel product rules (40)-(41) within the GHR calculus in (65).

2) Quaternion Affine Projection Algorithm: This section rederives the quaternion affine projection algorithm (QAPA) [32] based on the GHR calculus. The aim of QAPA is to minimize adaptively the squared Euclidean norm of the change in the weight vector $\boldsymbol{w}(n) \in \mathbb{H}^{N\times 1}$, that is

$$\begin{aligned}
&\text{minimise} \quad \|\Delta\boldsymbol{w}(n)\|^2 = \|\boldsymbol{w}(n+1) - \boldsymbol{w}(n)\|^2 \\
&\text{subject to } \boldsymbol{d}^T(n) = \boldsymbol{w}^H(n+1)\boldsymbol{Q}(n)
\end{aligned} \tag{70}$$

where $\boldsymbol{d}(n) = [d(n), \ldots, d(n-S+1)]^T \in \mathbb{H}^{S\times 1}$ denotes the desired signal vector and $\boldsymbol{Q}(n) = [\boldsymbol{q}(n), \ldots, \boldsymbol{q}(n-S+1)] \in \mathbb{H}^{N\times S}$ denotes the matrix of S past input vectors. Using the Lagrange multipliers, the constrained optimisation problem (70) can be solved by the following cost function

$$\begin{aligned}
J(n) &= \|\boldsymbol{w}(n+1) - \boldsymbol{w}(n)\|^2 \\
&\quad + \Re\left\{ \left(\boldsymbol{d}^T(n) - \boldsymbol{w}^H(n+1)\boldsymbol{Q}(n)\right)\boldsymbol{\lambda}^* \right\} \\
&= (\boldsymbol{w}(n+1) - \boldsymbol{w}(n))^H (\boldsymbol{w}(n+1) - \boldsymbol{w}(n)) \\
&\quad + \frac{1}{2}\left(\boldsymbol{d}^T(n) - \boldsymbol{w}^H(n+1)\boldsymbol{Q}(n)\right)\boldsymbol{\lambda}^* \\
&\quad + \frac{1}{2}\boldsymbol{\lambda}^T\left(\boldsymbol{d}^*(n) - \boldsymbol{Q}^H(n)\boldsymbol{w}(n+1)\right)
\end{aligned} \tag{71}$$

where $\boldsymbol{\lambda} \in \mathbb{H}^{S\times 1}$ denotes the Lagrange multipliers. Using the results in Table IV, we have

$$\begin{aligned}
\mathcal{D}_{\boldsymbol{w}(n+1)} J(n) &= \frac{1}{2}(\boldsymbol{w}(n+1) - \boldsymbol{w}(n))^H \\
&\quad + \frac{1}{2}\left(\frac{1}{2}(\boldsymbol{Q}(n)\boldsymbol{\lambda}^*)^H - \boldsymbol{\lambda}^T\boldsymbol{Q}^H(n)\right) \\
&= \frac{1}{2}(\boldsymbol{w}(n+1) - \boldsymbol{w}(n))^H - \frac{1}{4}\boldsymbol{\lambda}^T\boldsymbol{Q}^H(n)
\end{aligned} \tag{72}$$

Setting (72) to zero, the weight update of QAPA can be obtained as

$$\boldsymbol{w}(k+1) - \boldsymbol{w}(k) = \frac{1}{2}\boldsymbol{Q}(n)\,\boldsymbol{\lambda}^* \tag{73}$$

Using the fact that $\boldsymbol{e}^T(n) = \boldsymbol{d}^T(n) - \boldsymbol{y}^T(n) = (\boldsymbol{w}^H(n+1) - \boldsymbol{w}^H(n))\boldsymbol{Q}(n)$, and based on (73), $\boldsymbol{\lambda}$ can be solved as

$$\boldsymbol{\lambda}^T = 2\boldsymbol{e}^T(n)\left(\boldsymbol{Q}^H(n)\boldsymbol{Q}(n)\right)^{-1} \tag{74}$$

which gives

$$\boldsymbol{w}(n+1) = \boldsymbol{w}(n) + \boldsymbol{Q}(n)\left(\boldsymbol{Q}^H(n)\boldsymbol{Q}(n)\right)^{-1}\boldsymbol{e}^*(n) \tag{75}$$

To prevent the normalization matrix $\boldsymbol{Q}^H(n)\boldsymbol{Q}(n)$ within (75) from becoming singular, a small regularization term $\varepsilon\boldsymbol{I} \in \mathbb{H}^{S\times S}$

TABLE V
MATRIX DERIVATIVES OF FUNCTIONS OF THE TYPE $f(\boldsymbol{Q})$

$f(\boldsymbol{Q})$	$\frac{\partial f}{\partial \boldsymbol{Q}}$	$\frac{\partial f}{\partial \boldsymbol{Q}^*}$
$\text{Tr}(\boldsymbol{Q})$	$\boldsymbol{I}_N$	$-\frac{1}{2}\boldsymbol{I}_N$
$\text{Tr}(\boldsymbol{Q}^H)$	$-\frac{1}{2}\boldsymbol{I}_N$	$\boldsymbol{I}_N$
$\text{Tr}(\boldsymbol{A}\boldsymbol{Q})$	$\boldsymbol{A}^T$	$-\frac{1}{2}\boldsymbol{A}^T$
$\text{Tr}(\boldsymbol{A}\boldsymbol{Q}^H)$	$-\frac{1}{2}\boldsymbol{A}$	$\boldsymbol{A}$
$\text{Tr}(\boldsymbol{Q}\boldsymbol{A})$	$\mathfrak{R}(\boldsymbol{A}^T)$	$-\frac{1}{2}\boldsymbol{A}^H$
$\text{Tr}(\boldsymbol{Q}^H\boldsymbol{A})$	$-\frac{1}{2}\boldsymbol{A}^*$	$\mathfrak{R}(\boldsymbol{A})$
$\text{Tr}(\boldsymbol{A}_1\boldsymbol{Q}\boldsymbol{A}_2)$	$\boldsymbol{A}_1^T\mathfrak{R}(\boldsymbol{A}_2^T)$	$-\frac{1}{2}\boldsymbol{A}_1^T\boldsymbol{A}_2^H$
$\text{Tr}(\boldsymbol{A}_1\boldsymbol{Q}^*\boldsymbol{A}_2)$	$-\frac{1}{2}\boldsymbol{A}_1^T\boldsymbol{A}_2^H$	$\boldsymbol{A}_1^T\mathfrak{R}(\boldsymbol{A}_2^T)$
$\text{Tr}(\boldsymbol{A}_1\boldsymbol{Q}^T\boldsymbol{A}_2)$	$\mathfrak{R}(\boldsymbol{A}_2)\boldsymbol{A}_1$	$-\frac{1}{2}(\boldsymbol{A}_1^T\boldsymbol{A}_2^H)^T$
$\text{Tr}(\boldsymbol{A}_1\boldsymbol{Q}^H\boldsymbol{A}_2)$	$-\frac{1}{2}(\boldsymbol{A}_1^T\boldsymbol{A}_2^H)^T$	$\mathfrak{R}(\boldsymbol{A}_2)\boldsymbol{A}_1$
$\text{Tr}(\boldsymbol{Q}^n)$	$\sum_{m=1}^{n}(\boldsymbol{Q}^T)^{n-m}\mathfrak{R}(\boldsymbol{Q}^{m-1})^T$	$-\frac{1}{2}\sum_{m=1}^{n}(\boldsymbol{Q}^T)^{n-m}(\boldsymbol{Q}^{m-1})^H$
$\text{Tr}(\boldsymbol{Q}^{-1})$	$-(\boldsymbol{Q}^T)^{-1}\mathfrak{R}(\boldsymbol{Q}^{-1})^T$	$\frac{1}{2}(\boldsymbol{Q}^T)^{-1}(\boldsymbol{Q}^{-1})^H$
$\text{Tr}(\boldsymbol{A}_1\boldsymbol{Q}\boldsymbol{A}_2\boldsymbol{Q}\boldsymbol{A}_3)$	$\boldsymbol{A}_1^T\mathfrak{R}(\boldsymbol{A}_2\boldsymbol{Q}\boldsymbol{A}_3)^T+(\boldsymbol{A}_1\boldsymbol{Q}\boldsymbol{A}_2)^T\mathfrak{R}(\boldsymbol{A}_3^T)$	$-\frac{1}{2}\boldsymbol{A}_1^T(\boldsymbol{A}_2\boldsymbol{Q}\boldsymbol{A}_3)^H-\frac{1}{2}(\boldsymbol{A}_1\boldsymbol{Q}\boldsymbol{A}_2)^T\boldsymbol{A}_3^H$
$\text{Tr}(\boldsymbol{A}_1\boldsymbol{Q}\boldsymbol{A}_2\boldsymbol{Q}^*\boldsymbol{A}_3)$	$\boldsymbol{A}_1^T\mathfrak{R}(\boldsymbol{A}_2\boldsymbol{Q}^*\boldsymbol{A}_3)^T-\frac{1}{2}(\boldsymbol{A}_1\boldsymbol{Q}\boldsymbol{A}_2)^T\boldsymbol{A}_3^H$	$-\frac{1}{2}\boldsymbol{A}_1^T(\boldsymbol{A}_2\boldsymbol{Q}^*\boldsymbol{A}_3)^H+(\boldsymbol{A}_1\boldsymbol{Q}\boldsymbol{A}_2)^T\mathfrak{R}(\boldsymbol{A}_3^T)$
$\text{Tr}(\boldsymbol{A}_1\boldsymbol{Q}\boldsymbol{A}_2\boldsymbol{Q}^T\boldsymbol{A}_3)$	$\boldsymbol{A}_1^T\mathfrak{R}(\boldsymbol{A}_2\boldsymbol{Q}^T\boldsymbol{A}_3)^T+\mathfrak{R}(\boldsymbol{A}_3)\boldsymbol{A}_1\boldsymbol{Q}\boldsymbol{A}_2$	$-\frac{1}{2}\boldsymbol{A}_1^T(\boldsymbol{A}_2\boldsymbol{Q}^T\boldsymbol{A}_3)^H-\frac{1}{2}((\boldsymbol{A}_1\boldsymbol{Q}\boldsymbol{A}_2)^T\boldsymbol{A}_3^H)^T$
$\text{Tr}(\boldsymbol{A}_1\boldsymbol{Q}\boldsymbol{A}_2\boldsymbol{Q}^H\boldsymbol{A}_3)$	$\boldsymbol{A}_1^T\mathfrak{R}(\boldsymbol{A}_2\boldsymbol{Q}^H\boldsymbol{A}_3)^T-\frac{1}{2}((\boldsymbol{A}_1\boldsymbol{Q}\boldsymbol{A}_2)^T\boldsymbol{A}_3^H)^T$	$-\frac{1}{2}\boldsymbol{A}_1^T(\boldsymbol{A}_2\boldsymbol{Q}^H\boldsymbol{A}_3)^H+\mathfrak{R}(\boldsymbol{A}_3)\boldsymbol{A}_1\boldsymbol{Q}\boldsymbol{A}_2$
$\text{Tr}(\boldsymbol{A}_1\boldsymbol{Q}^H\boldsymbol{A}_2\boldsymbol{Q}\boldsymbol{A}_3)$	$-\frac{1}{2}(\boldsymbol{A}_1^T(\boldsymbol{A}_2\boldsymbol{Q}\boldsymbol{A}_3)^H)^T+(\boldsymbol{A}_1\boldsymbol{Q}^H\boldsymbol{A}_2)^T\mathfrak{R}(\boldsymbol{A}_3^T)$	$\mathfrak{R}(\boldsymbol{A}_2\boldsymbol{Q}\boldsymbol{A}_3)\boldsymbol{A}_1-\frac{1}{2}(\boldsymbol{A}_1\boldsymbol{Q}^H\boldsymbol{A}_2)^T\boldsymbol{A}_3^H$
$\text{Tr}(\boldsymbol{A}_1\boldsymbol{Q}^H\boldsymbol{A}_2\boldsymbol{Q}^*\boldsymbol{A}_3)$	$-\frac{1}{2}(\boldsymbol{A}_1^T(\boldsymbol{A}_2\boldsymbol{Q}^*\boldsymbol{A}_3)^H)^T-\frac{1}{2}(\boldsymbol{A}_1\boldsymbol{Q}^H\boldsymbol{A}_2)^T\boldsymbol{A}_3^H$	$\mathfrak{R}(\boldsymbol{A}_2\boldsymbol{Q}^*\boldsymbol{A}_3)\boldsymbol{A}_1+(\boldsymbol{A}_1\boldsymbol{Q}^H\boldsymbol{A}_2)^T\mathfrak{R}(\boldsymbol{A}_3^T)$
$\text{Tr}(\boldsymbol{A}_1\boldsymbol{Q}^H\boldsymbol{A}_2\boldsymbol{Q}^T\boldsymbol{A}_3)$	$-\frac{1}{2}(\boldsymbol{A}_1^T(\boldsymbol{A}_2\boldsymbol{Q}^T\boldsymbol{A}_3)^H)^T+\mathfrak{R}(\boldsymbol{A}_3)\boldsymbol{A}_1\boldsymbol{Q}^H\boldsymbol{A}_2$	$\mathfrak{R}(\boldsymbol{A}_2\boldsymbol{Q}^T\boldsymbol{A}_3)\boldsymbol{A}_1-\frac{1}{2}((\boldsymbol{A}_1\boldsymbol{Q}^H\boldsymbol{A}_2)^T\boldsymbol{A}_3^H)^T$
$\text{Tr}(\boldsymbol{A}_1\boldsymbol{Q}^H\boldsymbol{A}_2\boldsymbol{Q}^H\boldsymbol{A}_3)$	$-\frac{1}{2}(\boldsymbol{A}_1^T(\boldsymbol{A}_2\boldsymbol{Q}^H\boldsymbol{A}_3)^H)^T-\frac{1}{2}((\boldsymbol{A}_1\boldsymbol{Q}^H\boldsymbol{A}_2)^T\boldsymbol{A}_3^H)^T$	$\mathfrak{R}(\boldsymbol{A}_2\boldsymbol{Q}^H\boldsymbol{A}_3)\boldsymbol{A}_1+\mathfrak{R}(\boldsymbol{A}_3)\boldsymbol{A}_1\boldsymbol{Q}^H\boldsymbol{A}_2$

is usually added, where $\boldsymbol{I}$ is the identity matrix, while a step size η controls the convergence and the steady state performance. The final weight update of QAPA becomes

$$\boldsymbol{w}(n+1)=\boldsymbol{w}(n)+\eta\,\boldsymbol{Q}(n)\left(\boldsymbol{Q}^H(n)\boldsymbol{Q}(n)+\varepsilon\boldsymbol{I}\right)^{-1}\boldsymbol{e}^*(n) \quad (76)$$

and is a generic extension from the real and complex case.

B. Derivatives of $f(\boldsymbol{Q})$

For scalar functions $f:\mathbb{H}^{N\times S}\rightarrow\mathbb{H}$, it is common to define the following matrix derivatives

$$\frac{\partial f}{\partial \boldsymbol{Q}^\mu}\triangleq\begin{bmatrix}\frac{\partial f}{\partial q_{11}^\mu} & \cdots & \frac{\partial f}{\partial q_{1S}^\mu}\\ \vdots & \ddots & \vdots\\ \frac{\partial f}{\partial q_{N1}^\mu} & \cdots & \frac{\partial f}{\partial q_{NS}^\mu}\end{bmatrix},\ \frac{\partial f}{\partial \boldsymbol{Q}^{\mu*}}\triangleq\begin{bmatrix}\frac{\partial f}{\partial q_{11}^{\mu*}} & \cdots & \frac{\partial f}{\partial q_{1S}^{\mu*}}\\ \vdots & \ddots & \vdots\\ \frac{\partial f}{\partial q_{N1}^{\mu*}} & \cdots & \frac{\partial f}{\partial q_{NS}^{\mu*}}\end{bmatrix} \quad (77)$$

which are referred to as the gradient of f with respect to $\boldsymbol{Q}^\mu$ and $\boldsymbol{Q}^{\mu*}$. Equations in (77) are generalizations of the real- and complex-valued cases given in [42], [46] to the quaternion case. A comparison of (37) and (77), gives the connection

$$\mathcal{D}_{\boldsymbol{Q}^\mu}f=\text{vec}^T\left(\frac{\partial f}{\partial \boldsymbol{Q}^\mu}\right),\quad \mathcal{D}_{\boldsymbol{Q}^{\mu*}}f=\text{vec}^T\left(\frac{\partial f}{\partial \boldsymbol{Q}^{\mu*}}\right) \quad (78)$$

Then, the steepest descent method (61) can be reformulated as

$$\boldsymbol{Q}_{n+1}=\boldsymbol{Q}_n-\eta\frac{\partial f}{\partial \boldsymbol{Q}^*},\quad f\in\mathbb{R} \quad (79)$$

where $\eta>0$ is the step size. Some important results of functions of the type $f(\boldsymbol{Q})$ are summarized in Table V, where $\boldsymbol{Q}\in\mathbb{H}^{N\times S}$ or possibly $\boldsymbol{Q}\in\mathbb{H}^{N\times N}$, and $\boldsymbol{A}_1,\boldsymbol{A}_2,\boldsymbol{A}_3,\boldsymbol{A}$ are chosen such that the functions are well defined.

1) Quaternion Matrix Least Squares: Given $\boldsymbol{A}\in\mathbb{H}^{R\times N}$, $\boldsymbol{B}\in\mathbb{H}^{S\times P}$ and $\boldsymbol{C}\in\mathbb{H}^{R\times P}$, the task is to find $\boldsymbol{Q}\in\mathbb{H}^{N\times S}$ such that the error of the overdetermined linear system of equations

$$F(\boldsymbol{Q})=\text{Tr}\left\{(\boldsymbol{C}-\boldsymbol{A}\boldsymbol{Q}\boldsymbol{B})^H(\boldsymbol{C}-\boldsymbol{A}\boldsymbol{Q}\boldsymbol{B})\right\} \quad (80)$$

is minimized. Using the results in Table V, the gradient of $F(\boldsymbol{Q})$ can be calculated as

$$\begin{aligned}\frac{\partial F(\boldsymbol{Q})}{\partial \boldsymbol{Q}^*}&=\frac{\partial\text{Tr}\left\{\boldsymbol{C}^H\boldsymbol{C}-\boldsymbol{C}^H\boldsymbol{A}\boldsymbol{Q}\boldsymbol{B}-\boldsymbol{B}^H\boldsymbol{Q}^H\boldsymbol{A}^H\boldsymbol{C}\right\}}{\partial \boldsymbol{Q}^*}\\ &\quad+\frac{\partial\text{Tr}\left\{\boldsymbol{B}^H\boldsymbol{Q}^H\boldsymbol{A}^H\boldsymbol{A}\boldsymbol{Q}\boldsymbol{B}\right\}}{\partial \boldsymbol{Q}^*}\\ &=\frac{1}{2}(\boldsymbol{C}^H\boldsymbol{A})^T\boldsymbol{B}^H-\mathfrak{R}(\boldsymbol{A}^H\boldsymbol{C})\boldsymbol{B}^H\\ &\quad+\mathfrak{R}(\boldsymbol{A}^H\boldsymbol{A}\boldsymbol{Q}\boldsymbol{B})\boldsymbol{B}^H-\frac{1}{2}(\boldsymbol{B}^H\boldsymbol{Q}^H\boldsymbol{A}^H\boldsymbol{A})^T\boldsymbol{B}^H\\ &=-\frac{1}{2}(\boldsymbol{A}^H\boldsymbol{C})\boldsymbol{B}^H+\frac{1}{2}(\boldsymbol{A}^H\boldsymbol{A}\boldsymbol{Q}\boldsymbol{B})\boldsymbol{B}^H\\ &=-\frac{1}{2}\boldsymbol{A}^H(\boldsymbol{C}-\boldsymbol{A}\boldsymbol{Q}\boldsymbol{B})\boldsymbol{B}^H\end{aligned} \quad (81)$$

Setting (81) to zero, we obtain a normal equation

$$\boldsymbol{A}^H\boldsymbol{A}\boldsymbol{Q}\boldsymbol{B}\boldsymbol{B}^H=\boldsymbol{A}^H\boldsymbol{C}\boldsymbol{B}^H \quad (82)$$

TABLE VI
MATRIX DERIVATIVES OF FUNCTIONS OF THE TYPE $\boldsymbol{F}(\boldsymbol{Q})$

$\boldsymbol{F}(\boldsymbol{Q})$	$\mathcal{D}_{\boldsymbol{Q}}\boldsymbol{F}$	$\mathcal{D}_{\boldsymbol{Q}^*}\boldsymbol{F}$
$\boldsymbol{Q}$	$\boldsymbol{I}_{NS}$	$-\frac{1}{2}\boldsymbol{I}_{NS}$
$\boldsymbol{Q}^H$	$-\frac{1}{2}\boldsymbol{K}_{N,S}$	$\boldsymbol{K}_{N,S}$
$\boldsymbol{AQ}$	$\boldsymbol{I}_S \otimes \boldsymbol{A}$	$-\frac{1}{2}\boldsymbol{I}_S \otimes \boldsymbol{A}$
$\boldsymbol{AQ}^*$	$-\frac{1}{2}\boldsymbol{I}_S \otimes \boldsymbol{A}$	$\boldsymbol{I}_S \otimes \boldsymbol{A}$
$\boldsymbol{AQ}^T$	$(\boldsymbol{I}_N \otimes \boldsymbol{A})\boldsymbol{K}_{N,S}$	$-\frac{1}{2}(\boldsymbol{I}_N \otimes \boldsymbol{A})\boldsymbol{K}_{N,S}$
$\boldsymbol{AQ}^H$	$-\frac{1}{2}(\boldsymbol{I}_N \otimes \boldsymbol{A})\boldsymbol{K}_{N,S}$	$(\boldsymbol{I}_N \otimes \boldsymbol{A})\boldsymbol{K}_{N,S}$
$\boldsymbol{QA}$	$\Re(\boldsymbol{A}^T) \otimes \boldsymbol{I}_N$	$-\frac{1}{2}\boldsymbol{A}^H \otimes \boldsymbol{I}_N$
$\boldsymbol{Q}^*\boldsymbol{A}$	$-\frac{1}{2}\boldsymbol{A}^H \otimes \boldsymbol{I}_N$	$\Re(\boldsymbol{A}^T) \otimes \boldsymbol{I}_N$
$\boldsymbol{Q}^T\boldsymbol{A}$	$(\Re(\boldsymbol{A}^T) \otimes \boldsymbol{I}_S)\boldsymbol{K}_{N,S}$	$-\frac{1}{2}(\boldsymbol{A}^H \otimes \boldsymbol{I}_S)\boldsymbol{K}_{N,S}$
$\boldsymbol{Q}^H\boldsymbol{A}$	$-\frac{1}{2}(\boldsymbol{A}^H \otimes \boldsymbol{I}_S)\boldsymbol{K}_{N,S}$	$(\Re(\boldsymbol{A}^T) \otimes \boldsymbol{I}_S)\boldsymbol{K}_{N,S}$
$\boldsymbol{A}_1\boldsymbol{Q}\boldsymbol{A}_2$	$(\boldsymbol{I}_P \otimes \boldsymbol{A}_1)(\Re(\boldsymbol{A}_2^T) \otimes \boldsymbol{I}_N)$	$-\frac{1}{2}(\boldsymbol{I}_P \otimes \boldsymbol{A}_1)(\boldsymbol{A}_2^H \otimes \boldsymbol{I}_N)$
$\boldsymbol{A}_1\boldsymbol{Q}^*\boldsymbol{A}_2$	$-\frac{1}{2}(\boldsymbol{I}_P \otimes \boldsymbol{A}_1)(\boldsymbol{A}_2^H \otimes \boldsymbol{I}_N)$	$(\boldsymbol{I}_P \otimes \boldsymbol{A}_1)(\Re(\boldsymbol{A}_2^T) \otimes \boldsymbol{I}_N)$
$\boldsymbol{A}_1\boldsymbol{Q}^T\boldsymbol{A}_2$	$(\boldsymbol{I}_P \otimes \boldsymbol{A}_1)(\Re(\boldsymbol{A}_2^T) \otimes \boldsymbol{I}_S)\boldsymbol{K}_{N,S}$	$-\frac{1}{2}(\boldsymbol{I}_P \otimes \boldsymbol{A}_1)(\boldsymbol{A}_2^H \otimes \boldsymbol{I}_S)\boldsymbol{K}_{N,S}$
$\boldsymbol{A}_1\boldsymbol{Q}^H\boldsymbol{A}_2$	$-\frac{1}{2}(\boldsymbol{I}_P \otimes \boldsymbol{A}_1)(\boldsymbol{A}_2^H \otimes \boldsymbol{I}_S)\boldsymbol{K}_{N,S}$	$(\boldsymbol{I}_P \otimes \boldsymbol{A}_1)(\Re(\boldsymbol{A}_2^T) \otimes \boldsymbol{I}_S)\boldsymbol{K}_{N,S}$
$\boldsymbol{Q}^n$	$\sum_{m=1}^{n}(\boldsymbol{I}_N \otimes \boldsymbol{Q})^{n-m}(\Re(\boldsymbol{Q}^{m-1})^T \otimes \boldsymbol{I}_N)$	$-\frac{1}{2}\sum_{m=1}^{n}(\boldsymbol{I}_N \otimes \boldsymbol{Q})^{n-m}((\boldsymbol{Q}^{m-1})^H \otimes \boldsymbol{I}_N)$
$\boldsymbol{Q}^{-1}$	$-(\boldsymbol{I}_N \otimes \boldsymbol{Q})^{-1}(\Re(\boldsymbol{Q}^{-1})^T \otimes \boldsymbol{I}_N)$	$\frac{1}{2}(\boldsymbol{I}_N \otimes \boldsymbol{Q})^{-1}((\boldsymbol{Q}^{-1})^H \otimes \boldsymbol{I}_N)$
$\boldsymbol{QAQ}^T$	$\Re(\boldsymbol{AQ}^T)^T \otimes \boldsymbol{I}_N + (\boldsymbol{I}_N \otimes (\boldsymbol{QA}))\boldsymbol{K}_{N,S}$	$-\frac{1}{2}(\boldsymbol{AQ}^T)^H \otimes \boldsymbol{I}_N - \frac{1}{2}(\boldsymbol{I}_N \otimes (\boldsymbol{QA}))\boldsymbol{K}_{N,S}$
$\boldsymbol{QAQ}^H$	$\Re(\boldsymbol{AQ}^H)^T \otimes \boldsymbol{I}_N - \frac{1}{2}(\boldsymbol{I}_N \otimes (\boldsymbol{QA}))\boldsymbol{K}_{N,S}$	$-\frac{1}{2}(\boldsymbol{AQ}^H)^H \otimes \boldsymbol{I}_N + (\boldsymbol{I}_N \otimes (\boldsymbol{QA}))\boldsymbol{K}_{N,S}$
$\boldsymbol{Q}^H\boldsymbol{AQ}$	$-\frac{1}{2}((\boldsymbol{AQ})^H \otimes \boldsymbol{I}_S)\boldsymbol{K}_{N,S} + \boldsymbol{I}_S \otimes (\boldsymbol{Q}^H\boldsymbol{A})$	$(\Re(\boldsymbol{AQ})^T \otimes \boldsymbol{I}_S)\boldsymbol{K}_{N,S} - \frac{1}{2}\boldsymbol{I}_S \otimes (\boldsymbol{Q}^H\boldsymbol{A})$
$\boldsymbol{Q}^H\boldsymbol{AQ}^*$	$-\frac{1}{2}(\boldsymbol{AQ}^*)^H \otimes \boldsymbol{I}_S)\boldsymbol{K}_{N,S} - \frac{1}{2}\boldsymbol{I}_S \otimes (\boldsymbol{Q}^H\boldsymbol{A})$	$(\Re(\boldsymbol{AQ}^*)^T \otimes \boldsymbol{I}_S)\boldsymbol{K}_{N,S} + \boldsymbol{I}_S \otimes (\boldsymbol{Q}^H\boldsymbol{A})$

If $\boldsymbol{A}^H\boldsymbol{A}$ and $\boldsymbol{BB}^H$ are invertible, then the system (80) has a unique solution

$$\boldsymbol{Q} = (\boldsymbol{A}^H\boldsymbol{A})^{-1}\boldsymbol{A}^H\boldsymbol{CB}^H(\boldsymbol{BB}^H)^{-1} \tag{83}$$

C. Derivatives of $\boldsymbol{F}(\boldsymbol{Q})$

We next present the derivatives of some elementary matrix functions which are often used in nonlinear adaptive filters and neural networks. Other useful examples of matrix functions can be obtained by simply applying the basic concepts of this work and the results summarized in Table VI, where $\boldsymbol{Q} \in \mathbb{H}^{N\times S}$ or possibly $\boldsymbol{Q} \in \mathbb{H}^{N\times N}$, and $\boldsymbol{A}_1, \boldsymbol{A}_2, \boldsymbol{A}$ are chosen such that the functions are well defined.

1) Derivatives of Power Function: Let $\boldsymbol{F} : \mathbb{H}^{N\times N} \to \mathbb{H}^{N\times N}$ be given by $\boldsymbol{F}(\boldsymbol{Q}) = \boldsymbol{Q}^n$, where n is a positive integer. Using the product rule in Theorem 3.1, we have

$$\begin{aligned}\mathcal{D}_{\boldsymbol{Q}}(\boldsymbol{Q}^n) &= (\boldsymbol{I}_N \otimes \boldsymbol{Q})\mathcal{D}_{\boldsymbol{Q}}(\boldsymbol{Q}^{n-1}) + \mathcal{D}_{\boldsymbol{Q}}(\boldsymbol{Q}\boldsymbol{Q}^{n-1})|_{\boldsymbol{Q}^{n-1}=const}\\ &= (\boldsymbol{I}_N \otimes \boldsymbol{Q})\mathcal{D}_{\boldsymbol{Q}}(\boldsymbol{Q}^{n-1}) + \Re(\boldsymbol{Q}^{n-1})^T \otimes \boldsymbol{I}_N \end{aligned} \tag{84}$$

where the term $\mathcal{D}_{\boldsymbol{Q}}(\boldsymbol{QA})$, given in Table VI, was used in the last equality. Note that the above expression is recurrent about $\mathcal{D}_{\boldsymbol{Q}}(\boldsymbol{Q}^n)$. Expanding this expression and using the initial condition $\mathcal{D}_{\boldsymbol{Q}}(\boldsymbol{Q}) = \boldsymbol{I}_{N^2}$, yields

$$\mathcal{D}_{\boldsymbol{Q}}(\boldsymbol{Q}^n) = \sum_{m=1}^{n}(\boldsymbol{I}_N \otimes \boldsymbol{Q})^{n-m}\left(\Re(\boldsymbol{Q}^{m-1})^T \otimes \boldsymbol{I}_N\right) \tag{85}$$

In a similar manner, we have

$$\mathcal{D}_{\boldsymbol{Q}^*}(\boldsymbol{Q}^n) = -\frac{1}{2}\sum_{m=1}^{n}(\boldsymbol{I}_N \otimes \boldsymbol{Q})^{n-m}\left((\boldsymbol{Q}^{m-1})^H \otimes \boldsymbol{I}_N\right) \tag{86}$$

2) Derivatives of Exponential Function: Let $\boldsymbol{F} : \mathbb{H}^{N\times N} \to \mathbb{H}^{N\times N}$ be given by $\boldsymbol{F}(\boldsymbol{Q}) = \sum_{n=0}^{+\infty}\frac{\boldsymbol{Q}^n}{n!}$. From (85), we have

$$\mathcal{D}_{\boldsymbol{Q}}\boldsymbol{F} = \sum_{n=0}^{+\infty}\sum_{m=1}^{n}\frac{1}{n!}(\boldsymbol{I}_N \otimes \boldsymbol{Q})^{n-m}\left(\Re(\boldsymbol{Q}^{m-1})^T \otimes \boldsymbol{I}_N\right) \tag{87}$$

In a similar manner, we have

$$\mathcal{D}_{\boldsymbol{Q}^*}\boldsymbol{F} = \sum_{n=0}^{+\infty}\sum_{m=1}^{n}\frac{-1}{2\cdot n!}(\boldsymbol{I}_N \otimes \boldsymbol{Q})^{n-m}\left((\boldsymbol{Q}^{m-1})^H \otimes \boldsymbol{I}_N\right) \tag{88}$$

The two examples are a generalization of the quaternion scalar variable case treated in [34] to the quaternion matrix variable case. Likewise, the derivatives of the trigonometric functions and hyperbolic functions can be derived in terms of the exponential function.

VI. CONCLUSIONS

A systematic framework for the calculation of derivatives of quaternion matrix functions of quaternion matrix variables has been proposed based on the GHR calculus. New matrix forms of product and chain rules have been introduced to conveniently calculate the derivatives of quaternion matrix functions, and several results have been developed for quaternion gradient optimisation, such as for the identification of stationary points, direction of maximum change problems, and the steepest descent methods. Furthermore, the usefulness of the presented method has been illustrated on some typical gradient based optimization problems in signal processing. For convenience, key results are given in a tabular form.

Appendix I
Proof of (4) and (5)

1) The proof of (4): From (3), we have $q^{\mu} = \mu q \mu^{-1} = \left(\frac{\mu}{|\mu|}\right) q \left(\frac{\mu}{|\mu|}\right)^{-1} = q^{\left(\frac{\mu}{|\mu|}\right)}$, where $\frac{\mu}{|\mu|}$ is an unit quaternion, i.e., $|\frac{\mu}{|\mu|}| = \frac{|\mu|}{|\mu|} = 1$; $(pq)^{\mu} = \mu(pq)\mu^{-1} = \mu p(\mu^{-1}\mu)q\mu^{-1} = (\mu p \mu^{-1})(\mu q \mu^{-1}) = p^{\mu}q^{\mu}$; $pq = pq(p^{-1}p) = (pqp^{-1})p = q^{p}p$; $pq = (qq^{-1})pq = q(q^{-1}pq) = q\left(\left(\frac{q^*}{|q|^2}\right) p \left(\frac{q^*}{|q|^2}\right)^{-1}\right) = q\left(q^* p (q^*)^{-1}\right) = qp^{(q^*)}$.

2) The Proof of (5): From (3), we have $(q)^{\mu\nu} = (\mu\nu)q(\mu\nu)^{-1} = (\mu\nu)q(\nu^{-1}\mu^{-1}) = \mu(\nu q \nu^{-1})\mu^{-1} = \mu(q^{\nu})\mu^{-1} = (q^{\nu})^{\mu}$; $q^{\mu *} \triangleq (q^*)^{\mu} = \mu q^* \mu^{-1} = (\mu^*)^* q^* \left(\frac{\mu^*}{|\mu|^2}\right) = (\mu^{-1}|\mu|^2)^* q^* \left(\frac{\mu^*}{|\mu|^2}\right) = (\mu^{-1})^* q^* (\mu^*) = (\mu q \mu^{-1})^* = (q^{\mu})^* \triangleq q^{*\mu}$.

Appendix II
Proof of Lemma 2.1

Proof: Let $\boldsymbol{Q} = \boldsymbol{Q}_a + i\boldsymbol{Q}_b + j\boldsymbol{Q}_c + k\boldsymbol{Q}_d \in \mathbb{H}^{N\times S}$, where $\boldsymbol{Q}_a, \boldsymbol{Q}_b, \boldsymbol{Q}_c, \boldsymbol{Q}_d \in \mathbb{R}^{N\times S}$. From (4) and (10), we have $d\boldsymbol{Q}^{\mu} = d\boldsymbol{Q}_a + i^{\mu}(d\boldsymbol{Q}_b) + j^{\mu}(d\boldsymbol{Q}_c) + k^{\mu}(d\boldsymbol{Q}_d)$, $d\boldsymbol{Q}^{\mu i} = d\boldsymbol{Q}_a + i^{\mu}(d\boldsymbol{Q}_b) - j^{\mu}(d\boldsymbol{Q}_c) - k^{\mu}(d\boldsymbol{Q}_d)$, $d\boldsymbol{Q}^{\mu j} = d\boldsymbol{Q}_a - i^{\mu}(d\boldsymbol{Q}_b) + j^{\mu}(d\boldsymbol{Q}_c) - k^{\mu}(d\boldsymbol{Q}_d)$ and $d\boldsymbol{Q}^{\mu k} = d\boldsymbol{Q}_a - i^{\mu}(d\boldsymbol{Q}_b) - j^{\mu}(d\boldsymbol{Q}_c) + k^{\mu}(d\boldsymbol{Q}_d)$. By substituting $d\mathrm{vec}(\boldsymbol{Q}^{\mu}), d\mathrm{vec}(\boldsymbol{Q}^{\mu i}), d\mathrm{vec}(\boldsymbol{Q}^{\mu j})$ and $d\mathrm{vec}(\boldsymbol{Q}^{\mu k})$ into (12), we have $(\boldsymbol{A}_1 + \boldsymbol{A}_2 + \boldsymbol{A}_3 + \boldsymbol{A}_4)d\mathrm{vec}(\boldsymbol{Q}_a) + (\boldsymbol{A}_1 + \boldsymbol{A}_2 - \boldsymbol{A}_3 - \boldsymbol{A}_4)i^{\mu}d\mathrm{vec}(\boldsymbol{Q}_b) + (\boldsymbol{A}_1 - \boldsymbol{A}_2 + \boldsymbol{A}_3 - \boldsymbol{A}_4)j^{\mu}d\mathrm{vec}(\boldsymbol{Q}_c) + (\boldsymbol{A}_1 - \boldsymbol{A}_2 - \boldsymbol{A}_3 + \boldsymbol{A}_4)k^{\mu}d\mathrm{vec}(\boldsymbol{Q}_d) = \boldsymbol{0}$. Since the differentials $d\boldsymbol{Q}_a, d\boldsymbol{Q}_b, d\boldsymbol{Q}_c$ and $d\boldsymbol{Q}_d$ are independent, then $\boldsymbol{A}_1 + \boldsymbol{A}_2 + \boldsymbol{A}_3 + \boldsymbol{A}_4 = \boldsymbol{0}$, $\boldsymbol{A}_1 + \boldsymbol{A}_2 - \boldsymbol{A}_3 - \boldsymbol{A}_4 = \boldsymbol{0}$, $\boldsymbol{A}_1 - \boldsymbol{A}_2 + \boldsymbol{A}_3 - \boldsymbol{A}_4 = \boldsymbol{0}$ and $\boldsymbol{A}_1 - \boldsymbol{A}_2 - \boldsymbol{A}_3 + \boldsymbol{A}_4 = \boldsymbol{0}$. Hence, it follows that $\boldsymbol{A}_1 = \boldsymbol{A}_2 = \boldsymbol{A}_3 = \boldsymbol{A}_4 = \boldsymbol{0}$. The second part can be proved in a similar way. ■

Appendix III
The Derivation of the GHR Calculus

For any quaternion-valued function $f(q) \in \mathbb{H}$, we can state (since the fields $\mathbb{H}$ and $\mathbb{R}^4$ are isomorphic) that

$$f(q) = f_a(q_a, q_b, q_c, q_d) + if_b(q_a, q_b, q_c, q_d) + jf_c(q_a, q_b, q_c, q_d) + kf_d(q_a, q_b, q_c, q_d) \quad (89)$$

where $f_a(\cdot), f_b(\cdot), f_c(\cdot), f_d(\cdot) \in \mathbb{R}$. Then, the function f can be equally seen as a function of the four independent real variables q_a, q_b, q_c and q_d, and the differential of f can be expressed as [24]:

$$[\text{Left}]: \quad df = \frac{\partial f}{\partial q_a}dq_a + \frac{\partial f}{\partial q_b}dq_b + \frac{\partial f}{\partial q_c}dq_c + \frac{\partial f}{\partial q_d}dq_d \quad (90)$$

$$[\text{Right}]: \quad df = dq_a\frac{\partial f}{\partial q_a} + dq_b\frac{\partial f}{\partial q_b} + dq_c\frac{\partial f}{\partial q_c} + dq_d\frac{\partial f}{\partial q_d} \quad (91)$$

where $\frac{\partial f}{\partial q_a}$, $\frac{\partial f}{\partial q_b}$, $\frac{\partial f}{\partial q_c}$ and $\frac{\partial f}{\partial q_d}$ are the partial derivatives of f with respect to q_a, q_b, q_c and q_d, respectively. Note that the two equations are identical since dq_a, dq_b, dq_c and dq_d are real quantities. As a result, both equations are equally valid as a starting point for the derivation of the GHR calculus.

A. The Derivation of the Left GHR Derivatives From (90)

There are two ways to link the real and quaternion differentials, based on the approach in (8) and its conjugate, which respectively induce the left GHR derivatives and conjugate left GHR derivatives.

1) The Left GHR Derivatives: By applying the differential operator to both sides of each expression in (8), we have

$$dq_a = \frac{1}{4}(dq^{\mu} + dq^{\mu i} + dq^{\mu j} + dq^{\mu k}) \quad (92)$$

$$dq_b = -\frac{i^{\mu}}{4}(dq^{\mu} + dq^{\mu i} - dq^{\mu j} - dq^{\mu k}) \quad (93)$$

$$dq_c = -\frac{j^{\mu}}{4}(dq^{\mu} - dq^{\mu i} + dq^{\mu j} - dq^{\mu k}) \quad (94)$$

$$dq_d = -\frac{k^{\mu}}{4}(dq^{\mu} - dq^{\mu i} - dq^{\mu j} + dq^{\mu k}) \quad (95)$$

By inserting (92)–(95) into (90), the differential of f becomes

$$\begin{aligned} df = &\frac{1}{4}\frac{\partial f}{\partial q_a}(dq^{\mu} + dq^{\mu i} + dq^{\mu j} + dq^{\mu k}) \\ &- \frac{1}{4}\frac{\partial f}{\partial q_b}i^{\mu}(dq^{\mu} + dq^{\mu i} - dq^{\mu j} - dq^{\mu k}) \\ &- \frac{1}{4}\frac{\partial f}{\partial q_b}j^{\mu}(dq^{\mu} - dq^{\mu i} + dq^{\mu j} - dq^{\mu k}) \\ &- \frac{1}{4}\frac{\partial f}{\partial q_b}k^{\mu}(dq^{\mu} - dq^{\mu i} - dq^{\mu j} + dq^{\mu k}) \end{aligned} \quad (96)$$

Grouping together $dq^{\mu}, dq^{\mu i}, dq^{\mu j}$ and $dq^{\mu k}$ in (96) yields

$$\begin{aligned} df = &\frac{1}{4}\left(\frac{\partial f}{\partial q_a} - \frac{\partial f}{\partial q_b}i^{\mu} - \frac{\partial f}{\partial q_c}j^{\mu} - \frac{\partial f}{\partial q_d}k^{\mu}\right)dq^{\mu} \\ &+ \frac{1}{4}\left(\frac{\partial f}{\partial q_a} - \frac{\partial f}{\partial q_b}i^{\mu} + \frac{\partial f}{\partial q_c}j^{\mu} + \frac{\partial f}{\partial q_d}k^{\mu}\right)dq^{\mu i} \\ &+ \frac{1}{4}\left(\frac{\partial f}{\partial q_a} + \frac{\partial f}{\partial q_b}i^{\mu} - \frac{\partial f}{\partial q_c}j^{\mu} + \frac{\partial f}{\partial q_d}k^{\mu}\right)dq^{\mu j} \\ &+ \frac{1}{4}\left(\frac{\partial f}{\partial q_a} + \frac{\partial f}{\partial q_b}i^{\mu} + \frac{\partial f}{\partial q_c}j^{\mu} - \frac{\partial f}{\partial q_d}k^{\mu}\right)dq^{\mu k} \end{aligned} \quad (97)$$

Now, we can define the formal left GHR derivatives $\frac{\partial_l f}{\partial q^{\mu}}$, $\frac{\partial_l f}{\partial q^{\mu i}}$, $\frac{\partial_l f}{\partial q^{\mu j}}$ and $\frac{\partial_l f}{\partial q^{\mu k}}$ so that

$$df = \frac{\partial_l f}{\partial q^{\mu}}dq^{\mu} + \frac{\partial_l f}{\partial q^{\mu i}}dq^{\mu i} + \frac{\partial_l f}{\partial q^{\mu j}}dq^{\mu j} + \frac{\partial_l f}{\partial q^{\mu k}}dq^{\mu k} \quad (98)$$

holds. Comparing (98) with (96) and applying Lemma 2.1, gives the left GHR derivatives in the form

$$\frac{\partial_l f}{\partial q^{\mu}} = \frac{1}{4}\left(\frac{\partial f}{\partial q_a} - \frac{\partial f}{\partial q_b}i^{\mu} - \frac{\partial f}{\partial q_c}j^{\mu} - \frac{\partial f}{\partial q_d}k^{\mu}\right) \quad (99)$$

2) The Left Conjugate GHR Derivatives: Apply the conjugate operator to equations in (8) and take the differential, to give

$$dq_a = \frac{1}{4}(dq^{\mu *} + dq^{\mu i*} + dq^{\mu j*} + dq^{\mu k*}) \quad (100)$$

$$dq_b = \frac{i^\mu}{4}(dq^{\mu *} + dq^{\mu i*} - dq^{\mu j*} - dq^{\mu k*}) \quad (101)$$

$$dq_c = \frac{j^\mu}{4}(dq^{\mu *} - dq^{\mu i*} + dq^{\mu j*} - dq^{\mu k*}) \quad (102)$$

$$dq_d = \frac{k^\mu}{4}(dq^{\mu *} + dq^{\mu i*} + dq^{\mu j*} + dq^{\mu k*}) \quad (103)$$

By inserting (100)–(103) into (90), the differential of f becomes

$$\begin{aligned} df =& \frac{1}{4}\frac{\partial f}{\partial q_a}(dq^{\mu *} + dq^{\mu i*} + dq^{\mu j*} + dq^{\mu k*}) \\ &+ \frac{1}{4}\frac{\partial f}{\partial q_b} i^\mu (dq^{\mu *} + dq^{\mu i*} - dq^{\mu j*} - dq^{\mu k*}) \\ &+ \frac{1}{4}\frac{\partial f}{\partial q_b} j^\mu (dq^{\mu *} - dq^{\mu i*} + dq^{\mu j*} - dq^{\mu k*}) \\ &+ \frac{1}{4}\frac{\partial f}{\partial q_b} k^\mu (dq^{\mu *} - dq^{\mu i*} - dq^{\mu j*} + dq^{\mu k*}) \quad (104) \end{aligned}$$

Grouping together $dq^{\mu *}, dq^{\mu i*}, dq^{\mu j*}$ and $dq^{\mu k*}$ in (104) yields

$$\begin{aligned} df =& \frac{1}{4}\left(\frac{\partial f}{\partial q_a} + \frac{\partial f}{\partial q_b} i^\mu + \frac{\partial f}{\partial q_c} j^\mu + \frac{\partial f}{\partial q_d} k^\mu\right) dq^{\mu *} \\ &+ \frac{1}{4}\left(\frac{\partial f}{\partial q_a} + \frac{\partial f}{\partial q_b} i^\mu - \frac{\partial f}{\partial q_c} j^\mu - \frac{\partial f}{\partial q_d} k^\mu\right) dq^{\mu i*} \\ &+ \frac{1}{4}\left(\frac{\partial f}{\partial q_a} - \frac{\partial f}{\partial q_b} i^\mu + \frac{\partial f}{\partial q_c} j^\mu - \frac{\partial f}{\partial q_d} k^\mu\right) dq^{\mu j*} \\ &+ \frac{1}{4}\left(\frac{\partial f}{\partial q_a} - \frac{\partial f}{\partial q_b} i^\mu - \frac{\partial f}{\partial q_c} j^\mu + \frac{\partial f}{\partial q_d} k^\mu\right) dq^{\mu k*} \quad (105) \end{aligned}$$

We can now define the formal conjugate left GHR derivatives $\frac{\partial_l f}{\partial q^{\mu *}}, \frac{\partial_l f}{\partial q^{\mu i*}}, \frac{\partial_l f}{\partial q^{\mu j*}}$ and $\frac{\partial_l f}{\partial q^{\mu k*}}$ so that

$$\begin{aligned} df =& \frac{\partial_l f}{\partial q^{\mu *}} dq^{\mu *} + \frac{\partial_l f}{\partial q^{\mu i*}} dq^{\mu i*} \\ &+ \frac{\partial_l f}{\partial q^{\mu j*}} dq^{\mu j*} + \frac{\partial_l f}{\partial q^{\mu k*}} dq^{\mu k*} \quad (106) \end{aligned}$$

holds. Upon comparing (106) with (104) and applying Lemma 2.1, the following left conjugate GHR derivatives are obtained

$$\frac{\partial_l f}{\partial q^{\mu *}} = \frac{1}{4}\left(\frac{\partial f}{\partial q_a} + \frac{\partial f}{\partial q_b} i^\mu + \frac{\partial f}{\partial q_c} j^\mu + \frac{\partial f}{\partial q_d} k^\mu\right) \quad (107)$$

B. The Derivation of the Right GHR Derivatives From (91)

The derivation of the right GHR derivatives is similar to that of the left GHR derivatives and is omitted for space consideration. We therefore only give the differential of f using the right GHR derivatives

$$df = dq^\mu \frac{\partial_r f}{\partial q^\mu} + dq^{\mu i} \frac{\partial_r f}{\partial q^{\mu i}} + dq^{\mu j} \frac{\partial_r f}{\partial q^{\mu j}} + \frac{\partial_r f}{\partial q^{\mu k}} dq^{\mu k} \quad (108)$$

$$\begin{aligned} df =& dq^{\mu *} \frac{\partial_r f}{\partial q^{\mu *}} + dq^{\mu i*} \frac{\partial_r f}{\partial q^{\mu i*}} \\ &+ dq^{\mu j*} \frac{\partial_r f}{\partial q^{\mu j*}} + dq^{\mu k*} \frac{\partial_r f}{\partial q^{\mu k*}} \quad (109) \end{aligned}$$

Note that the terms $\frac{\partial_r f}{\partial q^\mu}, \frac{\partial_r f}{\partial q^{\mu i}}, \frac{\partial_r f}{\partial q^{\mu j}}, \frac{\partial_r f}{\partial q^{\mu k}}$ cannot swap position with the differentials $dq^\mu, dq^{\mu i}, dq^{\mu j}, dq^{\mu k}$ because of the non-commutative nature of quaternion product. By comparing (98) and (106) with (108) and (109), we notice that the left GHR derivatives stand on the left side of the quaternion differential (98) and (106), which is consistent with our common sense.

Appendix IV
Proofs of Lemma 2.2 and Lemma 2.3

1) The Proof of (24): Since f is real-valued, its partial derivatives $\frac{\partial f}{\partial q_a}, \frac{\partial f}{\partial q_b}, \frac{\partial f}{\partial q_c}$ and $\frac{\partial f}{\partial q_d}$ are real numbers, that is the partial derivatives $\frac{\partial f}{\partial q_b}, \frac{\partial f}{\partial q_c}$ and $\frac{\partial f}{\partial q_d}$ can swap positions with the imaginary units i^μ, j^μ, k^μ. Hence, the first part of (24) follows from (20) and (22), and the second part can be proved in a similar way.

2) The Proof of (25): Since f is real-valued, its partial derivatives $\frac{\partial f}{\partial q_a}, \frac{\partial f}{\partial q_b}, \frac{\partial f}{\partial q_c}$ and $\frac{\partial f}{\partial q_d}$ are real numbers, so that $\left(\frac{\partial f}{\partial \xi}\right)^\mu = \frac{\partial f}{\partial \xi}$, where $\xi \in \{q_a, q_b, q_c, q_d\}$. From (5) and (20), we have

$$\begin{aligned} \left(\frac{\partial f}{\partial q^\mu}\right)^\nu =& \frac{1}{4}\left(\left(\frac{\partial f}{\partial q_a}\right)^\nu - \left(\frac{\partial f}{\partial q_b}\right)^\nu i^{\nu\mu} \right. \\ &\left. - \left(\frac{\partial f}{\partial q_c}\right)^\nu j^{\nu\mu} - \left(\frac{\partial f}{\partial q_d}\right)^\nu k^{\nu\mu}\right) \\ =& \frac{1}{4}\left(\frac{\partial f}{\partial q_a} - \frac{\partial f}{\partial q_b} i^{\nu\mu} - \frac{\partial f}{\partial q_c} j^{\nu\mu} - \frac{\partial f}{\partial q_d} k^{\nu\mu}\right) = \frac{\partial f}{\partial q^{\nu\mu}} \end{aligned}$$

Hence, the first part of (25) follows, and the second part can be proved in a similar way.

3) The Proof of (26): Since f is real-valued, its partial derivatives $\frac{\partial f}{\partial q_a}, \frac{\partial f}{\partial q_b}, \frac{\partial f}{\partial q_c}$ and $\frac{\partial f}{\partial q_d}$ are real numbers, which yields $\left(\frac{\partial f}{\partial \xi}\right)^* = \frac{\partial f}{\partial \xi}$, where $\xi \in \{q_a, q_b, q_c, q_d\}$. From (5) and (20), we have $\left(\frac{\partial f}{\partial q^\mu}\right)^* = \frac{1}{4}\left(\frac{\partial f}{\partial q_a} - \frac{\partial f}{\partial q_b} i^{\mu *} - \frac{\partial f}{\partial q_c} j^{\mu *} - \frac{\partial f}{\partial q_d} k^{\mu *}\right) = \frac{\partial f}{\partial q^{\mu *}}$. Hence, the first part of (25) follows, and the second part can be proved in a similar way.

4) The Proof of (27): By the definition of the left GHR derivative (20) and (5), we have

$$\begin{aligned} &\frac{\partial(\alpha f \beta)}{\partial q^\mu} \\ &= \frac{1}{4}\left(\frac{\partial(\alpha f \beta)}{\partial q_a} - \frac{\partial(\alpha f \beta)}{\partial q_b} i^\mu - \frac{\partial(\alpha f \beta)}{\partial q_c} j^\mu - \frac{\partial(\alpha f \beta)}{\partial q_d} k^\mu\right) \\ &= \frac{1}{4}\left(\alpha\frac{\partial f}{\partial q_a}\beta - \alpha\frac{\partial f}{\partial q_b}\beta i^\mu - \alpha\frac{\partial f}{\partial q_c}\beta j^\mu - \alpha\frac{\partial f}{\partial q_d}\beta k^\mu\right) \\ &= \frac{1}{4}\alpha\left(\frac{\partial f}{\partial q_a} - \frac{\partial f}{\partial q_b}\beta i^\mu \beta^{-1} - \frac{\partial f}{\partial q_c}\beta j^\mu \beta^{-1} - \frac{\partial f}{\partial q_d}\beta k^\mu \beta^{-1}\right)\beta \\ &= \alpha\frac{1}{4}\left(\frac{\partial f}{\partial q_a} - \frac{\partial f}{\partial q_b} i^{\beta\mu} - \frac{\partial f}{\partial q_c} j^{\beta\mu} - \frac{\partial f}{\partial q_d} k^{\beta\mu}\right)\beta \\ &= \alpha\frac{\partial f}{\partial q^{\beta\mu}}\beta \end{aligned}$$

Hence, the first part of (27) follows, and the second part can be proved in a similar way.

5) The Proof of (28): By the definition of the right GHR derivative (22) and (5), we have

$$
\begin{aligned}
&\frac{\partial_r(\alpha f \beta)}{\partial q^{\mu}} \\
&= \frac{1}{4}\left(\frac{\partial(\alpha f \beta)}{\partial q_a} - i^{\mu}\frac{\partial(\alpha f \beta)}{\partial q_b} - j^{\mu}\frac{\partial(\alpha f \beta)}{\partial q_c} - k^{\mu}\frac{\partial(\alpha f \beta)}{\partial q_d}\right) \\
&= \frac{1}{4}\left(\alpha\frac{\partial f}{\partial q_a}\beta - i^{\mu}\alpha\frac{\partial f}{\partial q_b}\beta - j^{\mu}\alpha\frac{\partial f}{\partial q_c}\beta - k^{\mu}\alpha\frac{\partial f}{\partial q_d}\beta\right) \\
&= \frac{1}{4}\alpha\left(\frac{\partial f}{\partial q_a} - \alpha^{-1}i^{\mu}\alpha\frac{\partial f}{\partial q_b} - \alpha^{-1}j^{\mu}\alpha\frac{\partial f}{\partial q_c} - \alpha^{-1}k^{\mu}\alpha\frac{\partial f}{\partial q_d}\right)\beta \\
&= \alpha\frac{1}{4}\left(\frac{\partial f}{\partial q_a} - i^{(\alpha^{-1}\mu)}\frac{\partial f}{\partial q_b} - j^{(\alpha^{-1}\mu)}\frac{\partial f}{\partial q_c} - k^{(\alpha^{-1}\mu)}\frac{\partial f}{\partial q_d}\right)\beta \\
&= \alpha\frac{\partial f}{\partial q^{(\alpha^{-1}\mu)}}\beta
\end{aligned}
$$

Hence, the first part of (28) follows, and the second part can be proved in a similar way.

ACKNOWLEDGMENT

The authors dedicate this work to the memory of Prof. Are Hjørungnes, a pioneer of complex-valued matrix derivatives. The authors also thank the anonymous referees who made numerous constructive suggestions.

REFERENCES

[1] 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, May 1999.

[2] M. Felsberg and G. Sommer, "The monogenic signal," *IEEE Trans. Signal Process.*, vol. 49, no. 12, pp. 3136–3144, Dec. 2001.

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

[4] A. J. Hanson, *Visualizing Quaternions*. San Francisco, CA, USA: Morgan Kaufmann, 2005.

[5] J. B. Kuipers, *Quaternions and Rotation Sequences: A Primer with Applications to Orbits, Aerospace and Virtual Reality*. Princeton, NJ, USA: Princeton Univ. Press, Aug. 2002.

[6] L. Fortuna, G. Muscato, and M. Xibilia, "A comparison between HMLP and HRBF for attitude control," *IEEE Trans. Neural Netw.*, vol. 12, no. 2, pp. 318–328, Mar. 2001.

[7] C. C. 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, Apr. 2009.

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

[9] C. C. Took and D. P. Mandic, "Quaternion-valued stochastic gradient-based adaptive IIR filtering," *IEEE Trans. Signal Process.*, vol. 58, no. 7, pp. 3895–3901, July 2010.

[10] 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.

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

[12] S. Buchholz and N. Le Bihan, "Polarized signal classification by complex and quaternionic multi-layer perceptrons," *Int. J. Neural Syst.*, vol. 18, no. 2, pp. 75–85, 2008.

[13] J. C. Belfiore, G. Rekaya, and E. Viterbo, "The Golden code: A 2 $\times$ 2 full-rate space-time code with non-vanishing determinant," *IEEE Trans. Inf. Theory*, vol. 51, no. 4, pp. 1432–1436, Apr. 2005.

[14] M. Y. Chen, H. C. Li, and S. C. Pei, "Algebraic identification for optimal nonorthogonality 4 $\times$ 4 complex space-time block codes using tensor product on quaternions," *IEEE Trans. Inf. Theory*, vol. 51, no. 1, pp. 324–330, Jan. 2005.

[15] C. Hollanti, J. Lahtonen, K. Ranto, R. Vehkalahti, and E. Viterbo, "On the algebraic structure of the Silver code: A 2 $\times$ 2 perfect space-time block code," in *Proc. IEEE Inf. Theory Workshop*, Porto, Portugal, May 2008, pp. 91–94.

[16] J. Seberry, K. Finlayson, S. Adams, T. Wysocki, T. Xia, and B. Wysocki, "The theory of quaternion orthogonal designs," *IEEE Trans. Signal Process.*, vol. 56, no. 1, pp. 256–265, Jan. 2008.

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

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

[19] J. Vía, D. Ramírez, and I. Santamaría, "Properness and widely linear processing of quaternion random vectors," *IEEE Trans. Inf. Theory*, vol. 56, no. 7, pp. 3502–3515, Jul. 2010.

[20] P. Ginzberg and A. T. Walden, "Testing for quaternion propriety," *IEEE Trans. Signal Process.*, vol. 59, no. 7, pp. 3025–3034, Jul. 2011.

[21] C. C. Took and D. P. Mandic, "Augmented second order statistics of quaternion random signals," *Signal Process.*, vol. 91, no. 2, pp. 214–224, Feb. 2011.

[22] J. Navarro-Moreno, R. M. Fernáandez-Alcalá, 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, Dec. 2012.

[23] R. E. S. Watson, "The generalized Cauchy-Riemann-Fueter equation and handedness," *Compl. Variab.*, vol. 48, no. 7, pp. 555–568, Jul. 2003.

[24] A. Sudbery, "Quaternionic analysis," *Math. Proc. Camb. Philos. Soc.*, vol. 85, no. 2, pp. 199–225, 1979.

[25] G. Gentili and D. C. Struppa, "A new approach to Cullen-regular functions of a quaternionic variable," *C. R. Math. Acad. Sci. Paris*, vol. 342, no. 10, pp. 741–744, May 2006.

[26] G. Gentili and D. C. Struppa, "A new theory of regular functions of a quaternionic variable," *Adv. Math.*, vol. 216, no. 1, pp. 279–301, Dec. 2007.

[27] C. A. Deavours, "The quaternion calculus," *Amer. Math. Month.*, vol. 80, pp. 995–1008, 1979.

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

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

[30] S. Javidi, C. Took, and D. P. Mandic, "Fast independent component analysis algorithm for quaternion valued signals," *IEEE Trans. Neural Netw.*, vol. 22, no. 12, pp. 1967–1978, Dec. 2011.

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

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

[33] C. Jahanchahi and D. P. Mandic, "A class of quaternion Kalman filters," *IEEE Trans. Neural Netw. Learn. Syst*, vol. 25, no. 3, pp. 533–544, Mar. 2014.

[34] D. Xu, C. Jahanchahi, C. C. Took, and D. P. Mandic, "Quaternion derivatives: The GHR calculus," Imperial College London, London, U.K., Tech. Rep., Jan. 2014, pp. 1–38.

[35] W. Wirtinger, "Zur formalen theorie der funktionen von mehr komplexen veränderlichen," *Mathematische Annalen*, vol. 97, pp. 357–375, 1927.

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

[37] K. Kreutz-Delgado, "The complex gradient operator and the CR calculus," 2009, pp. 1–74, ArXiv preprint arXiv:0906.4835 [Online]. Available: http://arxiv.org/abs/0906.4835

[38] H. Li and T. Adali, "A class of complex ICA algorithms based on the kurtosis cost function," *IEEE Trans. Neural Netw.*, vol. 19, no. 3, pp. 408–420, Mar. 2008.

[39] B. Demissie and S. Kreuzer, "Performance predictions for parameter estimators that minimize cost-functions using Wirtinger calculus with application to CM blind equalization," *IEEE Trans. Signal Process.*, vol. 59, no. 8, pp. 3685–3698, Aug. 2011.

[40] D. P. Mandic and V. S. L. Goh, *Complex Valued Nonlinear Adaptive Filters: Noncircularity, Widely Linear and Neural Models*. New York, NY, USA: Wiley, 2009.

[41] L. Sorber, M. Van Barel, and L. De Lathauwer, "Unconstrained optimization of real functions in complex variables," *SIAM J. Optim.*, vol. 22, no. 3, pp. 879–898, July 2012.

[42] J. R. Magnus and H. Neudecker, *Matrix Differential Calculus With Application in Statistics and Econometrics*. Essex, U.K.: Wiley, 1988.

[43] J. W. Brewer, "Kronecker products and matrix calculus in system theory," *IEEE Trans. Circuits Syst.*, vol. CAS-25, no. 9, pp. 772–781, Sep. 1978.

[44] D. A. Harville, *Matrix Algebra From a Statistician's Perspective*. New York, NY, USA: Springer-Verlag, 1997, corrected 2nd printing, 1999.

[45] A. van den Bos, "Complex gradient and Hessian," *Proc. Inst. Electr. Eng.—Vis, Image Signal Process.*, vol. 141, no. 6, pp. 380–383, Dec. 1994.

[46] A. Hørungnes and D. Gesbert, "Complex-valued matrix differentiation: Techniques and key results," *IEEE Trans. Signal Process.*, vol. 55, no. 6, pp. 2740–2746, Jun. 2007.

[47] A. Hørungnes, *Complex-Valued Matrix Derivatives: With Applications in Signal Processing and Communications*. Cambridge, U.K.: Cambridge Univ. Press, 2011.

[48] J. P. Ward, *Quaternions and Cayley Numbers: Algebra and Applications*. Boston, MA, USA: Kluwer, 1997.

[49] F. Zhang, "Quaternions and matrices of quaternions," *Linear Algebra Appl*, vol. 251, pp. 21–57, Jan. 1997.

[50] T. A. Ell and S. J. Sangwine, "Quaternion involutions and anti-involutions," *Comput, Math. Appl.*, vol. 53, pp. 137–143, 2007.

[51] Q. Barthelemy, A. Larue, and J. I. Mars, "About QLMS derivations," *IEEE Signal Process. Lett.*, vol. 21, no. 2, pp. 240–243, Feb. 2014.

[52] M. Jiang, W. Liu, and Y. Li, "A general quaternion-valued gradient operator and its applications to computational fluid dynamics and adaptive beamforming," in *Int. Conf. Dig. Signal Process.*, Aug. 2014, pp. 821–826.

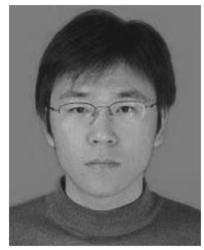
[53] M. Jiang, Y. Li, and W. Liu, "Properties and applications of a restricted HR gradient operator," 2014, pp. 1–8, ArXiv preprint arXiv:1407.5178 [Online]. Available: http://arxiv.org/abs/1407.5178

[54] Y. Xia, C. Jahanchahi, D. Xu, and D. P. Mandic, "The HC calculus quaternion derivatives and Caylay-Dickson form of quaternion adaptive filters and learning systems," in *Proc. Int. Joint Conf. Neural Netw.*, Jul. 2014, pp. 3395–3401.

[55] D. Xu and D. P. Mandic, "Quaternion gradient and Hessian," 2014, pp. 1–23, ArXiv preprint arXiv:1406.3587 [Online]. Available: http://arxiv.org/abs/1406.3587

[56] C. C. Took, C. Jahanchahi, and D. P. Mandic, "A unifying framework for the analysis of quaternion valued adaptive filters," *Conf. Rec. Asilomar Conf. Signals, Syst., Comput.*, pp. 1771–1774, 2011.

[57] B. Widrow, J. McCool, and M. Ball, "The complex LMS algorithm," *Proc. IEEE*, vol. 63, pp. 719–720, 1975.



Dongpo Xu received the B.S. degree in applied mathematics from Harbin Engineering University, Harbin, China, in 2004, and the Ph.D. degree in computational mathematics from Dalian University of Technology, Dalian, China, in 2009.

He was a Lecturer with the College of Science, Harbin Engineering University, from 2009 to 2014. From 2013 to 2014, he was a visiting scholar with the Department of Electrical and Electronic Engineering, Imperial College London. He is currently a Lecturer with the School of Mathematics and Statistics, Northeast Normal University, Changchun, China. His current research interests include linear and nonlinear adaptive filters, and complex valued and quaternion valued statistical analysis.

Danilo P. Mandic (M'99–SM'03–F'12) is a Professor of Signal Processing with Imperial College London, London, U.K., where he has been involved in nonlinear adaptive signal processing and nonlinear dynamics. He has been a Guest Professor with Katholieke Universiteit Leuven, Leuven, Belgium, the Tokyo University of Agriculture and Technology, Tokyo, Japan, and Westminster University, London, U.K., and a Frontier Researcher with RIKEN, Wako, Japan. He has two research monographs title *Recurrent Neural Networks for Prediction: Learning Algorithms, Architectures and Stability* (Wiley, 2001) and *Complex Valued Nonlinear Adaptive Filters: Noncircularity,Widely Linear and Neural Models* (Wiley, 2009), an edited book titled *Signal Processing for Information Fusion* (Springer, 2008), and more than 200 publications on signal and image processing.

Prof. Mandic has been a member of the IEEE Technical Committee on Signal Processing Theory and Methods, and an Associate Editor of the IEEE SIGNAL PROCESSING MAGAZINE, the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS II, the IEEE TRANSACTIONS ON SIGNAL PROCESSING, and the IEEE TRANSACTIONS ON NEURAL NETWORKS. He has produced award winning papers and products resulting from his collaboration with the industry.