Properties of a general quaternion-valued gradient operator and its applications to signal processing[*#]

Meng-di JIANG[1], Yi LI[2], Wei LIU[†‡1]

(*1Department of Electronic and Electrical Engineering, University of Sheffield, Sheffield S1 3JD, UK*)

(*2School of Mathematics and Statistics, University of Sheffield, Sheffield S3 7RH, UK*)

[†]E-mail: w.liu@sheffield.ac.uk

Abstract: The gradients of a quaternion-valued function are often required for quaternionic signal processing algorithms. The HR gradient operator provides a viable framework and has found a number of applications. However, the applications so far have been limited to mainly real-valued quaternion functions and linear quaternion-valued functions. To generalize the operator to nonlinear quaternion functions, we define a restricted version of the HR operator, which comes in two versions, the left and the right ones. We then present a detailed analysis of the properties of the operators, including several different product rules and chain rules. Using the new rules, we derive explicit expressions for the derivatives of a class of regular nonlinear quaternion-valued functions, and prove that the restricted HR gradients are consistent with the gradients in the real domain. As an application, the derivation of the least mean square algorithm and a nonlinear adaptive algorithm is provided. Simulation results based on vector sensor arrays are presented as an example to demonstrate the effectiveness of the quaternion-valued signal model and the derived signal processing algorithm.

Key words: Quaternion, Gradient operator, Signal processing, Least mean square (LMS) algorithm, Nonlinear adaptive filtering, Adaptive beamforming

http://dx.doi.org/10.1631/FITEE.1500334   **CLC number:** TN911.7; O29

1 Introduction

Quaternion calculus has been introduced in signal processing with application areas involving three- or four-dimensional signals, such as color image processing (Pei and Cheng, 1999; Sangwine and Ell, 2000; Parfieniuk and Petrovsky, 2010; Ell *et al.*, 2014; Liu *et al.*, 2014), vector-sensor array systems (Le Bihan and Mars, 2004; Miron *et al.*, 2006; Le Bihan *et al.*, 2007; Tao, 2013; Tao and Chang, 2014; Zhang *et al.*, 2014; Hawes and Liu, 2015), three-phase power

systems (Talebi and Mandic, 2015), quaternion-valued wireless communications (Liu, 2014), and wind profile prediction (Jiang *et al.*, 2014b). Several quaternion-valued adaptive filtering algorithms have been proposed in Barthelemy *et al.* (2014), Jiang *et al.* (2014c), Talebi *et al.* (2014), Tao and Chang (2014), and Zhang *et al.* (2014). Notwithstanding the advantages of the quaternionic algorithms, extra care has to be taken in their developments, in particular when the derivatives of quaternion-valued functions are involved, due to the fact that quaternion algebra is non-commutative. A so-called HR gradient operator was proposed in Mandic *et al.* (2011) and the interesting formulation appears to provide a general and flexible framework that could potentially have wide applications. However, it has

been applied only to real-valued functions and linear quaternion-valued functions. To consider more general quaternion-valued functions, we propose a pair of restricted HR gradient operators, the left and the right restricted HR gradient operators, based on the previous work on the HR gradient operator (Mandic *et al.*, 2011) and our recent work (Jiang *et al.*, 2014b).

To summarize, we make the following main contributions. First, we give a detailed derivation of the relation between the gradients and the increment of a quaternion function, highlighting the difference between the left and the right gradients due to the non-commutativity of quaternion algebra. Second, we document several properties of the operators that have not been reported before, in particular several different versions of product rules and chain rules. Third, we derive a general formula for the restricted HR derivatives of a wide class of regular quaternion-valued nonlinear functions, among which are the exponential, logarithmic, and the hyperbolic tangent functions. Finally, we prove that the restricted HR gradients are consistent with the usual definition for the gradient of a real function of a real variable. Its application to the derivation of a quaternion-valued least mean squares (QLMS) adaptive algorithm and a nonlinear adaptive algorithm based on the hyperbolic tangent function is also briefly discussed. As an example for quaternion-valued signal processing, we consider the reference signal based adaptive beamforming problem for vector sensor arrays consisting of multiple crossed-dipoles and provide some simulation results.

2 Restricted HR gradient operators

2.1 Introduction of quaternion

Quaternions are a non-commutative extension of complex numbers. A quaternion q is composed of four parts, i.e., $q = q_a + q_b\mathrm{i} + q_c\mathrm{j} + q_d\mathrm{k}$, where q_a is the real part, also denoted as $\mathrm{R}(q)$. The other three terms constitute the imaginary part $\mathrm{I}(q)$, where i, j, and k are the three imaginary units, satisfying the following rules: ij=k, jk=i, ki=j, $\mathrm{i}^2 = \mathrm{j}^2 = \mathrm{k}^2 = -1$, and $\mathrm{ij} = -\mathrm{ji}, \mathrm{ki} = -\mathrm{ik}, \mathrm{kj} = -\mathrm{jk}$. As a result, in general the product of two quaternions p and q depends on the order, i.e., $qp \neq pq$, and when one of the factors is real, we have $qp = pq$.

Let $v = |\mathrm{I}(q)|$ and $\hat{\mathbf{v}} = \mathrm{I}(q)/v$. The quaternion q can also be written as $q = q_a + v\hat{\mathbf{v}}$. Here, $\hat{\mathbf{v}}$ is a pure unit quaternion, which has the convenient property $\hat{\mathbf{v}}^2 := \hat{\mathbf{v}}\hat{\mathbf{v}} = -1$. The quaternionic conjugate of q is $q^* = q_a - q_b\mathrm{i} - q_c\mathrm{j} - q_d\mathrm{k}$, or $q^* = q_a - v\hat{\mathbf{v}}$. It is easy to show that $qq^* = q^*q = |q|^2$, and hence $q^{-1} = q^*/|q|^2$.

2.2 Definition of the restricted HR gradient operators

Let $f : H \to H$ be a quaternion-valued function of a quaternion q, where H is the non-commutative algebra of quaternions. We use the notation $f(q) = f_a + f_b\mathrm{i} + f_c\mathrm{j} + f_d\mathrm{k}$, where f_a, f_b, f_c, f_d are the components of f. Here, f can also be viewed as a function of the four components of q, i.e., $f = f(q_a, q_b, q_c, q_d)$. In this view f is a quaternion-valued function on $\mathbb{R}^4$: $f : \mathbb{R}^4 \to H$. To express the four real components of q, it is convenient to use its involutions $q^\nu := -\nu q\nu$ where $\nu \in \{\mathrm{i}, \mathrm{j}, \mathrm{k}\}$ (Ell and Sangwine, 2007). Explicitly, we have

$$q^\mathrm{i} = -\mathrm{i}q\mathrm{i} = q_a + q_b\mathrm{i} - q_c\mathrm{j} - q_d\mathrm{k}, \qquad (1)$$

$$q^\mathrm{j} = -\mathrm{j}q\mathrm{j} = q_a - q_b\mathrm{i} + q_c\mathrm{j} - q_d\mathrm{k}, \qquad (2)$$

$$q^\mathrm{k} = -\mathrm{k}q\mathrm{k} = q_a - q_b\mathrm{i} - q_c\mathrm{j} + q_d\mathrm{k}, \qquad (3)$$

$$q_a = \frac{1}{4}(q + q^\mathrm{i} + q^\mathrm{j} + q^\mathrm{k}), \qquad (4)$$

$$q_b = \frac{1}{4\mathrm{i}}(q + q^\mathrm{i} - q^\mathrm{j} - q^\mathrm{k}), \qquad (5)$$

$$q_c = \frac{1}{4\mathrm{j}}(q - q^\mathrm{i} + q^\mathrm{j} - q^\mathrm{k}), \qquad (6)$$

$$q_d = \frac{1}{4\mathrm{k}}(q - q^\mathrm{i} - q^\mathrm{j} + q^\mathrm{k}). \qquad (7)$$

Two useful relations are

$$\begin{cases} q^* = 1/2(q^\mathrm{i} + q^\mathrm{j} + q^\mathrm{k} - q), \\ q + q^\mathrm{i} + q^\mathrm{j} + q^\mathrm{k} = 4\mathrm{R}(q). \end{cases} \qquad (8)$$

A so-called HR gradient of $f(q)$ was introduced in Mandic *et al.* (2011), which has been applied to real-valued functions and linear quaternion-valued functions. To find the gradients of more general quaternion-valued functions, we follow a similar approach to propose a 'restricted' HR gradient operator (some of the derivation was first presented in Jiang *et al.* (2014b)). To motivate the definitions, we consider the differential $\mathrm{d}f(q)$ with respect to differential $\mathrm{d}q := \mathrm{d}q_a + \mathrm{d}q_b\mathrm{i} + \mathrm{d}q_c\mathrm{j} + \mathrm{d}q_d\mathrm{k}$. We observe that $\mathrm{d}f = \mathrm{d}f_a + \mathrm{i}\mathrm{d}f_b + \mathrm{j}\mathrm{d}f_c + \mathrm{k}\mathrm{d}f_d$, where

$$\mathrm{d}f_a = \frac{\partial f_a}{\partial q_a}\mathrm{d}q_a + \frac{\partial f_a}{\partial q_b}\mathrm{d}q_b + \frac{\partial f_a}{\partial q_c}\mathrm{d}q_c + \frac{\partial f_a}{\partial q_d}\mathrm{d}q_d. \quad (9)$$

We have $\mathrm{d}q_a = (\mathrm{d}q + \mathrm{d}q^{\mathrm{i}} + \mathrm{d}q^{\mathrm{j}} + \mathrm{d}q^{\mathrm{k}})/4$ according to Eq. (4). Making use of this and similar expressions for $\mathrm{d}q_b$, $\mathrm{d}q_c$, and $\mathrm{d}q_d$, we find an expression for $\mathrm{d}f_a$ in terms of the differentials $\mathrm{d}q$, $\mathrm{d}q^{\mathrm{i}}$, $\mathrm{d}q^{\mathrm{j}}$, and $\mathrm{d}q^{\mathrm{k}}$. Repeating the calculation for $\mathrm{i}\mathrm{d}f_b$, $\mathrm{j}\mathrm{d}f_c$, and $\mathrm{k}\mathrm{d}f_d$, we finally arrive at

$$\mathrm{d}f = D\mathrm{d}q + D_{\mathrm{i}}\mathrm{d}q^{\mathrm{i}} + D_{\mathrm{j}}\mathrm{d}q^{\mathrm{j}} + D_{\mathrm{k}}\mathrm{d}q^{\mathrm{k}}, \qquad (10)$$

where

$$D := \frac{1}{4}\left(\frac{\partial f}{\partial q_a} - \frac{\partial f}{\partial q_b}\mathrm{i} - \frac{\partial f}{\partial q_c}\mathrm{j} - \frac{\partial f}{\partial q_d}\mathrm{k}\right), \qquad (11)$$

$$D_{\mathrm{i}} := \frac{1}{4}\left(\frac{\partial f}{\partial q_a} - \frac{\partial f}{\partial q_b}\mathrm{i} + \frac{\partial f}{\partial q_c}\mathrm{j} + \frac{\partial f}{\partial q_d}\mathrm{k}\right), \qquad (12)$$

$$D_{\mathrm{j}} := \frac{1}{4}\left(\frac{\partial f}{\partial q_a} + \frac{\partial f}{\partial q_b}\mathrm{i} - \frac{\partial f}{\partial q_c}\mathrm{j} + \frac{\partial f}{\partial q_d}\mathrm{k}\right), \qquad (13)$$

$$D_{\mathrm{k}} := \frac{1}{4}\left(\frac{\partial f}{\partial q_a} + \frac{\partial f}{\partial q_b}\mathrm{i} + \frac{\partial f}{\partial q_c}\mathrm{j} - \frac{\partial f}{\partial q_d}\mathrm{k}\right). \qquad (14)$$

More details are given in Appendix A. Thus, one may define the partial derivatives of $f(q)$ as follows:

$$\frac{\partial f}{\partial q} := D, \ \frac{\partial f}{\partial q^{\mathrm{i}}} := D_{\mathrm{i}}, \ \frac{\partial f}{\partial q^{\mathrm{j}}} := D_{\mathrm{j}}, \ \frac{\partial f}{\partial q^{\mathrm{k}}} := D_{\mathrm{k}}. \ (15)$$

Introducing operators

$$\nabla_q := (\partial/\partial q, \partial/\partial q^{\mathrm{i}}, \partial/\partial q^{\mathrm{j}}, \partial/\partial q^{\mathrm{k}})$$

and

$$\nabla_r := (\partial/\partial q_a, \partial/\partial q_b, \partial/\partial q_c, \partial/\partial q_d),$$

Eqs. (11)–(15) can be written as

$$\nabla_q f = \nabla_r f \boldsymbol{J}^{\mathrm{H}}, \qquad (16)$$

where the Jacobian matrix is

$$\boldsymbol{J} = \frac{1}{4}\begin{bmatrix} 1 & \mathrm{i} & \mathrm{j} & \mathrm{k} \\ 1 & \mathrm{i} & -\mathrm{j} & -\mathrm{k} \\ 1 & -\mathrm{i} & \mathrm{j} & -\mathrm{k} \\ 1 & -\mathrm{i} & -\mathrm{j} & \mathrm{k} \end{bmatrix}, \qquad (17)$$

and $\boldsymbol{J}^{\mathrm{H}}$ is the Hermitian transpose of $\boldsymbol{J}$ (Mandic *et al.*, 2011). Using $\boldsymbol{J}\boldsymbol{J}^{\mathrm{H}} = \boldsymbol{J}^{\mathrm{H}}\boldsymbol{J} = 1/4\boldsymbol{I}$ where $\boldsymbol{I}$ is the identity matrix, we can also write

$$\nabla_q f \boldsymbol{J} = \frac{1}{4}\nabla_r f, \qquad (18)$$

which is the inverse formula for the derivatives.

We call the gradient operator defined by Eq. (16) the restricted HR gradient operator. The operator is closely related to the HR operator introduced in Mandic *et al.* (2011). However, in the original definition of the HR operator, the Jacobian $\boldsymbol{J}$ appears on the left-hand side of $\nabla_r f$, whereas in our definition it appears on the right (as the Hermitian transpose).

The differential $\mathrm{d}f$ is related to $\nabla_q f$ by

$$\mathrm{d}f = \frac{\partial f}{\partial q}\mathrm{d}q + \frac{\partial f}{\partial q^{\mathrm{i}}}\mathrm{d}q^{\mathrm{i}} + \frac{\partial f}{\partial q^{\mathrm{j}}}\mathrm{d}q^{\mathrm{j}} + \frac{\partial f}{\partial q^{\mathrm{k}}}\mathrm{d}q^{\mathrm{k}}. \quad (19)$$

Due to the non-commutativity of quaternion products, the order of the factors in the products of Eq. (19) (as well as Eqs. (11)–(14)) cannot be swapped. In fact, one may call the above operator the left restricted HR gradient operator. As is shown in Appendix A, one can also define a right restricted HR gradient operator by

$$(\nabla_q^{\mathrm{R}} f)^{\mathrm{T}} := \boldsymbol{J}^*(\nabla_r f)^{\mathrm{T}}, \qquad (20)$$

where

$$\nabla_q^{\mathrm{R}} := (\partial^{\mathrm{R}}/\partial q, \partial^{\mathrm{R}}/\partial q^{\mathrm{i}}, \partial^{\mathrm{R}}/\partial q^{\mathrm{j}}, \partial^{\mathrm{R}}/\partial q^{\mathrm{k}}),$$

and

$$\frac{\partial^{\mathrm{R}} f}{\partial q} := \frac{1}{4}\left(\frac{\partial f}{\partial q_a} - \mathrm{i}\frac{\partial f}{\partial q_b} - \mathrm{j}\frac{\partial f}{\partial q_c} - \mathrm{k}\frac{\partial f}{\partial q_d}\right), \quad (21)$$

$$\frac{\partial^{\mathrm{R}} f}{\partial q^{\mathrm{i}}} := \frac{1}{4}\left(\frac{\partial f}{\partial q_a} - \mathrm{i}\frac{\partial f}{\partial q_b} + \mathrm{j}\frac{\partial f}{\partial q_c} + \mathrm{k}\frac{\partial f}{\partial q_d}\right), \quad (22)$$

$$\frac{\partial^{\mathrm{R}} f}{\partial q^{\mathrm{j}}} := \frac{1}{4}\left(\frac{\partial f}{\partial q_a} + \mathrm{i}\frac{\partial f}{\partial q_b} - \mathrm{j}\frac{\partial f}{\partial q_c} + \mathrm{k}\frac{\partial f}{\partial q_d}\right), \quad (23)$$

$$\frac{\partial^{\mathrm{R}} f}{\partial q^{\mathrm{k}}} := \frac{1}{4}\left(\frac{\partial f}{\partial q_a} + \mathrm{i}\frac{\partial f}{\partial q_b} + \mathrm{j}\frac{\partial f}{\partial q_c} - \mathrm{k}\frac{\partial f}{\partial q_d}\right). \quad (24)$$

The right restricted HR gradient operator is related to the differential $\mathrm{d}f$ by

$$\mathrm{d}f = \mathrm{d}q\frac{\partial^{\mathrm{R}} f}{\partial q} + \mathrm{d}q^{\mathrm{i}}\frac{\partial^{\mathrm{R}} f}{\partial q^{\mathrm{i}}} + \mathrm{d}q^{\mathrm{j}}\frac{\partial^{\mathrm{R}} f}{\partial q^{\mathrm{j}}} + \mathrm{d}q^{\mathrm{k}}\frac{\partial^{\mathrm{R}} f}{\partial q^{\mathrm{k}}}. \ (25)$$

In general, the left and right restricted HR gradients are not the same. For example, even for the simplest linear function $f(q) = q_0 q$ with $q_0 \in H$ a constant, we have

$$\frac{\partial q_0 q}{\partial q} = q_0, \ \frac{\partial^{\mathrm{R}} q_0 q}{\partial q} = \mathrm{R}(q_0). \qquad (26)$$

However, we will show later that the two gradients coincide for a class of functions. In particular, they are the same for real-valued quaternion functions. The relationship between the gradients and the differential is an important ingredient of gradient-based methods, which we will discuss further later.

3 Properties and rules of the operator

We will now focus on the left restricted HR gradient and simply call it the restricted HR gradient unless stated otherwise. It can be easily calculated from the definitions that

$$\frac{\partial q}{\partial q} = 1, \ \frac{\partial q^\nu}{\partial q} = 0, \ \frac{\partial q^*}{\partial q} = -\frac{1}{2}, \qquad (27)$$

where $\nu \in \{i, j, k\}$. However, to find the derivatives for more complex quaternion functions, it is useful to first establish the rules of the gradient operators. We will see that some of the usual rules do not apply due to the non-commutativity of quaternion products.

1. Left-linearity: For arbitrary constant quaternions α and β, and functions $f(q)$ and $g(q)$, we have

$$\frac{\partial(\alpha f + \beta g)}{\partial q^\nu} = \alpha \frac{\partial f}{\partial q^\nu} + \beta \frac{\partial g}{\partial q^\nu} \qquad (28)$$

for $\nu \in \{1, i, j, k\}$ with $q^1 := q$. However, linearity does not hold for right multiplications, i.e., in general

$$\frac{\partial f\alpha}{\partial q} \neq \frac{\partial f}{\partial q}\alpha. \qquad (29)$$

This is because, according to Eq. (11),

$$\frac{\partial f\alpha}{\partial q} = \frac{1}{4}\left(\frac{\partial f}{\partial q_a}\alpha - \frac{\partial f}{\partial q_b}\alpha i - \frac{\partial f}{\partial q_c}\alpha j - \frac{\partial f}{\partial q_d}\alpha k\right), \tag{30}$$

where α is an arbitrary constant quaternion. However, $\alpha\nu \neq \nu\alpha$ in general. Therefore, it is different from $(\partial f/\partial q)\alpha$, which is

$$\frac{1}{4}\left(\frac{\partial f}{\partial q_a} - \frac{\partial f}{\partial q_b}i - \frac{\partial f}{\partial q_c}j - \frac{\partial f}{\partial q_d}k\right)\alpha. \qquad (31)$$

2. The first product rule: The following product rule holds:

$$\nabla_q(fg) = f\nabla_q g + [(\nabla_r f)g]\,\boldsymbol{J}^{\mathrm{H}}. \qquad (32)$$

For example,

$$\frac{\partial fq}{\partial q} = f\frac{\partial g}{\partial q} + \frac{1}{4}\left(\frac{\partial f}{\partial q_a}g - \frac{\partial f}{\partial q_b}gi - \frac{\partial f}{\partial q_c}gj - \frac{\partial f}{\partial q_d}gk\right). \tag{33}$$

Thus, the product rule in general is different from the usual one.

3. The second product rule: However, the usual product rule applies to differentiation with respect to real variables, i.e.,

$$\frac{\partial fg}{\partial q_\phi} = \frac{\partial f}{\partial q_\phi}g + f\frac{\partial g}{\partial q_\phi} \qquad (34)$$

for $\phi = a, b, c,$ or d.

4. The third product rule: The usual product rule also applies if at least one of the two functions $f(q)$ and $g(q)$ is real-valued, i.e.,

$$\frac{\partial fq}{\partial q} = f\frac{\partial g}{\partial q} + \frac{\partial f}{\partial q}g. \qquad (35)$$

5. The first chain rule: For a composite function $f(g(q))$, $g(q) := g_a + g_b i + g_c j + g_d k$ being a quaternion-valued function, we have the following chain rule:

$$\nabla_q f = (\nabla_q^g f)\boldsymbol{M}, \qquad (36)$$

where $\nabla_q^g := (\partial/\partial g, \partial/\partial g^i, \partial/\partial g^j, \partial/\partial g^k)$ and $\boldsymbol{M}$ is a 4×4 matrix with element $M_{\mu\nu} = \partial g^\mu/\partial q^\nu$ for $\mu, \nu \in \{1, i, j, k\}$ and $g^\mu = -\mu g\mu$ (g^1 is understood the same as g). Explicitly, we can write

$$\frac{\partial f}{\partial q^\nu} = \sum_\mu \frac{\partial f}{\partial g^\mu}\frac{\partial g^\mu}{\partial q^\nu}. \qquad (37)$$

The proof is outlined in Appendix C.

6. The second chain rule: The above chain rule uses g and its involutions as the intermediate variables. It is sometimes convenient to use the real components of g for that purpose instead. In this case, the following chain rule may be used:

$$\nabla_q f = (\nabla_r^g f)\boldsymbol{O}, \qquad (38)$$

where $\boldsymbol{O}$ is a 4×4 matrix with entry $O_{\phi\nu} = \partial g_\phi/\partial q^\nu$ with $\phi \in \{a, b, c, d\}$ and $\nu \in \{1, i, j, k\}$, and $\nabla_r^g := (\partial/\partial g_a, \partial/\partial g_b, \partial/\partial g_c, \partial/\partial g_d)$. Explicitly, we have

$$\frac{\partial f}{\partial q^\nu} = \sum_\phi \frac{\partial f}{\partial g_\phi}\frac{\partial g_\phi}{\partial q^\nu}. \qquad (39)$$

7. The third chain rule: If the intermediate function $g(q)$ is real-valued, i.e., $g = g_a$, then from the second chain rule, we obtain

$$\frac{\partial f}{\partial q^\nu} = \frac{\partial f}{\partial g}\frac{\partial g}{\partial q^\nu}. \qquad (40)$$

8. $f(q)$ is not independent of q^i, q^j, or q^k in the sense that, in general,

$$\frac{\partial f(q)}{\partial q^i} \neq 0, \ \frac{\partial f(q)}{\partial q^j} \neq 0, \ \frac{\partial f(q)}{\partial q^k} \neq 0. \qquad (41)$$

This can be illustrated by $f(q) = q^2$. Using the first product rule (Eq. (32)), we have

$$\frac{\partial q^2}{\partial q^i} = q\frac{\partial q}{\partial q^i} + \frac{1}{4}\sum_{(\phi,\nu)}\frac{\partial q}{\partial q_\phi}q\nu$$

for $(\phi, \nu) \in \{(a, 1), (b, \mathrm{i}), (c, -\mathrm{j}), (d, -\mathrm{k})\}$. It can then be shown that

$$\frac{\partial q^2}{\partial q^{\mathrm{i}}} = q_b \mathrm{i}, \ \ \frac{\partial q^2}{\partial q^{\mathrm{j}}} = q_c \mathrm{j}, \ \ \frac{\partial q^2}{\partial q^{\mathrm{k}}} = q_d \mathrm{k}. \qquad (42)$$

This property demonstrates the intriguing difference between the HR derivative and the usual derivatives, although we can indeed show that

$$\frac{\partial q}{\partial q^\nu} = 0. \qquad (43)$$

One implication of this observation is that, for a nonlinear algorithm involving simultaneously more than one gradient $\partial f/\partial q^\nu$, we have to take care to include all the terms.

4 Restricted HR derivatives for a class of regular functions

Using the above operation rules, we can find explicit expressions for the derivatives for a whole range of functions. We first introduce the following lemma:

Lemma 1   The derivative of the power function $f(q) = (q - q_0)^n$, with integer n and constant quaternion q_0, is

$$\frac{\partial f(q)}{\partial q} = \frac{1}{2} \left(n\tilde{q}^{n-1} + \frac{\tilde{q}^n - \tilde{q}^{*n}}{\tilde{q} - \tilde{q}^*} \right) \qquad (44)$$

with $\tilde{q} = q - q_0$.

Remark 1   The division in $(\tilde{q}^n - \tilde{q}^{*n})/(\tilde{q} - \tilde{q}^*)$ is understood as $(\tilde{q}^n - \tilde{q}^{*n})(\tilde{q} - \tilde{q}^*)^{-1}$ or $(\tilde{q} - \tilde{q}^*)^{-1}(\tilde{q}^n - \tilde{q}^{*n})$ which are the same since the two factors commute. The division operations in what follows are understood in the same way.

Proof   The lemma is obviously true for $n = 0$. Letting $n \geq 1$, we apply the first product rule, and find

$$\frac{\partial (q - q_0)^n}{\partial q} = \tilde{q} \frac{\partial \tilde{q}^{n-1}}{\partial q} + \mathrm{R}(\tilde{q}^{n-1}), \qquad (45)$$

where $\mathrm{R}(\tilde{q}^{n-1})$ is the real part of $\tilde{q}^{n-1}$. We then obtain by induction

$$\frac{\partial (q - q_0)^n}{\partial q} = \sum_{m=0}^{n-1} \tilde{q}^m \mathrm{R}(\tilde{q}^{n-1-m}). \qquad (46)$$

Using $\mathrm{R}(\tilde{q}^{n-1-m}) = (\tilde{q}^{n-1-m} + \tilde{q}^{*(n-1-m)})/2$, the summations can be evaluated explicitly, leading to Eq. (44).

For $n < 0$, we use the recurrent relation

$$\frac{\partial ((q - q_0)^{-n})}{\partial q} = \tilde{q}^{-1} \left[\frac{\partial \tilde{q}^{-(n-1)}}{\partial q} - \mathrm{R}(\tilde{q}^{-n}) \right] \quad (47)$$

and the result

$$\frac{\partial (q - q_0)^{-1}}{\partial q} = -\tilde{q}^{-1} \mathrm{R}(\tilde{q}^{-1}). \qquad (48)$$

Eq. (44) is proven by using induction as for $n > 0$. More details are given in Appendix B.

Theorem 1   Assuming $f : H \rightarrow H$ admits a power series representation $f(q) := g(\tilde{q}) := \sum_{n=-\infty}^{\infty} a_n \tilde{q}^n$, with a_n being a quaternion constant and $\tilde{q} = q - q_0$, for $R_1 \leq |\tilde{q}| \leq R_2$ with $R_1, R_2 > 0$ being some constants, then we have

$$\frac{\partial f(q)}{\partial q} = \frac{1}{2} \left[f'(q) + (g(\tilde{q}) - g(\tilde{q}^*))(\tilde{q} - \tilde{q}^*)^{-1} \right], \qquad (49)$$

where $f'(q)$ is the derivative in the usual sense, i.e.,

$$f'(q) := \sum_{n=-\infty}^{\infty} n a_n \tilde{q}^{n-1} = \sum_{n=-\infty}^{\infty} n a_n (q - q_0)^{n-1}. \qquad (50)$$

Proof   Using Lemma 1 and the restricted left-linearity of HR gradients, we have

$$\begin{aligned}
\frac{\partial f}{\partial q} &= \frac{1}{2} \sum_{n=-\infty}^{\infty} a_n [n\tilde{q}^{n-1} + (\tilde{q}^n - \tilde{q}^{*n})(\tilde{q} - \tilde{q}^*)^{-1}] \\
&= f'(q) + \frac{1}{2} \left[\sum_{n=\infty}^{\infty} a_n (\tilde{q}^n - \tilde{q}^{*n}) \right] (\tilde{q} - \tilde{q}^*)^{-1} \\
&= \frac{1}{2} [f'(q) + (g(\tilde{q}) - g(\tilde{q}^*))(\tilde{q} - \tilde{q}^*)^{-1}],
\end{aligned}$$

which proves the theorem.

The functions $f(q)$ form a class of regular functions on H. A full discussion of such functions is beyond the scope of this paper. However, we note that a similar class of functions have been discussed in Gentili and Struppa (2007). A parallel development for the former is possible, and will be the topic of a future paper. Meanwhile, we observe that many useful elementary functions satisfy the conditions in Theorem 1. To illustrate the application of the theorem, we list below the derivatives of a number of such functions.

Example 1   Exponential function $f(q) = \mathrm{e}^q$ has representation

$$\mathrm{e}^q := \sum_{n=0}^{\infty} \frac{q^n}{n!}. \qquad (51)$$

Applying Theorem 1 with $a_n = 1/n!$ and $q_0 = 0$, we have

$$\frac{\partial e^q}{\partial q} = \frac{1}{2}\left(e^q + \frac{e^q - e^{q*}}{q - q^*}\right). \tag{52}$$

Making use of $e^q = e^{q_a + \hat{\mathbf{v}}v} = e^{q_a}e^{\hat{\mathbf{v}}v} = e^{q_a}(\cos v + \hat{\mathbf{v}}\sin v)$ with the representation of $q = q_a + \hat{\mathbf{v}}v$ and $\hat{\mathbf{v}}^2 = -1$, respectively, we have

$$\frac{\partial e^q}{\partial q} = \frac{1}{2}\left(e^q + e^{q_a}v^{-1}\sin v\right). \tag{53}$$

Example 2 The logarithmic function $f(q) = \ln q$ has representation

$$\ln q = \sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n}(q-1)^n, \tag{54}$$

with $a_n = (-1)^{n-1}/n$ and $q_0 = 1$. Since q_0 is a real number, $g(\tilde{q}^*) = f(q^*)$. Therefore, from Theorem 1 we have

$$\frac{\partial \ln q}{\partial q} = \frac{1}{2}\left(q^{-1} + \frac{\ln q - \ln q^*}{q - q^*}\right). \tag{55}$$

Using representation $\ln q = \ln|q| + \hat{\mathbf{v}}\arccos(q_a/|q|)$, the expression can be simplified as

$$\frac{\partial \ln q}{\partial q} = \frac{1}{2}\left(q^{-1} + \frac{1}{v}\arccos\frac{q_a}{|q|}\right), \tag{56}$$

where $v = |\mathrm{I}(q)|$.

Example 3 Hyperbolic tangent function $f(q) = \tanh q$ is defined as

$$\tanh q := \frac{e^q - e^{-q}}{e^q + e^{-q}} = q - \frac{q^3}{3} + \frac{2q^5}{15} - \cdots. \tag{57}$$

Therefore, Theorem 1 applies. On the other hand, using the relation $e^q = e^{q_a}(\cos v + \hat{\mathbf{v}}\sin v)$, we can show that

$$\tanh q = \frac{1}{2}\frac{\sinh(2q_a) + \hat{\mathbf{v}}\sin(2v)}{\sinh^2 q_a + \cos^2 v}. \tag{58}$$

Then the second term in the expression given by Theorem 1 can be simplified. The final expression can be written as

$$\frac{\partial \tanh q}{\partial q} = \frac{1}{2}\left(\mathrm{sech}^2 q + \frac{v^{-1}\sin(2v)}{\cosh(2q_a) + \cos(2v)}\right), \tag{59}$$

where $\mathrm{sech}q := 1/\cosh q$ is the quaternionic hyperbolic secant function.

Remark 2 Apparently, the derivatives for these functions can also be found by direct calculations without resorting to Theorem 1.

We now turn to a question of more theoretical interests. Even though it might not be obvious from the definitions, the following theorem shows that the restricted HR derivative is consistent with the derivative in the real domain for a class of functions, including those in the above examples:

Theorem 2 For the function $f(q)$ in Theorem 1, if q_0 is a real number, then

$$\frac{\partial f(q)}{\partial q} \to f'(q), \tag{60}$$

when $q \to \mathrm{R}(q)$, i.e., when q approaches a real number.

Proof Using the polar representation, we write $\tilde{q} = |\tilde{q}|\exp(\hat{\mathbf{v}}\theta)$, where $\theta = \arcsin(v/|\tilde{q}|)$ is the argument of $\tilde{q}$ with $v = |\mathrm{I}(\tilde{q})|$. Then $\tilde{q}^n = |\tilde{q}|^n\exp(n\hat{\mathbf{v}}\theta)$, and

$$(\tilde{q}^n - \tilde{q}^{*n})(\tilde{q} - \tilde{q}^*)^{-1} = \frac{\mathrm{I}(\tilde{q}^n)}{\mathrm{I}(\tilde{q})} = \frac{|\tilde{q}|^{n-1}\sin(n\theta)}{\sin\theta}. \tag{61}$$

For real q_0, $\tilde{q} \to q_a - q_0$ and $v \to 0$ when $q \to \mathrm{R}(q)$. There are two possibilities. First, if $q_a - q_0 \geq 0$, then $\theta \to 0$ at the limit. Thus,

$$\frac{\sin(n\theta)}{\sin\theta} \sim \frac{\sin(n\theta)}{\theta} \to n, \quad |\tilde{q}|^{n-1} \to (q_a - q_0)^{n-1}. \tag{62}$$

Therefore,

$$(\tilde{q}^n - \tilde{q}^{*n})(\tilde{q} - \tilde{q}^*)^{-1} \to n\tilde{q}^{n-1}, \tag{63}$$

$$[g(\tilde{q}) - g(\tilde{q}^*)](\tilde{q} - \tilde{q}^*)^{-1} \to \sum_{n=-\infty}^{\infty} na_n\tilde{q}^{n-1} = f'(q). \tag{64}$$

Thus,

$$\frac{\partial f(q)}{\partial q} \to \frac{1}{2}[f'(q) + f'(q)] = f'(q). \tag{65}$$

Second, if $q_a - q_0 < 0$, then $\theta \to \pi$. Thus,

$$\frac{\sin(n\theta)}{\sin\theta} \sim \frac{\sin(n\theta)}{\pi - \theta}. \tag{66}$$

Noting $\sin(n\theta) = \sin[n\pi - n(\pi - \theta)] = (-1)^{n-1}\sin[n(\pi - \theta)]$, we have

$$\frac{\sin(n\theta)}{\sin\theta} \sim \frac{(-1)^{n-1}\sin(n(\pi - \theta))}{\pi - \theta} \to (-1)^{n-1}n. \tag{67}$$

On the other hand, in this case $|\tilde{q}| \to -(q_a - q_0)$, and hence $|\tilde{q}|^{n-1} \to (-1)^{n-1}(q_a - q_0)^{n-1}$. Since $\tilde{q} \to q_a - q_0$, as a consequence, we have

$$(\tilde{q}^n - \tilde{q}^{*n})(\tilde{q} - \tilde{q}^*)^{-1} \to n\tilde{q}^{n-1}, \tag{68}$$

which is the same as Eq. (63). The proof then follows from the first case.

The functions in above three examples all satisfy the conditions in Theorem 2. Hence, we expect Theorem 2 applies. One can easily verify by direct calculations that the theorem indeed holds.

5 Right restricted HR gradients

In this section, we briefly summarize the results for the right restricted HR gradients, and highlight the difference from the left restricted HR gradients.

1. Right-linearity: For arbitrary quaternion constants α and β, and functions $f(q)$ and $g(q)$, we have

$$\frac{\partial^{\mathrm{R}}(f\alpha + g\beta)}{\partial q^{\nu}} = \frac{\partial^{\mathrm{R}} f}{\partial q^{\nu}}\alpha + \frac{\partial^{\mathrm{R}} g}{\partial q^{\nu}}\beta. \qquad (69)$$

However, linearity does not hold for left multiplications, i.e., in general $\partial^{\mathrm{R}}\alpha f/\partial q \neq \alpha\partial^{\mathrm{R}} f/\partial q$.

2. The first product rule: For the right restricted HR operator, the following product rule holds:

$$[\nabla_q^{\mathrm{R}}(fg)]^{\mathrm{T}} = [(\nabla_q^{\mathrm{R}} f)g]^{\mathrm{T}} + \boldsymbol{J}^*[f(\nabla_r g)^{\mathrm{T}}] . \qquad (70)$$

The second and third product rules are the same as those of the left restricted operator.

3. The first chain rule: For the composite function $f(g(q))$, we have

$$(\nabla_q^{\mathrm{R}} f)^{\mathrm{T}} = \boldsymbol{M}^{\mathrm{T}}(\nabla_q^{g\mathrm{R}} f)^{\mathrm{T}}. \qquad (71)$$

4. The second chain rule becomes $(\nabla_q^{\mathrm{R}} f)^{\mathrm{T}} = \boldsymbol{O}^{\mathrm{T}}(\nabla_r^g f)^{\mathrm{T}}$.

5. The third chain rule becomes $\partial^{\mathrm{R}} f/\partial q^{\nu} = (\partial g/\partial q^{\nu})(\partial f/\partial g)$. Note that $\partial g/\partial q^{\nu} = \partial^{\mathrm{R}} g/\partial q^{\nu}$ since g is real-valued. We thus have omitted the superscript 'R'. Also, $\partial f/\partial g$ is a real derivative, so there is no distinction between left and right derivatives.

We can also find the right restricted HR gradients for common quaternion functions. First of all, Lemma 1 is also true for right derivatives.

Lemma 2   For $f(q) = (q - q_0)^n$ with n an integer and q_0 a constant quaternion, we have

$$\frac{\partial^{\mathrm{R}} f(q)}{\partial q} = \frac{1}{2}\left(n\tilde{q}^{n-1} + \frac{\tilde{q}^n - \tilde{q}^{*n}}{\tilde{q} - \tilde{q}^*}\right), \qquad (72)$$

with $\tilde{q} = q - q_0$.

Remark 3   To prove the lemma, we use the following recurrent relations:

$$\frac{\partial (q - q_0)^n}{\partial q} = \frac{\partial \tilde{q}^{n-1}}{\partial q}\tilde{q} + \mathrm{R}(\tilde{q}^{n-1}), \qquad (73)$$

$$\frac{\partial((q - q_0)^{-n})}{\partial q} = \left[\frac{\partial \tilde{q}^{-(n-1)}}{\partial q} - \mathrm{R}(\tilde{q}^{-n})\right]\tilde{q}^{-1}. \quad (74)$$

Using Lemma 2, We can prove the following result:

Theorem 3   Assuming $f : H \to H$ admits a power series representation $f(q) := g(\tilde{q}) := \sum_{n=-\infty}^{\infty} \tilde{q}^n a_n$, with a_n being a quaternion constant and $\tilde{q} = q - q_0$, for $R_1 \leq |\tilde{q}| \leq R_2$ with $R_1, R_2 > 0$ being some constants, then we have

$$\frac{\partial^{\mathrm{R}} f(q)}{\partial q} = \frac{1}{2}\left[f'(q) + (\tilde{q} - \tilde{q}^*)^{-1}(g(\tilde{q}) - g(\tilde{q}^*))\right], \tag{75}$$

where $f'(q)$ is the derivative in the usual sense, i.e.,

$$f'(q) := \sum_{n=-\infty}^{\infty} n\tilde{q}^{n-1} a_n = \sum_{n=-\infty}^{\infty} n(q - q_0)^{n-1} a_n. \tag{76}$$

Note that, the functions $f(q)$ in Theorem 3 in general form a class of functions different from the one in Theorem 1, because in the series representation a_n appears on the right-hand side of the powers. However, if a_n is a real number, then the two classes of functions coincide. Therefore, we have the following result:

Theorem 4   If a_n is real, then the left and right restricted HR gradients of $f(q)$ coincide.

Remark 4   As a consequence, we can see immediately that the right derivatives for the exponential, logarithmic, and hyperbolic tangent functions are the same as the left ones.

Apparently, Theorem 2 is also true for the right derivatives. Hence, we have:

Theorem 5   The right-restricted HR gradient is consistent with the real gradient in the sense of Theorem 2.

6 Increment of a quaternion function

When $f(q)$ is a real-valued quaternion function, both left and right restricted HR gradients are coincident with the HR gradients. Besides, we have

$$\frac{\partial^{\mathrm{R}} f}{\partial q^{\nu}} = \frac{\partial f}{\partial q^{\nu}} = \left(\frac{\partial f}{\partial q}\right)^{\nu}, \qquad (77)$$

where $\nu \in \{1, \mathrm{i}, \mathrm{j}, \mathrm{k}\}$. Thus, only $\partial f/\partial q$ is independent. As a consequence (see also Mandic *et al.*

(2011)),

$$\mathrm{d}f = \sum_\nu \frac{\partial f}{\partial q^\nu}\mathrm{d}q^\nu = \sum_\nu \left(\frac{\partial f}{\partial q}\right)^\nu \mathrm{d}q^\nu$$
$$= \sum_\nu \left(\frac{\partial f}{\partial q}\mathrm{d}q\right)^\nu = 4\mathrm{R}\left(\frac{\partial f}{\partial q}\mathrm{d}q\right), \qquad (78)$$

where Eq. (77) has been used. Hence, $-(\partial f/\partial q)^*$ gives the steepest descent direction for f, and the increment is determined by $\partial f/\partial q$.

On the other hand, if f is a quaternion-valued function, the increment will depend on all four derivatives. Taking $f(q) = q^2$ as an example, we have (see Eqs. (42) and (44))

$$\mathrm{d}q^2 = (q + q_a)\mathrm{d}q + q_b\mathrm{i}\mathrm{d}q^\mathrm{i} + q_c\mathrm{j}\mathrm{d}q^\mathrm{j} + q_d\mathrm{k}\mathrm{d}q^\mathrm{k}, \quad (79)$$

even though $f(q)$ appears to be independent of q^i, q^j, and q^k. It can be verified that the above expression is the same as the differential form given in terms of $\mathrm{d}q_a$, $\mathrm{d}q_b$, $\mathrm{d}q_c$, and $\mathrm{d}q_d$. Thus, it is essential to include the contributions from $\partial f/\partial q^\mathrm{i}$, etc.

We also note that, if the right gradient is used consistently, the same increment would be produced, since the basis of the definitions is the same, namely, the differential form in terms of $\mathrm{d}q_a$, $\mathrm{d}q_b$, $\mathrm{d}q_c$, and $\mathrm{d}q_d$.

6.1 Quaternion-valued LMS algorithm

As an application, we now apply the quaternion-valued restricted HR gradient operator to develop the QLMS algorithm. Different versions of the QLMS algorithm have been derived in Barthelemy *et al.* (2014), Jiang *et al.* (2014b), and Tao and Chang (2014). However, with the rules we have derived that, some of the calculations can be simplified, as we will show below.

In terms of a standard adaptive filter, the output $y[n]$ and error $e[n]$ can be expressed as

$$y[n] = \boldsymbol{w}^\mathrm{T}[n]\boldsymbol{x}[n], \ e[n] = d[n] - \boldsymbol{w}^\mathrm{T}[n]\boldsymbol{x}[n], \quad (80)$$

where $\boldsymbol{w}[n]$ is the adaptive weight coefficient vector, $d[n]$ the reference signal, and $\boldsymbol{x}[n]$ the input sample vector. The conjugate $e^*[n]$ of the error signal $e[n]$ is

$$e^*[n] = d^*[n] - \boldsymbol{x}^\mathrm{H}[n]\boldsymbol{w}^*[n]. \qquad (81)$$

The cost function is defined as $J[n] = e[n]e^*[n]$, which is real-valued. According to the discussion above and Brandwood (1983) and Mandic *et al.*

(2011), the conjugate gradient $(\nabla_{\boldsymbol{w}} J[n])^*$ gives the maximum steepness direction for the optimization surface. Therefore, it is used to update the weight vector. Specifically,

$$\boldsymbol{w}[n+1] = \boldsymbol{w}[n] - \mu(\nabla_{\boldsymbol{w}} J[n])^*, \qquad (82)$$

where μ is the step size. To find $\nabla_{\boldsymbol{w}} J$, we use the first product rule:

$$\nabla_{\boldsymbol{w}} J = \frac{\partial e[n]e^*[n]}{\partial \boldsymbol{w}}$$
$$= e[n]\frac{\partial e^*[n]}{\partial \boldsymbol{w}} + \frac{1}{4}\left(\frac{\partial e[n]}{\partial \boldsymbol{w}_a}e^*[n] - \frac{\partial e[n]}{\partial \boldsymbol{w}_b}e^*[n]\mathrm{i}\right.$$
$$\left. - \frac{\partial e[n]}{\partial \boldsymbol{w}_c}e^*[n]\mathrm{j} - \frac{\partial e[n]}{\partial \boldsymbol{w}_d}e^*[n]\mathrm{k}\right).$$
$$(83)$$

After some algebra, we find $\nabla_{\boldsymbol{w}} J[n] = -\boldsymbol{x}[n]e^*[n]/2$, which leads to the following update equation for the QLMS algorithm:

$$\boldsymbol{w}[n+1] = \boldsymbol{w}[n] + \mu(e[n]\boldsymbol{x}^*[n]). \qquad (84)$$

6.2 Quaternion-valued nonlinear adaptive algorithm

Another application is the derivation of quaternion-valued adaptive filtering algorithms. We use the quaternion-valued hyperbolic tangent function as an example (Roberts and Jayabalan, 2015), so that the output $s[n]$ of the adaptive filter can be given by $s[n] = \tanh(y[n]) = \tanh(\boldsymbol{w}^\mathrm{T}[n]\boldsymbol{x}[n])$. The cost function is given by $J[n] = e[n]e^*[n]$, with $e[n] = d[n] - \tanh(\boldsymbol{w}^\mathrm{T}[n]\boldsymbol{x}[n])$.

Using the product rules in Eq. (83) and chain rules, and letting $y[n] = \boldsymbol{w}^\mathrm{T}[n]\boldsymbol{x}[n]$, we have

$$\frac{\partial e^*[n]}{\partial \boldsymbol{w}[n]} = -\left(\frac{\partial \tanh(y^*[n])}{\partial(y^*[n])_a}\frac{\partial(y^*[n])_a}{\partial \boldsymbol{w}[n]}\right.$$
$$+ \frac{\partial \tanh(y^*[n])}{\partial(y^*[n])_b}\frac{\partial(y^*[n])_b}{\partial \boldsymbol{w}[n]}$$
$$+ \frac{\partial \tanh(y^*[n])}{\partial(y^*[n])_c}\frac{\partial(y^*[n])_c}{\partial \boldsymbol{w}[n]}$$
$$\left. + \frac{\partial \tanh(y^*[n])}{\partial(y^*[n])_d}\frac{\partial(y^*[n])_d}{\partial \boldsymbol{w}[n]}\right).$$
$$(85)$$

Let $u = |\mathrm{I}(y)|$ and $\hat{\mathbf{u}} = \mathrm{I}(y)/u$. Then the quaternion $y = y_a + \mathrm{I}(y)$ can also be written as $y = y_a + u\hat{\mathbf{u}}$. Here, $\hat{\mathbf{u}}$ is a pure unit quaternion. Finally, the gradient can

be expressed as follows by using Eq. (58):

$$
\nabla_{\boldsymbol{w}} J[n] = \frac{1}{4(\sinh^2 y_a + \cos^2 u)^2}
$$

$$
\cdot \Bigg[\Big(2\sin(2u)(e_a \sin^2 y_a + \sin(2u)(e\hat{\mathbf{u}})_a)
$$

$$
+ (\cos u - \frac{\sin u}{u})(\sinh^2 y_a + \cos^2 u)(e\hat{\mathbf{u}})_a \Big) \boldsymbol{x}\hat{\mathbf{u}}
$$

$$
+ e_a \Big((\sinh^2 y_a + \cos^2 u)(\frac{\sin u}{u} - 4\cosh(2y_a))
$$

$$
+ \sinh(2y_a)(\sinh^2 y_a - \sin(2u)(e\hat{\mathbf{u}})_a) \Big) \boldsymbol{x}
$$

$$
+ 2\frac{\sin u}{u}(\sinh^2 y_a + \cos^2 u)(e\boldsymbol{x}_a + e^*\boldsymbol{x})_a) \Bigg].
\tag{86}
$$

Substituting the above result into Eq. (82) we can then obtain the update equation for the nonlinear adaptive algorithm.

On the other hand, if we use the series representation of $\tanh q$, we can obtain another form of the gradient function and the corresponding update equation becomes

$$
\boldsymbol{w}[n+1]
$$

$$
= \boldsymbol{w}[n] + \frac{1}{2}\mu \sum_{m=0}^{\infty} \sum_{r=0}^{m-1} a_m (\boldsymbol{x}^{\mathrm{H}}[n]\boldsymbol{w}^*[n])^{m-1-r}
$$

$$
\cdot e[n](\boldsymbol{x}^{\mathrm{H}}[n]\boldsymbol{w}^*[n])^r \boldsymbol{x}^*[n],
\tag{87}
$$

where a_m is the coefficient in the series representation of $\tanh(y[n])$, i.e., $\tanh(y[n]) = \sum_{m=0}^{\infty} a_m (y[n])^m$. It can be shown that if the items in the gradient part of the above expression are commutative, it will be reduced to the same form as in the real or complex domain.

7 Application to adaptive beamforming based on vector sensor arrays

As an example for the application of quaternion-valued signal processing, we here consider the reference signal based adaptive beamforming problem for vector sensor arrays consisting of multiple crossed-dipoles, where the earlier derived QLMS algorithm can be employed for beamforming.

7.1 Vector sensor arrays with a quaternion model

A general structure for a uniform linear array (ULA) with M crossed-dipole pairs is shown in

Fig. 1, where these pairs are located along the y-axis with an adjacent distance d, and at each location the two crossed components are parallel to the x-axis and y-axis, respectively. For a far-field incident signal with a direction of arrival (DOA) defined by the angles θ and ϕ, its spatial steering vector is given by

$$
\boldsymbol{S}_{\mathrm{c}}(\theta,\phi) = \begin{bmatrix} 1 \\ \exp(-\mathrm{j}2\pi d\sin\theta\sin\phi/\lambda) \\ \vdots \\ \exp(-\mathrm{j}2\pi(M-1)d\sin\theta\sin\phi/\lambda) \end{bmatrix},
\tag{88}
$$

where λ is the wavelength of the incident signal. For a crossed-dipole the spatial-polarization coherent vector can be given by Compton (1981), Li and Compton (1991), Zhang et al. (2014), and Hawes and Liu (2015):

$$
\boldsymbol{S}_{\mathrm{p}}(\theta,\phi,\gamma,\eta) = \begin{cases} [-\cos\gamma, \cos\theta\sin\gamma\mathrm{e}^{\mathrm{j}\eta}], & \phi = \dfrac{\pi}{2}, \\ [\cos\gamma, -\cos\theta\sin\gamma\mathrm{e}^{\mathrm{j}\eta}], & \phi = -\dfrac{\pi}{2}, \end{cases}
\tag{89}
$$

where γ is the auxiliary polarization angle with $\gamma \in [0, \pi/2]$, and $\eta \in [-\pi, \pi]$ is the polarization phase difference.
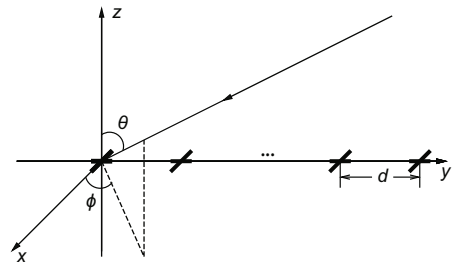

Fig. 1 A uniform linear array (ULA) with crossed-dipoles

The array structure can be divided into two sub-arrays: one parallel to the x-axis and one to the y-axis. The complex-valued steering vector of the x-axis sub-array is given by

$$
\boldsymbol{S}_x(\theta,\phi,\gamma,\eta) = \begin{cases} -\cos\gamma\, \boldsymbol{S}_{\mathrm{c}}(\theta,\phi), & \phi = \dfrac{\pi}{2}, \\ \cos\gamma\, \boldsymbol{S}_{\mathrm{c}}(\theta,\phi), & \phi = \dfrac{-\pi}{2}, \end{cases}
\tag{90}
$$

and for the y-axis it is expressed as

$$
\boldsymbol{S}_y(\theta,\phi,\gamma,\eta) = \begin{cases} \cos\theta\sin\gamma\mathrm{e}^{\mathrm{j}\eta}\boldsymbol{S}_{\mathrm{c}}(\theta,\phi), & \phi = \dfrac{\pi}{2}, \\ -\cos\theta\sin\gamma\mathrm{e}^{\mathrm{j}\eta}\boldsymbol{S}_{\mathrm{c}}(\theta,\phi), & \phi = \dfrac{-\pi}{2}. \end{cases}
\tag{91}
$$

Combining the two complex-valued subarray steering vectors together, an overall quaternion-valued steering vector with one real part and three imaginary parts can be constructed as

$$
\begin{aligned}
\boldsymbol{S}_{\mathrm{q}}&(\theta, \phi, \gamma, \eta)\\
&= \Re\{\boldsymbol{S}_x(\theta, \phi, \gamma, \eta)\} + \mathrm{i}\Re\{\boldsymbol{S}_y(\theta, \phi, \gamma, \eta)\}\\
&\quad + \mathrm{j}\Im\{\boldsymbol{S}_x(\theta, \phi, \gamma, \eta)\} + \mathrm{k}\Im\{\boldsymbol{S}_y(\theta, \phi, \gamma, \eta)\},
\end{aligned} \tag{92}
$$

where $\Re\{\cdot\}$ and $\Im\{\cdot\}$ are the real and imaginary parts of a complex number/vector, respectively. Given a set of coefficients, the response of the array is given by

$$
r(\theta, \phi, \gamma, \eta) = \boldsymbol{w}^{\mathrm{H}}\boldsymbol{S}_{\mathrm{q}}(\theta, \phi, \gamma, \eta), \tag{93}
$$

where $\boldsymbol{w}$ is the quaternion-valued weight vector.

7.2 Reference signal based adaptive beamforming

Suppose one of the incident signals to the array is the desired one and the remaining signals are interferences. Then the aim of beamforming is to receive the desired signal while suppressing the interferences at the output of the beamformer (Liu and Weiss, 2010). When a reference signal $d[n]$ is available, adaptive beamforming can be implemented by the standard adaptive filtering structure (Fig. 2), where $x_m[n]$ $(m = 1, 2, \cdots, M)$ are the received quaternion-valued input signals through the M pairs of crossed dipoles, and $w_m[n]$ $(m = 1, 2, \cdots, M)$ the corresponding quaternion-valued weight coefficients. The beamformer output $y[n]$ and the error signal $e[n]$ are

$$
\begin{cases}
y[n] = \boldsymbol{w}^{\mathrm{T}}[n]\boldsymbol{x}[n],\\
e[n] = d[n] - \boldsymbol{w}^{\mathrm{T}}[n]\boldsymbol{x}[n],
\end{cases} \tag{94}
$$

where

$$
\begin{cases}
\boldsymbol{w}[n] = [w_1[n], w_2[n], \ldots, w_M[n]]^{\mathrm{T}},\\
\boldsymbol{x}[n] = [x_1[n], x_2[n], \ldots, x_M[n]]^{\mathrm{T}}.
\end{cases} \tag{95}
$$

Simulations are performed based on such an array with 16 crossed dipoles and half-wavelength spacing using the QLMS algorithm in Eq. (84). The stepsize μ is set to 2×10^{-4}. A desired signal with a 20 dB signal-to-noise ratio (SNR) impinges from the broadside of the array ($\theta = 15°$) and two interfering signals with a signal-to-interference ratio (SIR) of -10 dB arrive from the directions $(30°, 90°)$ and $(15°, -90°)$, respectively. All the signals have the same polarisation of $(\gamma, \eta) = (30°, 0)$. The learning

curve obtained by averaging results from 200 simulation runs is shown in Fig. 3 and the resultant beam pattern is shown in Fig. 4, where for convenience positive values of θ indicate the value range $\theta \in [0°, 90°]$ for $\phi = 90°$, while negative values of $\theta \in [-90°, 0°]$ indicate an equivalent range of $\theta \in [0°, 90°]$ with $\phi = -90°$. We can see that the ensemble mean square error has reached almost -30 dB and two nulls have been formed successfully in the two interference directions, demonstrating the effectiveness of the quaternion-valued signal model and the derived QLMS algorithm.
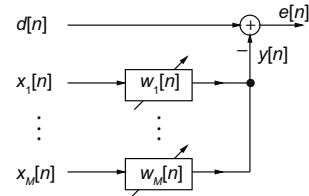

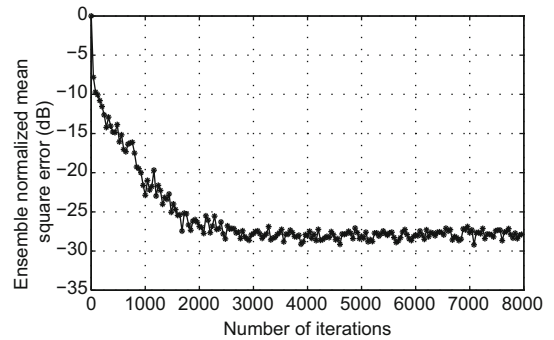**Fig. 2 Reference signal based adaptive beamforming structure**

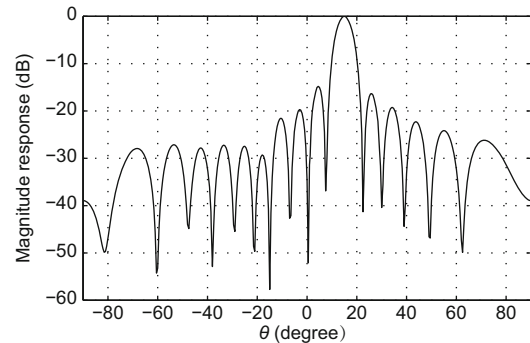**Fig. 3 Learning curve of the QLMS algorithm**

Fig. 4 Resultant beam pattern of the QLMS algorithm

8 Conclusions

We have proposed a restricted HR gradient operator and discussed its properties, in particular, several different versions of product rules and chain rules. Using the rules that we have established, we derived a general formula for the derivative of a large class of nonlinear quaternion-valued functions. The class includes the common elementary functions such as the exponential function and the logarithmic function. We also proved that, for a wide class of functions, the restricted HR gradient becomes the usual derivatives for real functions with respect to real variables, when the independent quaternion variable tends to the real axis, thus showing the consistency of the definition. Both linear and nonlinear adaptive filtering algorithms are derived to show the applications of the operator. An adaptive beamforming example based on vector sensor arrays has also been provided to demonstrate the effectiveness of the quaternion-valued signal model and the derived signal processing algorithm.

References

Barthelemy, Q., Larue, A., Mars, J.I., 2014. About QLMS derivations. *IEEE Signal Process. Lett.*, **21**(2):240-243.
http://dx.doi.org/10.1109/LSP.2014.2299066

Brandwood, D.H., 1983. A complex gradient operator and its application in adaptive array theory. *IEE Proc. F*, **130**(1):11-16.
http://dx.doi.org/10.1049/ip-f-1.1983.0003

Compton, R.T., 1981. On the performance of a polarization sensitive adaptive array. *IEEE Trans. Antenn. Propag.*, **29**(5):718-725.
http://dx.doi.org/10.1109/TAP.1981.1142651

Ell, T.A., Sangwine, S.J., 2007. Quaternion involutions and anti-involutions. *Comput. Math. Appl.*, **53**(1):137-143.
http://dx.doi.org/10.1016/j.camwa.2006.10.029

Ell, T.A., Le Bihan, N., Sangwine, S.J., 2014. Quaternion Fourier Transforms for Signal and Image Processing. Wiley, UK.

Gentili, G., Struppa, D.C., 2007. A new theory of regular functions of a quaternionic variable. *Adv. Math.*, **216**(1):279-301.
http://dx.doi.org/10.1016/j.aim.2007.05.010

Hawes, M., Liu, W., 2015. Design of fixed beamformers based on vector-sensor arrays. *Int. J. Antenn. Propag.*, **2015**:181937.1-181937.9.
http://dx.doi.org/10.1155/2015/181937

Jiang, M.D., Li, Y., Liu, W., 2014a. Properties and applications of a restricted HR gradient operator. arXiv:1407.5178.

Jiang, M.D., Liu, W., Li, Y., 2014b. A general quaternion-valued gradient operator and its applications to computational fluid dynamics and adaptive beamforming. Proc. 19th Int. Conf. on Digital Signal Processing,

p.821-826.
http://dx.doi.org/10.1109/ICDSP.2014.6900781

Jiang, M.D., Liu, W., Li, Y., 2014c. A zero-attracting quaternion-valued least mean square algorithm for sparse system identification. Proc. 9th Int. Symp. on Communication Systems, Networks and Digital Signal Processing, p.596-599.
http://dx.doi.org/10.1109/CSNDSP.2014.6923898

Le Bihan, N., Mars, J., 2004. Singular value decomposition of quaternion matrices: a new tool for vector-sensor signal processing. *Signal Process.*, **84**(7):1177-1199.
http://dx.doi.org/10.1016/j.sigpro.2004.04.001

Le Bihan, N., Miron, S., Mars, J.I., 2007. MUSIC algorithm for vector-sensors array using biquaternions. *IEEE Trans. Signal Process.*, **55**(9):4523-4533.
http://dx.doi.org/10.1109/TSP.2007.896067

Li, J., Compton, R.T., 1991. Angle and polarization estimation using ESPRIT with a polarization sensitive array. *IEEE Trans. Antenn. Propag.*, **39**:1376-1383.
http://dx.doi.org/10.1109/8.99047

Liu, H., Zhou, Y.L., Gu, Z.P., 2014. Inertial measurement unit-camera calibration based on incomplete inertial sensor information. *J. Zhejiang Univ.-Sci. C (Comput. & Electron.)*, **15**(11):999-1008.
http://dx.doi.org/10.1631/jzus.C1400038

Liu, W., 2014. Antenna array signal processing for a quaternion-valued wireless communication system. Proc. IEEE Benjamin Franklin Symp. on Microwave and Antenna Sub-systems, arXiv:1504.02921.

Liu, W., Weiss, S., 2010. Wideband Beamforming: Concepts and Techniques. Wiley, UK.

Mandic, D.P., Jahanchahi, C., Took, C.C., 2011. A quaternion gradient operator and its applications. *IEEE Signal Process. Lett.*, **18**(1):47-50.
http://dx.doi.org/10.1109/LSP.2010.2091126

Miron, S., Le Bihan, N., Mars, J.I., 2006. Quaternion-MUSIC for vector-sensor array processing. *IEEE Trans. Signal Process.*, **54**(4):1218-1229.
http://dx.doi.org/10.1109/TSP.2006.870630

Parfieniuk, M., Petrovsky, A., 2010. Inherently lossless structures for eight- and six-channel linear-phase paraunitary filter banks based on quaternion multipliers. *Signal Process.*, **90**(6):1755-1767.
http://dx.doi.org/10.1016/j.sigpro.2010.01.008

Pei, S.C., Cheng, C.M., 1999. Color image processing by using binary quaternion-moment-preserving thresholding technique. *IEEE Trans. Image Process.*, **8**(5):614-628.
http://dx.doi.org/10.1109/83.760310

Roberts, M.K., Jayabalan, R., 2015. An improved low-complexity sum-product decoding algorithm for low-density parity-check codes. *Front. Inform. Technol. Electron. Eng.*, **16**(6):511-518.
http://dx.doi.org/10.1631/FITEE.1400269

Sangwine, S.J., Ell, T.A., 2000. The discrete Fourier transform of a colour image. Proc. Image Processing II: Mathematical Methods, Algorithms and Applications, p.430-441.

Talebi, S.P., Mandic, D.P., 2015. A quaternion frequency estimator for three-phase power systems. Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing, p.3956-3960.
http://dx.doi.org/10.1109/ICASSP.2015.7178713

Talebi, S.P., Xu, D.P., Kuh, A., *et al.*, 2014. A quaternion least mean phase adaptive estimator. Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing, p.6419-6423.
http://dx.doi.org/10.1109/ICASSP.2014.6854840

Tao, J.W., 2013. Performance analysis for interference and noise canceller based on hypercomplex and spatio-temporal-polarisation processes. *IET Radar Sonar Navig.*, **7**(3):277-286.
http://dx.doi.org/10.1049/iet-rsn.2012.0151

Tao, J.W., Chang, W.X., 2014. Adaptive beamforming based on complex quaternion processes. *Math. Prob. Eng.*, **2014**:291249.1-291249.10.
http://dx.doi.org/10.1155/2014/291249

Zhang, X.R., Liu, W., Xu, Y.G., *et al.*, 2014. Quaternion-valued robust adaptive beamformer for electromagnetic vector-sensor arrays with worst-case constraint. *Signal Process.*, **104**:274-283.
http://dx.doi.org/10.1016/j.sigpro.2014.04.006

Appendix A: Definition of the operators

We consider $\mathrm{d}f = \mathrm{d}f_a + \mathrm{i}\mathrm{d}f_b + \mathrm{j}\mathrm{d}f_c + \mathrm{k}\mathrm{d}f_d$. By definition, we have $\mathrm{d}f_\gamma = \sum_\phi (\partial f_\gamma / \partial q_\phi)\mathrm{d}q_\phi$, with $\gamma, \phi \in \{a, b, c, d\}$. Using the relations

$$\mathrm{d}q_a = \frac{1}{4}\left(\mathrm{d}q + \mathrm{d}q^{\mathrm{i}} + \mathrm{d}q^{\mathrm{j}} + \mathrm{d}q^{\mathrm{k}}\right), \tag{A1}$$

$$\mathrm{d}q_b = \frac{1}{4\mathrm{i}}\left(\mathrm{d}q + \mathrm{d}q^{\mathrm{i}} - \mathrm{d}q^{\mathrm{j}} - \mathrm{d}q^{\mathrm{k}}\right), \tag{A2}$$

$$\mathrm{d}q_c = \frac{1}{4\mathrm{j}}\left(\mathrm{d}q - \mathrm{d}q^{\mathrm{i}} + \mathrm{d}q^{\mathrm{j}} - \mathrm{d}q^{\mathrm{k}}\right), \tag{A3}$$

$$\mathrm{d}q_d = \frac{1}{4\mathrm{k}}\left(\mathrm{d}q - \mathrm{d}q^{\mathrm{i}} - \mathrm{d}q^{\mathrm{j}} + \mathrm{d}q^{\mathrm{k}}\right), \tag{A4}$$

we can rewrite $\mathrm{d}f_\gamma$ as follows:

$$\begin{aligned}
\mathrm{d}f_\gamma = \;& \frac{1}{4}\left(\frac{\partial f_\gamma}{\partial q_a} - \mathrm{i}\frac{\partial f_\gamma}{\partial q_b} - \mathrm{j}\frac{\partial f_\gamma}{\partial q_c} - \mathrm{k}\frac{\partial f_\gamma}{\partial q_d}\right)\mathrm{d}q \\
&+ \frac{1}{4}\left(\frac{\partial f_\gamma}{\partial q_a} - \mathrm{i}\frac{\partial f_\gamma}{\partial q_b} + \mathrm{j}\frac{\partial f_\gamma}{\partial q_c} + \mathrm{k}\frac{\partial f_\gamma}{\partial q_d}\right)\mathrm{d}q^{\mathrm{i}} \\
&+ \frac{1}{4}\left(\frac{\partial f_\gamma}{\partial q_a} + \mathrm{i}\frac{\partial f_\gamma}{\partial q_b} - \mathrm{j}\frac{\partial f_\gamma}{\partial q_c} + \mathrm{k}\frac{\partial f_\gamma}{\partial q_d}\right)\mathrm{d}q^{\mathrm{j}} \\
&+ \frac{1}{4}\left(\frac{\partial f_\gamma}{\partial q_a} + \mathrm{i}\frac{\partial f_\gamma}{\partial q_b} + \mathrm{j}\frac{\partial f_\gamma}{\partial q_c} - \mathrm{k}\frac{\partial f_\gamma}{\partial q_d}\right)\mathrm{d}q^{\mathrm{k}},
\end{aligned}$$

which can be written as

$$\mathrm{d}f_\gamma = \frac{1}{4}\sum_\nu \left(\sum_{(\phi,\mu)} \frac{\partial f_\gamma}{\partial q_\phi}\mu^\nu\right)\mathrm{d}q^\nu, \tag{A5}$$

where $(\phi, \mu) \in \{(a,1), (b,-\mathrm{i}), (c,-\mathrm{j}), (d,-\mathrm{k})\}$, $\nu \in \{1, \mathrm{i}, \mathrm{j}, \mathrm{k}\}$, and μ^ν is the ν-involution of μ. Therefore,

$$\begin{aligned}
\mathrm{d}f &= \mathrm{d}f_a + \mathrm{i}\mathrm{d}f_b + \mathrm{j}\mathrm{d}f_c + \mathrm{k}\mathrm{d}f_d \\
&= \frac{1}{4}\sum_\nu \left(\sum_{(\phi,\mu)} \frac{\partial(f_a + \mathrm{i}f_b + \mathrm{j}f_c + \mathrm{k}f_d)}{\partial q_\phi}\mu^\nu\right)\mathrm{d}q^\nu \\
&= \frac{1}{4}\sum_\nu \left(\sum_{(\phi,\mu)} \frac{\partial f}{\partial q_\phi}\mu^\nu\right)\mathrm{d}q^\nu, \tag{A6}
\end{aligned}$$

which leads to Eqs. (11)–(19) in the main text. Note that, because μ^ν and $\mathrm{d}q^\nu$ are quaternions, to obtain the last equation, we need to multiply $\mathrm{d}f_b$, $\mathrm{d}f_c$, and $\mathrm{d}f_d$ by i, j, and k from the left.

On the other hand, we notice that the prefactors in Eqs. (A2)–(A4) can be moved to the right-hand side of the other factors; i.e., we can write

$$\mathrm{d}q_a = (\mathrm{d}q + \mathrm{d}q^{\mathrm{i}} + \mathrm{d}q^{\mathrm{j}} + \mathrm{d}q^{\mathrm{k}})\frac{1}{4}, \tag{A7}$$

$$\mathrm{d}q_b = (\mathrm{d}q + \mathrm{d}q^{\mathrm{i}} - \mathrm{d}q^{\mathrm{j}} - \mathrm{d}q^{\mathrm{k}})\frac{1}{4\mathrm{i}}, \tag{A8}$$

$$\mathrm{d}q_c = (\mathrm{d}q - \mathrm{d}q^{\mathrm{i}} + \mathrm{d}q^{\mathrm{j}} - \mathrm{d}q^{\mathrm{k}})\frac{1}{4\mathrm{j}}, \tag{A9}$$

$$\mathrm{d}q_d = (\mathrm{d}q - \mathrm{d}q^{\mathrm{i}} - \mathrm{d}q^{\mathrm{j}} + \mathrm{d}q^{\mathrm{k}})\frac{1}{4\mathrm{k}}. \tag{A10}$$

Using these relations, we can find another expression for $\mathrm{d}f_\gamma$ following the procedure above:

$$\mathrm{d}f_\gamma = \frac{1}{4}\sum_\nu \mathrm{d}q^\nu \left(\sum_{(\phi,\mu)} \mu^\nu \frac{\partial f_\gamma}{\partial q_\phi}\right). \tag{A11}$$

The expression is different from Eq. (A5), in that the differentials $\mathrm{d}q^\nu$ are on the left of μ^ν. Therefore, we derive

$$\begin{aligned}
\mathrm{d}f &= \mathrm{d}f_a + \mathrm{d}f_b\mathrm{i} + \mathrm{d}f_c\mathrm{j} + \mathrm{d}f_d\mathrm{k} \\
&= \frac{1}{4}\sum_\nu \mathrm{d}q^\nu \left(\sum_{(\phi,\mu)} \mu^\nu \frac{\partial(f_a + f_b\mathrm{i} + f_c\mathrm{j} + f_d\mathrm{k})}{\partial q_\phi}\right) \\
&= \frac{1}{4}\sum_\nu \mathrm{d}q^\nu \left(\sum_{(\phi,\mu)} \mu^\nu \frac{\partial f}{\partial q_\phi}\right), \tag{A12}
\end{aligned}$$

which is the basis for the definitions for the right restricted HR derivatives as given in the main text.

Appendix B: Additional details for the proof of Lemma 1

To prove Lemma 1, we have used the following relation:

$$\frac{\partial q^{-1}}{\partial q} = -q^{-1}\mathrm{R}(q^{-1}). \tag{B1}$$

To show this result, we note $\partial(qq^{-1})/\partial q = \partial 1/\partial q = 0$. Thus,

$$\begin{aligned}
0 &= q\frac{\partial q^{-1}}{\partial q} + \frac{1}{4}\left(q^{-1} - \mathrm{i}q^{-1}\mathrm{i} - \mathrm{j}q^{-1}\mathrm{j} - \mathrm{k}q^{-1}\mathrm{k}\right) \\
&= q\frac{\partial q^{-1}}{\partial q} + \mathrm{R}(q^{-1}),
\end{aligned} \tag{B2}$$

from which the result follows. We have used Eq. (11) and the fact that

$$\frac{\partial q}{\partial q_a} = 1, \ \frac{\partial q}{\partial q_b} = \mathrm{i}, \ \frac{\partial q}{\partial q_c} = \mathrm{j}, \ \frac{\partial q}{\partial q_d} = \mathrm{k}. \tag{B3}$$

The proof also uses the following recurrent relation:

$$\frac{\partial q^{-n}}{\partial q} = q^{-1}\left[\frac{\partial q^{-(n-1)}}{\partial q} - \mathrm{R}(q^{-n})\right], \tag{B4}$$

which can be shown as follows. By using the first product rule, we have

$$\begin{aligned}
\frac{\partial q^{-n}}{\partial q} = {}&q^{-1}\frac{\partial q^{-(n-1)}}{\partial q} + \frac{1}{4}\Bigg(\frac{\partial q^{-1}}{\partial q_a}q^{-(n-1)} \\
&-\frac{\partial q^{-1}}{\partial q_b}q^{-(n-1)}\mathrm{i} - \frac{\partial q^{-1}}{\partial q_c}q^{-(n-1)}\mathrm{j} - \frac{\partial q^{-1}}{\partial q_d}q^{-(n-1)}\mathrm{k}\Bigg).
\end{aligned} \tag{B5}$$

Using the fact $\partial qq^{-1}/\partial q_\phi = 0$ for $\phi \in \{a, b, c, d\}$, and the second product rule, we find

$$\frac{\partial q^{-1}}{\partial q_\phi} = -q^{-1}\frac{\partial q}{\partial q_\phi}q^{-1}. \tag{B6}$$

Thus,

$$\begin{aligned}
\frac{\partial q^{-n}}{\partial q} = {}&q^{-1}\frac{\partial q^{-(n-1)}}{\partial q} \\
&- \frac{q^{-1}}{4}\left(q^{-n} - \mathrm{i}q^{-n}\mathrm{i} - \mathrm{j}q^{-n}\mathrm{j} - \mathrm{k}q^{-n}\mathrm{k}\right) \\
= {}&q^{-1}\frac{\partial q^{-(n-1)}}{\partial q} - q^{-1}\mathrm{R}(q^{-n}).
\end{aligned} \tag{B7}$$

Appendix C: Derivations of the first chain rule

The function $f(g(q))$ can be viewed as a function of intermediate variables g_a, g_b, g_c, and g_d. Using the usual chain rule, we have

$$\frac{\partial f}{\partial q_\beta} = \sum_\phi \frac{\partial f}{\partial g_\phi}\frac{\partial g_\phi}{\partial q_\beta}, \tag{C1}$$

with $\beta \in \{a, b, c, d\}$, which gives $\nabla_r f = (\nabla_r^g f)\boldsymbol{P}$, where $\boldsymbol{P}$ is a 4×4 matrix with $P_{\phi\beta} = \partial g_\phi/\partial q_\beta$. With $(\nabla_r f)\boldsymbol{J}^{\mathrm{H}} = \nabla_q f$ and $\nabla_r^g f = 4(\nabla_q^g f)\boldsymbol{J}$, the above equation leads to

$$\nabla_q f = 4(\nabla_q^g f)\boldsymbol{JPJ}^{\mathrm{H}}, \tag{C2}$$

where it is easy to show that $4\boldsymbol{JPJ}^{\mathrm{H}} = \boldsymbol{M}$.

Dr. Wei LIU, corresponding author of this invited paper, received his B.Sc. in 1996 and LL.B. in 1997, both from Peking University, China, master degree from University of Hong Kong, China, in 2001, and PhD degree in the School of Electronics and Computer Science from University of Southampton, UK, in 2003. He later worked as a postdoctor at Imperial College London. Since Sept. 2005, he has been with the Department of Electronic and Electrical Engineering, University of Sheffield, UK, first as a lecturer, and then a senior lecturer. His research interests are in sensor array signal processing, blind signal processing, multirate signal processing, and their various applications in wireless communications, radar, sonar, satellite navigation, human computer interface, renewable energy exploitation, etc. He has now published more than 180 journal and conference papers, three book chapters, and a research monograph about wideband beamforming. He is a senior member of IEEE, an elected member of the Digital Signal Processing Technical Committee of the IEEE Circuits and Systems Society, currently associate editor of *IEEE Transactions on Signal Processing* and editorial board member of *Frontiers of Information Technology & Electronic Engineering*.