The Quaternion LMS Algorithm for Adaptive Filtering of Hypercomplex Processes

Clive Cheong Took, *Member, IEEE*, and Danilo P. Mandic, *Senior Member, IEEE*

***Abstract*—The quaternion least mean square (QLMS) algorithm is introduced for adaptive filtering of three- and four-dimensional processes, such as those observed in atmospheric modeling (wind, vector fields). These processes exhibit complex nonlinear dynamics and coupling between the dimensions, which make their component-wise processing by multiple univariate LMS, bivariate complex LMS (CLMS), or multichannel LMS (MLMS) algorithms inadequate. The QLMS accounts for these problems naturally, as it is derived directly in the quaternion domain. The analysis shows that QLMS operates inherently based on the so called "augmented" statistics, that is, both the covariance $E\{\mathbf{x}\mathbf{x}^H\}$ and pseudocovariance $E\{\mathbf{x}\mathbf{x}^T\}$ of the tap input vector $\mathbf{x}$ are taken into account. In addition, the operation in the quaternion domain facilitates fusion of heterogeneous data sources, for instance, the three vector dimensions of the wind field and air temperature. Simulations on both benchmark and real world data support the approach.**

***Index Terms*—Adaptive multistep ahead prediction, data fusion via vector spaces, multidimensional adaptive filters, quaternion signal processing, wind modeling.**

I. Introduction

DUE to its simplicity and robustness, the least mean square (LMS) algorithm has been at the core of adaptive filtering applications [1], [2], and its online adaptive mode of operation makes it suited for the processing of nonstationary real-world signals. These attractive properties have led to its applications in noise reduction, radar/sonar signal processing, channel equalization for cellular mobile phones, echo cancelation, and low delay speech coding [3]. The LMS update can be expressed as

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \mu e(n)\mathbf{x}(n) \tag{1}$$

where $\mathbf{w}(n)$, $e(n)$, μ, and $\mathbf{x}(n)$ denote, respectively, the adaptive weight vector, instantaneous output error, step size, and the input data vector of length L. Extensions proposed to improve the performance of LMS include those based on the optimization of the step size [4], filtering of error gradients [5], and adaptive filter length [6].

To process *bi-variate* signals, such as those in digital communications, chaotic maps [7], and vector fields [8], the LMS algorithm was extended to the complex domain [9]. Recently, Mandic *et al.* have exploited the bivariate model of wind [8], [10] in this context; this was achieved by using the so-called augmented statistics [11]. In many other fields, the simultaneous processing of the two dimensions of a signal (radar, sonar) can lead to a more efficient signal processing algorithm than processing each dimension separately. As the quaternion domain represents an extension of the complex field, it is natural to ask whether we can extend the class of LMS algorithms to cater for adaptive filtering of three- and four- dimensional (hyper-complex) signals.

Quaternions can be regarded as a noncommutative extension of complex numbers, and comprise at most four variables [12], [13]. A quaternion variable $q \in \mathbb{H}$ which has a real/scalar part $\Re\{q\}$ (denoted with subscript a), and a vector part $\Im\{q\}$ comprising of three imaginary parts (denoted with subscripts b, c, and d), can be expressed as

$$\begin{aligned} q &= [\Re\{q\}, \Im\{q\}] = [q_a, \mathbf{q}] \in \mathbb{H} \\ &= [q_a, (q_b, q_c, q_d)] \\ &= q_a + q_b\imath + q_c\jmath + q_d\kappa \;\; \{q_a, q_b, q_c, q_d \in \mathbb{R}\}. \end{aligned} \tag{2}$$

Quaternions have been used for more than 150 years (conceived by Hamilton in 1843) and have found applications in computer graphics, for the modeling of three-dimensional (3-D) rotations [14], in robotics [15], and molecular modeling [16]. Within the image and signal processing community, Pei and Cheng employed quaternions to process color images [17], Toyoshima implemented efficient hyper-complex digital filters [18], Bülow and Sommer used a hyper-complex representation in texture segmentation [19], whereas Zarzoso utilized quaternions to help solve source separation problems [20]. Le Bihan *et al.* used quaternions in watermarking [21]; they also proposed quaternion algorithms for spectrum estimation, such as a fast complexified quaternion Fourier transform [22], quaternion singular value decomposition (QSVD) and MUSIC algorithm to process polarized waves [23], [24]. Although the standard least squares problem has also been addressed in the quaternion domain [16], [25], [26], adaptive filtering algorithms for the processing of quaternion valued signals are lacking.

The recent progress in technology, environmental sciences, robotics, and biomedicine, has highlighted the need for adaptive filtering of several important classes of multidimensional signals, for instance, 3-D wind field measured by three axis

Manuscript received April 10, 2008; revised October 10, 2008. First published December 09, 2008; current version published March 11, 2009. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Roberto Lopez-Valcarce.

The authors are with the Department of Electrical and Electronic Engineering, Imperial College London, London SW7 2AZ, U.K. (e-mail: c.cheong-took@ic.ac.uk; d.mandic@ic.ac.uk).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSP.2008.2010600

anemometers. By processing those data directly in the multidimensional domain where they reside, we can exploit the correlation and coupling between each dimension and therefore provide enhanced modeling. To this end, we propose the quaternion least mean square (QLMS) algorithm. To cater for noncircularly symmetric distributions in $\mathbb{H}$, similarly to the widely linear model in $\mathbb{C}$ [27], and augmented CLMS (ACLMS) [28], [29], we also investigate the benefits of so called augmented statistics.

The organization of the paper is as follows; in Section II we briefly review the elements of quaternion algebra necessary for the development of QLMS adaptive filters. In Section III, QLMS and its augmented version are derived. This is followed by a statistical analysis on both QLMS and AQLMS algorithms in Section IV. Section V compares the performances of the proposed approaches against the univariate LMS, bivariate CLMS and multichannel LMS [30]. Simulations are based on both benchmark data and real world three-dimensional wind field data. Section VI concludes the paper.

II. Quaternion Algebra

The properties of the orthogonal unit vectors, $\imath$, $\jmath$, κ describing the three vector dimensions of a quaternion are

$$\begin{aligned} \imath\jmath &= \kappa \\ \jmath\kappa &= \imath \\ \kappa\imath &= \jmath \\ \imath\jmath\kappa &= \imath^2 = \jmath^2 = \kappa^2 = -1. \end{aligned} \tag{3}$$

Due to the noncommutativity of the quaternion, for example, $\jmath\imath \neq \imath\jmath$, instead $\jmath\imath = -\kappa$. Other elements of quaternion algebra that are used in this work include the multiplication given by

$$\begin{aligned} q_1q_2 &= [q_{a,1}, \mathbf{q}_1][q_{a,2}, \mathbf{q}_2] \\ &= [q_{a,1}q_{a,2} - \mathbf{q}_1 \cdot \mathbf{q}_2,\ q_{a,1}\mathbf{q}_2 + q_{a,2}\mathbf{q}_1 + \mathbf{q}_1 \times \mathbf{q}_2] \end{aligned} \tag{4}$$

where $q = q_a + q_b\imath + q_c\jmath + q_d\kappa = [q_a, \mathbf{q}]$. Symbols "$\cdot$" and "$\times$" denote respectively the dot-product and the cross-product, the conjugate of a quaternion $q^* = [q_a, \mathbf{q}]^* = [q_a, -\mathbf{q}]$, and the norm $||q||_2^2 = qq^*$. Note, that quaternion conjugation is anti-involution, that is, $(q_1q_2)^* = q_2^*q_1^*$. A quaternion is said to be pure, if its real part vanishes. The quaternion vector space $\mathbb{H}$ forms a noncommutative normed division algebra, that is

$$q_1q_2 \neq q_2q_1.$$

For an introduction to quaternions we refer to [31]; for a more advanced reading, we recommend [32]. In this paper, unless otherwise stated, all the quantities are quaternion-valued, for instance $\mathbf{w}(n)$ is a vector of quaternions, while $e(n)$ is a quaternion variable.

III. Derivation of the QLMS and Its Variant AQLMS

Based on the quaternion algebra, and standard stochastic gradient approximation, we shall now derive the quaternion LMS (QLMS) algorithm for quaternion-valued linear adaptive finite-impulse-response (FIR) filters.

A. The Quaternion LMS

The same real-valued quadratic cost function (the quaternion norm) as in LMS and CLMS is used, that is

$$\begin{aligned} \mathcal{J}(n) &= e(n)e^*(n) \\ &= e_a^2(n) + e_b^2(n) + e_c^2(n) + e_d^2(n) \end{aligned} \tag{5}$$

where the error[1] $e(n) = d(n) - \mathbf{w}^T(n)\mathbf{x}(n)$, with $d(n)$, $\mathbf{w}(n)$, and $\mathbf{x}(n)$ denoting respectively the desired signal, the adaptive weight vector, and the filter input. Symbols $(\cdot)^T$, $(\cdot)^H$, and $(\cdot)^*$ denote respectively the transpose, Hermitian, and quaternion conjugate operator. Based on the cost function (5), within the steepest descent optimization, the following gradients need to be calculated[2]

$$\begin{aligned} \nabla_{\mathbf{w}}\big(e(n)e^*(n)\big) = &\nabla_{\mathbf{w}_a}\big(e(n)e^*(n)\big) + \nabla_{\mathbf{w}_b}\big(e(n)e^*(n)\big)\imath \\ &+\nabla_{\mathbf{w}_c}\big(e(n)e^*(n)\big)\jmath + \nabla_{\mathbf{w}_d}\big(e(n)e^*(n)\big)\kappa \end{aligned} \tag{7}$$

where $\mathbf{w} = \mathbf{w}_a + \mathbf{w}_b\imath + \mathbf{w}_c\jmath + \mathbf{w}_d\kappa$ and

$$\begin{aligned} \nabla_{\mathbf{w}_a}\big(e(n)e^*(n)\big) &= e(n)\nabla_{\mathbf{w}_a}\big(e^*(n)\big) + \nabla_{\mathbf{w}_a}\big(e(n)\big)e^*(n) \\ \nabla_{\mathbf{w}_b}\big(e(n)e^*(n)\big) &= e(n)\nabla_{\mathbf{w}_b}\big(e^*(n)\big) + \nabla_{\mathbf{w}_b}\big(e(n)\big)e^*(n) \\ \nabla_{\mathbf{w}_c}\big(e(n)e^*(n)\big) &= e(n)\nabla_{\mathbf{w}_c}\big(e^*(n)\big) + \nabla_{\mathbf{w}_c}\big(e(n)\big)e^*(n) \\ \nabla_{\mathbf{w}_d}\big(e(n)e^*(n)\big) &= e(n)\nabla_{\mathbf{w}_d}\big(e^*(n)\big) + \nabla_{\mathbf{w}_d}\big(e(n)\big)e^*(n). \end{aligned} \tag{8}$$

Subsequently, the update of the adaptive weight vector of QLMS can be expressed as (the full derivation of the gradient can be found in Appendix VIII-A)

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \mu\big(2e(n)\mathbf{x}^*(n) - \mathbf{x}^*(n)e^*(n)\big). \tag{9}$$

Due to the noncommutativity of the quaternion product, there are two terms within the gradient of the cost function, that is, $2e(n)\mathbf{x}^*(n)$ and $\mathbf{x}^*(n)e^*(n)$. The QLMS update includes the term $\mu e(n)\mathbf{x}^*(n)$ which is similar to that within the complex LMS [9], together with an additional term $\mathbf{x}^*(n)e^*(n)$ which

[1]In this work we use the formulation $y(n) = \mathbf{w}^T(n)\mathbf{x}(n)$ following on Widrow's CLMS [9], however $y(n) = \mathbf{x}^T(n)\mathbf{w}(n)$ can also be used as a starting point. Unlike the LMS and CLMS, due to the noncommutativity of quaternion product, the formulation $y(n) = \mathbf{x}^T(n)\mathbf{w}(n)$ requires separate derivation of QLMS, and can be addressed similarly to the correspondence between (9) and (39), when the cost function is written as $\mathcal{J}(n) = e^*(n)e(n)$ instead of $\mathcal{J}(n) = e(n)e^*(n)$.

[2]The noncommutativity of the quaternion product and the conjugation properties are taken into account, that is

$$\begin{aligned} \nabla_{\mathbf{w}}\mathcal{J}(n) = &e(n)\Big(\nabla_{\mathbf{w}_a}\Big(e^*(n)\Big) + \nabla_{\mathbf{w}_b}\Big(e^*(n)\Big)\imath \\ &\quad + \nabla_{\mathbf{w}_c}\Big(e^*(n)\Big)\jmath + \nabla_{\mathbf{w}_d}\Big(e^*(n)\Big)\kappa\Big) \\ &+ \Big(\nabla_{\mathbf{w}_a}\Big(e(n)\Big) + \nabla_{\mathbf{w}_b}\Big(e(n)\Big)\imath \\ &\quad + \nabla_{\mathbf{w}_c}\Big(e(n)\Big)\jmath + \nabla_{\mathbf{w}_d}\Big(e(n)\Big)\kappa\Big)e^*(n) \\ \neq &e(n)\Big(\nabla_{\mathbf{w}_a}\Big(e^*(n)\Big) + \nabla_{\mathbf{w}_b}\Big(e^*(n)\Big)\imath \\ &\quad + \nabla_{\mathbf{w}_c}\Big(e^*(n)\Big)\jmath + \nabla_{\mathbf{w}_d}\Big(e^*(n)\Big)\kappa\Big) \\ &+ e^*(n)\Big(\nabla_{\mathbf{w}_a}\Big(e(n)\Big) + \nabla_{\mathbf{w}_b}\Big(e(n)\Big)\imath \\ &\quad + \nabla_{\mathbf{w}_c}\Big(e(n)\Big)\jmath + \nabla_{\mathbf{w}_d}\Big(e(n)\Big)\kappa\Big) \end{aligned} \tag{6}$$

is specific to the quaternion domain. To answer whether QLMS simplify exactly into CLMS when quaternion-valued signals are limited to two dimensions, let the imaginary parts $\jmath$ and κ of quaternions $\mathbf{x}(n)$ and $e(n)$ vanish. The QLMS for such a case simplifies into

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \mu\big([e_a(n)\mathbf{x}_a(n) + 3e_b(n)\mathbf{x}_b(n)] + \imath[-e_a(n)\mathbf{x}_b(n) + 3e_b(n)\mathbf{x}_a(n)]\big).$$

A comparison with the CLMS update

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \mu\big([e_a(n)\mathbf{x}_a(n) + e_b(n)\mathbf{x}_b(n)] + \imath[-e_a(n)\mathbf{x}_b(n) + e_b(n)\mathbf{x}_a(n)]\big)$$

shows that QLMS does not simplify exactly into CLMS, highlighting the direct multidimensional mode of operation. This is also the case if any other combination of two dimensions of a quaternion are made to vanish, however, it can be shown that if the quaternion data are in the "isomorphic" form of $q = q_a + Qi_r$, where $Q = \sqrt{q_b^2 + q_c^2 + q_d^2}$, and $i_r = (q_b\imath + q_c\jmath + q_d\kappa)/Q$, the derivation of QLMS is exactly identical to that of CLMS.

B. The Augmented QLMS (AQLMS)

Motivated by some recent developments in so-called augmented complex statistics, we now derive the augmented QLMS algorithm, which is capable of dealing with the generality of quaternion data. It is usually assumed that the statistics in $\mathbb{C}$ are a simple extension of the statistics in $\mathbb{R}$, obtained by replacing the $(\cdot)^T$ operator by the $(\cdot)^H$ operator in the corresponding second order statistical moments. For example, the covariance $E\{\mathbf{x}\mathbf{x}^T\}$ in the real domain is replaced by $E\{\mathbf{x}\mathbf{x}^H\}$ in the complex domain. This is, however, not adequate for noncircular data (for more detail, see [10] and [33]). Recently, for the processing of real world data which can be made complex by convenience of representation, Goh and Mandic (for wind processing [10] and [28]), Novey and Adali (for source separation applications [34] and [35]) and, Schreier and Scharf (for communication applications [11]) have highlighted the need to adopt the so-called "augmented statistics," a concept introduced by Picinbono [36] and [37]. The use of augmented statistics is crucially important when processing noncircular complex signals; noncircularity or improperness is a second order statistical property, which can be defined as

$$\mathcal{C}_{\mathbf{xx}} = E\{\mathbf{x}\mathbf{x}^H\} \neq 0 \quad \mathcal{P}_{\mathbf{xx}} = E\{\mathbf{x}\mathbf{x}^T\} \neq 0 \tag{10}$$

that is, the pseudocovariance $\mathcal{P}_{\mathbf{xx}}$ does not vanish for a noncircular complex signal.[3] In the context of quaternion statistics, properness (known as $\mathbb{Q}$-properness) is defined as the invariance of the probability density function (pdf) under some special angle rotations [38], [39]. More recently, Amblard and Le Bihan extended the definition of properness by proposing $\mathbb{C}$-properness and $\mathbb{H}$-properness [39]. There is, however, no explicit mention of the role of pseudo-covariance in quantifying properness of quaternions. More specifically, the pseudo-covariance does not vanish even for a $\mathbb{Q}$-proper signal.[4] Motivated by the augmented CLMS [28], augmented CRTRL [40], and augmented statistics for wind profile [8], we now investigate the benefits of including the pseudocovariance $\mathcal{P}_{\mathbf{xx}}$ into the QLMS algorithm. In order for QLMS to cater for *general* quaternion processes, we employ a quaternion-valued widely linear model [37], given by

$$\mathbf{y}(n) = \mathbf{w}^T(n)\mathbf{x}(n) + \mathbf{g}^T(n)\mathbf{x}^*(n). \tag{11}$$

This model incorporates both the information contained in the covariance and pseudocovariance (for more detail, see [27]). For the quaternion scenario, the update for vector $\mathbf{g}$ in (11) can be found similarly to that for QLMS, and is given by (full derivation can be found in Appendix VIII-B)

$$\mathbf{g}(n+1) = \mathbf{g}(n) + \mu\big(2e(n)\mathbf{x}(n) - \mathbf{x}(n)e^*(n)\big). \tag{12}$$

Again, the noncommutativity of the quaternion products must be taken into account during the derivation of the update (12). Finally, (9) and (12) can be combined into a compact "augmented" form as

$$\mathbf{h}^a(n) = [\,\mathbf{w}^T(n) \quad \mathbf{g}^T(n)\,]^T \tag{13}$$

and the weight update of the augmented QLMS (AQLMS), can be expressed as

$$\mathbf{h}^a(n+1) = \mathbf{h}^a(n) + \mu\big[2e^a(n)\mathbf{x}^{a^*}(n) - \mathbf{x}^{a^*}(n)e^{a^*}(n)\big] \tag{14}$$

where the augmented error and input vector are given by

$$e^a(n) = d(n) - \mathbf{h}^{a^T}(n)\mathbf{x}^a(n) \quad \mathbf{x}^a(n) = [\,\mathbf{x}^T(n) \quad \mathbf{x}^H(n)\,]^T. \tag{15}$$

In the complex domain, it has been shown that adaptive algorithms based on augmented statistics exhibit advantages over standard algorithms, for data which are not circularly symmetric [40], [41].

IV. Properties of QLMS Algorithms

The noncommutativity of quaternion product make their algebraic manipulation demanding. One way to circumvent this problem is to treat the real/scalar $\Re\{\cdot\}$ and the vector $\Im\{\cdot\}$ part of a quaternion separately, similarly to [42]. The analysis will be based on the following two observations:

[3]For a complex circular signal, the pseudo-covariance $\mathcal{P}_{\mathbf{xx}}$ vanishes, that is, $E\{\mathbf{x}\mathbf{x}^T\} = E\{\mathbf{x}_r\mathbf{x}_r^T\} - E\{\mathbf{x}_i\mathbf{x}_i^T\} + \imath E\{\mathbf{x}_r\mathbf{x}_i^T\} + \imath E\{\mathbf{x}_i\mathbf{x}_r^T\} = 0$. This is because for a circular signal, the real part $\mathbf{x}_r$ and the imaginary part $\mathbf{x}_i$ have the same covariance, but are uncorrelated.

[4]For a $\mathbb{Q}$-proper signal, the pseudo-covariance $\mathcal{P}_{\mathbf{xx}}$ does not vanish, that is, $E\{\mathbf{x}\mathbf{x}^T\} = E\{\mathbf{x}_a\mathbf{x}_a^T\} - E\{\mathbf{x}_b\mathbf{x}_b^T\} - E\{\mathbf{x}_c\mathbf{x}_c^T\} - E\{\mathbf{x}_d\mathbf{x}_d^T\} + \imath 2E\{\mathbf{x}_a\mathbf{x}_b^T\} + \jmath 2E\{\mathbf{x}_a\mathbf{x}_c^T\} + \kappa 2E\{\mathbf{x}_a\mathbf{x}_d^T\} \neq 0$. This is because for a $\mathbb{Q}$-proper signal, the scalar/real part $\mathbf{x}_a$ and the vector/imaginary part $\mathbf{x}_{b,c,d}$ have equal variances, but are uncorrelated. This follows from conditions (4) and (5) of [38, Theorem 2].

1) *Property 1:*

$$y = -y^* \text{ iff } \Re\{y\} = 0. \quad (16)$$

2) *Property 2:*

$$y = y^* \text{ iff } \Im\{y\} = 0. \quad (17)$$

To analyze the QLMS algorithms, we shall now make the standard assumption in adaptive filtering that $d(n) = \mathbf{w}_{\text{opt}}^T \mathbf{x}(n)$ [43]. Following the standard analysis of the convergence in the mean [43], the weight error vector is defined as

$$\mathbf{v}(n) = \mathbf{w}(n) - \mathbf{w}_{\text{opt}} \quad (18)$$

where $\mathbf{w}_{\text{opt}}$ is the optimal weight vector, while the error $e(n)$ between the desired signal $d(n)$ and its estimate $y(n)$ is given by

$$\begin{aligned} e(n) &= d(n) - y(n) \\ &= \mathbf{w}_{\text{opt}}^T \mathbf{x}(n) - \mathbf{w}^T(n)\mathbf{x}(n) \\ &= \left(\mathbf{w}_{\text{opt}}^T - \mathbf{w}^T(n)\right)\mathbf{x}(n) = -\mathbf{v}^T(n)\mathbf{x}(n). \end{aligned} \quad (19)$$

A. Analysis of QLMS

From the QLMS update (9), the real part can be computed as

$$\begin{aligned} \Re\{\mathbf{w}(n+1)\} = \Re\{\mathbf{w}(n)\} \\ +\Re\left\{\mu\left(2e(n)\mathbf{x}^*(n) - \mathbf{x}^*(n)e^*(n)\right)\right\}. \end{aligned} \quad (20)$$

By employing Property 2, it can be shown that (20) is equivalent to

$$\begin{aligned} \Re\{\mathbf{w}(n+1)\} = \Re\{\mathbf{w}(n)\} \\ +2\Re\{\mu(e(n)\mathbf{x}^*(n)\} - \Re\{\mu e(n)\mathbf{x}(n)\}. \end{aligned} \quad (21)$$

Substitute (19) into (21) to yield

$$\begin{aligned} \Re\{\mathbf{w}(n+1)\} &= \Re\{\mathbf{w}(n)\} - 2\Re\{\mu \mathbf{v}^T(n)\mathbf{x}(n)\mathbf{x}^*(n)\} \\ &\quad + \Re\{\mu \mathbf{v}^T(n)\mathbf{x}(n)\mathbf{x}(n)\} \\ &= \Re\{\mathbf{w}(n)\} - 2\Re\left\{\mu\left(\mathbf{v}^T(n)\mathbf{x}(n)\mathbf{x}^H(n)\right)^T\right\} \\ &\quad + \Re\left\{\mu\left(\mathbf{v}^T(n)\mathbf{x}(n)\mathbf{x}^T(n)\right)^T\right\}. \end{aligned} \quad (22)$$

Substract $\mathbf{w}_{\text{opt}}$ from both sides of (22) to give

$$\begin{aligned} &\Re\{\mathbf{v}(n+1)\} \\ &\quad = \Re\{\mathbf{v}(n)\} - 2\Re\left\{\mu\left(\mathbf{v}^T(n)\mathbf{x}(n)\mathbf{x}^H(n)\right)^T\right\} \\ &\quad + \Re\left\{\mu\left(\mathbf{v}^T(n)\mathbf{x}(n)\mathbf{x}^T(n)\right)^T\right\} \\ &\Re\{\mathbf{v}^T(n+1)\} \\ &\quad = \Re\left\{\mathbf{v}^T(n)\left(\mathbf{I} + \mu \underbrace{\left[\mathbf{x}(n)\mathbf{x}^T(n) - 2\mathbf{x}(n)\mathbf{x}^H(n)\right]}_{\text{Real QLMS statistics}}\right)\right\}. \end{aligned} \quad (23)$$

From (23), we can see that in terms of statistics, QLMS includes both the pseudocovariance $\mathcal{P}_{\mathbf{xx}} = E\{\mathbf{x}\mathbf{x}^T\}$ and the covariance $\mathcal{C}_{\mathbf{xx}} = E\{\mathbf{x}\mathbf{x}^H\}$. This is a major difference as compared with CLMS, and therefore, it is expected that the QLMS and augmented QLMS will have similar performance.

The vector part $\Im\{\cdot\}$ of the QLMS update (9) can be analyzed using Property 1 and (19), that is

$$\begin{aligned} \Im\{\mathbf{w}(n+1)\} &= \Im\{\mathbf{w}(n)\} \\ &\quad + \Im\{\mu[2e(n)\mathbf{x}^*(n) - \mathbf{x}^*(n)e^*(n)]\} \quad (24) \\ &= \Im\{\mathbf{w}(n)\} + \Im\{\mu 2e(n)\mathbf{x}^*(n)\} \\ &\quad + \Im\{\mu e(n)\mathbf{x}(n)\} \\ &= \Im\{\mathbf{w}(n)\} - \Im\{\mu[2\mathbf{v}^T(n)\mathbf{x}(n)\mathbf{x}^H(n)]^T\} \\ &\quad -\Im\{\mu[\mathbf{v}^T(n)\mathbf{x}(n)\mathbf{x}^T(n)]^T\} \quad (25) \end{aligned}$$

yielding

$$\begin{aligned} &\Im\{\mathbf{w}^T(n+1)\} = \Im\{\mathbf{w}^T(n)\} \\ &-\Im\{\mu(2\mathbf{v}^T(n)\mathbf{x}(n)\mathbf{x}^H(n)\} - \Im\{\mu \mathbf{v}^T(n)\mathbf{x}(n)\mathbf{x}^T(n)\}. \end{aligned} \quad (26)$$

Subtract $\mathbf{w}_{\text{opt}}$ from both sides of (26) to give

$$\begin{aligned} &\Im\{\mathbf{v}^T(n+1)\} \\ &\quad = \Im\{\mathbf{v}^T(n)\} - \Im\{\mu(2\mathbf{v}^T(n)\mathbf{x}(n)\mathbf{x}^H(n)\} \\ &\quad - \Im\{\mu \mathbf{v}^T(n)\mathbf{x}(n)\mathbf{x}^T(n)\} \\ &\quad = \Im\left\{\mathbf{v}^T(n)\left(\mathbf{I} + \mu \underbrace{\left[-\mathbf{x}(n)\mathbf{x}^T(n) - 2\mathbf{x}(n)\mathbf{x}^H(n)\right]}_{\text{Vector QLMS statistics}}\right)\right\}. \end{aligned} \quad (27)$$

Again, both the pseudocovariance and the covariance estimates are involved in the weight update of QLMS. This indicates that the "augmented" statistics is inherent to the QLMS, which is a unique property of this class of algorithms. We next proceed to establish the extent to which AQLMS has advantages over QLMS.

B. Analysis of AQLMS

To investigate statistical properties of AQLMS, define the error $e(n)$ in terms of the "augmented" weight error vectors $\mathbf{v}_{\mathbf{w}}(n)$ and $\mathbf{v}_{\mathbf{g}}(n)$ to give

$$\begin{aligned} e(n) &= d(n) - \left[\mathbf{w}^T(n)\mathbf{x}(n) + \mathbf{g}^T(n)\mathbf{x}^*(n)\right] \\ &= \left[\mathbf{w}_{\text{opt}}^T \mathbf{x}(n) + \mathbf{g}_{\text{opt}}^T \mathbf{x}^*(n)\right] \\ &\quad - \left[\mathbf{w}^T(n)\mathbf{x}(n) + \mathbf{g}^T(n)\mathbf{x}^*(n)\right] \\ &= -\left[\mathbf{v}_{\mathbf{w}}^T(n)\mathbf{x}(n) + \mathbf{v}_{\mathbf{g}}^T(n)\mathbf{x}^*(n)\right] \end{aligned} \quad (28)$$

where $\mathbf{v_w}(n) = \mathbf{w}(n) - \mathbf{w}_{\rm opt}$ and $\mathbf{v_g}(n) = \mathbf{g}(n) - \mathbf{g}_{\rm opt}$. Based on Property 2, the real/scalar part $\Re\{\cdot\}$ of the AQLMS update of $\mathbf{w}(n)$ (9) can now be written as

$$\Re\{\mathbf{w}(n+1)\} = \Re\{\mathbf{w}(n)\} + \mu\Re\{e(n)\left[2\mathbf{x}^*(n) - \mathbf{x}(n)\right]\}. \quad (29)$$

Replace the error $e(n)$ with its augmented counterpart, to give

$$\begin{aligned}\Re\{\mathbf{w}(n+1)\} &= \Re\{\mathbf{w}(n)\} - \mu\Re\{\left[\mathbf{v}_\mathbf{w}^T(n)\mathbf{x}(n) + \mathbf{v}_\mathbf{g}^T(n)\mathbf{x}^*(n)\right] \left[2\mathbf{x}^*(n) - \mathbf{x}(n)\right]\} \\ &= \Re\{\mathbf{w}(n)\} - \mu\Re\{\mathbf{v}_\mathbf{w}^T(n)\mathbf{x}(n)\left[2\mathbf{x}^*(n) - \mathbf{x}(n)\right]\} \\ &\quad - \mu\Re\{\mathbf{v}_\mathbf{g}^T(n)\mathbf{x}^*(n)\left[2\mathbf{x}^*(n) - \mathbf{x}(n)\right]\}. \end{aligned} \quad (30)$$

Substract $\mathbf{w}_{\rm opt}(n)$ from both sides of (30) to obtain

$$\begin{aligned}\Re\left\{\mathbf{v}_\mathbf{w}^T(n+1)\right\} &= \Re\{\mathbf{v}_\mathbf{w}^T(n)\left[\mathbf{I} - 2\mu\mathbf{x}(n)\mathbf{x}^H(n) - \mu\mathbf{x}(n)\mathbf{x}^T(n)\right]\} \\ &\quad - \mu\Re\{\mathbf{v}_\mathbf{g}^T(n)\left[2\mathbf{x}^*(n)\mathbf{x}^H(n) - \mathbf{x}^*(n)\mathbf{x}^T(n)\right]\}. \end{aligned} \quad (31)$$

Observe that the statistics of AQLMS include the covariance $\mathbf{x}(n)\mathbf{x}^H(n)$, the pseudocovariance $\mathbf{x}(n)\mathbf{x}^T(n)$ and their conjugates. Similarly to the analysis of QLMS, we have

$$\begin{aligned}\Re\left\{\mathbf{v}_\mathbf{g}^T(n+1)\right\} &= \Re\{\mathbf{v}_\mathbf{g}^T(n)\left[\mathbf{I} - 2\mu\mathbf{x}^*(n)\mathbf{x}^T(n) - \mu\mathbf{x}^*(n)\mathbf{x}^H(n)\right]\} \\ &\quad - \mu\Re\{\mathbf{v}_\mathbf{w}^T(n)\left[2\mathbf{x}(n)\mathbf{x}^T(n) - \mathbf{x}(n)\mathbf{x}^H(n)\right]\} \quad (32) \\ \Im\left\{\mathbf{v}_\mathbf{w}^T(n+1)\right\} &= \Im\{\mathbf{v}_\mathbf{w}^T(n)\left[\mathbf{I} - 2\mu\mathbf{x}(n)\mathbf{x}^H(n) - \mu\mathbf{x}(n)\mathbf{x}^T(n)\right]\} \\ &\quad - \mu\Im\{\mathbf{v}_\mathbf{g}^T(n)\left[2\mathbf{x}^*(n)\mathbf{x}^H(n) + \mathbf{x}^*(n)\mathbf{x}^T(n)\right]\} \quad (33) \\ \Im\left\{\mathbf{v}_\mathbf{g}^T(n+1)\right\} &= \Im\{\mathbf{v}_\mathbf{g}^T(n)\left[\mathbf{I} - 2\mu\mathbf{x}^*(n)\mathbf{x}^T(n) - \mu\mathbf{x}^*(n)\mathbf{x}^H(n)\right]\} \\ &\quad - \mu\Im\{\mathbf{v}_\mathbf{w}^T(n)\left[2\mathbf{x}(n)\mathbf{x}^T(n) + \mathbf{x}(n)\mathbf{x}^H(n)\right]\}. \quad (34)\end{aligned}$$

It is shown in the simulations that the use of augmented statistics provides minor improvement in the performance, (due to the *deterministic* relationship between $\mathcal{C}_\mathbf{xx}$ ($\mathcal{P}_\mathbf{xx}$) and $\mathcal{C}_\mathbf{xx}^*$ ($\mathcal{P}_\mathbf{xx}^*$)).

C. Choice of Parameters of QLMS

The choice of parameters of QLMS is crucial to its performance, however, e.g., determining the range of step size is not trivial, as it requires eigendecomposition of the correlation matrix. The difficulty arises from the noncommutativity of the quaternion product, which gives rise to the notion of the left and right eigenvalue decomposition [13]. Furthermore, the left eigenvalue decomposition of a quaternion is still an ongoing research topic [44]. Variants of the proposed class of QLMS algorithms are pretty much along those introduced for the LMS, this is however beyond the scope of this paper.

V. Simulations

Since prediction is at the core of adaptive filtering, our simulation was conducted in the prediction setting, for M-step ahead prediction. For a quantitative assessment of the prediction performance, we employ the prediction gain R_p [45], given by

$$R_p = 10\log\frac{\sigma_\mathbf{x}^2}{\sigma_e^2}\ (\text{dB}) \quad (35)$$

where $\sigma_\mathbf{x}^2$ and σ_e^2 denote respectively the estimated variances of the input and the error. The prediction gain was measured at the steady state. Two input processes were considered, the well-known atmospheric motion inspired, chaotic signal-the Lorenz attractor, and a real-world three-dimensional wind field. For rigor, the performances of QLMS and AQLMS were compared with multiple univariate LMS applied component-wise, CLMS [9], and multichannel LMS [30]. Within the four-channel LMS setting, the jth output of the multichannel adaptive filter is given by [30]

$$y_j(n) = \sum_{i=1}^{4}\mathbf{h}_{ij}^T(n)\mathbf{x}_i(n) \quad j = 1,\ldots,4 \quad (36)$$

where the adaptive weight vector $\mathbf{h}_{ij}(n) = [h_{ij}(n),\ldots,h_{ij}(n-L+1)]^T$ corresponds to the ith input vector $\mathbf{x}_i(n) = [x_i(n),\ldots,x_i(n-L+1)]^T$ and jth output $y_j(n)$ channel. The update for each coefficient vector $\mathbf{h}_{ij}(n)$ is given by [30]

$$\begin{aligned}\mathbf{h}_{ij}(n+1) &= \mathbf{h}_{ij}(n) + \Delta\mathbf{h}_{ij}(n) \\ &= \mathbf{h}_{ij}(n) + \mu e_j(n)\mathbf{x}_i(n) \quad i,j = 1,\ldots,4. \end{aligned} \quad (37)$$

The error $e_j(n) = d_j(n) - y_j(n)$ is a scalar instantaneous output error corresponding to the jth channel $y_j(n)$.

A. Experiment 1: Lorenz Attractor—Atmospheric Convection Rolls Prediction

The Lorenz attractor is a three-dimensional nonlinear system used originally to model atmospheric turbulence, but also to model lasers, dynamos, and the motion of waterwheel [7]. Mathematically, the Lorenz system can be expressed as a system of coupled differential equations

$$\begin{aligned}\frac{\partial x}{\partial t} &= \alpha(y-x) \\ \frac{\partial y}{\partial t} &= x(\rho - z) - y \\ \frac{\partial z}{\partial t} &= xy - \beta z \end{aligned} \quad (38)$$

where $\alpha,\rho,\beta > 0$. Lorenz attractor, shown in Fig. 1, can be regarded as a pure quaternion, and is used as a benchmark signal to test the performance of the QLMS algorithms. For a chaotic behavior of Lorenz attractor, the parameters were selected as

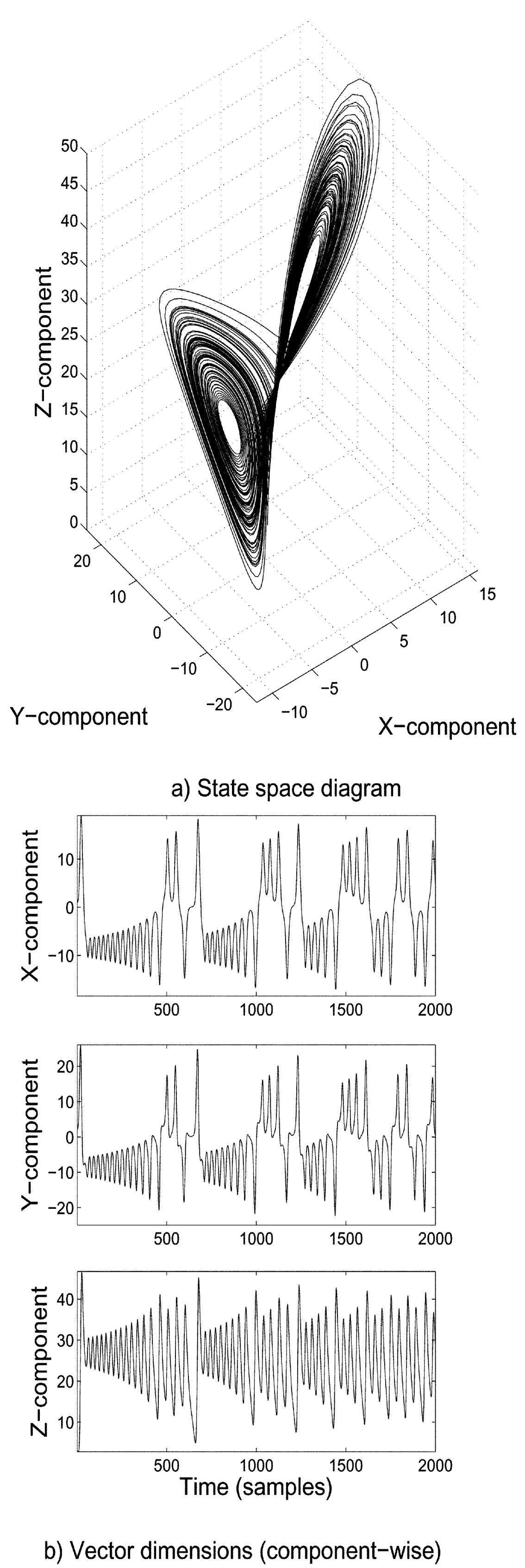

Fig. 1. The 3-D Lorenz attractor with parameters $\alpha = 10$, $\rho = 28$, and $\beta = 8/3$ in (38).

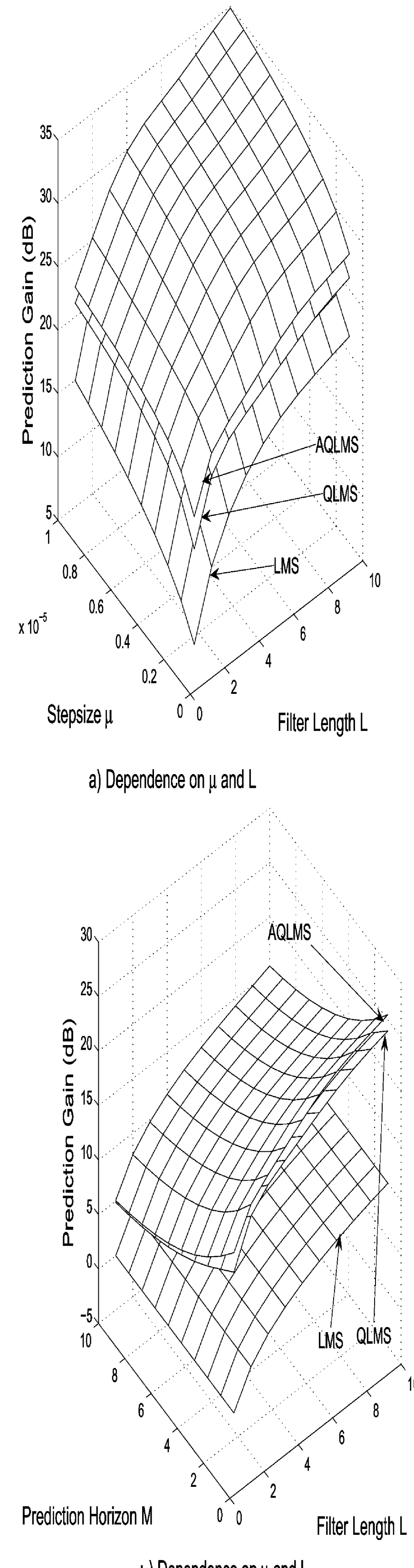

Fig. 2. Performance of LMS (applied componentwise), QLMS, and AQLMS on the prediction of 3-D Lorenz signal.

$\alpha = 10$, $\rho = 28$, and $\beta = 8/3$. Fig. 2 demonstrates the performance of LMS, QLMS, and AQLMS, as a function of the prediction horizon M (with $\mu = 10^{-6}$ constant) and the step size ($M = 1$ constant) for varying filter length.

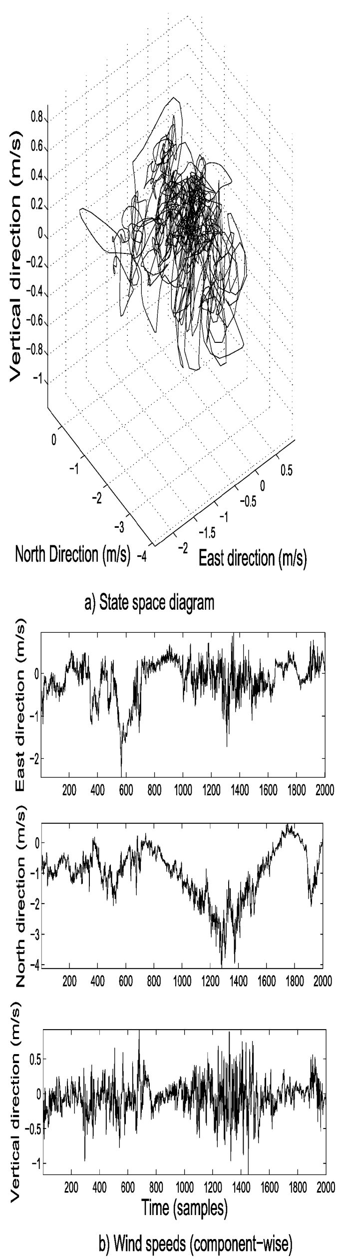

Fig. 3. Three-dimensional wind signal.

The strong correlation between dimensions of the Lorenz attractor explains the better performance of the QLMS approaches over multiple univariate LMS applied componentwise, as shown in Fig. 2.

B. Experiment 2: Wind Forecasting

Wind forecasting at short scales plays an important role in renewable energy, air pollution modeling and aviation safety [46]. In the simulations, 3-D wind speed data (a segment shown in Fig. 3) were used,[5] together with air temperature measurements.

[5]The wind data were recorded by Prof. K. Aihara and his team at the University of Tokyo, in an urban environment. The wind data was initially sampled at 50 Hz, but resampled at 5 Hz for simulation purposes.

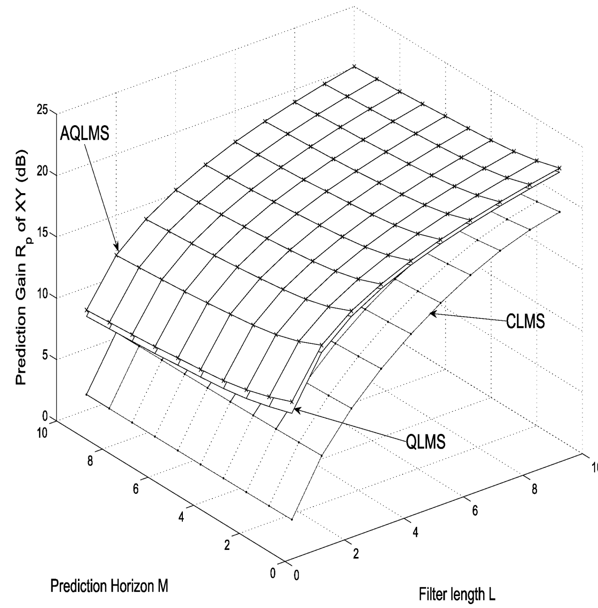

Fig. 4. Performance of CLMS, QLMS and AQLMS for the prediction of wind. The R_p was calculated based on the east (X) and north (Y) dimensions.

In the first set of experiments, wind was considered as a pure quaternion, that is, the real part was zero. For a fair comparison between QLMS and CLMS, the prediction gain R_p was calculated based on two dimensions at a time and over a range of the prediction horizons M and filter lengths L, keeping the step size constant $\mu = 10^{-3}$. Figs. 4–6 illustrate the performances when the prediction gain R_p was computed based on east–north, east–vertical, and north–vertical directions, respectively. In all the cases, QLMS and AQLMS outperformed standard CLMS. Figs. 4–6 illustrate that both QLMS and AQLMS have similar performance, with AQLMS outperforming QLMS, conforming with the analysis in Sections IV-A and IV-B. The significant performance advantage over CLMS is due to the fact that QLMS and AQLMS fully exploit the information in the three dimensions of the wind data. Another factor which contributes to the enhanced performance of QLMS algorithms is that owing to their quaternionic nature, the so called the "augmented" statistics is inherent to the weight updates.

C. Experiment 3: Data Fusion Via Quaternion Spaces

To demonstrate the ability of quaternion models in the fusion of heterogeneous data sources [47], the air temperature was used as a scalar/real part of the quaternion, whereas the three wind directions were the vector part. An experiment was conducted to investigate whether the joint quaternionic model of the temperature and 3-D wind vector would lead to improved performance. Fig. 7 shows that the 4D quaternion model of wind provided enhanced performance for both the QLMS and AQLMS. The next experiment comparing the performance of two CLMS (combined into a quaternion output) against QLMS approaches is illustrated in Fig. 8. This was achieved for the best empirical choice of the parameters of the QLMS and a pair of CLMS.

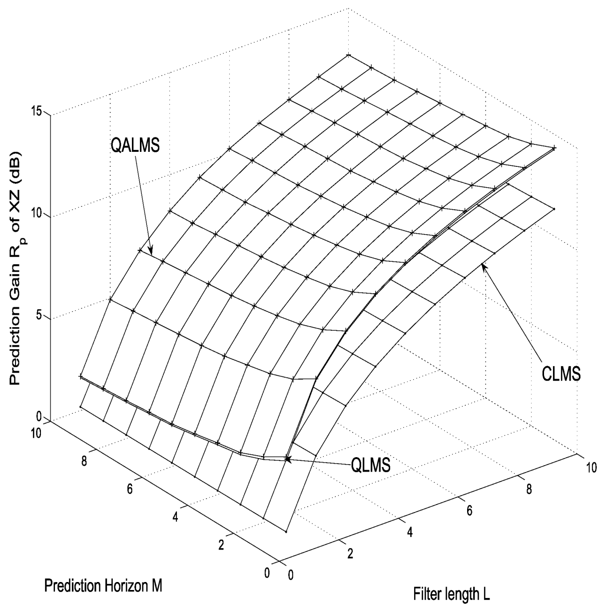

Fig. 5. Performance of CLMS, QLMS and AQLMS for M step ahead prediction of wind. The R_p was calculated based on the east (X) and vertical (Z) dimensions.

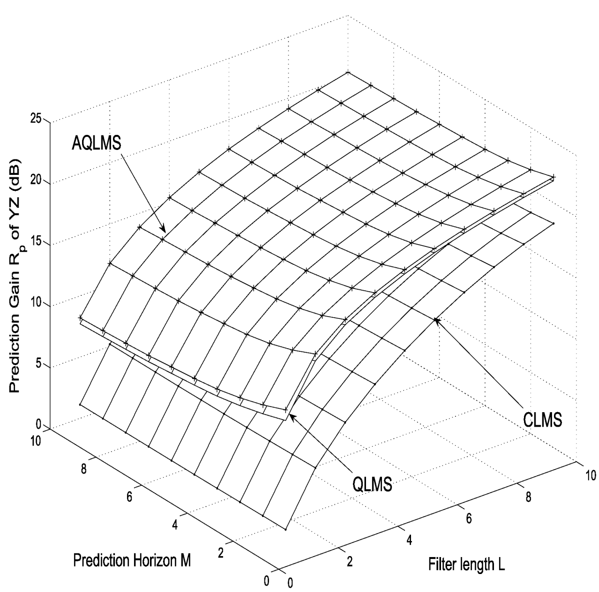

Fig. 6. Performance of CLMS, QLMS, and AQLMS for the M step ahead prediction of wind. The R_p was calculated based on the north (Y) and vertical (Z) dimensions.

The CLMS did not yield satisfactory prediction of the temperature dimension, whereas both the QLMS and AQLMS algorithms exhibited excellent performance. In the context of filtering a *quartet* of signals, another aspect that needs to be addressed is the computational complexity, summarized in Table I. The computational complexity of QLMS is seven times that of LMS, three times that of CLMS, and less than two times that of multichannel LMS [30]. For a fair comparison in terms of computational complexity, the bottom plot of Fig. 9 compares performances of QLMS ($\mathcal{O}(56L)$) and the combined CLMS ($\mathcal{O}(60L)$), with the filter length L of CLMS three times that of QLMS. The QLMS exhibited superior performance. In the

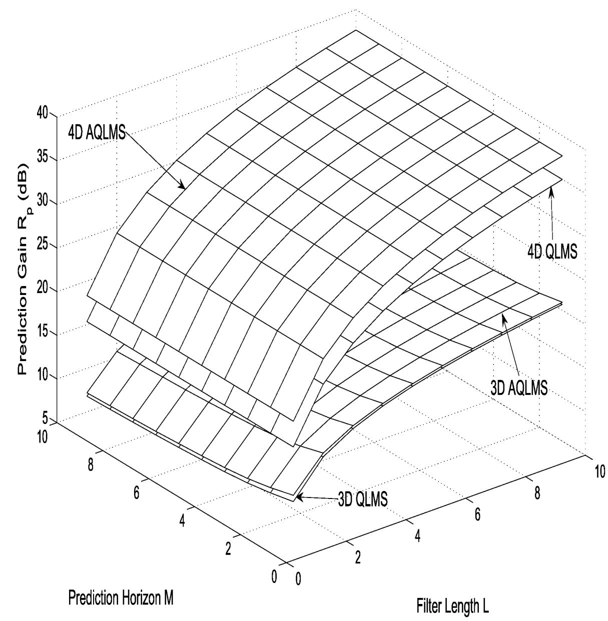

Fig. 7. Comparison of the performances of 3-D QLMS, 3-D AQLMS, 4D QLMS, and 4D AQLMS for the "data fusion" model of wind. The use of temperature as a scalar/real dimension and the 3-D wind data as the vector dimensions resulted in a much improved performance.

TABLE I
COMPUTATIONAL COMPLEXITIES OF THE ALGORITHMS

Algorithms	Additions	Multiplications
4× LMS	8L	8L+4
2× CLMS	12L	20L
1× MLMS	32L	32L+16
1× QLMS	48L	56L
1× AQLMS	112L	96L

next set of simulations, the performance of multichannel LMS (MLMS) [30] was compared against the QLMS algorithms, as shown in Fig. 10. The performances were similar for small step sizes; however, both QLMS and AQLMS outperformed MLMS with an increase in filter length.

VI. Concluding Remarks

A class of quaternion least mean square (QLMS) stochastic gradient adaptive filtering algorithms has been designed for adaptive filtering of hyper-complex processes. Such three- and four- dimensional processes (e.g., the 3-D Lorenz attractor, and 4D wind model) exhibit complex nonlinear dynamics, together with the coupling between their components, which makes their processing by the multiple univariate LMS and a pair of complex LMS (CLMS) inadequate. A rigorous analysis has shown that QLMS incorporates both the covariance and pseudocovariance terms within its update and can therefore cater for noncircularly symmetric quaternion data. For rigor, the augmented QLMS (AQLMS) has also been derived by taking into account the so-called augmented second order statistics. Further, it has been shown that the operation in the quaternion domain allows for the fusion of heterogenous data sources. Simulation results on both the benchmark 3-D data (Lorenz attractor) and real world wind data support the approach.

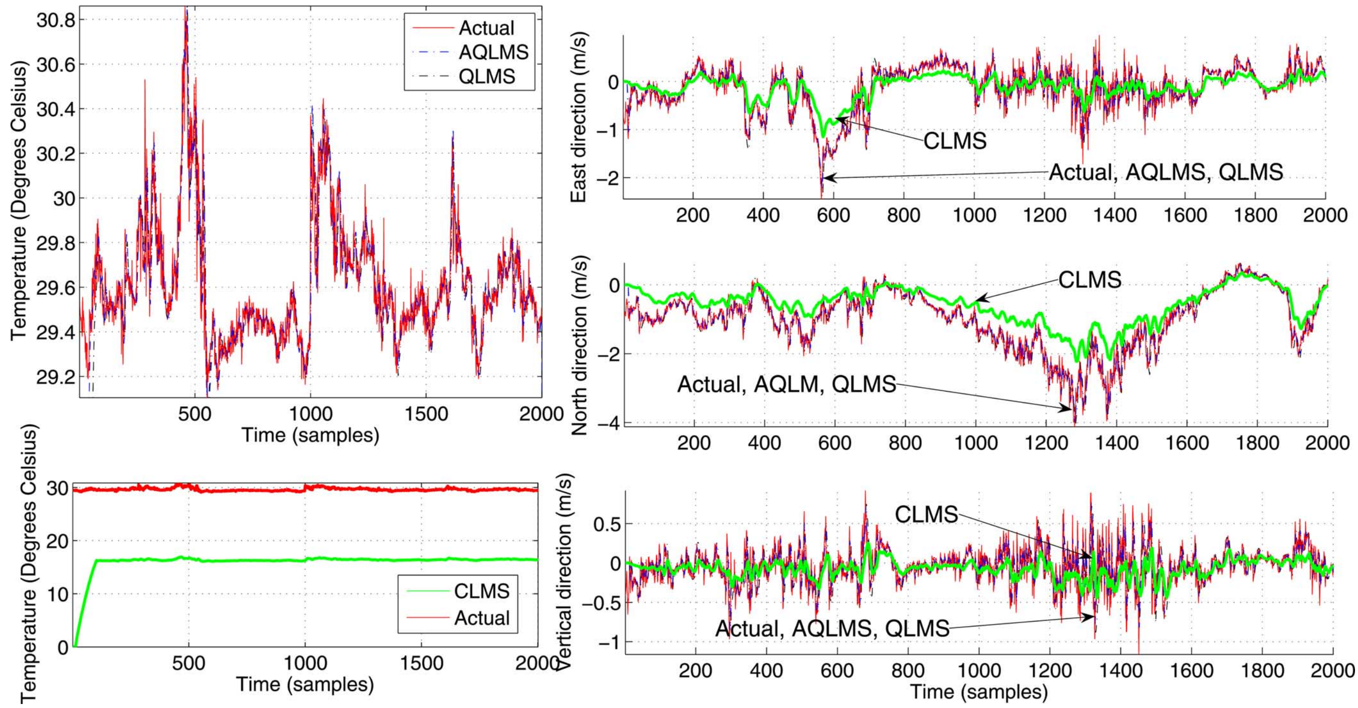

Fig. 8. One-step ahead prediction of the 4D wind using CLMS, QLMS and AQLMS. Note that the CLMS performance was not satisfactory when modeling the air temperature. (Color version available online at http://ieeexplore.ieee.org.).

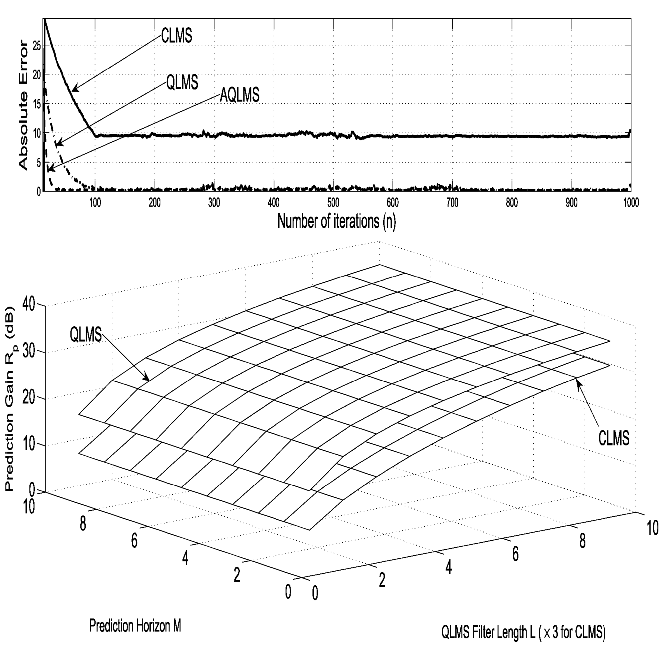

Fig. 9. Convergence of QLMS algorithms. Top: Absolute error curves. One CLMS filter predicts the temperature and east wind direction, and another CLMS filter predicts the north and vertical wind direction. Bottom: Performances of QLMS ($\mathcal{O}(56L)$) and CLMS ($\mathcal{O}(60L)$) on the 4D wind signal. The filter length of CLMS is three times that of QLMS to ensure they have approximately the same computational complexity.

In this work, we have considered the cost function $\mathcal{J}(n) = \|e(n)\|_2^2 = e(n)e^*(n)$, however, the same cost function can be expressed in a different way, that is, $\mathcal{J}(n) = \|e(n)\|_2^2 = e^*(n)e(n)$. Due to the noncommutativity of quaternion product, this gives rise to a variant of QLMS given by

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \mu\big(2\mathbf{x}^*(n)e(n) - e^*(n)\mathbf{x}^*(n)\big). \quad (39)$$

Since the same cost function is being minimized, this results in identical performance. Future work on the class of QLMS algorithms will include algorithms with an optimal adaptive step size, infinite impulse response (IIR) adaptive filters in $\mathbb{H}$, and algorithms with a time-varying filter length [6].

APPENDIX

Derivation of the Stochastic Gradient Update Within QLMS: To calculate the derivatives of the error $e(n)$ and its conjugate with respect to the weight vector $\mathbf{w}(n)$, the terms $\mathbf{w}^T(n)\mathbf{x}(n)$ and $\mathbf{x}^H(n)\mathbf{w}^*(n)$ that appear in the calculations, can be expanded as (for space limitation, the time index 'n' is omitted)

$$\mathbf{w}^T(n)\mathbf{x}(n) = \begin{bmatrix} \mathbf{w}_a^T\mathbf{x}_a - \mathbf{w}_b^T\mathbf{x}_b - \mathbf{w}_c^T\mathbf{x}_c - \mathbf{w}_d^T\mathbf{x}_d \\ \mathbf{w}_a^T\mathbf{x}_b + \mathbf{w}_b^T\mathbf{x}_a + \mathbf{w}_c^T\mathbf{x}_d - \mathbf{w}_d^T\mathbf{x}_c \\ \mathbf{w}_a^T\mathbf{x}_c + \mathbf{w}_c^T\mathbf{x}_a + \mathbf{w}_d^T\mathbf{x}_b - \mathbf{w}_b^T\mathbf{x}_d \\ \mathbf{w}_a^T\mathbf{x}_d + \mathbf{w}_d^T\mathbf{x}_a + \mathbf{w}_b^T\mathbf{x}_c - \mathbf{w}_c^T\mathbf{x}_b \end{bmatrix} \quad (40)$$

$$\mathbf{x}^H(n)\mathbf{w}^*(n) = \begin{bmatrix} \mathbf{w}_a^T\mathbf{x}_a - \mathbf{w}_b^T\mathbf{x}_b - \mathbf{w}_c^T\mathbf{x}_c - \mathbf{w}_d^T\mathbf{x}_d \\ -\mathbf{w}_a^T\mathbf{x}_b - \mathbf{w}_b^T\mathbf{x}_a - \mathbf{w}_c^T\mathbf{x}_d + \mathbf{w}_d^T\mathbf{x}_c \\ -\mathbf{w}_a^T\mathbf{x}_c - \mathbf{w}_c^T\mathbf{x}_a - \mathbf{w}_d^T\mathbf{x}_b + \mathbf{w}_b^T\mathbf{x}_d \\ -\mathbf{w}_a^T\mathbf{x}_d - \mathbf{w}_d^T\mathbf{x}_a - \mathbf{w}_b^T\mathbf{x}_c + \mathbf{w}_c^T\mathbf{x}_b \end{bmatrix}. \quad (41)$$

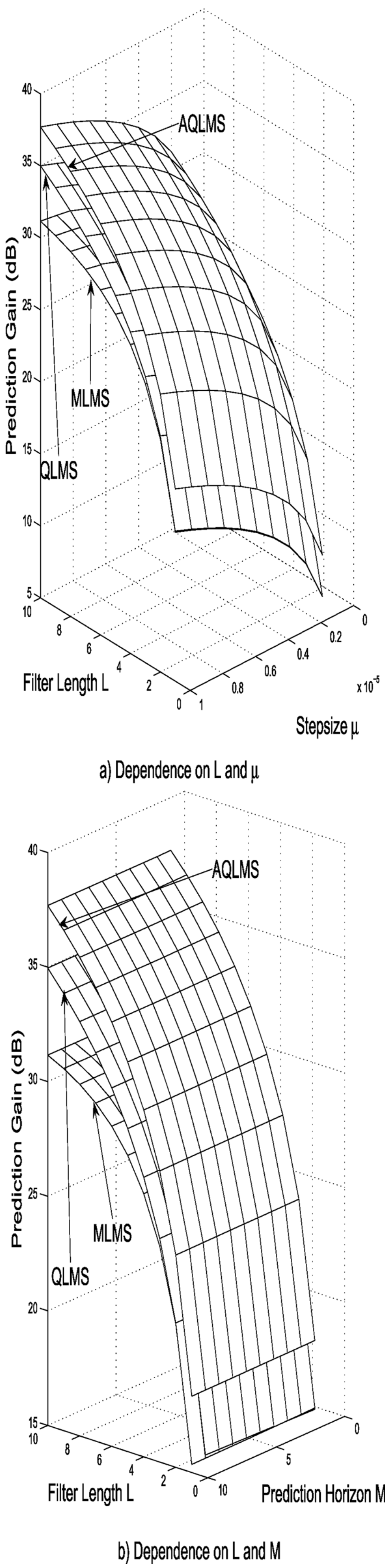

Fig. 10. Dependence of the performance of multichannel LMS (MLMS), QLMS and AQLMS on the choice of parameters. The experiments were conducted on a 4D wind signal (3D wind speed and temperature).

Based on (40) and (41), the derivatives from (8) can be computed as

$$\begin{aligned}
\nabla_{\mathbf{w}_a}\big(e(n)e^*(n)\big) &= e(n)\big(-\mathbf{x}^*(n)\big) + \big(-\mathbf{x}(n)\big)e^*(n) \\
&= -e(n)\mathbf{x}^*(n) - \mathbf{x}(n)e^*(n) &\quad (42)\\
\nabla_{\mathbf{w}_b}\big(e(n)e^*(n)\big)\imath &= e(n)\big(\mathbf{x}_b + \mathbf{x}_a\imath - \mathbf{x}_d\jmath + \mathbf{x}_c\kappa\big)\imath \\
&\quad + \big(\mathbf{x}_b - \mathbf{x}_a\imath + \mathbf{x}_d\jmath - \mathbf{x}_c\kappa\big)e^*(n)\imath \\
&= e(n)\big(-\mathbf{x}_a + \mathbf{x}_b\imath + \mathbf{x}_c\jmath + \mathbf{x}_d\kappa\big) \\
&\quad + \big(\mathbf{x}_a + \mathbf{x}_b\imath - \mathbf{x}_c\jmath - \mathbf{x}_d\kappa\big)e^*(n) &\quad (43)\\
\nabla_{\mathbf{w}_c}\big(e(n)e^*(n)\big)\jmath &= e(n)\big(\mathbf{x}_c + \mathbf{x}_d\imath + \mathbf{x}_a\jmath - \mathbf{x}_b\kappa\big)\jmath \\
&\quad + \big(\mathbf{x}_c - \mathbf{x}_d\imath - \mathbf{x}_a\jmath + \mathbf{x}_b\kappa\big)e^*(n)\jmath \\
&= e(n)\big(-\mathbf{x}_a + \mathbf{x}_b\imath + \mathbf{x}_c\jmath + \mathbf{x}_d\kappa\big) \\
&\quad + \big(\mathbf{x}_a - \mathbf{x}_b\imath + \mathbf{x}_c\jmath - \mathbf{x}_d\kappa\big)e^*(n) &\quad (44)\\
\nabla_{\mathbf{w}_d}\big(e(n)e^*(n)\big)\kappa &= e(n)\big(\mathbf{x}_d - \mathbf{x}_c\imath + \mathbf{x}_b\jmath + \mathbf{x}_a\kappa\big)\kappa \\
&\quad + \big(\mathbf{x}_d + \mathbf{x}_c\imath - \mathbf{x}_b\jmath - \mathbf{x}_a\kappa\big)e^*(n)\kappa \\
&= e(n)\big(-\mathbf{x}_a + \mathbf{x}_b\imath + \mathbf{x}_c\jmath + \mathbf{x}_d\kappa\big) \\
&\quad + \big(\mathbf{x}_a - \mathbf{x}_b\imath - \mathbf{x}_c\jmath + \mathbf{x}_d\kappa\big)e^*(n). &\quad (45)
\end{aligned}$$

Substituting (42)–(45) into (7), we obtain the final expression for the gradient of the cost function (5) in the form [see equation (46) at the top of the next page], where $\Re\{\mathbf{x}\}$ denotes the real/scalar part of $\mathbf{x}$.

Derivation of the Stochastic Gradient Update Within AQLMS: Similarly to the update for $\mathbf{w}$ in (9), the update for $\mathbf{g}$ is derived by firstly expanding terms $\mathbf{g}(n)^T\mathbf{x}^*(n)$ and $\mathbf{x}^T(n)\mathbf{g}^*(n)$, to yield

$$\mathbf{g}^T(n)\mathbf{x}^*(n) = \begin{bmatrix} \mathbf{g}_a^T\mathbf{x}_a + \mathbf{g}_b^T\mathbf{x}_b + \mathbf{g}_c^T\mathbf{x}_c + \mathbf{g}_d^T\mathbf{x}_d \\ -\mathbf{g}_a^T\mathbf{x}_b + \mathbf{g}_b^T\mathbf{x}_a - \mathbf{g}_c^T\mathbf{x}_d + \mathbf{g}_d^T\mathbf{x}_c \\ -\mathbf{g}_a^T\mathbf{x}_c + \mathbf{g}_c^T\mathbf{x}_a - \mathbf{g}_d^T\mathbf{x}_b + \mathbf{g}_b^T\mathbf{x}_d \\ -\mathbf{g}_a^T\mathbf{x}_d + \mathbf{g}_d^T\mathbf{x}_a - \mathbf{g}_b^T\mathbf{x}_c + \mathbf{g}_c^T\mathbf{x}_b \end{bmatrix} \quad (47)$$

$$\mathbf{x}^T(n)\mathbf{g}^*(n) = \begin{bmatrix} \mathbf{g}_a^T\mathbf{x}_a + \mathbf{g}_b^T\mathbf{x}_b + \mathbf{g}_c^T\mathbf{x}_c + \mathbf{g}_d^T\mathbf{x}_d \\ \mathbf{g}_a^T\mathbf{x}_b - \mathbf{g}_b^T\mathbf{x}_a + \mathbf{g}_c^T\mathbf{x}_d - \mathbf{g}_d^T\mathbf{x}_c \\ \mathbf{g}_a^T\mathbf{x}_c - \mathbf{g}_c^T\mathbf{x}_a + \mathbf{g}_d^T\mathbf{x}_b - \mathbf{g}_b^T\mathbf{x}_d \\ \mathbf{g}_a^T\mathbf{x}_d - \mathbf{g}_d^T\mathbf{x}_a + \mathbf{g}_b^T\mathbf{x}_c - \mathbf{g}_c^T\mathbf{x}_b \end{bmatrix}. \quad (48)$$

From (47) and (48), the quaternion gradients of the cost function (5) with respect to $\mathbf{g}$ are computed as

$$\begin{aligned}
\nabla_{\mathbf{g}_a}\big(e(n)e^*(n)\big) &= e(n)\big(-\mathbf{x}(n)\big) + \big(-\mathbf{x}^*(n)\big)e^*(n) \\
&= -e(n)\mathbf{x}(n) - \mathbf{x}^*(n)e^*(n) &\quad (49)\\
\nabla_{\mathbf{g}_b}\big(e(n)e^*(n)\big)\imath &= e(n)\big(-\mathbf{x}_b + \mathbf{x}_a\imath + \mathbf{x}_d\jmath - \mathbf{x}_c\kappa\big)\imath \\
&\quad + \big(-\mathbf{x}_b - \mathbf{x}_a\imath - \mathbf{x}_d\jmath + \mathbf{x}_c\kappa\big)e^*(n)\imath \\
&= e(n)\big(-\mathbf{x}_a - \mathbf{x}_b\imath - \mathbf{x}_c\jmath - \mathbf{x}_d\kappa\big) \\
&\quad + \big(\mathbf{x}_a - \mathbf{x}_b\imath + \mathbf{x}_c\jmath + \mathbf{x}_d\kappa\big)e^*(n) &\quad (50)\\
\nabla_{\mathbf{g}_c}\big(e(n)e^*(n)\big)\jmath &= e(n)\big(-\mathbf{x}_c - \mathbf{x}_d\imath + \mathbf{x}_a\jmath + \mathbf{x}_b\kappa\big)\jmath \\
&\quad + \big(-\mathbf{x}_c + \mathbf{x}_d\imath - \mathbf{x}_a\jmath - \mathbf{x}_b\kappa\big)e^*(n)\jmath \\
&= e(n)\big(-\mathbf{x}_a - \mathbf{x}_b\imath - \mathbf{x}_c\jmath - \mathbf{x}_d\kappa\big) \\
&\quad + \big(\mathbf{x}_a + \mathbf{x}_b\imath - \mathbf{x}_c\jmath + \mathbf{x}_d\kappa\big)e^*(n) &\quad (51)
\end{aligned}$$

$$
\begin{aligned}
\nabla_{\mathbf{w}}\big(e(n)e^*(n)\big) &= \nabla_{\mathbf{w}_a}\big(e(n)e^*(n)\big) + \big[\nabla_{\mathbf{w}_b}\big(e(n)e^*(n)\big)\imath + \nabla_{\mathbf{w}_c}\big(e(n)e^*(n)\big)\jmath + \nabla_{\mathbf{w}_d}\big(e(n)e^*(n)\big)\kappa\big] \\
&= -e(n)\mathbf{x}^*(n) - \mathbf{x}(n)e^*(n) + \big[-3e(n)\mathbf{x}^*(n) + \big(3\mathbf{x}_a - \mathbf{x}_b\imath - \mathbf{x}_c\jmath - \mathbf{x}_d\kappa\big)e^*(n)\big] \\
&= -e(n)\mathbf{x}^*(n) - \mathbf{x}(n)e^*(n) + \big[-3e(n)\mathbf{x}^*(n) + \mathbf{x}(n)e^*(n) + 2\Re\{\mathbf{x}(n)\}e^*(n)\big] \\
&= -4e(n)\mathbf{x}^*(n) + \mathbf{x}^*(n)e^*(n) - \mathbf{x}(n)e^*(n) + 2\Re\{\mathbf{x}(n)\}e^*(n) \\
&= -4e(n)\mathbf{x}^*(n) + \big[\mathbf{x}^*(n) - \mathbf{x}(n) + 2\Re\{\mathbf{x}(n)\}\big]e^*(n) \\
&= -4e(n)\mathbf{x}^*(n) + 2\mathbf{x}^*(n)e^*(n)
\end{aligned}
\tag{46}
$$

$$
\begin{aligned}
\nabla_{\mathbf{g}}\big(e(n)e^*(n)\big) &= \nabla_{\mathbf{g}_a}\big(e(n)e^*(n)\big) + \big[\nabla_{\mathbf{g}_b}\big(e(n)e^*(n)\big)\imath + \nabla_{\mathbf{g}_c}\big(e(n)e^*(n)\big)\jmath + \nabla_{\mathbf{g}_d}\big(e(n)e^*(n)\big)\kappa\big] \\
&= -e(n)\mathbf{x}(n) - \mathbf{x}^*(n)e^*(n) + \big[-3e(n)\mathbf{x}(n) + \big(3\mathbf{x}_a + \mathbf{x}_b\imath + \mathbf{x}_c\jmath + \mathbf{x}_d\kappa\big)e^*(n)\big] \\
&= -e(n)\mathbf{x}(n) - \mathbf{x}^*(n)e^*(n) + \big[-3e(n)\mathbf{x}(n) + \mathbf{x}(n)e^*(n) + 2\Re\{\mathbf{x}(n)\}e^*(n)\big] \\
&= -4e(n)\mathbf{x}(n) + \big[\mathbf{x}(n) - \mathbf{x}^*(n) + 2\Re\{\mathbf{x}(n)\}\big]e^*(n) \\
&= -4e(n)\mathbf{x}(n) + 2\mathbf{x}(n)e^*(n)
\end{aligned}
\tag{53}
$$

$$
\begin{aligned}
\nabla_{\mathbf{g}_d}\big(e(n)e^*(n)\big)\kappa &= e(n)\big(-\mathbf{x}_d + \mathbf{x}_c\imath - \mathbf{x}_b\jmath + \mathbf{x}_a\kappa\big)\kappa \\
&\quad + \big(-\mathbf{x}_d - \mathbf{x}_c\imath + \mathbf{x}_b\jmath - \mathbf{x}_a\kappa\big)e^*(n)\kappa \\
&= e(n)\big(-\mathbf{x}_a - \mathbf{x}_b\imath - \mathbf{x}_c\jmath - \mathbf{x}_d\kappa\big) \\
&\quad + \big(\mathbf{x}_a + \mathbf{x}_b\imath + \mathbf{x}_c\jmath - \mathbf{x}_d\kappa\big)e^*(n)
\end{aligned}
\tag{52}
$$

which yields [see equation (53) at the top of the page].

ACKNOWLEDGMENT

The authors would like to express our gratitude to Dr. Sangwine and Dr. Le Bihan for providing the quaternion toolbox [48]. The authors also acknowledge the valuable input by anonymous reviewers.

REFERENCES

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

[2] B. Farhang-Boroujeny, *Adaptive Filters Theory and Applications*. New York: Wiley, 1999.

[3] S. V. Vaseghi, *Advanced Digital Signal Processing and Noise Reduction*, 2nd ed. New York: Wiley, 2002.

[4] V. J. Mathews and Z. Xie, "A stochastic gradient adaptive filter with gradient adaptive stepsize," *IEEE Trans. Signal Process.*, vol. 41, no. 6, pp. 2075–2087, Jun. 1993.

[5] T. Aboulnasr and K. Mayyas, "A robust variable step-size LMS-type algorithm: analysis and simulations," *IEEE Trans. Signal Process.*, vol. 45, no. 3, pp. 631–639, Mar. 1997.

[6] Y. Gong and C. F. N. Cowan, "An LMS style variable tap-length algorithm for structure adaptation," *IEEE Trans. Signal Process.*, vol. 53, no. 7, pp. 2400–2407, Jul. 2005.

[7] S. H. Strogatz, *Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry and Engineering (Studies in Nonlinearity)*, 1st ed. Boulder, CO: Westview Press, 2001.

[8] S. L. Goh, M. Chen, D. H. Popovic, K. Aihara, D. Obradovic, and D. P. Mandic, "Complex valued forecasting of wind profile," *Renewable Energy*, vol. 31, pp. 1733–1750, 2006.

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

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

[11] P. J. Schreier and L. L. Scharf, "Second-order analysis of improper complex random vectors and processes," *IEEE Trans. Signal Process.*, vol. 51, no. 3, pp. 714–725, Mar. 2003.

[12] W. Hamilton, *Elements of Quaternions*, 2nd ed. London, U.K.: Longmans, Green, & Company, 1899.

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

[14] S. B. Choe and J. J. Faraway, "Modeling head and hand orientation during motion using quaternions," *J. Aerosp.*, vol. 113, no. 1, pp. 186–192, 2004.

[15] D. Biamino, G. Cannata, M. Maggiali, and A. Piazza, "Mac-eye: a tendon driven fully embedded robot eye," in *Proc. IEEE-RAS Int. Conf. Humanoid Robots*, 2005, pp. 62–67.

[16] C. F. F. Karney, "Quaternions in molecular modelling," *J. Molecul. Graphics Model.*, vol. 25, no. 5, pp. 595–604, 2007.

[17] 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, N\May 1999.

[18] H. Toyoshima, "Computationally efficient implementation of hypercomplex digital filters," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, 1998, vol. 3, pp. 1761–1764.

[19] T. Bülow and G. Sommer, "Hypercomplex signals—A novel extension of the analytic signal to the multidimensional case," *IEEE Trans. Signal Process.*, vol. 49, no. 11, pp. 2844–2852, Nov. 2001.

[20] V. Zarzoso and A. K. Nandi, "Closed-form semi-blind separation of three sources from three real-valued instantaneous linear mixtures via quaternions," in *Proc. IEEE 6th Symp. Signal Processing Applications*, 2001, vol. 1, pp. 1–4.

[21] P. Bas, N. L. Bihan, and J. M. Chassery, "Color image watermarking using quaternion Fourier transform," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, 2003, vol. 3, pp. III–521.

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

[23] N. L. 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.

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

[25] T. Jiang and L. Chen, "Algebraic algorithms for least squares next term problem in quaternionic quantum theory," *Comput. Phys. Commun.*, vol. 176, no. 7, pp. 481–485, 2007.

[26] A. Ude, "Nonlinear least squares optimisation of unit quaternion functions for pose estimation from corresponding features," in *Proc. IEEE 14th Int. Conf. Pattern Recognit.*, 1998, vol. 1, pp. 425–427.

[27] B. Picinbono and P. Chevalier, "Widely linear estimation with complex data," *IEEE Trans. Signal Process.*, vol. 43, no. 8, pp. 2030–2033, Aug. 1995.

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

[29] D. P. Mandic, S. Javidi, S. L. Goh, A. Kuh, and K. Aihara, "Complex-valued prediction of wind profile using augmented complex statistics," *Renew. Energy*, vol. 34, pp. 196–201, 2009.

[30] Y. Huang and J. Benesty, *Audio Signal Processing for Next Generation Multimedia Communication Systems*. Boston, MA: Kluwer, 2004.
[31] I. Kantor and A. Solodovnikov, *Hypercomplex Numbers: An Elementary Introduction to Algebras*. New York: Springer, 1989.
[32] J. P. Ward, *Quaternions and Cayley Numbers: Algebra and Applications*. Boston, MA: Kluwer, 1997.
[33] D. P. Mandic and V. S. L. Goh, *Complex Valued Nonlinear Adaptive Filters: A Neural Network Approach*. New York: Wiley, 2008.
[34] M. Novey and T. Adali, "Complex ICA by negentropy maximization," *IEEE Trans. Neural Netw.*, vol. 19, no. 4, pp. 596–609, 2008.
[35] M. Novey and T. Adali, "On extending the complex fastICA algorithm to non-circular sources," *IEEE Trans. Signal Process.*, vol. 56, no. 5, pp. 2148–2154, May 2008.
[36] B. Picinbono, "On circularity," *IEEE Trans. Signal Process.*, vol. 42, no. 12, pp. 3473–3482, Dec. 1994.
[37] B. Picinbono and P. Bondon, "Second-order statistics of complex signals," *IEEE Trans. Signal Process.*, vol. 45, no. 2, pp. 411–420, Feb. 1997.
[38] N. N. Vakhania, "Random vectors with values in quaternion Hilbert spaces," *Theory Probab. Appl.*, vol. 43, no. 1, pp. 99–115, 1999.
[39] P. O. Amblard and N. L. Bihan, "On properness of quaternion valued random variables," in *Proc. IMA Conf. Mathematics in Signal Processing*, Cirencester, U.K., Dec. 2004, pp. 23–26.
[40] S. L. Goh and D. P. Mandic, "An augmented ACRTRL for complex valued recurrent neural networks," *Elsevier Neural Netw.*, vol. 20, no. 10, pp. 1061–1066, 2007.
[41] S. C. Douglas and D. P. Mandic, "Mean and mean-square analysis of the complex LMS algorithm for non-circular Gaussian signals," presented at the IEEE Digital Signal Processing and Signal Processing Education Workshop, Marco Island, FL, Jan. 2008 [Online]. Available: http://www.2009dsp.org/index.html
[42] J. Seberry, K. Finlayson, S. S. Adams, T. A. Wysocki, T. Xia, and B. J. Wysocki, "The theory of quaternion orthogonal designs," *IEEE Trans. Signal Process.*, vol. 56, no. 1, pp. 256–265, Jan. 2008.
[43] S. Haykin, *Adaptive Filter Theory*, 4th ed. Englewood Cliffs, NJ: Prentice-Hall, 2001.
[44] L. Huang and W. So, "On left eigenvalues of a quaternionic matrix," *Linear Algebra Appl.*, vol. 323, pp. 105–116, 2001.
[45] S. Haykin and L. Li, "Nonlinear adaptive prediction of nonstationary signals," *IEEE Trans. Signal Process.*, vol. 43, no. 2, pp. 526–535, Feb. 1995.
[46] J. G. M. J. F. Manwell and A. L. Rogers, *Wind Energy Explained: Theory, Design and Application*. New York: Wiley, 2002.
[47] D. Mandic, M. Golz, A. Kuh, D. Obradovic, and T. Tanaka, *Signal Processing Techniques for Knowledge Extraction and Information Fusion*. New York: Springer, 2008.
[48] S. Sangwine and N. L. Bihan, Quaternion Toolbox for Matlab 2005 [Online]. Available: http://qtfm.sourceforge.net/

Clive Cheong Took (S'06–M'07) was born in the Island of Mauritius in 1982. He received the B.S. degree in telecommunication engineering from King's College London University, London, U.K., where he was the top departmental graduate in 2004 and the Ph.D. degree in blind signal processing from Cardiff University, Cardiff, U.K., in 2007.

Currently, he is a Research Associate at the Imperial College London, London, U.K. His research interests include adaptive, blind, and multidimensional signal processing.

Danilo P. Mandic (M'99–SM'03) is a Reader in Signal Processing at Imperial College London. He has been working in the area of nonlinear adaptive signal processing and nonlinear dynamics. His publication record includes two research monographs titled *Recurrent Neural Networks for Prediction* (1st ed., Aug. 2001) and *Complex Valued Nonlinear Adaptive Filters: Noncircularity, Widely Linear and Neural Models* (1st ed., Wiley, Apr. 2009), an edited book titled *Signal Processing for Information Fusion* (Springer, 2008) and more than 200 publications on signal and image processing. He has been a Guest Professor at K.U. Leuven Belgium, TUAT Tokyo, Japan, and Westminster University, U.K., and a Frontier Researcher in RIKEN Japan.

Dr. Mandic has been a Member of the IEEE Technical Committee on Machine Learning for Signal Processing, Associate Editor for the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS II, IEEE TRANSACTIONS ON SIGNAL PROCESSING, IEEE TRANSACTIONS ON NEURAL NETWORKS, and the *International Journal of Mathematical Modelling and Algorithms*. He has produced award winning papers and products resulting from his collaboration with Industry. He is a Member of the London Mathematical Society.