Signal Processing 93 (2013) 1712–1723

Contents lists available at SciVerse ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

A class of quaternion valued affine projection algorithms

Cyrus Jahanchahi [a,*], Clive Cheong Took [b], Danilo P. Mandic [a]

[a] Department of Electrical and Electronic Engineering, Imperial College, London SW7 2AZ, United Kingdom
[b] Department of Computing, University of Surrey, GU7 2XH, United Kingdom

ARTICLE INFO

Article history:
Received 25 January 2012
Received in revised form
13 November 2012
Accepted 26 December 2012
Available online 4 January 2013

Keywords:
Quaternion affine projection
Widely linear model
Improperness
Quaternion noncircularity
HR-calculus
Quaternion gradient
3D wind modeling

ABSTRACT

The strictly linear quaternion valued affine projection algorithm (QAPA) and its widely linear counterpart (WLQAPA) are introduced, in order to provide fast converging stochastic gradient learning in the quaternion domain, for the processing of both second order circular (proper) and second order noncircular (improper) signals. This is achieved based on the recent advances in augmented quaternion statistics, which employs all second order information available, together with the associated widely linear models and through performing rigorous gradient calculation ($\mathbb{HR}$-calculus). Further, mean square error analysis is performed based on the energy conservation principle, which provides a theoretical justification for the WLQAPA offering enhanced steady state performance for quaternion noncircular (improper) signals, a typical case in real world scenarios. Simulations on benchmark circular and noncircular signals, and on noncircular real world 4D wind and 3D body motion data support the analysis.

© 2013 Elsevier B.V. All rights reserved.

1. Introduction

Recent advances in sensing technology have brought to light data sources which are almost invariably two-, three-, and four-dimensional, such as measurements from inertial bodysensors and ultrasonic anemometers. The quaternion domain offers a convenient means to process three- and four-dimensional signals under the same umbrella, as a generalization of the complex domain which is a natural choice for processing bivariate (two-dimensional) signals. Several adaptive filtering algorithms have recently emerged in the quaternion domain, taking advantage of the power of its division algebra and the convenience of data representation offered. In particular, a number of applications involving rotations in three-dimensional spaces have benefited, since quaternions offer a simultaneous and accurate model of the axis of rotation and rotation angle [1,2]. Other areas where quaternions have become prominent include color image processing [3,4], robotics [5], renewable energy [6], and blind source separation [7,8].

The quaternion least mean square (QLMS) [9] was recently introduced for adaptive filtering of quaternion valued data, however, it also highlighted the need for faster converging practical algorithms. The stochastic gradient based normalized QLMS can solve this issue only partially whereas the fast converging Quaternion Recursive Least Squares (QRLS) [10] is computationally demanding. To that end, our aim is to introduce a fast converging and computationally scalable affine projection scheme for adaptive filtering of quaternion valued data.

Ozeki and Umeda [11] employed affine subspace projections to develop the affine projection algorithm (APA) for real valued finite impulse response (FIR) adaptive filters, thus ensuring fast convergence for colored inputs. Structurally, the APA spans the range between the normalized least mean square (NLMS) and recursive least squares (RLS), both in terms of performance and computational requirements. In practical terms, whereas the

* Corresponding author.
E-mail addresses: cyrus.jahanchahi@imperial.ac.uk (C. Jahanchahi), c.cheongtook@surrey.ac.uk (C. Cheong Took), d.mandic@ic.ac.uk (D.P. Mandic).

0165-1684/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.sigpro.2012.12.019

NLMS updates the weight vector based only on the current input vector, the APA employs N past input vectors, thus making use of the history of weight update evolutions, and achieving faster convergence than the NLMS. For large N, the performance of APA compares with RLS, which updates the weight vectors based on the cumulative error powers calculated from all the past input vectors, therefore exhibiting a high computational complexity. These algorithms were developed for a single channel (univariate) scenario, and were only recently extended to the widely linear complex scenarios [12], thus catering for the generality of complex signals, both circular and noncircular.

The development of real time adaptive filtering algorithms in the quaternion domain has long been hampered by the lack of a rigorous definition of quaternion gradient. This is not surprising, as the gradient calculation is critical in division algebras (complex, quaternion), where standard differentiability conditions (e.g. Cauchy Riemann) do not allow for the gradient of real function of complex and quaternion variables to be calculated; yet typical cost functions are real valued error power. For instance, it is only the $\mathbb{CR}$-calculus [13,14], that made it possible for the real APA to be extended to the complex domain [12], and for the augmented complex APA (AAPA) to be derived for the processing of noncircular complex signals. Similarly, until the recent development of the $\mathbb{HR}$-calculus [15] and the augmented quaternion statistics [16–18], stochastic gradient algorithms in the quaternion domain lacked a formal treatment of the gradient, while statistics were only suited for the processing of second order circular signals.[1]

Unlike the standard, strictly linear, quaternion valued algorithms where the covariance matrix $E[\mathbf{q}\mathbf{q}^H]$ is assumed to be sufficient to model second order quaternion statistics, widely linear quaternion statistics use the additional pseudocovariance matrices $E[\mathbf{q}\mathbf{q}^{\imath H}]$, $E[\mathbf{q}\mathbf{q}^{\jmath H}]$ and $E[\mathbf{q}\mathbf{q}^{\kappa H}]$ to fully describe second order information in $\mathbb{H}$. This offers more degrees of freedom in the modeling and the possibility to perform optimal second order filtering of noncircular (improper) signals, for which the powers in the components of the quaternion signal are different, a typical case in real world scenarios. Building upon those advantages, the augmented statistics have recently been used to derive the widely linear quaternion recursive least squares (WLQRLS) [10] and the widely linear quaternion least mean square (WLQLMS) [18], together with the widely linear quaternion Kalman filter [19].

In this paper, we introduce a general class of quaternion APA algorithms comprising both the strictly linear (QAPA) and widely linear (WLQAPA) cases, in order to provide a rigorous framework for fast converging adaptive filtering of the generality of quaternion valued signals. Based on the principle of energy conservation [20], expressions for the mean square error (MSE) of both the QAPA and WLQAPA are also derived. Simulations show that the WLQAPA offers lower MSE when processing noncircular data, and is thus second order optimal for noncircular signals. These advantages are illustrated in the context of renewable energy (wind modeling) and human centered computing (inertial bodysensors).

The paper is organized as follows. We first provide an overview of quaternion algebra. In Section 3 we give the mathematical foundations for the gradient calculation through the $\mathbb{HR}$-calculus. Section 4 introduces the background necessary for the modeling of improper signals via the quaternion widely linear model, both necessary for the derivation of QAPA and WLQAPA. Sections 6 and 7 provide theoretical analysis for the MSE of the QAPA and WLQAPA for noncircular inputs. In Section 8, the performances of the QAPA and WLQAPA are illustrated on both circular (proper) and second order noncircular (improper) real world signals.

2. Elements of quaternion algebra

Quaternions are a skew field over $\mathbb{R}^4$ defined as

$$\{q_r, q_\imath, q_\jmath, q_\kappa\} \in \mathbb{R}^4 \rightarrow q_r + \imath q_\imath + \jmath q_\jmath + \kappa q_\kappa \in \mathbb{H}$$

The unit axis vectors $\imath$, $\jmath$ and κ are also the imaginary units, and obey the following rules:

$$\imath\jmath = \kappa, \quad \jmath\kappa = \imath, \quad \kappa\imath = \jmath$$
$$\imath^2 = \jmath^2 = \kappa^2 = \imath\jmath\kappa = -1$$

Note that quaternion multiplication is not commutative, that is, $\imath\jmath \neq \jmath\imath = -\kappa$. The product of quaternions $q_1, q_2 \in \mathbb{H}$ is given by

$$\begin{aligned} q_1 q_2 &= (Sq_1 + Vq_1)(Sq_2 + Vq_2) \\ &= Sq_1 Sq_2 - Vq_1 \bullet Vq_2 + Sq_2 Vq_1 + Sq_1 Vq_2 + Vq_1 \times Vq_2 \end{aligned}$$

where $Sq = q_r$ and $Vq = \imath q_\imath + \jmath q_\jmath + \kappa q_\kappa$ are respectively the scalar and vector part of a quaternion q, the symbol '$\bullet$' denotes the dot-product and '$\times$' the cross-product. It is the cross-product above that makes the quaternion multiplication noncommutative. The norm $\|q\|$ is defined as

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

while the quaternion conjugate, denoted by q^*, is given by

$$q^* = Sq - Vq = q_r - \imath q_\imath - \jmath q_\jmath - \kappa q_\kappa \tag{1}$$

In addition to the standard quaternion conjugate, we can also define the three involutions (self-inverse mappings) as [21]

$$\begin{aligned} q^\imath &= -\imath q \imath = q_r + \imath q_\imath - \jmath q_\jmath - \kappa q_\kappa \\ q^\jmath &= -\jmath q \jmath = q_r - \imath q_\imath + \jmath q_\jmath - \kappa q_\kappa \\ q^\kappa &= -\kappa q \kappa = q_r - \imath q_\imath - \jmath q_\jmath + \kappa q_\kappa \end{aligned} \tag{2}$$

These perpendicular involutions have the following properties (for $inv_3 \neq inv_2 \neq inv_1$):

$$P1: \quad (q^{inv})^{inv} = q \quad \text{for } inv \in \{\imath, \jmath, \kappa\} \tag{3}$$

$$P2: \quad (q_1 q_2)^{inv} = q_1^{inv} q_2^{inv} \tag{4}$$

$$P3: \quad (q_1 + q_2)^{inv} = q_1^{inv} + q_2^{inv} \tag{5}$$

[1] Second order circular signals (proper) have rotation invariant distributions and equal powers in all the signal components.

$$P4: \quad (q^{inv_1})^{inv_2} = (q^{inv_2})^{inv_1} = q^{inv_3} \tag{6}$$

Involutions can be seen as a counterpart of the complex conjugate, as they allow for the components of a quaternion variable q to be expressed in terms of the actual variable q and its 'partial conjugates' $q^\imath$, $q^\jmath$, q^κ, that is[2]

$$q_r = \frac{1}{4}[q+q^\imath+q^\jmath+q^\kappa], \quad q_\imath = \frac{1}{4\imath}[q+q^\imath-q^\jmath-q^\kappa]$$
$$q_\jmath = \frac{1}{4\jmath}[q-q^\imath+q^\jmath-q^\kappa], \quad q_\kappa = \frac{1}{4\kappa}[q-q^\imath-q^\jmath+q^\kappa] \tag{7}$$

The above representation via the intrinsic elementary involutions will be instrumental in introducing quaternion gradients, widely linear models, and the associated adaptive filtering algorithms.

2.1. The advantages of quaternion algebra

To highlight some of the benefits offered by quaternion algebra, consider a matrix $\mathbf{R} \in \mathbb{R}^{3\times 3}$ that maps a point $\mathbf{x} \in \mathbb{R}^3$ to a point $\mathbf{y} \in \mathbb{R}^3$ and a quaternion $q_R \in \mathbb{H}$ that relates a pure quaternion $q_x \in \mathbb{H}$ with a pure quaternion $q_y \in \mathbb{H}$, that is

$$\mathbf{y} = \mathbf{R}\mathbf{x} \in \mathbb{R}^3 \sim q_y = q_R q_x q_R^* \in \mathbb{H} \tag{8}$$

Remark 1. The mapping $\mathbf{R}$ is described by nine coefficients, although physically only four parameters are needed (two for the axis of rotation, one for the angle of rotation and one for the scaling factor). The four components of a quaternion offer this physical insight, and express straightforwardly the axis of rotation, angle of rotation, and scaling factor.

Remark 2. For the mapping $\mathbf{R}$ that represents a succession of rotations in the x, y, z directions (using Euler angles), a degree of freedom can be lost if any two axis are aligned, resulting in the so called gimbal lock phenomenon. This cannot happen in the quaternion domain, where the quaternion transformation in (8) is expressed as $q_y = q_R q_x q_R^{-1}$, where q_R is a unit quaternion. This property has made quaternions an invaluable tool in computer graphics [1].

Remark 3. The quaternion rotation q_R is better conditioned than the real rotation matrix $\mathbf{R}$, as q_R is only required to be a unit quaternion whereas $\mathbf{R}$ must satisfy $\mathbf{R}^T\mathbf{R} = \mathbf{I}$ and $det(\mathbf{R}) = 1$. Computer graphics often require many rotations to be performed successively making it necessary to re-normalize periodically to mitigate the effect of finite precision and ensure that the conditions $\mathbf{R}^T\mathbf{R} = \mathbf{I}$ and $det(\mathbf{R}) = 1$ are satisfied. This is computationally intensive and is much more efficiently achieved using quaternions. This has led to the use of quaternions in e.g. spacecraft orientation problems where they allow for convenient closed form solutions [22–24].

[2] Compare this with the complex domain where the real and imaginary parts of the complex numbers $z = x + \imath y$ are expressed by $x = \frac{1}{2}(z+z^*)$ and $y = (1/2i)(z-z^*)$.

3. The $\mathbb{HR}$-calculus

In gradient based optimization in adaptive filtering, the goal is to minimize a measure of error power, typically a real scalar function of quaternion variables, that is

$$J = ee^* = |e|^2$$

However, the Cauchy–Riemann–Fueter (CRF) differentiability condition

$$\frac{\partial J}{\partial \mathbf{w}^*} = \frac{1}{4}\left(\frac{\partial J}{\partial \mathbf{w}_r} + \imath\frac{\partial J}{\partial \mathbf{w}_\imath} + \jmath\frac{\partial J}{\partial \mathbf{w}_\jmath} + \kappa\frac{\partial J}{\partial \mathbf{w}_\kappa}\right) = \mathbf{0} \tag{9}$$

where $\mathbf{w}$ is a vector parameter does not admit the calculation of such gradients, as (9) is not defined for real functions. Indeed, the CRF conditions are very restrictive and only allow for the differentiation of linear functions; a way to bypass this problem in nonlinear adaptive filtering is given in [25].

Owing to the isomorphism between the fields $\mathbb{H}$ and $\mathbb{R}^4$ a quaternion function $f(q)$:$\mathbb{H} \to \mathbb{H}$ has its dual real quadrivariate function $g(q_r,q_\imath,q_\jmath,q_\kappa)$: $\mathbb{R}^4 \to \mathbb{R}^4$. This was the basis for the development of the recently introduced $\mathbb{HR}$-calculus [26], which removed the restrictions imposed by the CRF condition for the class of functions in (9). Therefore, to calculate the gradients necessary to derive the affine projection algorithm in the quaternion domain, we resort to the $\mathbb{HR}^*$-derivatives, given by

$$\begin{bmatrix} \frac{\partial f(q^*,q^{\imath*},q^{\jmath*},q^{\kappa*})}{\partial q^*} \\ \frac{\partial f(q^*,q^{\imath*},q^{\jmath*},q^{\kappa*})}{\partial q^{\imath*}} \\ \frac{\partial f(q^*,q^{\imath*},q^{\jmath*},q^{\kappa*})}{\partial q^{\jmath*}} \\ \frac{\partial f(q^*,q^{\imath*},q^{\jmath*},q^{\kappa*})}{\partial q^{\kappa*}} \end{bmatrix} = \frac{1}{4}\begin{bmatrix} 1 & \imath & \jmath & \kappa \\ 1 & \imath & -\jmath & -\kappa \\ 1 & -\imath & \jmath & -\kappa \\ 1 & -\imath & -\jmath & \kappa \end{bmatrix} \begin{bmatrix} \frac{\partial f(q_r,q_\imath,q_\jmath,q_\kappa)}{\partial q_r} \\ \frac{\partial f(q_r,q_\imath,q_\jmath,q_\kappa)}{\partial q_\imath} \\ \frac{\partial f(q_r,q_\imath,q_\jmath,q_\kappa)}{\partial q_\jmath} \\ \frac{\partial f(q_r,q_\imath,q_\jmath,q_\kappa)}{\partial q_\kappa} \end{bmatrix} \tag{10}$$

where the symbol $q^{inv*} = (q^{inv})^*$ for inv $\in \{\imath,\jmath,\kappa\}$ and $f(\cdot)$ is a general quaternion-valued function, linear or nonlinear. For more detail on the $\mathbb{HR}$-calculus, see [26].

Remark 4. The maximum rate of change of f with respect to q occurs in the direction of $\partial f/\partial q^*$, making the conjugate gradient $\partial f/\partial q^*$ a natural choice of gradient in the optimization of real valued quaternion functions [26].

Remark 5. The $\mathbb{HR}^*$-derivative $\partial f(q^*,q^{\imath*},q^{\jmath*},q^{\kappa*})/\partial q\imath^*)$ is equivalent to the quaternion derivative operator introduced by Fueter [27], however, unlike the CRF derivative in (9), the derivative $\partial f(q^*,q^{\imath*},q^{\jmath*},q^{\kappa*})/\partial q\imath^*)$ also introduces a condition on the argument of the function $f(\cdot)$.

Remark 6. The $\mathbb{HR}^*$-derivatives in $\mathbb{H}$ are a natural generalization of the $\mathbb{R}^*$-derivative within the $\mathbb{CR}$-calculus in the complex domain [14,28]. For instance, to perform a direct $\mathbb{HR}^*$ differentiation of a function written in terms of q, it must first be written in terms of q^*, $q^{\imath*}$, $q^{\jmath*}$ and $q^{\kappa*}$, using the substitution

$$q = \tfrac{1}{2}(q^{\imath*} + q^{\jmath*} + q^{\kappa*} - q^*) \tag{11}$$

This way, the $\mathbb{HR}$-calculus in (10) provides a universal tool for differentiating quaternion functions directly, rather than employing partial derivatives with respect to the real valued q_r, $q_\imath$, $q_\jmath$, q_κ, as is current practice (within the pseudogradient). This also provides an opportunity to

obtain closed form solutions of stochastic gradient learning algorithms.

4. Widely linear quaternion modeling

4.1. Quaternion widely linear model

Consider a real valued mean square error (MSE) estimator given by

$$\hat{y} = E[y|\mathbf{x}]$$

where $\hat{y}$ is the estimated process, $\mathbf{x}$ the observed variable and $E[\cdot]$ the statistical expectation operator. For jointly Gaussian $\mathbf{x}$ and y, the optimal solution is a linear estimator, given by

$$\hat{y} = \mathbf{h}^T\mathbf{x} \tag{12}$$

where $\mathbf{h}$ and $\mathbf{x}$ are respectively the coefficient and regressor vector. For the standard complex domain MSE estimator it is also assumed that $\hat{y} = E[y|\mathbf{x}]$, leading to the *strictly linear* model

$$\hat{y} = \mathbf{h}^T\mathbf{x} \tag{13}$$

However, observe that

$$\hat{y}_r = E[y_r|\mathbf{x}_r,\mathbf{x}_i], \quad \hat{y}_i = E[y_i|\mathbf{x}_r,\mathbf{x}_i]$$

and since

$$\mathbf{x}_r = \frac{\mathbf{x}+\mathbf{x}^*}{2} \quad \text{and} \quad \mathbf{x}_i = \frac{\mathbf{x}-\mathbf{x}^*}{2\imath}$$

we have e.g. $\hat{y} = E[y|\mathbf{x},\mathbf{x}^*]$, and the complex *widely linear* [29] model is given by [30]

$$\hat{y} = E[y|\mathbf{x},\mathbf{x}^*] \Rightarrow y = \mathbf{h}^T\mathbf{x}+\mathbf{g}^T\mathbf{x}^*$$

and comprises both the strictly linear part $\mathbf{h}^T\mathbf{x}$ and the conjugate part $\mathbf{g}^T\mathbf{x}^*$, where $\mathbf{g}$ is a coefficient vector. Similarly, the existing strictly linear quaternion model is given by

$$\hat{y} = \mathbf{h}^T\mathbf{x} \tag{14}$$

Observe, however, that for all the components, r, $\imath$, $\jmath$, κ we have

$$\hat{y}_\eta = E[y_\eta|\mathbf{x}_r,\mathbf{x}_\imath,\mathbf{x}_\jmath,\mathbf{x}_\kappa], \quad \eta \in \{r,\imath,\jmath,\kappa\}$$

and similarly to the complex domain, by using the involutions in (2), we can express each element of a quaternion variable as in (7). This gives, for instance, for the real component of a quaternion $\mathbf{x}_r = \frac{1}{4}(\mathbf{x}+\mathbf{x}^\imath+\mathbf{x}^\jmath+\mathbf{x}^\kappa)$, leading to the general expression for all the components

$$\hat{y}_\eta = E[y_\eta|\mathbf{x},\mathbf{x}^\imath,\mathbf{x}^\jmath,\mathbf{x}^\kappa] \quad \text{and} \quad \hat{y} = E[y|\mathbf{x},\mathbf{x}^\imath,\mathbf{x}^\jmath,\mathbf{x}^\kappa]$$

In other words, to capture the full second order information available we should use the original quaternion and its $\imath$, $\jmath$, κ involutions, allowing us to arrive at the *widely linear* model

$$y = \mathbf{u}^T\mathbf{x}+\mathbf{v}^T\mathbf{x}^\imath+\mathbf{g}^T\mathbf{x}^\jmath+\mathbf{h}^T\mathbf{x}^\kappa = \mathbf{w}^{aT}\mathbf{x}^a \tag{15}$$

where the augmented coefficient vector $\mathbf{w}^a = [\mathbf{u}^T, \mathbf{v}^T,\mathbf{g}^T,\mathbf{h}^T]^T$ and the augmented regressor vector $\mathbf{x}^a = [\mathbf{x}^T, \mathbf{x}^{\imath T},\mathbf{x}^{\jmath T},\mathbf{x}^{\kappa T}]^T$ comprise all the relevant terms (for more detail see [16]).

4.2. Augmented quaternion statistics

Unlike the real domain where complete second order statistics of $\mathbf{x}(k)$ are described by the covariance matrix, in the complex and quaternion domains the covariance matrix is sufficient to describe only second order circular (proper) signals. For the general second order noncircular (improper) signals, which exhibit unequal powers in the quaternion components, the additional pseudocovariance matrices $E[\mathbf{x}\mathbf{x}^{\imath H}]$, $E[\mathbf{x}\mathbf{x}^{\jmath H}]$ and $E[\mathbf{x}\mathbf{x}^{\kappa H}]$ are needed to describe complete second order statistics. This is achieved based on the quaternion widely linear model in (15), where the augmented vector $\mathbf{x}^a = [\mathbf{x}^T,\mathbf{x}^{\imath H},\mathbf{x}^{\jmath H},\mathbf{x}^{\kappa H}]^T$ is used to produce the augmented covariance matrix $\mathbf{R}_{xx} = E[\mathbf{x}^a\mathbf{x}^{aH}]$, which comprises information from both the covariance matrix and the three pseudocovariance matrices, and is given by

$$\mathbf{R}_{xx} = \begin{bmatrix} \mathbf{R} & \mathbf{P} & \mathbf{S} & \mathbf{T} \\ \mathbf{P}^\imath & \mathbf{R}^\imath & \mathbf{T}^\imath & \mathbf{S}^\imath \\ \mathbf{S}^\jmath & \mathbf{T}^\jmath & \mathbf{R}^\jmath & \mathbf{P}^\jmath \\ \mathbf{T}^\kappa & \mathbf{S}^\kappa & \mathbf{P}^\kappa & \mathbf{R}^\kappa \end{bmatrix} \tag{16}$$

where $\mathbf{R} = E[\mathbf{x}\mathbf{x}^H]$, $\mathbf{P} = E[\mathbf{x}\mathbf{x}^{\imath H}]$, $\mathbf{S} = E[\mathbf{x}\mathbf{x}^{\jmath H}]$ and $\mathbf{T} = E[\mathbf{x}\mathbf{x}^{\kappa H}]$.

Notice that for proper signals, the pseudocovariance matrices $\mathbf{P}$, $\mathbf{S}$ and $\mathbf{T}$ vanish—a signal that obeys this structure has a probability distribution that is rotation invariant with respect to all the six possible pairs of axes [16,17]. In this case the scatter graphs for each of the six pairs of axes $\{q_r,q_\imath\}$, $\{q_r,q_\jmath\}$, $\{q_r,q_\kappa\}$, $\{q_\imath,q_\jmath\}$, $\{q_\imath,q_\kappa\}$, $\{q_\jmath,q_\kappa\}$ describe a rotation invariant (circular) distribution. However, in most real world applications, probability density functions are rotation dependent, and require the use of the augmented covariance matrix.

Remark 7. The processing in $\mathbb{R}^4$ requires 10 covariance and cross-covariance matrices, as opposed to the four corresponding matrices in the quaternion domain. Although both representations convey the same information, the quaternion representation offers a more compact notation, enhanced physical insight, and more degrees of freedom in estimation [31].

4.3. Augmented statistics and the quaternion affine projection algorithm

To further illustrate that to process noncircular data, an adaptive filtering algorithm should incorporate the augmented statistics, consider the real APA which can be seen as an approximation to the steepest descent weight update [32], given by

$$\mathbf{w}(k+1) = \mathbf{w}(k)+\mu(E[\mathbf{x}(k)d(k)]-E[\mathbf{x}(k)\mathbf{x}^T(k)]\mathbf{w}(k)) \tag{17}$$

where $\mathbf{w}$ is the filter coefficient vector, μ the stepsize, $\mathbf{x}$ the input vector and d the teaching signal. The weight update is a function of the second order statistics of the input vector $\mathbf{x}(k)$ as exemplified by the term $\mathbf{x}\mathbf{x}^T$. To deal with both circular and noncircular signals the quaternion valued APA must therefore employ the augmented statistics based on the augmented covariance matrix in (16), a subject of this work.

5. A class of quaternion affine projection algorithms

To provide a unifying approach to the derivation of both the QAPA and WLQAPA, we start by deriving an expression for the WLQAPA and show that for circular signals it simplifies into the QAPA. The aim of WLQAPA class of algorithms is to minimize adaptively the squared Euclidean norm of the change in the augmented weight vector $\mathbf{w}^a \in \mathbb{H}^{M\times 1}$, that is

$$\begin{aligned}&\text{minimize} \quad \|\delta\mathbf{w}^a(k+1)\|^2 = \|\mathbf{w}^a(k+1)-\mathbf{w}^a(k)\|^2\\ &\text{subject to } d(k-n) = \mathbf{w}^{aH}(k+1)\mathbf{x}^a(k-n)\end{aligned} \tag{18}$$

for $n=0,\ldots,N-1$, where the augmented weight vector $\mathbf{w}^a=[\mathbf{u}^T,\mathbf{v}^T,\mathbf{g}^T,\mathbf{h}^T]^T \in \mathbb{H}^{4\,M\times 1}$, the augmented input vector $\mathbf{x}^a=[\mathbf{x}^T,\mathbf{x}^{iT},\mathbf{x}^{jT},\mathbf{x}^{kT}] \in \mathbb{H}^{4M\times 1}$ and $\|\cdot\|^2$ is the Euclidean norm.

Using the Lagrange multipliers to solve the constrained optimization problem, the cost function to be minimized can be written as

$$\begin{aligned}J(k) = &\|\mathbf{u}(k+1)-\mathbf{u}(k)\|^2+\|\mathbf{v}(k+1)-\mathbf{v}(k)\|^2+\|\mathbf{g}(k+1)-\mathbf{g}(k)\|^2\\ &+\|\mathbf{h}(k+1)-\mathbf{h}(k)\|^2+\Re[(\mathbf{d}^T(k)-\mathbf{u}^H(k+1)\mathbf{A}(k)-\mathbf{v}^H(k+1)\mathbf{A}^i(k)\\ &-\mathbf{g}^H(k+1)\mathbf{A}^j(k)-\mathbf{h}^H(k+1)\mathbf{A}^\kappa(k))\boldsymbol{\lambda}^*]\end{aligned}$$

where the symbol $\Re[\cdot]$ denotes the real part of a quaternion variable and

$$\begin{aligned}&\mathbf{A}(k)=[\mathbf{x}(k),\mathbf{x}(k-1),\ldots,\mathbf{x}(k-N+1)]\\ &\mathbf{d}(k)=[d(k),d(k-1),\ldots,d(k-N+1)]^T\\ &\boldsymbol{\lambda}=[\lambda_0,\lambda_1,\ldots,\lambda_{k-N+1}]^T\end{aligned}$$

are past values in the filter memory. Standard quaternion differentiation does not allow for the calculation of the gradient of $J(k)$, however, using the $\mathbb{HR}^*$-gradient in (10), the gradient of $J(k)$ with respect to the augmented weight vector $\mathbf{w}^{a*}(k+1)$ can be obtained by employing the vector derivatives

$$\begin{aligned}\frac{\partial J(k)}{\partial \mathbf{u}^*(k+1)} &= \mathbf{u}(k+1)-\mathbf{u}(k)-\frac{1}{2}(\mathbf{u}(k+1)-\mathbf{u}(k))^*\\ &\quad+\mathbf{A}(k)\boldsymbol{\lambda}^*-\frac{1}{2}(\mathbf{A}(k)\boldsymbol{\lambda}^*)^*\end{aligned}$$

$$\begin{aligned}\frac{\partial J(k)}{\partial \mathbf{v}^*(k+1)} &= \mathbf{v}(k+1)-\mathbf{v}(k)-\frac{1}{2}(\mathbf{v}(k+1)-\mathbf{v}(k))^*\\ &\quad+\mathbf{A}^i(k)\boldsymbol{\lambda}^*-\frac{1}{2}(\mathbf{A}^i(k)\boldsymbol{\lambda}^*)^*\end{aligned}$$

$$\begin{aligned}\frac{\partial J(k)}{\partial \mathbf{g}^*(k+1)} &= \mathbf{g}(k+1)-\mathbf{g}(k)-\frac{1}{2}(\mathbf{g}(k+1)-\mathbf{g}(k))^*\\ &\quad+\mathbf{A}^j(k)\boldsymbol{\lambda}^*-\frac{1}{2}(\mathbf{A}^j(k)\boldsymbol{\lambda}^*)^*\end{aligned}$$

$$\begin{aligned}\frac{\partial J(k)}{\partial \mathbf{h}^*(k+1)} &= \mathbf{h}(k+1)-\mathbf{h}(k)-\frac{1}{2}(\mathbf{h}(k+1)-\mathbf{h}(k))^*\\ &\quad+\mathbf{A}^\kappa(k)\boldsymbol{\lambda}^*-\frac{1}{2}(\mathbf{A}^\kappa(k)\boldsymbol{\lambda}^*)^*\end{aligned}$$

where

$$\frac{\partial J(k)}{\partial \mathbf{w}^*}=\left[\frac{\partial J(k)}{\partial w_0^*},\frac{\partial J(k)}{\partial w_1^*},\ldots,\frac{\partial J(k)}{\partial w_M^*}\right]^T$$

for $\mathbf{w}=\{\mathbf{u},\mathbf{v},\mathbf{g},\mathbf{h}\}$. Setting the above derivatives to $\mathbf{0}$ and solving for $\boldsymbol{\lambda}$ gives the weight updates

$$\mathbf{w}(k+1)=\mathbf{w}(k)+\mathbf{A}^a(k)(\mathbf{A}^{aH}(k)\mathbf{A}^a(k))^{-1}\mathbf{e}^*(k) \tag{19}$$

where

$$\begin{aligned}&\mathbf{A}^a=[\mathbf{A}^T,\mathbf{A}^{iT},\mathbf{A}^{jT},\mathbf{A}^{kT}]^T\\ &\mathbf{e}(k)=\mathbf{d}(k)-(\mathbf{w}^{aH}(k+1)\mathbf{A}^a(k))^T\end{aligned}$$

To prevent the normalization matrix $\mathbf{A}(k)^{aH}\mathbf{A}^a(k)$ from becoming singular, a small regularization term $\varepsilon\mathbf{I}\in\mathbb{H}^{4N\times 4N}$ is typically added, where $\mathbf{I}$ is the identity matrix. A real step size μ can also be incorporated to improve steady state performance, giving the final weight update of the widely linear quaternion valued affine projection algorithm (WLQAPA) in the form

$$\mathbf{w}^a(k+1)=\mathbf{w}^a(k)+\mu\mathbf{A}^a(k)(\mathbf{A}^{aH}(k)\mathbf{A}^a(k)+\epsilon\mathbf{I})^{-1}\mathbf{e}^*(k) \tag{20}$$

where ε takes a small value, and for stability $\mu<2$ (shown later).

When the input vector is strictly linear, then $\mathbf{x}^a=\mathbf{x}$, and the WLQAPA simplifies into the strictly linear QAPA, for which the weight update takes the form

$$\mathbf{w}(k+1)=\mathbf{w}(k)+\mu\mathbf{A}(k)(\mathbf{A}^H(k)\mathbf{A}(k)+\epsilon\mathbf{I})^{-1}\mathbf{e}^*(k) \tag{21}$$

where the error term $\mathbf{e}(k)=\mathbf{d}(k)-(\mathbf{w}^H(k+1)\mathbf{A}(k))^T$ and the regularization term $\varepsilon\mathbf{I}\in\mathbb{H}^{N\times N}$.

6. Mean square error analysis of the strictly linear QAPA on noncircular data

To investigate the MSE performance, we next evaluate the MSE of the strictly linear QAPA, given by

$$MSE=\lim_{k\to\infty}E[|e(k)|^2] \tag{22}$$

where

$$e(k)=d(k)-\mathbf{w}^H(k)\mathbf{x}(k) \tag{23}$$

For generality, the teaching signal $d(k)$ is assumed to be coming from a widely linear system, taking the form

$$d(k)=\mathbf{w}_o^H\mathbf{x}(k)+\mathbf{v}_o^H\mathbf{x}^i(k)+\mathbf{g}_o^H\mathbf{x}^j(k)+\mathbf{h}_o^H\mathbf{x}^\kappa(k)+\nu(k) \tag{24}$$

where $\nu(k)$ is zero mean quadruply white Gaussian noise.

To evaluate the mean square error we substitute for $\tilde{\mathbf{w}}(k)=\mathbf{w}_o-\mathbf{w}(k)$ into the weight update (21) to yield

$$\tilde{\mathbf{w}}(k+1)=\tilde{\mathbf{w}}(k)-\mu\mathbf{A}(k)(\mathbf{A}^H(k)\mathbf{A}(k)+\epsilon\mathbf{I})^{-1}\mathbf{e}^*(k)$$

Upon applying the Hermitian transpose operator to both sides

$$\tilde{\mathbf{w}}^H(k+1)=\tilde{\mathbf{w}}^H(k)-\mu\mathbf{e}^T(k)(\mathbf{A}^H(k)\mathbf{A}(k)+\epsilon\mathbf{I})^{-1}\mathbf{A}^H(k) \tag{25}$$

and after post-multiplying by $\mathbf{A}(k)$ we obtain

$$\tilde{\mathbf{w}}^H(k+1)\mathbf{A}(k)=\tilde{\mathbf{w}}^H(k)\mathbf{A}(k)-\mu\mathbf{e}^T(k)(\mathbf{A}^H(k)\mathbf{A}(k)+\epsilon\mathbf{I})^{-1}\mathbf{A}^H(k)\mathbf{A}(k) \tag{26}$$

The a priori error, $\mathbf{e}_a(k)$, and a posteriori error $\mathbf{e}_p(k)$ can now be defined as

$$\mathbf{e}_a^T(k) = \mathbf{w}^H(k)\mathbf{A}(k), \quad \mathbf{e}_p^T(k) = \mathbf{w}^H(k+1)\mathbf{A}(k)$$

which in combination with (26) gives

$$\mu\mathbf{e}^T(k)(\mathbf{A}^H(k)\mathbf{A}(k)+\epsilon\mathbf{I})^{-1} = (\mathbf{e}_a^T(k)-\mathbf{e}_p^T(k))(\mathbf{A}^H(k)\mathbf{A}(k))^{-1}$$

allowing us to rewrite (25) in the form

$$\tilde{\mathbf{w}}^H(k+1) = \tilde{\mathbf{w}}^H(k)-(\mathbf{e}_a^T(k)-\mathbf{e}_p^T(k))(\mathbf{A}^H(k)\mathbf{A}(k))^{-1}\mathbf{A}^H(k)$$

and to re-arrange the above terms and evaluate the energy ($\|\cdot\|^2$), to give

$$\begin{aligned}&\|\mathbf{w}(k+1)\|^2+\mathbf{e}_a^T(k)(\mathbf{A}^H(k)\mathbf{A}(k))^{-1}\mathbf{e}_a^*(k)\\ &\quad = \|\mathbf{w}(k)\|^2+\mathbf{e}_p^T(k)(\mathbf{A}^H(k)\mathbf{A}(k))^{-1}\mathbf{e}_p^*(k)\end{aligned}$$

This expression is known as the energy conservation relationship [20]. Since we are interested in the mean square error at the steady state, that is, in the limit as $k\to\infty$, upon application of the statistical expectation operator we have

$$E[\mathbf{e}_a^T(k)(\mathbf{A}^H(k)\mathbf{A}(k))^{-1}\mathbf{e}_a^*(k)] = E[\mathbf{e}_p^T(k)(\mathbf{A}^H(k)\mathbf{A}(k))^{-1}\mathbf{e}_p^*(k)]$$

Substitute for $\mathbf{e}_p^T(k) = \mathbf{e}_a^T(k)-\mu\mathbf{e}^T(k)(\mathbf{A}^H(k)\mathbf{A}(k)+\epsilon\mathbf{I})^{-1}\mathbf{A}^H(k)\mathbf{A}(k)$ to arrive at

$$\begin{aligned}&E[\mathbf{e}_a^T(k)(\mathbf{A}^H(k)\mathbf{A}(k))^{-1}\mathbf{e}_a^*(k)] = E[\mathbf{e}_a^T(k)(\mathbf{A}^H(k)\mathbf{A}(k))^{-1}\mathbf{e}_a^*(k)]\\ &\quad -\mu E[\mathbf{e}_a^T(k)(\mathbf{A}^H(k)\mathbf{A}(k)+\epsilon\mathbf{I})^{-1}\mathbf{e}^*(k)]\\ &\quad -\mu E[\mathbf{e}^T(k)(\mathbf{A}^H(k)\mathbf{A}(k)+\epsilon\mathbf{I})^{-1}\mathbf{e}_a^*(k)]+\mu^2E[\mathbf{e}^T(k)(\mathbf{A}^H(k)\mathbf{A}(k)\\ &\quad +\epsilon\mathbf{I})^{-1}\mathbf{A}^H(k)\mathbf{A}(k)(\mathbf{A}^H(k)\mathbf{A}(k)+\epsilon\mathbf{I})^{-1}\mathbf{e}^*(k)]\end{aligned}$$

For this to be satisfied, it must hold that

$$\mu E[\mathbf{e}^T(k)\mathbf{D}(k)\mathbf{e}^*(k)] = E[\mathbf{e}_a^T(k)\mathbf{C}(k)\mathbf{e}^*(k)]+E[\mathbf{e}^T(k)\mathbf{C}(k)\mathbf{e}_a^*(k)] \tag{27}$$

where

$$\begin{aligned}&\mathbf{C}(k) = (\mathbf{A}^H(k)\mathbf{A}(k)+\epsilon\mathbf{I})^{-1}\\ &\quad \mathbf{D}(k) = (\mathbf{A}^H(k)\mathbf{A}(k)+\epsilon\mathbf{I})^{-1}\mathbf{A}^H(k)\mathbf{A}(k)(\mathbf{A}^H(k)\mathbf{A}(k)+\epsilon\mathbf{I})^{-1}\end{aligned}$$

From (23), it can be shown that the MSE and the a priori error $\mathbf{e}_a(k)$ are related by

$$\mathbf{e}^T(k) = \mathbf{e}_a^T(k)+\mathbf{w}^{cH}(k)\mathbf{A}^c(k)+\mathbf{v}^T(k) \tag{28}$$

where

$$\mathbf{A}^c(k) = [\mathbf{A}^{iT}(k),\mathbf{A}^{jT}(k),\mathbf{A}^{\kappa T}(k)]^T \tag{29}$$

$$\mathbf{w}^c(k) = [\mathbf{v}_o^T(k),\mathbf{g}_o^T(k),\mathbf{h}_o^T(k)]^T \tag{30}$$

From the definition of the error vector $\mathbf{e}(k) = [e(k), e(k-1),\ldots,e(k-N+1)]^T$ and using the symbol $(\cdot)_1$ to define the first element of the term in hand, the mean square error can now be expressed as

$$MSE = \lim_{k\to\infty} E\|e(k)\| = \lim_{k\to\infty} E\|e_a(k)+\nu(k)+(\mathbf{w}^{cH}(k)\mathbf{A}^c(k))_1\|^2 \tag{31}$$

Using the usual 'independence assumptions' and assuming that $\lim_{k\to\infty}E[e_a(k)] = E[\nu(k)] = 0$, we arrive at

$$MSE = \lim_{k\to\infty} E[\|e_a(k)\|^2]+\|(\mathbf{w}^{cH}(k)\mathbf{A}^c(k))_1\|^2+E\|\nu(k)\| \tag{32}$$

$$MSE = EMSE+\|(\mathbf{w}^{cH}(k)\mathbf{A}^c(k))_1\|^2+\sigma_\nu^2 \tag{33}$$

where the acronym EMSE is the excess mean square error, that is, the deviation from the optimal theoretical MSE, defined as

$$EMSE = \lim_{k\to\infty} E[\|e_a(k)\|^2] \tag{34}$$

To find an expression for $\lim_{k\to\infty}E[\|e_a(k)\|^2]$ using (27), we proceed by simplifying the three components in (27). Using (28) to substitute for $\mathbf{e}(k)$ in $\mu E[\mathbf{e}^T(k)\mathbf{D}(k)\mathbf{e}^*(k)]$ and taking the limit as $k\to\infty$ ($E[\mathbf{e}_a(k)] = 0$]) we have[3]

$$\begin{aligned}&\mu E[\mathbf{e}^T(k)\mathbf{D}(k)\mathbf{e}^*(k)] = \mu E[\mathbf{e}_a^T(k)\mathbf{D}(k)\mathbf{e}_a^*(k)]+\mu E[\mathbf{v}^T(k)\mathbf{D}(k)\mathbf{v}^*(k)]\\ &\quad +\mu E[\mathbf{w}^{cH}(k)\mathbf{A}^c(k)\mathbf{D}(k)\mathbf{A}^{cH}(k)\mathbf{w}^c(k)]\end{aligned}$$

Repeating this procedure for the other two terms in (27) gives

$$\begin{aligned}&E[\mathbf{e}_a^T(k)\mathbf{C}(k)\mathbf{e}^*(k)] = E[\mathbf{e}_a^T(k)\mathbf{C}(k)\mathbf{e}_a^*(k)]\\ &\quad E[\mathbf{e}^T(k)\mathbf{C}(k)\mathbf{e}_a^*(k)] = E[\mathbf{e}_a^T(k)\mathbf{C}(k)\mathbf{e}_a^*(k)]\end{aligned}$$

allowing us to re-write (27) as

$$\begin{aligned}&\mu E[\mathbf{e}_a^T(k)\mathbf{D}(k)\mathbf{e}_a^*(k)]+\mu E[\mathbf{v}^T(k)\mathbf{D}(k)\mathbf{v}^*(k)]\\ &\quad +\mu E[\mathbf{w}^{cH}(k)\mathbf{A}^c(k)\mathbf{D}(k)\mathbf{A}^{cH}(k)\mathbf{w}^c] = 2E[\mathbf{e}_a^T(k)\mathbf{C}(k)\mathbf{e}_a^*(k)(k)]\end{aligned} \tag{35}$$

This form cannot be used to obtain an expression for EMSE, and to this end, we shall now rewrite each of the three terms in (35) in a more convenient form, starting with $E[\mathbf{e}_a(k)\mathbf{D}(k)\mathbf{e}_a^H(k)]$.

The matrix $\mathbf{D}(k)$ is Hermitian and hence positive definite, and therefore the above term is a positive scalar and we therefore have[4]

$$\begin{aligned}&\mu E[\mathbf{e}_a^T(k)\mathbf{D}(k)\mathbf{e}_a^*(k)] = \mu\Re\left[\mathrm{Tr}(E\left[\mathbf{e}_a^*(k)\mathbf{e}_a^T(k)\mathbf{D}(k)\right])\right]\\ &\quad = \mu\Re\left[\mathrm{Tr}(E[\mathbf{e}_a^*(k)\mathbf{e}_a^T(k)]E[\mathbf{D}(k)])\right]\end{aligned}$$

It can be easily shown that at high signal to noise ratio (SNR), at steady state we have $E[\mathbf{e}_a^*(k)\mathbf{e}_a(k)] = E\left|e_a(k)\right|^2\mathbf{S}$, where for a step size μ close to unity, $\mathbf{S}\approx[\mathbf{1}\mathbf{1}^T]$ where$\mathbf{1} = [1,0,\ldots,0]^T$. Thus, for a step size approaching zero $\mathbf{S}\approx\mathbf{I}$ and using the property $\mathrm{Tr}(c\mathbf{A}) = c\,\mathrm{Tr}(\mathbf{A})$ we have

$$\mu E[\mathbf{e}_a^T(k)\mathbf{D}(k)\mathbf{e}_a^*(k)] = \mu E\left|e_a(k)\right|^2\Re[\mathrm{Tr}(\mathbf{S}E[\mathbf{D}(k)])]$$

Following the same approach for the other two terms in (35) gives

$$\begin{aligned}&E[\mathbf{e}_a^T(k)\mathbf{C}(k)(\mathbf{e}_a^*(k)] = E\left|e_a(k)\right|^2\Re[\mathrm{Tr}(\mathbf{S}E[\mathbf{C}(k)])]\\ &\quad \mu E[\mathbf{v}^T(k)\mathbf{D}(k)\mathbf{v}^*(k)] = \mu\sigma_\nu^2\Re[\mathrm{Tr}(E[\mathbf{D}(k)])]\end{aligned}$$

allowing us to re-write (35) as

$$\begin{aligned}&E\left|e_a(k)\right|^2(2\Re[\mathrm{Tr}(\mathbf{S}E[\mathbf{C}(k)])]-\mu\Re[\mathrm{Tr}(\mathbf{S}E[\mathbf{D}(k)])])\\ &\quad = \mu\sigma_\nu^2\Re[\mathrm{Tr}(E[\mathbf{D}(k)])]+\mu E[\mathbf{w}^{cH}(k)\mathbf{A}^c(k)\mathbf{D}(k)\mathbf{A}^{cH}(k)\mathbf{w}^c(k)]\end{aligned}$$

[3] We make the usual 'independence' assumptions that $\mathbf{e}_a(k)$ and $\mathbf{v}(k)$ are statistically independent from $\mathbf{A}(k)$.

[4] The real part operator $\Re[\cdot]$ arises because although quaternion multiplication is noncommutative, the real component of a quaternion product is commutative for cyclic permutations of the product.

to give the expression for EMSE in the final form

$$EMSE = \frac{\sigma_v^2 \mu \Re[\mathrm{Tr}(E[\mathbf{D}(k)])] + \mu E[\mathbf{w}^{cH}(k)\mathbf{A}^c(k)\mathbf{D}(k)\mathbf{A}^{cH}(k)\mathbf{w}^c(k)]}{2\Re[\mathrm{Tr}(\mathbf{S}E[\mathbf{C}(k)])] - \mu\Re[\mathrm{Tr}(\mathbf{S}E[\mathbf{D}(k)])]} \tag{36}$$

Finally, for the theoretical MSE we have

$$MSE = \frac{\sigma_v^2 \mu \Re[\mathrm{Tr}(E[\mathbf{D}(k)])] + \mu E[\mathbf{w}^{cH}(k)\mathbf{A}^c(k)\mathbf{D}(k)\mathbf{A}^{cH}(k)\mathbf{w}^c(k)]}{2\Re[\mathrm{Tr}(\mathbf{S}E[\mathbf{C}(k)])] - \mu\Re[\mathrm{Tr}(\mathbf{S}E[\mathbf{D}(k)])]} + \left|(\mathbf{w}^{cH}(k)\mathbf{A}^c(k))_1\right|^2 + \sigma_v^2 \tag{37}$$

Remark 8. For a small value of the regularization parameter ϵ, we can assume $\mathbf{C}(k) = \mathbf{D}(k)$ and the EMSE simplifies into

$$EMSE = \frac{\sigma_v^2 \mu \Re[\mathrm{Tr}(E[\mathbf{C}(k)])] + \mu E[\mathbf{w}^{cH}(k)\mathbf{A}^c(k)\mathbf{C}(k)\mathbf{A}^{cH}(k)\mathbf{w}^c(k)]}{(2-\mu)\Re[\mathrm{Tr}(\mathbf{S}E[\mathbf{C}(k)])]} \tag{38}$$

Observe that due to the term in the denominator the EMSE diverges for $\mu \approx 2$, imposing a bound on the stepsize $0 < \mu < 2$. The analysis of APA class of methods is routinely performed separately for the small and big stepsize cases. The small stepsize cases are addressed in Remark 9 whereas the big stepsize case is addressed in Remark 10.

Remark 9 (*Small stepsize analysis*). For a small value of the regularization parameter ϵ and a small step size μ, we can assume $\mathbf{C}(k) = \mathbf{D}(k)$ and $\mathbf{S} = \mathbf{I}$, allowing us to simplify the EMSE as

$$EMSE = \frac{\sigma_v^2 \mu}{(2-\mu)} + \frac{\mu E[\mathbf{w}^{cH}(k)\mathbf{A}^c(k)\mathbf{C}(k)\mathbf{A}^{cH}(k)\mathbf{w}^c(k)]}{(2-\mu)\Re[\mathrm{Tr}(E[\mathbf{C}(k)])} \tag{39}$$

Remark 10 (*Large stepsize analysis*). For a small value of the regularization parameter ϵ and a large step size $\mu \approx 1$, we can assume $\mathbf{C}(k) = \mathbf{D}(k)$ and $\mathbf{S} = \mathbf{1}\mathbf{1}^T$, allowing us to simplify the EMSE into

$$EMSE = \frac{\sigma_v^2 \mu \Re[\mathrm{Tr}(E[\mathbf{C}(k)])] + \mu E[\mathbf{w}^{cH}(k)\mathbf{A}^c(k)\mathbf{C}(k)\mathbf{A}^{cH}(k)\mathbf{w}^c(k)]}{(2-\mu)\Re[(E[\mathbf{C}(k)])_{1,1}]} \tag{40}$$

where $(\cdot)_{x,y}$ is the element in the xth row and yth column of the bracketed term.

Remark 11 (*MSE for circular signals*). The term $E[\mathbf{w}^{cH}(k)\mathbf{A}^c(k)\mathbf{C}(k)\mathbf{A}^{cH}(k)\mathbf{w}^c(k)]$ in (39) and (40) vanishes, giving the EMSE of strictly linear QAPA for proper signals. For improper data, however, this term can have a high value, illustrating the inadequacy of the strictly linear QAPA for the processing of second order noncircular signals.

7. Mean square error analysis of the widely linear QAPA on noncircular data

Following the same approach as for the QAPA, we can re-write the weight updates of the widely linear QAPA based on the augmented weight error vector $\tilde{\mathbf{w}}^a(k) = \mathbf{w}_o^a - \mathbf{w}^a(k)$, to give

$$\tilde{\mathbf{w}}^a(k+1) = \tilde{\mathbf{w}}^a(k) - \mathbf{A}^a(k)(\mathbf{A}^{aH}(k)\mathbf{A}^a(k) + \delta\mathbf{I})^{-1}\mathbf{e}^*(k)$$

In this way, the expression for the conservation of energy of the weight update takes the form

$$\|\mathbf{w}^a(k+1)\|^2 + \mathbf{e}_a^T(k)(\mathbf{A}^{aH}(k)\mathbf{A}^a(k))^{-1}\mathbf{e}_a^*(k) = \|\mathbf{w}^a(k)\|^2 + \mathbf{e}_p^T(k)(\mathbf{A}^{aH}(k)\mathbf{A}^a(k))^{-1}\mathbf{e}_p^*(k)$$

where the a posteriori error $\mathbf{e}_p(k)$ and a priori error $\mathbf{e}_a(k)$ are defined as

$$\mathbf{e}_p^T(k) = \tilde{\mathbf{w}}^{aH}(k+1)\mathbf{A}^a(k) \tag{41}$$

$$\mathbf{e}_a^T(k) = \tilde{\mathbf{w}}^{aH}(k)\mathbf{A}^a(k) \tag{42}$$

and

$$\mathbf{e}(k) = \mathbf{e}_a(k) + \mathbf{v}(k) \tag{43}$$

Comparing with the error in (28) for QAPA, we observe that the expression for WLQAPA does not comprise the term, $\mathbf{w}^{cH}\mathbf{A}^c$, and MSE takes the form

$$MSE = \lim_{k\to\infty} E[\|e_a(k)\|^2] + E\|v(k)\| \tag{44}$$

$$MSE = EMSE + \sigma_v^2 \tag{45}$$

Following on the analysis of EMSE for the strictly linear QAPA in (34)–(36), we obtain the EMSE for WLQAPA in the form

$$EMSE = E\left|e_a(k)\right|^2 = \frac{\sigma_v^2 \mu \Re[\mathrm{Tr}(E[\mathbf{D}(k)])]}{2\Re[\mathrm{Tr}(\mathbf{S}E[\mathbf{C}(k)])] - \mu\Re[\mathrm{Tr}(\mathbf{S}E[\mathbf{D}(k)])]}$$

where

$$\mathbf{C}(k) = (\mathbf{A}^{aH}(k)\mathbf{A}^a(k) + \epsilon\mathbf{I})^{-1} \tag{46}$$

$$\mathbf{D}(k) = (\mathbf{A}^{aH}(k)\mathbf{A}^a(k) + \epsilon\mathbf{I})^{-1}\mathbf{A}^{aH}(k)\mathbf{A}^a(k)(\mathbf{A}^{aH}(k)\mathbf{A}^a(k) + \epsilon\mathbf{I})^{-1} \tag{47}$$

Remark 12. For a small value of the regularization parameter ϵ, we can assume that $\mathbf{C}(k) = \mathbf{D}(k)$, allowing us to simplify the EMSE into

$$EMSE = E\left|e_a(k)\right|^2 = \frac{\sigma_v^2 \mu \Re[\mathrm{Tr}(E[\mathbf{C}(k)])]}{(2-\mu)\Re[\mathrm{Tr}(\mathbf{S}E[\mathbf{C}(k)])]} \tag{48}$$

Similarly to the QAPA, the bounds $0 < \mu < 2$ are imposed on the step size μ. Below we look at the cases with a small and big stepsize.

Remark 13 (*Small stepsize analysis*). For a small value of the regularization parameter ϵ and a small step size μ, we can assume $\mathbf{C}(k) = \mathbf{D}(k)$ and $\mathbf{S} = \mathbf{I}$, allowing us to simplify the EMSE into

$$EMSE = E\left|e_a(k)\right|^2 = \frac{\sigma_v^2 \mu}{(2-\mu)} \tag{49}$$

Note that this expression conforms to that obtained in the complex domain in [20].

Remark 14 (*Large stepsize analysis*). For a small value of the regularization parameter ϵ and a step size $\mu \approx 1$, we can assume $\mathbf{C}(k) = \mathbf{D}(k)$ and $\mathbf{S} = \mathbf{1}\mathbf{1}^T$, allowing us to

simplify the EMSE into

$$EMSE = E\left|e_a(k)\right|^2 = \frac{\sigma_v^2 \mu \Re[\mathrm{Tr}(E[\mathbf{C}(k)])]}{(2-\mu)(E[\mathbf{C}(k)])_{1,1}} \tag{50}$$

When the cross-diagonal terms of $E[\mathbf{C}(k)]$ are small compared to the diagonal terms (i.e. the regression vectors are orthogonal or close to being orthogonal), we can further assume that

$$\frac{1}{(E[\mathbf{C}(k)])_{1,1}} \approx \mathrm{Tr}(\mathbf{R}_{uu}), \quad \Re[\mathrm{Tr}(E[\mathbf{C}(k)])] \approx E\left[\frac{N}{\|\mathbf{u}(k)\|^2}\right] \tag{51}$$

where $\mathbf{R}_{uu} = E[\mathbf{u}(k)\mathbf{u}^H(k)]$, allowing us to simplify the expression for EMSE as

$$EMSE = \frac{\sigma_v^2 \mu}{(2-\mu)} \mathrm{Tr}(\mathbf{R}_{uu}) E\left[\frac{N}{\|\mathbf{u}(k)\|^2}\right] \tag{52}$$

Notice that, similarly to the complex domain [20], the EMSE is proportional to the number of constraints N.

Remark 15. Compared to the EMSE expression in (40), the EMSE of WLQAPA in (50) does not depend on the noncircularity of the signal, and is second order optimal for the generality of quaternion signals, both proper and improper. WLQAPA therefore achieves the same MSE as the strictly linear QAPA for proper signals and greatly enhanced performance for improper signals. The QAPA offers a lower computational cost and faster convergence for circular signals due to the fewer weights that are to be adjusted but is inadequate for general noncircular signals.

8. Simulations

The quaternion affine projection algorithm and its widely linear extension are next validated by comprehensive simulations over a number of scenarios. All the simulations were conducted in the prediction setting (one step ahead), with filter length $M=10$ and the regularization parameter $\epsilon = 0.01$. The test signals employed in the simulations were:

- The circular AR(1) process driven by both quadruply white circular Gaussian noise given by

$$y(k) = 0.9y(k-1) + n(k) \tag{53}$$

- The noncircular MA(3) process driven by quadruply white noncircular Gaussian noise

$$y(k) = ax(k-1) + bx^{\imath}(k-2) + cx^{\kappa}(k-3) + n(k) \tag{54}$$

  where $\{a,b,c\}$ are quaternion valued coefficients.
- The widely linear autoregressive moving average process driven by quadruply white circular Gaussian noise, giving the noncircular signal

$$y(k) = y(k-1) + 2n(k) + 0.5n^{\imath}(k) + n(k-1) + 0.9n^{\jmath}(k-1) \tag{55}$$

- The 4D noncircular wind signal[5] with three different dynamic regions, identified as low, medium and high dynamics, based on the changes in the wind intensity.
- The 3D noncircular body motion signal. Two gyroscopes were placed on the left hand and right hand of an athlete performing Tai Chi movements, recording two three-dimensional signals.[6]

Fig. 1 illustrates a geometric notion of noncircularity, by showing the scatter plots of the quaternion-valued signals considered. Observe that only the AR(1) signal had a rotation invariant distribution (circular) and that all the real world signals (wind, body motion) were noncircular.

In the first experiment we addressed by simulations the theoretical stability bound on the learning rate μ for the class of APA algorithms derived in Sections 6 and 7. Figs. 2 and 3 show that both QAPA and WLQAPA have a stability bound of $\mu < 2$ for all values of the constraint N, conforming with the corresponding bound in the complex domain [33] and Remarks 8 and 12. Fig. 4 shows the learning curves of QAPA and WLQAPA for the circular AR(1) process. Observe that, for the second order circular (proper) process both the strictly linear QAPA and the widely linear WLQAPA algorithms offered the same steady state performance and that QAPA exhibited faster initial convergence due to having fewer coefficients to adjust, conforming with Remark 15.

Fig. 5 repeats the previous experiment for the noncircular MA process; the WLQAPA was able to track the underlying dynamics of the signal in a second order optimal manner and therefore offered better steady state performance compared to the QAPA, validating the analysis in Section 7. In Fig. 6 we show how the performance of the WLQAPA on the noncircular MA process varies as the number of constraints N is increased, when the driving noise in (54) was colored. This is inline with the analysis in (52) (see Remark 14). The steady state performance decreased as the number of constraints N increased from $N=1$ (where the filter is equivalent to the normalized QLMS) to $N=4$. Also, as is the case with all APA algorithms [11] the convergence rate was faster as the number of constraints increased; this is expected as the driving noise to the model (54) was colored. Fig. 7 shows the learning curves for the noncircular ARMA process, with WLQAPA being able to better track the

[5] The wind speed measurement in the North, East and vertical direction formed the imaginary part of the full quaternion while the temperature was incorporated in the real part to form a full quaternion. The dataset was recorded using the WindMaster, a 3D Gill Instruments ultrasonic anemometer, which was resampled at 5 Hz for simulation purposes.

[6] The motion data was recorded using the XSense MTx 3DOF Orientation Tracker.

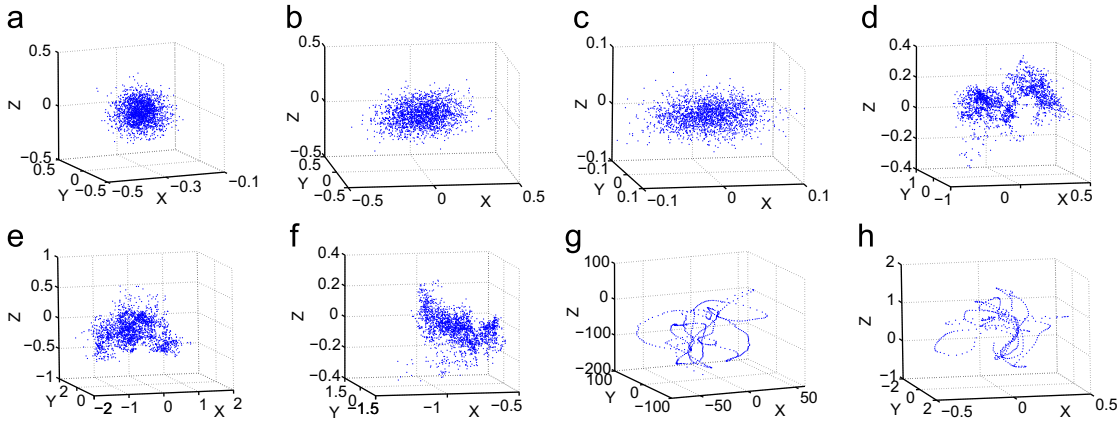

Fig. 1. Geometric view of circularity. To visualize the circularity on a three dimensional plot, only the i-, j- and k-components of the quaternion signals are plotted within a scatter diagram. Observe that only the AR(1) signal in (a) is circular Gaussian. The MA(3) and ARMA signals are both noncircular Gaussian signals. The real world wind signal and Tai Chi body motion signal both exhibit noncircular distributions. (a) AR(1) signal, (b) MA(3) signal, (c) ARMA signal, (d) Wind signal (low), (e) Wind signal (medium), (f) Wind signal (high), (g) Tai Chi body motion (left) and (h) Tai Chi body motion (right).

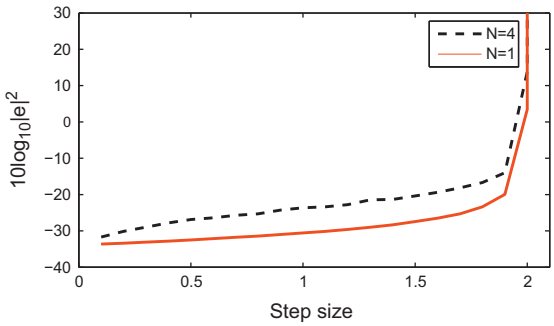

Fig. 2. Simulated MSE of QAPA as a function of the step size μ, for the improper ARMA process in (55).

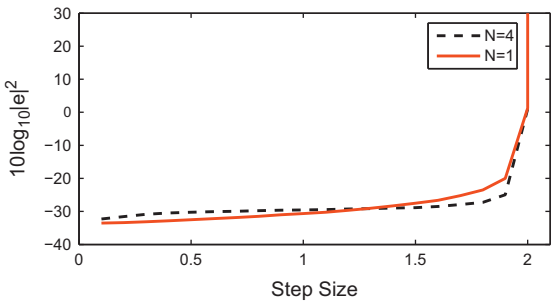

Fig. 3. Simulated MSE of WLQAPA as a function of the step size μ, for the improper ARMA process in (55).

underlying dynamics of the signal, therefore offering enhanced steady state performance.

In order to verify the theoretical result in (52) for the steady state performance of WLQAPA, we next performed one step ahead prediction of the improper ARMA signal in (55) and improper MA(3) process in (54) and compared the measured MSEs to the theoretical MSEs in (52). Fig. 8 compares the theoretical MSE derived in (52) to the simulated MSE for a step size in the range $\mu \in [0.05-1]$.

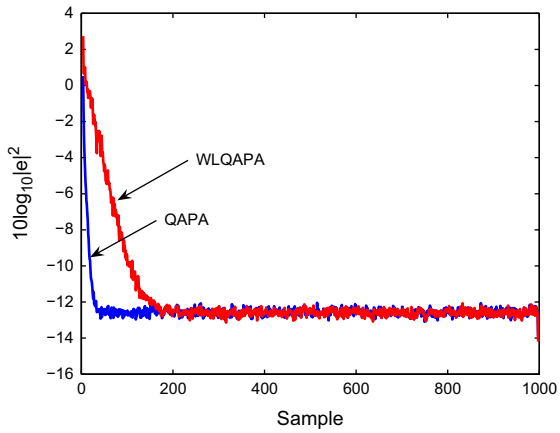

Fig. 4. Learning curves for QAPA and WLQAPA for the circular AR process in (53), with $\mu=1$ and $N=2$.

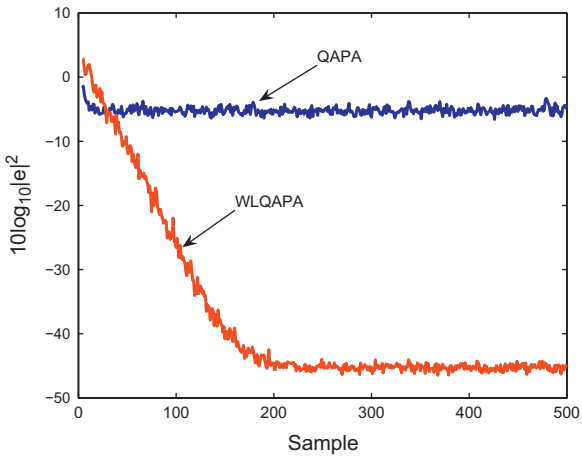

Fig. 5. Learning curves of QAPA and WLQAPA for the noncircular MA process in (54), with $\mu=1$ and $N=2$.

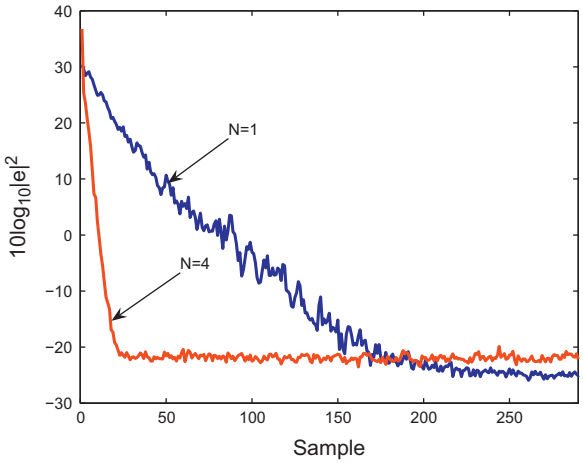

Fig. 6. Learning curves of WLQAPA for the noncircular MA process in (54), for $N=1$ and $N=4$.

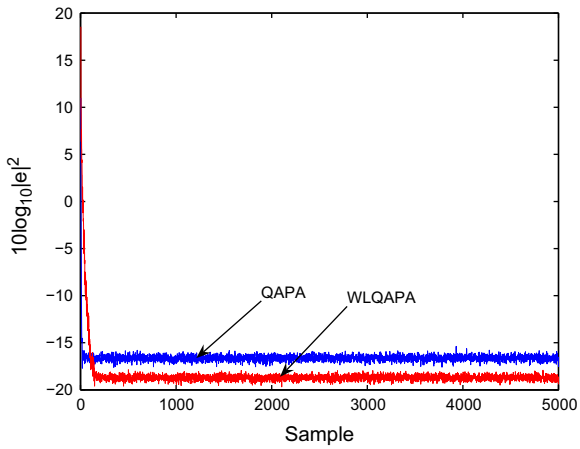

Fig. 7. Learning curves of QAPA and WLQAPA for the noncircular ARMA process in (55), with $\mu=1$ and $N=2$.

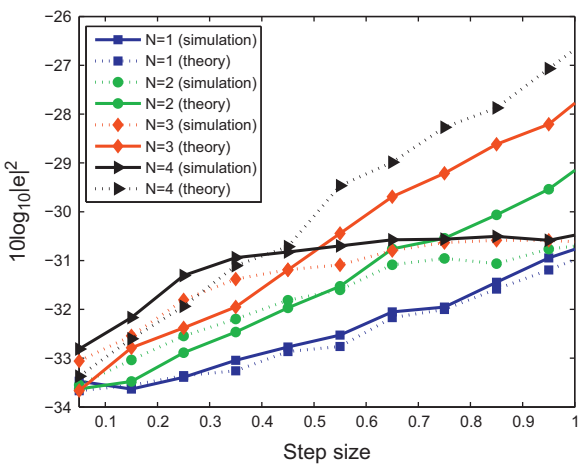

Fig. 8. Comparison of the theoretical bound in (50) and simulated steady state MSEs for WLQAPA, based on noncircular ARMA signal in (55), when $N=1,2,3$ and 4.

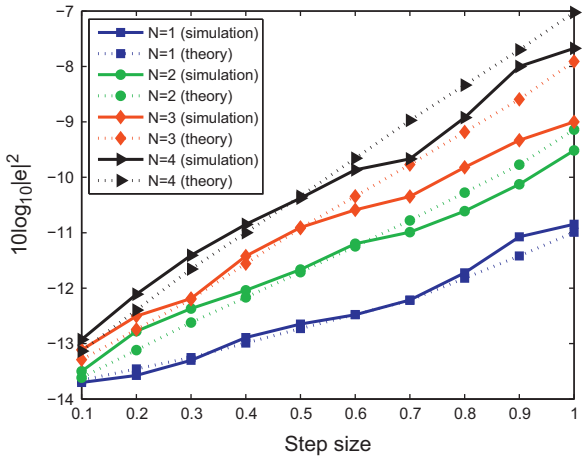

Fig. 9. Comparison of the theoretical bound in (50) and simulated steady state MSE of WLQAPA, on noncircular MA signal in (54) when $N=1,2,3$ and 4.

This allows us to make two observations. First, the theoretical MSE closely matches the simulated MSE when the number of constraints $N=1$. Second, for large step-sizes ($\mu>0.5$), as the number of constraints increases, the theoretical bound in (52) is no longer an accurate estimate of MSE. We can explain the discrepancy that exists for $N>1$ by the assumptions made in deriving the theoretical MSE expression in (52) as elaborated in Appendix.

In Fig. 9 we compare the theoretical MSE to the simulated MSE for the MA signal in (54). Observe that in this case the theoretical MSE closely matches the simulated MSE for every value of the constraint N. Note also that as the step size approaches 0 the performance difference between the number of constraints $N=1$ and $N=4$ decreases. This is expected from Eq. (49) which shows that for a small step size the MSE is independent of the number of constraints.

Table 1 illustrates the average MSE of the proposed algorithms in all the scenarios considered. Columns 2–4 show the results for the synthetic data. For the proper AR(1) model the performances of QAPA and WLQAPA were nearly identical, illustrating the ability of WLQAPA to deal with both proper and improper data. The MA(3) model had by design a very high degree of noncircularity (improper noise driving a widely linear model) and consequently the strictly linear QAPA was inadequate, the fact also confirmed in Fig. 5. The behavior for the improper ARMA process was along the same lines. Columns 5–9 show the MSE for both the QAPA and WLQAPA for the one step ahead prediction of the real world 4D wind signal and two 3D Tai Chi body motion signals, for $M=10$, $N=4$ and $\mu=1$. For the three wind regimes (low, medium, high) of different dynamics, due to the noncircular nature of wind (see Fig. 1), the WLQAPA offered better steady state performance for all the three wind regimes considered. As shown in Fig. 1 the Tai Chi body motion signal for both the left hand and right hand was also highly noncircular, consequently, the WLQAPA offered better tracking performance than the QAPA.

Table 1
MSE of QAPA and WLQAPA for different processes. Key: 1. AR(1), 2. MA(3), 3. ARMA, 4. Tai Chi (left), 5. Tai Chi (right), 6. Wind (low), 7. Wind (medium), and 8. Wind (high).

MSE (dB)	1	2	3	4	5	6	7	8
QAPA	−12.48	−4.19	−16.11	−20.38	−20.86	−13.72	−13.90	−11.13
WLQAPA	−12.49	−45.86	−17.39	−22.31	−23.26	−15.24	−15.31	−12.49

9. Conclusion

A class of quaternion affine projection algorithms (QAPA) have been introduced in order to provide a unified platform for fast and accurate adaptive filtering of both second order circular (proper) and noncircular (improper) real world signals. This has been achieved using the recent advances in the quaternion statistics (augmented quaternion statistics) and the emergence of the $\mathbb{HR}$-calculus for gradient calculation. Expressions for the MSE of the QAPA and WLQAPA have been obtained for the general case of improper signals, highlighting the advantage offered by the widely linear WLQAPA over the strictly linear QAPA in real world scenarios. Simulations on both circular and noncircular signals support the analysis.

Appendix A. The discrepancy between the measured MSE of the ARMA process in (55) and theoretical MSE in (52)

Because the affine projection algorithm operates on an adaptive FIR filter, it was assumed that the signal is generated by an FIR filter. This then allowed us to assume that the noise vector $\mathbf{v}(k)$ in (28) is independent of the data matrix $\mathbf{A}(k)$ in the analysis (see footnote 3). While this assumption is valid for an FIR filter when $\mathbf{v}(k)$ is white noise, it is only valid for an AR model (and also ARMA model) when $N=1$. To illustrate this, take the simple AR(1) model

$$x(k)=0.9x(k-1)+v(k) \tag{56}$$

where $v(k)$ is white Gaussian noise. When the filter length $M=3$ and the number of constraints $N=1$, then the data matrix $\mathbf{A}(k)=[x(k-1),x(k-2),x(k-3)]^T$ and $\mathbf{v}(k)=[v(k)]$. In this case we see that it is accurate to assume that $v(k)$ is independent from $\mathbf{A}(k)$. However, For $N=2$ we have

$$\mathbf{A}(k)=\begin{bmatrix} x(k-1) & x(k-2) \\ x(k-2) & x(k-3) \\ x(k-3) & x(k-4) \end{bmatrix}$$

and $\mathbf{v}(k)=[v(k),v(k-1)]$. In this case $\mathbf{v}(k)$ and $\mathbf{A}(k)$ are correlated because of the presence of the term $v(k-1)$ in $\mathbf{v}(k)$ and $x(k-1)$ in $\mathbf{A}(k)$. As the number of constraints N increases there are more and more terms in $\mathbf{A}(k)$ that are correlated with $\mathbf{v}(k)$ and so the assumption of independence between the noise vector $\mathbf{v}(k)$ and data matrix $\mathbf{A}(k)$ is no longer valid.

References

[1] K. Shoemake, Animating rotation with quaternion curves, SIGGRAPH Computer Graphics 19 (1985) 245–254.
[2] C.F. Karney, Quaternions in molecular modeling, Journal of Molecular Graphics and Modelling 25 (5) (2007) 595–604.
[3] S.-C. Pei, C.-M. Cheng, Color image processing by using binary quaternion-moment-preserving thresholding technique, IEEE Transactions on Image Processing 8 (5) (1999) 614–628.
[4] N. Le Bihan, S. Sangwine, Quaternion principal component analysis of color images, in: Proceedings of the International Conference on Image Processing, vol. 1, no. 9, 2003, pp. 1, 809–812.
[5] R. Campa, K. Camarillo, L. Arias, Kinematic modeling and control of robot manipulators via unit quaternions: application to a spherical wrist, in: Proceedings of the 45th IEEE Conference on Decision and Control, no. 12, 2006, pp. 6474–6479.
[6] C. Cheong Took, D. Mandic, and K. Aihara, Quaternion-valued short term forecasting of wind profile, in: Proceedings of the International Joint Conference on Neural Networks (IJCNN), no. 7, 2010, pp. 1–6.
[7] J. Via, D. Palomar, L. Vielva, I. Santamaria, Quaternion ICA from second-order statistics, IEEE Transactions on Signal Processing 59 (4) (2011) 1586–1600.
[8] S. Javidi, C. Took, D. Mandic, Fast independent component analysis algorithm for quaternion valued signals, IEEE Transactions on Neural Networks 22 (12) (2011) 1967–1978.
[9] C. Cheong Took, D.P. Mandic, The quaternion LMS algorithm for adaptive filtering of hypercomplex processes, IEEE Transactions on Signal Processing 57 (3) (2009) 1316–1327.
[10] C. Jahanchahi, C. Cheong Took, D. Mandic, The widely linear quaternion recursive least squares filter, in: Proceedings of the 2nd International Workshop on Cognitive Information Processing (CIP), no. 6, 2010, pp. 87–92.
[11] K. Ozeki, T. Umeda, An adaptive filtering algorithm using an orthogonal projection to an affine subspace and its properties, Electronics and Communications in Japan 67-A (5) (1984) 19–27.
[12] Y. Xia, C. Cheong Took, D.P. Mandic, An augmented affine projection algorithm for the filtering of noncircular complex signals, Signal Processing 90 (6) (2010) 1788–1799.
[13] B.D.H. Brandwood, A complex gradient operator and its applications in adaptive array theory, IEE Proceedings: Communications, Radar and Signal Proceedings 130 (1) (1983) 11–16.
[14] K. Kreutz-Delgado, The Complex Gradient Operator and the $\mathbb{CR}$-Calculus, Technical Report, University of California, San Diego, 2006.
[15] C. Jahanchahi, C. Cheong-Took, D.P. Mandic, $\mathbb{HR}$ calculus and quaternion analyticity, Technical Report, Imperial College, London, 2010.
[16] C. Cheong Took, D.P. Mandic, Augmented second-order statistics of quaternion random signals, Signal Processing 91 (2) (2011) 214–224.
[17] J. Via, D. Ramirez, I. Santamaria, Properness and widely linear processing of quaternion random vectors, IEEE Transactions on Information Theory 56 (7) (2010) 3502–3515.
[18] C. Cheong Took, D.P. Mandic, A quaternion widely linear adaptive filter, IEEE Transactions on Signal Processing 58 (8) (2010) 4427–4431.
[19] C. Jahanchahi, D. Mandic, D. Dini, Widely linear complex and quaternion valued bearings only tracking, IET Signal Processing 6 (5) (2012) 435–445.
[20] H.-C. Shin, A. Sayed, Mean-square performance of a family of affine projection algorithms, IEEE Transactions on Signal Processing 52 (1) (2004) 90–102.
[21] T.A. Ell, S.J. Sangwine, Quaternion involutions and anti-involutions, Computers and Mathematics with Applications 53 (1) (2007) 137–143.

[22] J. Marins, X. Yun, E. Bachmann, R. McGhee, M. Zyda, An extended Kalman filter for quaternion-based orientation estimation using MARG sensors, in: Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems, vol. 4, 2001, pp. 2003–2011.
[23] D. Choukroun, I. Bar-Itzhack, Y. Oshman, Novel quaternion Kalman filter, IEEE Transactions on Aerospace and Electronic Systems 42 (1) (2006) 174–190.
[24] B.K.P. Horn, Closed-form solution of absolute orientation using unit quaternions, Journal of the Optical Society of America A 4 (4) (1987) 629–642.
[25] C. Ujang, C. Cheong Took, D.P. Mandic, Quaternion valued nonlinear adaptive filtering, IEEE Transactions on Neural Networks 22 (8) (2011) 1193–1206.
[26] D.P. Mandic, C. Jahanchahi, C. Cheong Took, A quaternion gradient operator and its applications, IEEE Signal Processing Letters 18 (1) (2011) 47–50.
[27] W. Krlikowski, On 4-dimensional locally conformally flat almost Kahler manifolds, Archivum Mathematicum 42 (3) (2006) 215–223.
[28] D.P. Mandic, V.S.L. Goh, Complex Valued Nonlinear Adaptive Filters: Noncircularity, Widely Linear and Neural Models, John Wiley and Sons Ltd, 2009.
[29] A. Oya, J. Navarro-Moreno, J. Ruiz-Molina, Widely linear simulation of continuous-time complex-valued random signals, IEEE Signal Processing Letters 18 (9) (2011) 513–516.
[30] B. Picinbono, P. Bondon, Second-order statistics of complex signals, IEEE Transactions on Signal Processing 45 (2) (1997) 411–420.
[31] S. Javidi, C. Cheong Took, C. Jahanchahi, N. Le Bihan, D.P. Mandic, Blind extraction of improper quaternion sources, in: Proceedings of the International Conference on Acoustics, Speech and Signal Processing (ICASSP), no. 5, 2011, pp. 3708–3711.
[32] S. Haykin, Adaptive Filter Theory, 3rd ed. Prentice Hall, 1996.
[33] N. Yousef, A. Sayed, A unified approach to the steady-state and tracking analyses of adaptive filters, IEEE Transactions on Signal Processing 49 (2) (2001) 314–324.