A Class of Quaternion Kalman Filters

Cyrus Jahanchahi and Danilo P. Mandic, *Fellow, IEEE*

***Abstract*— The existing Kalman filters for quaternion-valued signals do not operate fully in the quaternion domain, and are combined with the real Kalman filter to enable the tracking in 3-D spaces. Using the recently introduced HR-calculus, we develop the fully quaternion-valued Kalman filter (QKF) and quaternion-extended Kalman filter (QEKF), allowing for the tracking of 3-D and 4-D signals directly in the quaternion domain. To consider the second-order noncircularity of signals, we employ the recently developed augmented quaternion statistics to derive the widely linear QKF (WL-QKF) and widely linear QEKF (WL-QEKF). To reduce computational requirements of the widely linear algorithms, their efficient implementation are proposed and it is shown that the quaternion widely linear model can be simplified when processing 3-D data, further reducing the computational requirements. Simulations using both synthetic and real-world circular and noncircular signals illustrate the advantages offered by widely linear over strictly linear quaternion Kalman filters.**

***Index Terms*— Extended Kalman filter, improperness, quaternion augmented statistics, quaternion Kalman filters, quaternion noncircularity, state space prediction, trajectory tracking, widely linear model, wind modeling.**

I. Introduction

THE Kalman filter is a standard tool in sequential state space estimation, and has found application in a number of areas, including space navigation [1], training of neural networks [2]–[6], and sensor fusion [7]. The extended Kalman filter enhances the scope of Kalman filtering to nonlinear state and observation models. It has proven invaluable in, for instance, nonlinear attitude estimation, and has also recently been combined with the quaternion domain representation to offer an effective way for solving 3-D attitude estimation problems [8], [9].

Quaternions have been used for a long time in mathematics, however, it is only recently that we have witnessed their resurgence in engineering and physics, where they have become prominent in computer graphics [10] and orientation modeling [11]–[15]. This is largely due to their ability to offer a compact representation for orientation in a 3-D space that is convenient, computationally efficient, and accurate. The quaternion domain modeling has also found applications in the processing of 3-D and 4-D rotational data, where the power of their division algebra helps mitigate the problems associated with real vector spaces, such as the gimbal lock. This also makes it possible to obtain the corresponding learning algorithms in a more elegant form than those in $\mathbb{R}^3$ or $\mathbb{R}^4$, together with rigorous physical interpretation, compact representations, and closed form solutions. Other applications where quaternions have attracted significant interest include color image processing [16], source separation [17], [18], and spectrum estimation algorithms for processing polarized waves [19].

The benefits of quaternions are conveniently illustrated in spacecraft orientation tracking where the aim is to find the mapping between the coordinate system on a reference-frame $\mathbf{r} \in \mathbb{R}^3$ and a local frame $\mathbf{b} \in \mathbb{R}^3$ on the spacecraft body-frame, such that $\mathbf{b} = \mathbf{A}\mathbf{r}$ where $\mathbf{r}$ is a reference-frame vector, $\mathbf{b}$ is the body-frame vector, and $\mathbf{A}$ is the attitude matrix (or rotation matrix). Quaternions provide a convenient closed form solution (Davenport's q-method [20]) to this mapping, which is generally referred to as Wahba's problem [21]. This is largely due to a more compact notation for rotations compared with that obtained with a 3×3 matrix, resulting in fewer constraints and a more mathematically tractable functional expression. Numerous authors have used this quaternion representation to track orientation in 3-D spaces [9], [22], [23], and for the training of quaternion-valued neural networks for time series prediction [24].

Existing quaternion Kalman filters are, however, not intrinsically quaternion valued as they use quaternions in the form of a quadrivariate real ordered vector, that is, $[q_a, q_b, q_c, q_d] \in \mathbb{R}$ rather than using their natural representation $q = q_a + \imath q_b + \jmath q_c + \kappa q_d \in \mathbb{H}$. It would therefore be analytically more appropriate and practically more physically meaningful for the algorithms to be derived directly in the quaternion domain, rather than having to switch between quaternions and the reals. This would also allow for a more intuitive approach to the concept of circularity and widely linear modeling, key notions for the optimal processing of the generality of quaternion-valued signals. Such a fully quaternion Kalman filter would not only be suited for orientation tracking but also for the processing of general 3-D and 4-D processes, and as an enabling tool in multidimensional learning systems.

One major obstacle in deriving a full quaternion Kalman filter is the inability of standard second-order statistics to capture full information in quaternion correlation matrices. It has recently been shown that to fully utilize the second-order statistical properties of quaternion-valued signals, three pseudocovariance matrices must be used in addition to the standard covariance matrix $E[\mathbf{x}\mathbf{x}^H]$ [25], [26]; this is achieved by employing the widely linear model. Conventional quaternion signal processing algorithms are strictly linear, and are therefore only suitable for the processing of second-order

Manuscript received August 2, 2012; revised February 25, 2013; accepted July 3, 2013. Date of publication October 2, 2013; date of current version February 14, 2014.

The authors are with Communications and Signal Processing Research Group, Imperial College London, London SW7 2AZ, U.K. (e-mail: cyrus.jahanchahi@imperial.ac.uk; d.mandic@imperial.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/TNNLS.2013.2277540

circular (proper) signals.[1] For orientation problems, a strictly linear model is usually sufficient because this transformation is described by a standard quaternion number. However, most real-world signals are noncircular and should therefore be processed using a widely linear model to capture complete second-order statistics available. Furthermore, as shown in this paper, the widely linear model is a prerequisite for the development of extended Kalman filter.

In this paper, we develop a class of linear and widely linear Kalman filters, that operate directly in the quaternion domain, allowing for a unified treatment of both circular and noncircular data. For the widely linear Kalman filter an efficient implementation is also proposed, reducing the computations by approximately a quarter. For nonlinear state space models, the quaternion-extended Kalman filter (QEKF) is introduced and is shown to be inherently widely linear. This allows us to bridge the gap in the literature on 3-D and 4-D Kalman filters and provide an algorithmic support for learning systems that operate on vector sensor data.

The organization of this paper is as follows: in Sections II and III, we briefly review the quaternion algebra and quaternion gradient ($\mathbb{HR}$-calculus) necessary for the development of the quaternion Kalman filter. In Section IV, we introduce the widely linear model and provide a more efficient implementation for 3-D signals. In Section V, the quaternion Kalman filter is derived. Section VI introduces the widely linear quaternion Kalman filter while in Section VII we provide its efficient implementation of the filter. To cater for nonlinear state models, Section VIII presents the extended quaternion Kalman filter. Section IX demonstrates the isomorphism that exists between the quaternion Kalman filter and the real Kalman filter in $\mathbb{R}^4$. In Section X, simulations on both benchmark and real-world signals illustrate the tracking capability of the algorithms derived.

II. Quaternion Algebra

Quaternions are a 4-D associative, noncommutative, normed division algebra over the real numbers, defined as

$$\{q_r, q_\imath, q_\jmath, q_\kappa\} \in \mathbb{R}^4 \rightarrow q_r + \imath q_\imath + \jmath q_\jmath + \kappa q_\kappa \in \mathbb{H} \quad (1)$$

where $\{\imath, \jmath, \kappa\}$ are the imaginary units obeying the following rules

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

Consider a quaternion $q = Sq + Vq$, where $Sq = q_r$ denotes the scalar part of q and $Vq = \imath q_\imath + \jmath q_\jmath + \kappa q_\kappa$ the vector part, then the product of quaternions $q_1, q_2 \in \mathbb{H}$ can be written as [27]

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

where the symbol "$\bullet$" denotes the dot-product and "$\times$" the cross-product. The 3-D vector part Vq is also called a pure quaternion, while the inclusion of the real part Sq gives a full quaternion. The algebraic structure of quaternions therefore enables a unified processing of both 3-D and 4-D vector processes. The quaternion conjugate, denoted by q^*, is given by

$$q^* = Sq - Vq = q_r - \imath q_\imath - \jmath q_\jmath - \kappa q_\kappa \quad (2)$$

and the norm $\| q \|$ is defined as

$$\| q \| = \sqrt{qq^*} = \sqrt{q_r^2 + q_\imath^2 + q_\jmath^2 + q_\kappa^2}.$$

A. Quaternion Involutions

Of great importance for this paper are quaternion involutions, which are similarity relations, or self-inverse mappings, defined as[2] [28]

$$\begin{aligned} q^\imath &= -\imath q \imath = q_r + \imath q_\imath - \jmath q_\jmath - \kappa q_\kappa \\ q^\jmath &= -\jmath q \jmath = q_r - \imath q_\imath + \jmath q_\jmath - \kappa q_\kappa \\ q^\kappa &= -\kappa q \kappa = q_r - \imath q_\imath - \jmath q_\jmath + \kappa q_\kappa \end{aligned} \quad (3)$$

and have the following properties (for $\text{inv}_3 \neq \text{inv}_2 \neq \text{inv}_1$):

$$P1: \ (q^{\text{inv}})^{\text{inv}} = q \ \text{ for inv} \in \{\imath, \jmath, \kappa\} \quad (4)$$
$$P2: \ (q_1 q_2)^{\text{inv}} = q_1^{\text{inv}} q_2^{\text{inv}} \quad (5)$$
$$P3: \ (q_1 + q_2)^{\text{inv}} = q_1^{\text{inv}} + q_2^{\text{inv}} \quad (6)$$
$$P4: \ (q^{\text{inv}_1})^{\text{inv}_2} = (q^{\text{inv}_2})^{\text{inv}_1} = q^{\text{inv}_3}. \quad (7)$$

Involutions can be seen as a counterpart of the complex conjugate, as they allow 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[3]

$$\begin{aligned} q_r &= \frac{1}{4}[q + q^\imath + q^\jmath + q^\kappa] & q_\imath &= \frac{1}{4\imath}[q + q^\imath - q^\jmath - q^\kappa] \\ q_\jmath &= \frac{1}{4\jmath}[q - q^\imath + q^\jmath - q^\kappa] & q_\kappa &= \frac{1}{4\kappa}[q - q^\imath - q^\jmath + q^\kappa]. \end{aligned} \quad (8)$$

B. Advantages of Quaternion Algebra Over 3-D Vector Algebra

Compare the real mapping that rotates and scales a point $\mathbf{x} \in \mathbb{R}^3$ into a point $\mathbf{y} \in \mathbb{R}^3$, to the quaternion mapping that rotates and scales a point $q_x \in \mathbb{H}$ into a point $q_y \in \mathbb{H}$

$$q_y = q_T q_x q_T^* \quad \in \mathbb{H} \quad \sim \quad \mathbf{y} = \mathbf{T}\mathbf{x} \quad \in \mathbb{R}^3. \quad (9)$$

Remark 1: The mapping $\mathbf{T} \in \mathbb{R}^{3\times 3}$ requires nine coefficients to relate two vectors in $\mathbb{R}^3$, however, physically only four parameters are needed (two for the axis of rotation, one for the scaling factor and one for the angle of rotation). The four elements of a quaternion offer this physical insight and compact representation, expressing straightforwardly the axis of rotation, scaling factor, and angle of rotation.

[1] Second-order circular signals have rotation invariant probability distributions.

[2] Note that the quaternion conjugate is also an involution.

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

Remark 2: When the mapping $\mathbf{T}$ describes a concatenation of multiple rotations in the x, y, z directions (using Euler angles), a degree of freedom (DoF) can be lost if any two axes are aligned, resulting in the gimbal lock phenomenon. No such phenomenon exists in the quaternion domain, where the quaternion transformation in (9) is expressed as $q_y = q_A q_x q_A^{-1}$, where q_A is a unit quaternion.

Remark 3: The quaternion rotation q_T is better conditioned than the real rotation matrix $\mathbf{T}$, as the only requirement imposed on q_T is for it to be a unit quaternion, whereas $\mathbf{T}$ must satisfy $\mathbf{T}^T\mathbf{T} = \mathbf{I}$ and $\det(\mathbf{T}) = 1$. This has led to the widespread use of quaternions in, e.g., spacecraft orientation problems where they provide convenient closed form solutions [8], [9], [29].

III. HR-Calculus

The rigorous calculation of quaternion gradient has so far been a major stumbling block in the derivation of quaternion-valued learning algorithms. This is mainly because the Cauchy–Riemann–Fueter equation (differentiability condition) for a quaternion function J with respect to a vector parameter $\mathbf{w}$, given by

$$\frac{\partial J}{\partial \mathbf{w}^*} = \frac{1}{4}\left(\frac{\partial J}{\partial \mathbf{w}_r} + \imath\frac{\partial J}{\partial \mathbf{w}_\imath} + \jmath\frac{\partial J}{\partial \mathbf{w}_\jmath} + \kappa\frac{\partial J}{\partial \mathbf{w}_\kappa}\right) = \mathbf{0} \tag{10}$$

imposes very strict constraints on quaternion functions, permitting only differentiability of linear functions. This is prohibitive to the development of quaternion adaptive filters, however, for real-valued functions of quaternion variables, we can use the duality that exists between the real vectors and quaternions to circumvent these stringent conditions, as exemplified in the recently developed $\mathbb{HR}$-calculus [30], an extension of the $\mathbb{CR}$-calculus [31] in the complex domain. The $\mathbb{HR}$-calculus comprises eight partial derivatives, split into two groups: the $\mathbb{HR}$-derivatives

$$\begin{bmatrix} \frac{\partial f(q,q^\imath,q^\jmath,q^\kappa)}{\partial q} \\ \frac{\partial f(q,q^\imath,q^\jmath,q^\kappa)}{\partial q^\imath} \\ \frac{\partial f(q,q^\imath,q^\jmath,q^\kappa)}{\partial q^\jmath} \\ \frac{\partial f(q,q^\imath,q^\jmath,q^\kappa)}{\partial q^\kappa} \end{bmatrix} = \frac{1}{4}\begin{bmatrix} 1 & -\imath & -\jmath & -\kappa \\ 1 & -\imath & \jmath & \kappa \\ 1 & \imath & -\jmath & \kappa \\ 1 & \imath & \jmath & -\kappa \end{bmatrix}\begin{bmatrix} \frac{\partial f(q_r,q_\imath,q_\jmath,q_\kappa)}{\partial q_r} \\ \frac{\partial f(q_r,q_\imath,q_\jmath,q_\kappa)}{\partial q_\imath} \\ \frac{\partial f(q_r,q_\imath,q_\jmath,q_\kappa)}{\partial q_\jmath} \\ \frac{\partial f(q_r,q_\imath,q_\jmath,q_\kappa)}{\partial q_\kappa} \end{bmatrix} \tag{11}$$

and the $\mathbb{HR}^*$-derivatives

$$\begin{bmatrix} \frac{\partial f(q^*,q^{\imath*},q^{\jmath*},q^{\kappa*})}{\partial q^*} \\ \frac{\partial f(q^*,q^{\imath*},q^{\jmath*},q^{\kappa*})}{\partial q^{\imath*}} \\ \frac{\partial f(q^*,q^{\imath*},q^{\jmath*},q^{\kappa*})}{\partial q^{\jmath*}} \\ \frac{\partial f(q^*,q^{\imath*},q^{\jmath*},q^{\kappa*})}{\partial q^{\kappa*}} \end{bmatrix} = \frac{1}{4}\begin{bmatrix} 1 & \imath & \jmath & \kappa \\ 1 & \imath & -\jmath & -\kappa \\ 1 & -\imath & \jmath & -\kappa \\ 1 & -\imath & -\jmath & \kappa \end{bmatrix}\begin{bmatrix} \frac{\partial f(q_r,q_\imath,q_\jmath,q_\kappa)}{\partial q_r} \\ \frac{\partial f(q_r,q_\imath,q_\jmath,q_\kappa)}{\partial q_\imath} \\ \frac{\partial f(q_r,q_\imath,q_\jmath,q_\kappa)}{\partial q_\jmath} \\ \frac{\partial f(q_r,q_\imath,q_\jmath,q_\kappa)}{\partial q_\kappa} \end{bmatrix} \tag{12}$$

where $q^{\text{inv}*} = (q^{\text{inv}})^*$ for $\text{inv} \in \{\imath, \jmath, \kappa\}$ are the conjugates of the involutions in (3).

Remark 4: The $\mathbb{HR}^*$-derivative $\partial f(q^*, q^{\imath*}, q^{\jmath*}, q^{\kappa*})/\partial q^*$ is equivalent to the quaternion derivative operator introduced by Fueter [32], however, unlike the CRF derivative in (10), the derivative $\partial f(q^*, q^{\imath*}, q^{\jmath*}, q^{\kappa*})/\partial q^*$ also introduces a restriction on the dependent variables that compose the function $f(\cdot)$, namely that $f(\cdot)$ must be expressed as a function of the involutions q^*, $q^{\imath*}$, $q^{\jmath*}$, or $q^{\kappa*}$.

Remark 5: The $\mathbb{HR}$- and $\mathbb{HR}^*$-derivatives can be used in a similar way to the $\mathbb{R}$- and $\mathbb{R}^*$-derivatives in the complex domain [33]. 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^κ, using the substitution

$$q^* = \frac{1}{2}(q^\imath + q^\jmath + q^\kappa - q). \tag{13}$$

Similarly, to differentiate a function of q using the $\mathbb{HR}^*$-derivatives, we can substitute for q using

$$q = \frac{1}{2}(q^{\imath*} + q^{\jmath*} + q^{\kappa*} - q^*). \tag{14}$$

This way, the $\mathbb{HR}$-calculus provides a 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).

Remark 6: The derivatives in (11) and (12) have the imaginary unit vectors placed on the left-hand side of the real partial derivatives $\partial f/\partial q_r$, $\partial f/\partial q_\imath$, $\partial f/\partial q_\jmath$, and $\partial f/\partial q_\kappa$ (termed the left-$\mathbb{HR}$- and left-$\mathbb{HR}^*$-derivatives). The unit vectors could equally well have been placed on the right-hand side of the real partial derivatives, giving rise to the right-$\mathbb{HR}$- and right-$\mathbb{HR}^*$-derivatives.

IV. Widely Linear Model

Another stumbling block that has been detrimental to a more widespread use of quaternions in learning systems has been that the standard strictly linear solutions are only optimal for a very restrictive class of so called circular data (or strictly linear systems), for which the powers in all the quaternion components are equal. The recent introduction of quaternion augmented statistics [26] has highlighted that for a general (improper) quaternion vector $\mathbf{x}$, second-order estimation based solely on the covariance matrix $\mathbf{R}_{qq} = E[\mathbf{x}\mathbf{x}^H]$ is inadequate, and to fully capture the second-order statistics the pseudocovariances $\mathbf{G} = E[\mathbf{x}\mathbf{x}^{\imath H}]$, $\mathbf{L} = E[\mathbf{x}\mathbf{x}^{\jmath H}]$, and $\mathbf{T} = E[\mathbf{x}\mathbf{x}^{\kappa H}]$ are also needed. To introduce an optimal second-order estimator for the generality of quaternion signals, consider first the mean square error (MSE) estimator of a real-valued scalar y from an observed real vector $\mathbf{x}$, that is $\hat{y} = E[y \mid \mathbf{x}]$. For jointly Gaussian $\mathbf{x}$ and y, the optimal solution is a strictly linear estimator, given by

$$\hat{y} = \mathbf{h}^T\mathbf{x} \tag{15}$$

where $\mathbf{h}$ and $\mathbf{x}$ are, respectively, the coefficient and regressor vector. For the standard complex domain MSE, the same form is assumed but this time with $\mathbf{h}$ and $\mathbf{x}$ complex valued. However, in terms of the real and imaginary parts of the complex variables, we have

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

and since $x_r = x + x^*/2$ and $x_i = x - x^*/2\imath$, the complex widely linear model is given by [34]

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

that is, it 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 also given by (15), with $\mathbf{h}$ and $\mathbf{x}$ quaternion valued. Observe that, for all the components

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

and using the involutions in (3), we have, e.g., $x_a = 1/4\,(x + x^\imath + x^\jmath + x^\kappa)$, leading to

$$\begin{aligned} \hat{y}_r &= E[y_r \,|\, \mathbf{x}, \mathbf{x}^\imath, \mathbf{x}^\jmath, \mathbf{x}^\kappa] \quad \hat{y}_\imath = E[y_\imath \,|\, \mathbf{x}, \mathbf{x}^\imath, \mathbf{x}^\jmath, \mathbf{x}^\kappa] \\ \hat{y}_\jmath &= E[y_\jmath \,|\, \mathbf{x}, \mathbf{x}^\imath, \mathbf{x}^\jmath, \mathbf{x}^\kappa] \quad \hat{y_\kappa} = E[y_\kappa \,|\, \mathbf{x}, \mathbf{x}^\imath, \mathbf{x}^\jmath, \mathbf{x}^\kappa] \end{aligned}$$

which can be merged as

$$\hat{y} = E[y \,|\, \mathbf{x}, \mathbf{x}^\imath, \mathbf{x}^\jmath, \mathbf{x}^\kappa].$$

In other words, to capture the full second-order information available, we should use the quaternion widely linear model

$$y = \mathbf{u}^T\mathbf{x} + \mathbf{v}^T\mathbf{x}^\imath + \mathbf{g}^T\mathbf{x}^\jmath + \mathbf{h}^T\mathbf{x}^\kappa = \mathbf{w}^T\mathbf{q} \tag{16}$$

where the augmented coefficient vector $\mathbf{w} = [\mathbf{u}^T, \mathbf{v}^T, \mathbf{g}^T, \mathbf{h}^T]^T$ and the augmented regressor vector $\mathbf{q} = [\mathbf{x}^T, \mathbf{x}^{\imath T}, \mathbf{x}^{\jmath T}, \mathbf{x}^{\kappa T}]^T$.

Current statistical signal processing solutions in $\mathbb{H}$ are strictly linear, with most of the algorithms drawing upon the covariance matrix $\mathbf{R} = E[\mathbf{q}\mathbf{q}^H]$. However, as shown in (16), to be able to model both the second-order circular (proper) and second-order noncircular (improper) signals, we need to employ the augmented covariance matrix, given by [26]

$$\mathbf{R}_{qq^*} = \begin{bmatrix} \mathbf{R} & \mathbf{G} & \mathbf{L} & \mathbf{T} \\ \mathbf{G}^\imath & \mathbf{R}^\imath & \mathbf{T}^\imath & \mathbf{L}^\imath \\ \mathbf{L}^\jmath & \mathbf{T}^\jmath & \mathbf{R}^\jmath & \mathbf{G}^\jmath \\ \mathbf{T}^\kappa & \mathbf{L}^\kappa & \mathbf{G}^\kappa & \mathbf{R}^\kappa \end{bmatrix} \tag{17}$$

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

Remark 7: For the second-order circular signals (proper), the pseudocovariance matrices $\mathbf{G}$, $\mathbf{L}$, and $\mathbf{T}$ vanish. A signal that obeys this structure has a probability distribution that is rotation invariant with respect to all the six pairs of axes (combinations of $\imath$, $\jmath$, and κ) [25], [26].

Remark 8: The processing in $\mathbb{R}^4$ would require 10 covariance and cross-covariance matrices, as opposed to four in the quaternion domain, and would not model the signal noncircularity straightforwardly.

A. Widely Linear Model for 3-D Signals

For many Kalman filtering applications, the aim is to track a 3-D quantity, as is the case in aeronautics. The widely linear model in (16) is the only rigorous model for general noncircular 3-D signals [35], and to reduce its computational complexity, we show that it is overparameterized and provide a simplified solution. To arrive at the more parsimonious form, we start from the linear model in (15) in its scalar form (i.e., $\hat{y} = hx \in \mathbb{H}$) and, using the duality between quaternions and real numbers, it can be rewrite in $\mathbb{R}^4$ as

$$\begin{bmatrix} \hat{y}_r \\ \hat{y}_\imath \\ \hat{y}_\jmath \\ \hat{y}_\kappa \end{bmatrix} = \mathbf{H}\mathbf{x} = \begin{bmatrix} u_r & u_\imath & u_\jmath & u_\kappa \\ -u_\imath & u_r & -u_\kappa & u_\jmath \\ -u_\jmath & u_\kappa & u_r & -u_\imath \\ -u_\kappa & -u_\jmath & u_\imath & u_r \end{bmatrix} \begin{bmatrix} x_r \\ x_\imath \\ x_\jmath \\ x_\kappa \end{bmatrix} \quad \in \mathbb{R}^4.$$

For the strictly linear model above, observe that the matrix $\mathbf{H}$ has only four out of the possible 16 DoF needed to describe the transformation from $\mathbf{x}$ to $\widehat{y}$. If the signal $\mathbf{x}$ is 3-D (i.e., a pure quaternion), then the first column of $\mathbf{H}$ vanishes and matrix $\mathbf{H}$ reduces to a 4×3 matrix. Hence, to fully describe the transformation between the 3-D vector $\mathbf{x}$ and the 4-D vector $\mathbf{y}$, 12 DoF are needed. We next show that for 3-D signals, only three of the four terms that constitute the widely linear model are needed to obtain the required 12 DoF. Indeed, consider the reduced widely linear model for the 3-D data, given by

$$\hat{y} = \mathbf{u}\mathbf{x} + \mathbf{v}\mathbf{x}^\imath + \mathbf{g}\mathbf{x}^\jmath \quad \in \mathbb{H} \tag{18}$$

and rewrite it in $\mathbb{R}^3$ as

$$\begin{aligned} \begin{bmatrix} \hat{y}_r \\ \hat{y}_\imath \\ \hat{y}_\jmath \\ \hat{y}_\kappa \end{bmatrix} &= \begin{bmatrix} u_r & u_\imath & u_\jmath & u_\kappa \\ -u_\imath & u_r & -u_\kappa & u_\jmath \\ -u_\jmath & u_\kappa & u_r & -u_\imath \\ -u_\kappa & -u_\jmath & u_\imath & u_r \end{bmatrix} \begin{bmatrix} x_r \\ x_\imath \\ x_\jmath \\ x_\kappa \end{bmatrix} \\ &+ \begin{bmatrix} v_r & v_\imath & v_\jmath & v_\kappa \\ -v_\imath & v_r & -v_\kappa & v_\jmath \\ -v_\jmath & v_\kappa & u_r & -v_\imath \\ -v_\kappa & -v_\jmath & u_\imath & v_r \end{bmatrix} \begin{bmatrix} x_r \\ x_\imath \\ -x_\jmath \\ -x_\kappa \end{bmatrix} \\ &+ \begin{bmatrix} g_r & g_\imath & g_\jmath & g_\kappa \\ -g_\imath & g_r & -g_\kappa & g_\jmath \\ -g_\jmath & g_\kappa & g_r & -g_\imath \\ -g_\kappa & -g_\jmath & g_\imath & g_r \end{bmatrix} \begin{bmatrix} x_r \\ -x_\imath \\ x_\jmath \\ -x_\kappa \end{bmatrix}. \end{aligned}$$

Upon absorbing the negative signs in the vector $\mathbf{x}$ into the coefficient matrices, we can rewrite the above as

$$\begin{aligned} \begin{bmatrix} \hat{y}_r \\ \hat{y}_\imath \\ \hat{y}_\jmath \\ \hat{y}_\kappa \end{bmatrix} = \mathbf{H}\mathbf{x} = \begin{bmatrix} u_r + v_r + g_r & u_\imath + v_\imath - g_\imath \\ -u_\imath - v_\imath - g_\imath & u_r + v_r - g_r \\ -u_\jmath - v_\jmath - g_\jmath & u_\kappa + v_\kappa - g_\kappa \\ -u_\kappa - v_\kappa - g_\kappa & -u_\jmath - v_\jmath + g_\jmath \end{bmatrix} \\ \begin{bmatrix} u_\jmath - v_\jmath + g_\jmath & u_\kappa - v_\kappa - g_\kappa \\ -u_\kappa + v_\kappa - g_\kappa & u_\jmath - v_\jmath - g_\jmath \\ u_r - v_r + g_r & -u_\imath + v_\imath + g_\imath \\ u_\imath - v_\imath - g_\imath & u_r - v_r - g_r \end{bmatrix} \begin{bmatrix} x_r \\ x_\imath \\ x_\jmath \\ x_\kappa \end{bmatrix}. \end{aligned}$$

For $\mathbf{x}$ a pure quaternion, the first column of $\mathbf{H}$ vanishes while the remaining 12 terms in $\mathbf{H}$ are linearly independent, and are described by

$$\begin{aligned} u_r + v_r - g_r &= \mathbf{H}_{22} & u_\imath + v_\imath - g_\imath &= \mathbf{H}_{12} \\ u_r - v_r + g_r &= \mathbf{H}_{33} & -u_\imath + v_\imath + g_\imath &= \mathbf{H}_{34} \\ u_r - v_r - g_r &= \mathbf{H}_{44} & u_\imath - v_\imath - g_\imath &= \mathbf{H}_{43} \end{aligned}$$

and

$$\begin{aligned} -u_\jmath - v_\jmath + g_\jmath &= \mathbf{H}_{42} & u_\kappa + v_\kappa - g_\kappa &= \mathbf{H}_{32} \\ u_\jmath - v_\jmath + g_\jmath &= \mathbf{H}_{13} & -u_\kappa + v_\kappa - g_\kappa &= \mathbf{H}_{23} \\ u_\jmath - v_\jmath - g_\jmath &= \mathbf{H}_{24} & u_\kappa - v_\kappa - g_\kappa &= \mathbf{H}_{14}. \end{aligned}$$

Observe that now the 12 equations above are linearly independent and therefore the reduced widely linear model is sufficient to model 3-D data (pure quaternions).

V. Quaternion Kalman Filter

Following on the derivation of the Kalman filter in the complex domain [36], we assume that the quaternion state $\mathbf{x}_k \in \mathbb{H}^{n\times 1}$ evolves according to the following model:

$$\mathbf{x}_k = \mathbf{A}_k\mathbf{x}_{k-1} + \mathbf{B}_k\mathbf{u}_k + \mathbf{w}_k \tag{19}$$

where $\mathbf{A}_k \in \mathbb{H}^{n\times n}$ is the state transition matrix, $\mathbf{B}_k \in \mathbb{H}^{n\times n}$ is the control input matrix for the control input $\mathbf{u}_k \in \mathbb{H}^{n\times 1}$, and $\mathbf{w}_k \in \mathbb{H}^{n\times 1}$ is the state noise.

The state $\mathbf{x}_k$ cannot be observed directly but we can measure the quantity $\mathbf{z}_k \in \mathbb{H}^{m\times 1}$ that relates to the state $\mathbf{x}_k$ through

$$\mathbf{z}_k = \mathbf{H}\mathbf{x}_k + \mathbf{v}_k \tag{20}$$

where $\mathbf{H} \in \mathbb{H}^{m\times n}$ is the observation matrix and $\mathbf{v}_k \in \mathbb{H}^{m\times 1}$ is the measurement noise. Both the model noise and measurement noise are zero mean and Gaussian, with properties

$$\mathbf{w}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_k) \tag{21}$$

$$\mathbf{v}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_k). \tag{22}$$

Since in the real-world noise normally comprises of a number of independent sources, by the central limit theorem this is a reasonable assumption.

The a priori state estimate $\hat{\mathbf{x}}_{k|k-1} \in \mathbb{H}^{n\times 1}$ (the estimate of the state $\mathbf{x}_k$ before obtaining the new measurement) can be obtained from the state model

$$\hat{\mathbf{x}}_{k|k-1} = \mathbf{A}_k\hat{\mathbf{x}}_{k-1|k-1} + \mathbf{B}_k\mathbf{u}_k \tag{23}$$

where $\hat{\mathbf{x}}_{k-1|k-1}$ is the previous state estimate.

Using the measurement $\mathbf{z}_k$, the estimate $\hat{\mathbf{x}}_{k|k-1}$ can be improved to obtain the a posteriori state estimate $\hat{\mathbf{x}}_{k|k} \in \mathbb{H}^{n\times 1}$, given by

$$\hat{\mathbf{x}}_{k|k} = \hat{\mathbf{x}}_{k|k-1} + \mathbf{K}_k\big(\mathbf{z}_k - \mathbf{H}\hat{\mathbf{x}}_{k|k-1}\big) \tag{24}$$

where $\mathbf{z}_k - \mathbf{H}\hat{\mathbf{x}}_{k|k-1}$ is the innovation term (the error between the estimate and the measurement). The aim is to find the Kalman gain $\mathbf{K}_k$ that minimizes MSE $E[\mathbf{e}_{k|k}^H\mathbf{e}_{k|k}]$, where the error is given by

$$\mathbf{e}_{k|k} = \mathbf{x}_k - \hat{\mathbf{x}}_{k|k} = \mathbf{x}_k - \big[\hat{\mathbf{x}}_{k|k-1} + \mathbf{K}_k(\mathbf{z}_k - \mathbf{H}\hat{\mathbf{x}}_{k|k-1})\big].$$

Minimizing the error $E[\mathbf{e}_{k|k}^H\mathbf{e}_{k|k}]$ is equivalent to minimizing the trace of the error covariance $\mathbf{P}_k = \mathrm{cov}(\mathbf{e}_{k|k}) = E[\mathbf{e}_{k|k}\mathbf{e}_{k|k}^H]$. This allows us to write MSE as

$$E[\mathbf{e}_{k|k}^H\mathbf{e}_{k|k}] = Tr[\mathrm{cov}(\mathbf{x}_k - (\hat{\mathbf{x}}_{k|k-1} + \mathbf{K}_k(\mathbf{z}_k - \mathbf{H}\hat{\mathbf{x}}_{k|k-1})))].$$

Substituting for $\mathbf{z}_k$, we have

$$\begin{aligned} E[\mathbf{e}_{k|k}^H\mathbf{e}_{k|k}] &= Tr\big[\mathrm{cov}\big(\mathbf{x}_k - \big(\hat{\mathbf{x}}_{k|k-1} \\ &\quad + \mathbf{K}_k(\mathbf{H}\mathbf{x}_k + \mathbf{v}_k - \mathbf{H}\hat{\mathbf{x}}_{k|k-1})\big)\big)\big] \\ &= Tr\big[\mathrm{cov}\big[(\mathbf{I} - \mathbf{K}_k\mathbf{H})(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1} - \mathbf{K}_k\mathbf{v}_k)\big]\big] \end{aligned}$$

and using the property $\mathrm{cov}(\mathbf{A}\mathbf{x}) = E[(\mathbf{A}\mathbf{x})(\mathbf{A}\mathbf{x})^H] = \mathbf{A}E[\mathbf{x}\mathbf{x}^H]\mathbf{A}^H$ (the Hermitian is commutative in the quaternion domain), we arrive at

$$\begin{aligned} E[\mathbf{e}_{k|k}^H\mathbf{e}_{k|k}] = Tr\big[&(\mathbf{I} - \mathbf{K}_k\mathbf{H})\mathrm{cov}(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1})(\mathbf{I} - \mathbf{K}_k\mathbf{H})^H \\ &+ \mathbf{K}_k\mathrm{cov}(\mathbf{v}_k)\mathbf{K}_k^H\big]. \end{aligned}$$

Denoting the *a priori* error covariance matrix $\mathrm{cov}(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1})$ by $\mathbf{P}_{k|k-1}$, we can now write

$$\begin{aligned} E[\mathbf{e}_{k|k}^H\mathbf{e}_{k|k}] &= Tr\big[(\mathbf{P}_{k|k-1} - \mathbf{K}_k\mathbf{H}\mathbf{P}_{k|k-1})(\mathbf{I} - \mathbf{K}_k\mathbf{H})^H \\ &\quad + \mathbf{K}_k\mathbf{R}_k\mathbf{K}_k^H\big] \\ &= Tr\big[\mathbf{P}_{k|k-1} - \mathbf{K}_k\mathbf{H}\mathbf{P}_{k|k-1} - \mathbf{P}_{k|k-1}\mathbf{H}^H\mathbf{K}_k^H \\ &\quad + \mathbf{K}_k\mathbf{S}_k\mathbf{K}_k^H\big] \end{aligned} \tag{25}$$

where $\mathbf{S}_k$ is independent of $\mathbf{K}_k$ and is given by $\mathbf{S}_k = \mathbf{H}\mathbf{P}_{k|k-1}\mathbf{H}^H + \mathbf{R}_k$. Upon calculating all the derivatives (found using $\mathbb{HR}$-calculus) with respect to $\mathbf{K}_k$ and setting to zero, we obtain

$$\begin{aligned} \frac{\partial E[\mathbf{e}_{k|k}^H\mathbf{e}_{k|k}]}{\partial \mathbf{K}_k} &= \frac{\partial Tr[\mathbf{K}_\mathbf{k}\mathbf{H}\mathbf{P}_{k|k-1}]}{\partial \mathbf{K}_k} - \frac{\partial Tr[\mathbf{P}_{k|k-1}\mathbf{H}^H\mathbf{K}_k^H]}{\partial \mathbf{K}_k} \\ &\quad + \frac{\partial \mathbf{K}_k\mathbf{S}_k\mathbf{K}_k^H}{\partial \mathbf{K}_k} = \mathbf{0} \end{aligned} \tag{26}$$

where the partial derivatives can be written as (observe that these derivatives are different to what would be obtained in the complex domain)

$$(\mathbf{S}_k\mathbf{K}_k^H)^T - \frac{1}{2}(\mathbf{K}_k\mathbf{S}_k) = (\mathbf{H}\mathbf{P}_{k|k-1})^T - \frac{1}{2}\mathbf{P}_{k|k-1}\mathbf{H}^H. \tag{27}$$

The following decomposition is important for the derivation:

$$\begin{aligned} ((\mathbf{S}_k\mathbf{K}_k^H)^T)^* &= (\mathbf{S}_k\mathbf{K}_k^H)^H = \mathbf{K}_k\mathbf{S}_k^H = \mathbf{K}_k\mathbf{S}_k \\ ((\mathbf{H}\mathbf{P}_{k|k-1})^T)^* &= (\mathbf{H}\mathbf{P}_{k|k-1})^H = \mathbf{P}_{k|k-1}^H\mathbf{H}^H = \mathbf{P}_{k|k-1}\mathbf{H}^H \end{aligned}$$

allowing us to obtain a solution to (27) in the form $(\mathbf{S}_k\mathbf{K}_k^H)^T = (\mathbf{H}\mathbf{P}_{k|k-1})^T$ (since this also implies $-1/2\mathbf{K}_k\mathbf{S}_k = -1/2\mathbf{P}_{k|k-1}\mathbf{H}^H$), finally giving

$$\mathbf{K}_k = \mathbf{P}_{k|k-1}\mathbf{H}^H\mathbf{S}_k^{-1}. \tag{28}$$

Therefore, the a priori error covariance matrix $\mathbf{P}_{k|k-1}$ can be solved recursively as

$$\begin{aligned} \mathbf{P}_{k|k-1} &= \mathrm{cov}\big[\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1}\big] \\ &= \mathbf{A}E\big[\mathbf{e}_{k|k-1}\mathbf{e}_{k|k-1}^H\big]\mathbf{A}^H + 2\mathbf{A}E\big[\mathbf{e}_{k|k-1}\mathbf{w}_k\big] + E\big[\mathbf{w}_k\mathbf{w}_k^H\big]. \end{aligned} \tag{29}$$

The assumption made that the noise is white means that $\mathbf{e}_{k|k-1}$ is uncorrelated with $\mathbf{w}(k)$ and so we can simplify the above expression as

$$\mathbf{P}_{k|k-1} = \mathbf{A}\mathbf{P}_{k-1|k-1}\mathbf{A}^H + \mathbf{Q}_k. \tag{30}$$

To obtain a recursive expression for $\mathbf{P}_{k|k}$ (used in obtaining $\mathbf{P}_{k|k-1}$), observe that from (25) the matrix $\mathbf{P}_k$ can be written as

$$\mathbf{P}_k = \mathbf{P}_{k|k-1} - \mathbf{K}_k\mathbf{H}\mathbf{P}_{k|k-1} - \mathbf{P}_{k|k-1}\mathbf{H}^H\mathbf{K}_k^H + \mathbf{K}_k\mathbf{S}_k\mathbf{K}_k^H. \tag{31}$$

Substituting for $\mathbf{K}_k$ in the above, the last two terms on the right cancel out, giving

$$\mathbf{P}_{k|k} = \mathbf{P}_{k|k-1} - \mathbf{K}_k\mathbf{H}\mathbf{P}_{k|k-1}. \tag{32}$$

This completes the derivation of the strictly linear quaternion Kalman filter, which, unlike the existing approaches, is derived directly in the quaternion domain, and is given in Algorithm 1.

Algorithm 1 Quaternion-Valued Kalman Filter

Model output

$$\hat{\mathbf{x}}_{k|k-1} = \mathbf{A}_k \hat{\mathbf{x}}_{k-1|k-1} + \mathbf{B}_k \mathbf{u}_k \tag{33}$$
$$\mathbf{P}_{k|k-1} = \mathbf{A}_k \mathbf{P}_{k-1|k-1} \mathbf{A}_k^H + \mathbf{Q}_k \tag{34}$$

Measurement output

$$\mathbf{S}_k = \mathbf{H} \mathbf{P}_{k|k-1} \mathbf{H}^H + \mathbf{R}_k \tag{35}$$
$$\mathbf{K}_k = \mathbf{P}_{k|k-1} \mathbf{H}^H \mathbf{S}_k^{-1} \tag{36}$$
$$\hat{\mathbf{x}}_{k|k} = \hat{\mathbf{x}}_{k|k-1} + \mathbf{K}_k [\mathbf{z}_k - \mathbf{H} \hat{\mathbf{x}}_{k|k-1}] \tag{37}$$
$$\mathbf{P}_{k|k} = [\mathbf{I} - \mathbf{K}_k \mathbf{H}] \mathbf{P}_{k|k-1} \tag{38}$$

VI. Widely Linear Quaternion Kalman Filter

The Kalman filter derived in the previous section is optimal for strictly linear models or for circular data. To account for the noncircularity associated with real-world data, the Kalman filter must propagate the augmented state estimates and augmented covariance matrix. Following on (16) and (19), to incorporate the widely linear model we can rewrite the state and observation equations as

$$\begin{aligned} \mathbf{x}_k^a &= \mathbf{A}_k^a \mathbf{x}_{k-1}^a + \mathbf{B}_k^a \mathbf{u}_k^a + \mathbf{w}_k^a \\ \mathbf{z}_k^a &= \mathbf{H}^a \mathbf{x}_{k-1}^a + \mathbf{v}_k^a \end{aligned} \tag{39}$$

where the augmented state vector $\mathbf{x}_k^a = [\mathbf{x}_k^T, \mathbf{x}_k^{\imath T}, \mathbf{x}_k^{\jmath T}, \mathbf{x}_k^{\kappa T}]^T$, the augmented state noise vector $\mathbf{w}_k^a = [\mathbf{w}_k^T, \mathbf{w}_k^{\imath T}, \mathbf{w}_k^{\jmath T}, \mathbf{w}_k^{\kappa T}]^T$, the augmented observation noise vector $\mathbf{v}_k^a = [\mathbf{v}_k^T, \mathbf{v}_k^{\imath T}, \mathbf{v}_k^{\jmath T}, \mathbf{v}_k^{\kappa T}]^T$, while the augmented state transition matrix and observation matrix are, respectively, given by

$$\mathbf{A}_k^a = \begin{bmatrix} \mathbf{A}_1 & \mathbf{A}_2 & \mathbf{A}_3 & \mathbf{A}_4 \\ \mathbf{A}_2^\imath & \mathbf{A}_1^\imath & \mathbf{A}_4^\imath & \mathbf{A}_3^\imath \\ \mathbf{A}_3 & \mathbf{A}_4^\jmath & \mathbf{A}_1^\jmath & \mathbf{A}_2^\jmath \\ \mathbf{A}_4^\kappa & \mathbf{A}_3^\kappa & \mathbf{A}_2^\kappa & \mathbf{A}_4^\kappa \end{bmatrix}$$
$$\mathbf{H}^a = \begin{bmatrix} \mathbf{H}_1 & \mathbf{H}_2 & \mathbf{H}_3 & \mathbf{H}_4 \\ \mathbf{H}_2^\imath & \mathbf{H}_1^\imath & \mathbf{H}_4^\imath & \mathbf{H}_3^\imath \\ \mathbf{H}_3^\jmath & \mathbf{H}_4^\jmath & \mathbf{H}_1^\jmath & \mathbf{H}_2^\jmath \\ \mathbf{H}_4^\kappa & \mathbf{H}_3^\kappa & \mathbf{H}_2^\kappa & \mathbf{H}_1^\kappa \end{bmatrix}. \tag{40}$$

Note that the augmented block matrix structures in $\mathbf{H}^a$ and $\mathbf{A}_k^a$ (the subscript k is dropped in the partitioned matrix for ease of representation) are necessary for the model in (39) to hold. Using the widely linear model described above, the widely linear quaternion Kalman filter algorithm becomes a generic extension of the quaternion Kalman filter described in (34) and (38), and provides a unified treatment of both the second-order circular and noncircular data, with the subscript a emphasizing the fact that all vectors and matrices are augmented. The widely linear quaternion-valued Kalman filter (WL-QKF) is summarized in Algorithm 2.

Observe from (39) that the widely linear quaternion Kalman filter is not only able to model the noncircularity of the state model and observation model via the widely linear $\mathbf{A}_k^a$ and $\mathbf{H}^a$, but also the noncircularity of the state and observation noise via the augmented covariance matrices $\mathbf{Q}_k^a$ and $\mathbf{R}_k^a$.

Algorithm 2 Widely Linear Quaternion-Valued Kalman Filter (WL-QKF)

Model output

$$\hat{\mathbf{x}}_{k|k-1}^a = \mathbf{A}_k^a \hat{\mathbf{x}}_{k-1|k-1}^a + \mathbf{B}_k^a \mathbf{u}_k^a \tag{41}$$
$$\mathbf{P}_{k|k-1}^a = \mathbf{A}_k^a \mathbf{P}_{k-1|k-1}^a \mathbf{A}_k^{aH} + \mathbf{Q}_k^a \tag{42}$$

Measurement output

$$\mathbf{S}_k^a = \mathbf{H}^a \mathbf{P}_{k|k-1}^a \mathbf{H}^{aH} + \mathbf{R}_k^a \tag{43}$$
$$\mathbf{K}_k^a = \mathbf{P}_{k|k-1}^a \mathbf{H}^{aH} \mathbf{S}_k^{a^{-1}} \tag{44}$$
$$\hat{\mathbf{x}}_{k|k}^a = \hat{\mathbf{x}}_{k|k-1}^a + \mathbf{K}_k^a [\mathbf{z}_k^a - \mathbf{H}^a \hat{\mathbf{x}}_{k|k-1}^a] \tag{45}$$
$$\mathbf{P}_k^a = [\mathbf{I} - \mathbf{K}_k^a \mathbf{H}^a] \mathbf{P}_{k|k-1}^a \tag{46}$$

Remark 9: Even for strictly linear state or observation model, the widely linear Kalman filter can still offer better performance than its strictly linear counterpart if the noise is noncircular. This is in contrast to the widely linear quaternion recursive least squares [37], which can only cater for the noncircularity of the model and not explicitly for the noise noncircularity.

VII. Efficient Implementation of the Quaternion Widely Linear Kalman Filter

Observe that matrices $\mathbf{A}_k^a$ and $\mathbf{H}_k^a$ in (40) are overparameterized since only one quarter of the entries are needed to fully describe each matrix. We shall next make use of this property to simplify each of the six equations describing the quaternion widely linear Kalman filter.

1) Equation (41):

$$\hat{\mathbf{x}}_{k|k-1}^a = \mathbf{A}_k^a \hat{\mathbf{x}}_{k-1|k-1}^a + \mathbf{B}_k^a \mathbf{u}_k^a. \tag{47}$$

Upon partitioning each matrix and array into its components we have

$$\begin{bmatrix} \hat{\mathbf{x}}_{k|k-1} \\ \hat{\mathbf{x}}_{k|k-1}^\imath \\ \hat{\mathbf{x}}_{k|k-1}^\jmath \\ \hat{\mathbf{x}}_{k|k-1}^\kappa \end{bmatrix} = \begin{bmatrix} \mathbf{A}_{1k} & \mathbf{A}_{2k} & \mathbf{A}_{3k} & \mathbf{A}_{4k} \\ \mathbf{A}_{2k}^\imath & \mathbf{A}_{1k}^\imath & \mathbf{A}_{4k}^\imath & \mathbf{A}_{3k}^\imath \\ \mathbf{A}_{3k}^\jmath & \mathbf{A}_{4k}^\jmath & \mathbf{A}_{1k}^\jmath & \mathbf{A}_{2k}^\jmath \\ \mathbf{A}_{4k}^\kappa & \mathbf{A}_{3k}^\kappa & \mathbf{A}_{2k}^\kappa & \mathbf{A}_{4k}^\kappa \end{bmatrix} \begin{bmatrix} \hat{\mathbf{x}}_{k-1|k-1} \\ \hat{\mathbf{x}}_{k-1|k-1}^\imath \\ \hat{\mathbf{x}}_{k-1|k-1}^\jmath \\ \hat{\mathbf{x}}_{k-1|k-1}^\kappa \end{bmatrix}. \tag{48}$$

Observe that there is only one linearly independent row above, and so (41) can be simplified into

$$\begin{aligned} \hat{\mathbf{x}}_{k|k-1} &= \mathbf{A}_{1k} \hat{\mathbf{x}}_{k-1|k-1} + \mathbf{A}_{2k} \hat{\mathbf{x}}_{k-1|k-1} + \mathbf{A}_{3k} \hat{\mathbf{x}}_{k-1|k-1} \\ &\quad + \mathbf{A}_{4k} \hat{\mathbf{x}}_{k-1|k-1} \\ &= \mathbf{A}_k^{\alpha T} \hat{\mathbf{x}}_{k-1|k-1}^a \end{aligned} \tag{49}$$

where $\mathbf{A}_k^{\alpha T} = [\mathbf{A}_{1k}, \mathbf{A}_{2k}, \mathbf{A}_{3k}, \mathbf{A}_{4k}]$.

Remark 10: The advantage of using this representation is twofold. The output has only one quarter of the terms and so less memory is needed in the implementation (this is also the case for matrix $\mathbf{A}_k^\alpha$ as compared with $\mathbf{A}_k^a$). Secondly, this form is four times less computationally intensive.

2) Equation (42):

$$\mathbf{P}^a_{k|k-1} = \mathbf{A}^a_k \mathbf{P}^a_{k-1|k-1} \mathbf{A}^{aH}_k + \mathbf{Q}^a_k. \tag{50}$$

To simplify the above, we start from the definition of $\mathbf{P}^a_{k|k-1}$, that is

$$\mathbf{P}^a_{k|k-1} = E[\mathbf{e}^a_{k|k-1}\mathbf{e}^{aH}_{k|k-1}] = E\left[\begin{bmatrix}\mathbf{e}_{k|k-1}\\ \mathbf{e}^\imath_{k|k-1}\\ \mathbf{e}^\jmath_{k|k-1}\\ \mathbf{e}^\kappa_{k|k-1}\end{bmatrix}\begin{bmatrix}\mathbf{e}_{k|k-1}\\ \mathbf{e}^\imath_{k|k-1}\\ \mathbf{e}^\jmath_{k|k-1}\\ \mathbf{e}^\kappa_{k|k-1}\end{bmatrix}^T\right]$$
$$= \begin{bmatrix}\mathbf{P}_{1k|k-1} & \mathbf{P}_{2k|k-1} & \mathbf{P}_{3k|k-1} & \mathbf{P}_{4k|k-1}\\ \mathbf{P}^\imath_{2k|k-1} & \mathbf{P}^\imath_{1k|k-1} & \mathbf{P}^\imath_{4k|k-1} & \mathbf{P}^\imath_{3k|k-1}\\ \mathbf{P}^\jmath_{3k|k-1} & \mathbf{P}^\jmath_{4k|k-1} & \mathbf{P}^\jmath_{1k|k-1} & \mathbf{P}^\jmath_{2k|k-1}\\ \mathbf{P}^\kappa_{4k|k-1} & \mathbf{P}^\kappa_{3k|k-1} & \mathbf{P}^\kappa_{2k|k-1} & \mathbf{P}^\kappa_{1k|k-1}\end{bmatrix} \tag{51}$$

where $\mathbf{P}_{1k|k-1} = E[\mathbf{e}_{k|k-1}\left(\mathbf{e}_{k|k-1}\right)^H]$, $\mathbf{P}_{2k|k-1} = E[\mathbf{e}_{k|k-1}\left(\mathbf{e}_{k|k-1}\right)^{\imath H}]$, $\mathbf{P}_{3k|k-1} = E[\mathbf{e}_{k|k-1}\left(\mathbf{e}_{k|k-1}\right)^{\jmath H}]$, and $\mathbf{P}_{4k|k-1} = E[\mathbf{e}_{k|k-1}\left(\mathbf{e}_{k|k-1}\right)^{\kappa H}]$.

Note that the product of two matrices with the augmented block matrix structure as above also has an augmented block matrix structure. Thus, only the first row of the product need to be computed to express the whole matrix. To formalize this product, we introduce the operator $f_{wl}(\cdot)$ that reconstructs a block matrix from its first row. The terms comprising $\mathbf{P}^a_{k|k-1}$ can now be calculated as

$$\mathbf{P}^\alpha_{k|k-1} = \mathbf{A}^\alpha f_{wl}\left(\mathbf{P}^\alpha_{k-1|k-1}\mathbf{A}^{aH}_k\right) + \mathbf{Q}^\alpha_k \tag{52}$$

where $\mathbf{P}^\alpha_{k|k-1} = [\mathbf{P}_{1k|k-1}, \mathbf{P}_{2k|k-1}, \mathbf{P}_{3k|k-1}, \mathbf{P}_{4k|k-1}]$ and $\mathbf{Q}^\alpha_k = [\mathbf{Q}_{1k}, \mathbf{Q}_{2k}, \mathbf{Q}_{3k}, \mathbf{Q}_{4k}]$.

Remark 11: This form requires a quarter of the mathematical operations and only one quarter of the terms need be stored.

3) Equation (43):

$$\mathbf{S}^a_k = \mathbf{H}^a \mathbf{P}^a_{k|k-1}\mathbf{H}^{aH} + \mathbf{R}^a_k. \tag{53}$$

Following the same approach as that for (42), since $\mathbf{S}^a$ assumes the form

$$\begin{bmatrix}\mathbf{S}_{1k} & \mathbf{S}_{2k} & \mathbf{S}_{3k} & \mathbf{S}_{4k}\\ \mathbf{S}^\imath_{2k} & \mathbf{S}^\imath_{1k} & \mathbf{S}^\imath_{4k} & \mathbf{S}^\imath_{3k}\\ \mathbf{S}^\jmath_{3k} & \mathbf{S}^\jmath_{4k} & \mathbf{S}^\jmath_{1k} & \mathbf{S}^\jmath_{2k}\\ \mathbf{S}^\kappa_{4k} & \mathbf{S}^\kappa_{3k} & \mathbf{S}^\kappa_{2k} & \mathbf{S}^\kappa_{4k}\end{bmatrix} \tag{54}$$

this allows us to write

$$\mathbf{S}^\alpha_k = \mathbf{H}^\alpha f_{wl}\left(\mathbf{P}^\alpha_{k|k-1}\mathbf{H}^{aH}\right) + \mathbf{R}^\alpha_k \tag{55}$$

where $\mathbf{S}^\alpha_k = [\mathbf{S}_{1k}, \mathbf{S}_{2k}, \mathbf{S}_{3k}\mathbf{S}_{4k}]$, $\mathbf{H}^\alpha = [\mathbf{H}_1, \mathbf{H}_2, \mathbf{H}_3, \mathbf{H}_4]$, and $\mathbf{R}^\alpha_k = [\mathbf{R}_{1k}, \mathbf{R}_{2k}, \mathbf{R}_{3k}, \mathbf{R}_{4k}]$.

Remark 12: This form requires a quarter of the mathematical operations and only one quarter of the memory requirement.

4) Equation (44):

$$\mathbf{K}^a_k = \mathbf{P}^a_{k|k-1}\mathbf{H}^{aH}\mathbf{S}^{a^{-1}}_k. \tag{56}$$

Knowing that $\mathbf{K}^a$ has an augmented block matrix structure we can follow the same approach as above and to arrive at a m more compact form given by

$$\mathbf{K}^\alpha_k = \mathbf{P}^\alpha_{k|k-1} f_{wl}\left(\mathbf{H}^{aH}\mathbf{S}^{a^{-1}}_k\right)$$

where $\mathbf{K}^\alpha_k = [\mathbf{K}_{1k}, \mathbf{K}_{2k}, \mathbf{K}_{3k}, \mathbf{K}_{4k}]$. The matrix inverse operation is computationally very intensive, however, we can take advantage of the structure of matrix $\mathbf{S}^a$ to simplify this expression. The inverse of $\mathbf{S}$ has the same structure as $\mathbf{S}$, and can therefore be written as

$$\mathbf{S}^{a^{-1}}_k = \mathbf{C}^a = \begin{bmatrix}\mathbf{C}_1 & \mathbf{C}_2 & \mathbf{C}_3 & \mathbf{C}_4\\ \mathbf{C}^\imath_2 & \mathbf{C}^\imath_1 & \mathbf{C}^\imath_4 & \mathbf{C}^\imath_3\\ \mathbf{C}^\jmath_3 & \mathbf{C}^\jmath_4 & \mathbf{C}^\jmath_1 & \mathbf{C}^\jmath_2\\ \mathbf{C}^\kappa_4 & \mathbf{C}^\kappa_3 & \mathbf{C}^\kappa_2 & \mathbf{C}^\kappa_4\end{bmatrix}. \tag{57}$$

By applying the block inversion lemma twice onto the matrix $\mathbf{S}$, the matrices $\mathbf{C}_1$, $\mathbf{C}_2$, $\mathbf{C}_3$, and $\mathbf{C}_4$ can be evaluated as

$$\mathbf{C}_1 = (\mathbf{U}_1 - \mathbf{V}\mathbf{U}_3)^{-1} \tag{58}$$
$$\mathbf{C}_2 = -\mathbf{C}_1 V \tag{59}$$
$$\mathbf{C}_3 = -(\mathbf{C}_1\mathbf{O}_1 + \mathbf{C}_2\mathbf{O}_3) \tag{60}$$
$$\mathbf{C}_4 = -(\mathbf{C}_1\mathbf{O}_2 + \mathbf{C}_1\mathbf{O}_4) \tag{61}$$

where

$$\mathbf{V} = \mathbf{U}_2\mathbf{U}_4^{-1} \tag{62}$$
$$\begin{bmatrix}\mathbf{U}_1 & \mathbf{U}_2\\ \mathbf{U}_3 & \mathbf{U}_4\end{bmatrix} = \mathbf{S}_a - \mathbf{O}\mathbf{S}_c \tag{63}$$
$$\begin{bmatrix}\mathbf{O}_1 & \mathbf{O}_2\\ \mathbf{O}_3 & \mathbf{O}_4\end{bmatrix} = \mathbf{O} = \mathbf{S}_b\mathbf{S}_d^{-1} \tag{64}$$
$$\begin{bmatrix}\mathbf{S}_a & \mathbf{S}_b\\ \mathbf{S}_c & \mathbf{S}_d\end{bmatrix} = \mathbf{S}^a_k. \tag{65}$$

Remark 13: The above approach to obtain the matrix inverse requires half the operations needed for inverting the whole of matrix $\mathbf{S}^a$.

5) Equation (45):

$$\hat{\mathbf{x}}^a_{k|} = \hat{\mathbf{x}}^a_{k|k-1} + \mathbf{K}^a_k[\mathbf{z}^a_k - \mathbf{H}^a\hat{\mathbf{x}}^a_{k|k-1}]. \tag{66}$$

This can be simplified to

$$\hat{\mathbf{x}}_{k|k} = \hat{\mathbf{x}}_{k|k-1} + \mathbf{K}^\alpha_k(\mathbf{z}^a - \mathbf{H}^a\mathbf{x}^a_k) \tag{67}$$

where $\mathbf{K}^\alpha_k = [\mathbf{K}_{1k}, \mathbf{K}_{2k}, \mathbf{K}_{3k}, \mathbf{K}_{4k}]$.

6) Equation 46:

$$\mathbf{P}^a_{k|k} = [\mathbf{I}^a - \mathbf{K}^a_k\mathbf{H}^a]\mathbf{P}^a_{k|k-1}. \tag{68}$$

We can simplify this to

$$\mathbf{P}^\alpha_{k|k} = [[\mathbf{I}, \mathbf{0}, \mathbf{0}, \mathbf{0}] - \mathbf{K}^\alpha_k\mathbf{H}^a]\mathbf{P}^a_{k|k-1}. \tag{69}$$

Algorithm 3 summarizes the efficient implementation of the quaternion widely linear Kalman filter.

VIII. Widely Linear Quaternion Extended Kalman Filter

For nonlinear systems, the state transition and the observation equations are nonlinear and are modeled using the extended Kalman filter, given by

$$\mathbf{x}_k = f[\mathbf{x}_{k-1}] + \mathbf{w}_k \tag{84}$$
$$\mathbf{z}_k = h[\mathbf{x}_k] + \mathbf{v}_k \tag{85}$$

where $f[\cdot]$ and $h[\cdot]$ are, respectively, the nonlinear state and observation functions. For their approximations, the extended Kalman filter uses the first-order Taylor series expansions

Algorithm 3 Efficient Implementation of the Quaternion Widely Linear Kalman Filter

Model output

$$\hat{\mathbf{x}}_{k|k-1} = \mathbf{A}_k^{\alpha T}\hat{\mathbf{x}}_{k-1|k-1}^{a} \tag{70}$$

$$\mathbf{P}_{k|k-1}^{\alpha} = \mathbf{A}_k^{\alpha} f_{wl}\left(\mathbf{P}_{k-1|k-1}^{\alpha}\mathbf{A}^{aH}\right) + \mathbf{Q}_k^{\alpha} \tag{71}$$

Measurement output

$$\mathbf{S}_k^{\alpha} = \mathbf{H}^{\alpha} f_{wl}\left(\mathbf{P}_{k|k-1}^{\alpha}\mathbf{H}^{aH}\right) + \mathbf{R}_k^{\alpha} \tag{72}$$

$$\mathbf{K}^{\alpha} = \mathbf{P}_{k|k-1}^{\alpha} f_{wl}\left(\mathbf{H}^{\alpha H}\mathbf{C}^{a}\right) \tag{73}$$

$$\hat{\mathbf{x}}_{k|k} = \hat{\mathbf{x}}_{k|k-1} + \mathbf{K}_k^{\alpha}(\mathbf{z}_k^{a} - \mathbf{H}^{a}\mathbf{x}_{k|k-1}^{a}) \tag{74}$$

$$\mathbf{P}_{k|k}^{\alpha} = [[\mathbf{I}, \mathbf{0}, \mathbf{0}, \mathbf{0}] - \mathbf{K}_k^{\alpha}\mathbf{H}^{a}]\mathbf{P}_{k|k-1}^{a} \tag{75}$$

where

$$\mathbf{C}_1 = (\mathbf{U}_1 - \mathbf{U}_2\mathbf{U}_4^{-1}\mathbf{U}_3)^{-1} \tag{76}$$

$$\mathbf{C}_2 = -\mathbf{C}_1\mathbf{U}_2\mathbf{U}_4^{-1} \tag{77}$$

$$\mathbf{C}_3 = -(\mathbf{C}_1\mathbf{O}_1 + \mathbf{C}_2\mathbf{O}_3) \tag{78}$$

$$\mathbf{C}_4 = -(\mathbf{C}_1\mathbf{O}_2 + \mathbf{C}_1\mathbf{O}_4) \tag{79}$$

$$\mathbf{U} = \begin{bmatrix} \mathbf{U}_1 & \mathbf{U}_2 \\ \mathbf{U}_3 & \mathbf{U}_4 \end{bmatrix} = \mathbf{S}_a - \mathbf{O}\mathbf{S}_c \tag{80}$$

$$\mathbf{O} = \begin{bmatrix} \mathbf{O}_1 & \mathbf{O}_2 \\ \mathbf{O}_3 & \mathbf{O}_4 \end{bmatrix} = \mathbf{S}_b\mathbf{S}_d^{-1} \tag{81}$$

while the superscript α in $\mathbf{M}^{\alpha}$ denots the augmented array of matrix $\mathbf{M}$.

$$\mathbf{M}^{\alpha} = [\mathbf{M}_1, \mathbf{M}_2, \mathbf{M}_3, \mathbf{M}_4]. \tag{82}$$

The superscript a in $\mathbf{M}^{a}$ defines the augmented matrix of matrix $\mathbf{M}$, given by

$$\mathbf{M}^{a} = \begin{bmatrix} \mathbf{M}_1 & \mathbf{M}_2 & \mathbf{M}_3 & \mathbf{M}_4 \\ \mathbf{M}_2^{\iota} & \mathbf{M}_1^{\iota} & \mathbf{M}_4^{\iota} & \mathbf{M}_3^{\iota} \\ \mathbf{M}_3^{\jmath} & \mathbf{M}_4^{\jmath} & \mathbf{M}_1^{\jmath} & \mathbf{M}_2^{\jmath} \\ \mathbf{M}_4^{\kappa} & \mathbf{M}_3^{\kappa} & \mathbf{M}_2^{\kappa} & \mathbf{M}_4^{\kappa} \end{bmatrix} = \begin{bmatrix} \mathbf{M}_a & \mathbf{M}_b \\ \mathbf{M}_c & \mathbf{M}_d \end{bmatrix} \tag{83}$$

where $f_{wl}(\mathbf{M}^{\alpha}) = \mathbf{M}^{a}$ for any matrix $\mathbf{M}$.

(TSEs) around the current state estimate. Calculating the quaternion derivative requires the function to be analytic in the CRF sense, however, this also imposes very strict conditions as only linear functions are analytic in the quaternion domain. Although nonlinear functions are not differentiable in the Cauchy–Riemann–Fueter sense, by exploiting the isomorphism between the quaternion domain $\mathbb{H}$ and real domain $\mathbb{R}^4$, the $\mathbb{HR}$-calculus allows for nonanalytic functions to be differentiated [30]. This also guarantees the existence of the first-order quaternion TSE of a nonanalytic quaternion function, which assumes the form

$$f(q+dq) = f(q) + \frac{\partial f}{\partial q}dq + \frac{\partial f}{\partial q^{\iota}}dq^{\iota} + \frac{\partial f}{\partial q^{\jmath}}dq^{\jmath} + \frac{\partial f}{\partial q^{\kappa}}dq^{\kappa} \tag{86}$$

where for analytic functions the involution derivatives vanish.

Remark 14: Because quaternion nonlinear functions are nonanalytic (a consequence of CRF condition), the first-order Taylor series in (86) will always be widely linear, due to the presence of the terms dq^{ι}, $dq^{\jmath}$, and dq^{κ}. Therefore the QEKF is always widely linear. This is in contrast to the complex domain where the extended Kalman filter can be strictly linear if the nonlinear function is analytic [38].

Upon expanding $f[\cdot]$ around $\hat{\mathbf{x}}_{k-1|k-1}$ and $h[\cdot]$ around $\hat{\mathbf{x}}_{k|k-1}$ using the first-order Taylor series in (86), the state and measurement equations become

$$\mathbf{x}_k = \mathbf{A}_k\mathbf{x}_{k-1} + \mathbf{B}_k\mathbf{x}_{k-1}^{\iota} + \mathbf{C}_k\mathbf{x}_{k-1}^{\jmath} + \mathbf{D}_k\mathbf{x}_{k-1}^{\kappa} + \mathbf{w}_k + \mathbf{d}_k$$

$$\mathbf{y}_k = \mathbf{E}\mathbf{x}_{k-1} + \mathbf{F}\mathbf{x}_{k-1}^{\iota} + \mathbf{G}\mathbf{x}_{k-1}^{\jmath} + \mathbf{H}\mathbf{x}_{k-1}^{\kappa} + \mathbf{w}_k + \mathbf{l}_k$$

where

$$\begin{aligned}\mathbf{d}_k &= f(\hat{\mathbf{x}}_{k-1|k-1}) - \mathbf{A}_k\hat{\mathbf{x}}_{k-1|k-1} - \mathbf{B}_k\hat{\mathbf{x}}_{k-1|k-1}^{\iota} \\ &\quad -\mathbf{C}_k\hat{\mathbf{x}}_{k-1|k-1}^{\jmath} - \mathbf{D}_k\hat{\mathbf{x}}_{k-1|k-1}^{\kappa} \\ \mathbf{l}_k &= h(\hat{\mathbf{x}}_{k|k-1}) - \mathbf{E}\hat{\mathbf{x}}_{k|k-1} - \mathbf{F}\hat{\mathbf{x}}_{k|k-1}^{\iota} - \mathbf{G}\hat{\mathbf{x}}_{k|k-1}^{\jmath} \\ &\quad -\mathbf{H}\hat{\mathbf{x}}_{k|k-1}^{\kappa}\end{aligned}$$

$$\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D} = \frac{\partial f}{\partial x_{k-1}^{\varphi}}\Big|_{\mathbf{x}_{k-1}^{\varphi} = \hat{\mathbf{x}}_{k-1|k-1}^{\varphi}} \text{ for } \varphi \in \{1, \iota, \jmath, \kappa\}$$

$$\mathbf{E}, \mathbf{F}, \mathbf{G}, \mathbf{H} = \frac{\partial h}{\partial x_{k}^{\varphi}}\Big|_{\mathbf{x}_{k}^{\varphi} = \hat{\mathbf{x}}_{k|k-1}^{\varphi}} \text{ for } \varphi \in \{1, \iota, \jmath, \kappa\}$$

Observing that both the linearized process and observation equations are widely linear, the state and observation model of the extended Kalman filter can now be written as

$$\mathbf{x}_k = \mathbf{F}_k^{a}\mathbf{x}_{k-1}^{a}\mathbf{w}_k^{a} + \mathbf{d}_k^{a} \tag{87}$$

$$\mathbf{y}_k = \mathbf{H}^{a}\mathbf{x}_{k-1}^{a} + \mathbf{w}_k^{a} + \mathbf{l}_k^{a} \tag{88}$$

where $\mathbf{d}_k^{a} = [\mathbf{d}_k^{T}, \mathbf{d}_k^{\iota T}, \mathbf{d}_k^{\jmath T}, \mathbf{d}_k^{\kappa T}]^{T}$, $\mathbf{l}_k^{a} = [\mathbf{l}_k^{T}, \mathbf{l}_k^{\iota T}, \mathbf{l}_k^{\jmath T}, \mathbf{l}_k^{\kappa T}]^{T}$

$$\mathbf{F}_k^{a} = \begin{bmatrix} \mathbf{A}_k & \mathbf{B}_k & \mathbf{C}_k & \mathbf{D}_k \\ \mathbf{B}_k^{\iota} & \mathbf{A}_k^{\iota} & \mathbf{D}_k^{\iota} & \mathbf{C}_k^{\iota} \\ \mathbf{C}_k^{\jmath} & \mathbf{D}_k^{\jmath} & \mathbf{A}_k^{\jmath} & \mathbf{B}_k^{\jmath} \\ \mathbf{D}_k^{\kappa} & \mathbf{C}_k^{\kappa} & \mathbf{B}_k^{\kappa} & \mathbf{A}_k^{\kappa} \end{bmatrix} \tag{89}$$

$$\mathbf{H}^{a} = \begin{bmatrix} \mathbf{E} & \mathbf{F} & \mathbf{G} & \mathbf{H} \\ \mathbf{F}^{\iota} & \mathbf{E}^{\iota} & \mathbf{H}^{\iota} & \mathbf{G}^{\iota} \\ \mathbf{G}^{\jmath} & \mathbf{H}^{\jmath} & \mathbf{E}^{\jmath} & \mathbf{F}^{\jmath} \\ \mathbf{H}^{\kappa} & \mathbf{G}^{\kappa} & \mathbf{F}^{\kappa} & \mathbf{E}^{\kappa} \end{bmatrix}. \tag{90}$$

The WL-QEKF is summarized in Algorithm 4.

IX. Duality Between the Quaternion- and Real-Valued Kalman Filter

Following on the definition of quaternions in (1), a one-to-one mapping exists between the points in $\mathbb{H}$ and $\mathbb{R}^4$. For the quaternion vector $\mathbf{q} = \mathbf{q}_a + \iota\mathbf{q}_b + \jmath\mathbf{q}_c + \kappa\mathbf{q}_d \in \mathbb{H}^n$, this duality is described by

$$\underbrace{\begin{bmatrix} \mathbf{q} \\ \mathbf{q}^{\iota} \\ \mathbf{q}^{\jmath} \\ \mathbf{q}^{\kappa} \end{bmatrix}}_{\mathbf{q}^{a} \in \mathbb{H}^{4n}} = \underbrace{\begin{bmatrix} \mathbf{I} & \iota\mathbf{I} & \jmath\mathbf{I} & \kappa\mathbf{I} \\ \mathbf{I} & \iota\mathbf{I} & -\jmath\mathbf{I} & -\kappa\mathbf{I} \\ \mathbf{I} & -\iota\mathbf{I} & \jmath\mathbf{I} & -\kappa\mathbf{I} \\ \mathbf{I} & -\iota\mathbf{I} & -\jmath\mathbf{I} & \kappa\mathbf{I} \end{bmatrix}}_{\mathbf{J}_n \in \mathbb{H}^{n\times n}} \underbrace{\begin{bmatrix} \mathbf{q}_r \\ \mathbf{q}_i \\ \mathbf{q}_j \\ \mathbf{q}_k \end{bmatrix}}_{\mathbf{q}^{r} \in \mathbb{R}^{4n}} \tag{97}$$

where $\mathbf{I}$ is the identity matrix and $\mathbf{J}$ describes the invertible mapping $\mathbf{J} : \mathbb{R}^4 \to \mathbb{H}$ (where $\mathbf{J}^{-1} = \frac{1}{4}\mathbf{J}^{H}$). Using this

Algorithm 4 Widely Linear Quaternion-Extended Kalman Filter

Model output

$$\hat{\mathbf{x}}^a_{k|k-1} = f^a_k[\hat{\mathbf{x}}^a_{k-1|k-1}] + \mathbf{B}^a_k \mathbf{u}^a_k \tag{91}$$

$$\mathbf{P}^a_{k|k-1} = \mathbf{F}^a_k \mathbf{P}^a_{k-1} \mathbf{F}^{aH}_k + \mathbf{Q}^a_k \tag{92}$$

Measurement output

$$\mathbf{S}^a = \mathbf{H}^a \mathbf{P}^a_{k|k-1} \mathbf{H}^{aH} + \mathbf{R}^a_k \tag{93}$$

$$\mathbf{K}^a = \mathbf{P}^a_{k|k-1} \mathbf{H}^{aH} \mathbf{S}^{a^{-1}} \tag{94}$$

$$\hat{\mathbf{x}}^a_{k|k} = \hat{\mathbf{x}}^a_{k|k-1} + \mathbf{K}^a[\mathbf{z}^a_k - h^a[\hat{\mathbf{x}}^a_{k|k-1}]] \tag{95}$$

$$\mathbf{P}^a_{k|k} = [\mathbf{I} - \mathbf{K}^a \mathbf{H}^a] \mathbf{P}^a_{k|k-1} \tag{96}$$

isomorphism, we can obtain a dual real quadrivariate version of the augmented quaternion state and measurement model in (39), in the form

$$\begin{aligned} \mathbf{x}^r_k &= \mathbf{A}^r_k \mathbf{x}^r_{k-1} + \mathbf{B}^r_k \mathbf{u}^r_k + \mathbf{w}^r_k \\ \mathbf{z}^r_k &= \mathbf{H}^r \mathbf{x}^r_{k-1} + \mathbf{v}^r_k \end{aligned} \tag{98}$$

where $\mathbf{x}^r_k = \mathbf{J}_n^{-1}\mathbf{x}^a_k$, $\mathbf{z}^r_k = \mathbf{J}_m^{-1}\mathbf{z}^a_k$, $\mathbf{F}^r_k = \mathbf{J}_n^{-1}\mathbf{F}^a_k\mathbf{J}_n$, $\mathbf{H}^r = \mathbf{J}_m^{-1}\mathbf{H}^a\mathbf{J}_n$, $\mathbf{w}^r_k = \mathbf{J}_n^{-1}\mathbf{w}^a_k$, and $\mathbf{v}^r_k = \mathbf{J}_n^{-1}\mathbf{v}^a_k$ while the error covariance matrix are related by $\mathbf{P}^r_k = \mathbf{J}_n^{-1}\mathbf{P}^a_k\mathbf{J}_n^{-H}$. We can now obtain an expression relating the MSEs of the quaternion and real-valued Kalman filter in the form

$$\begin{aligned} Tr(\mathbf{P}^r_k) &= Tr(\mathbf{J}_n^{-1}\mathbf{P}^a_k\mathbf{J}_n^{-H}) \\ &= Tr(\mathbf{P}^a_k\mathbf{J}_n^{-H}\mathbf{J}_n^{-1}) \\ &= \frac{1}{4}Tr(\mathbf{P}^a_k). \end{aligned} \tag{99}$$

Because of the redundant information in the augmented pseudocovariance matrix $\mathbf{P}^a_k$ [see (51)], the MSE for the widely linear Kalman filter is $(1/4)Tr(\mathbf{P}^a_k)$ and not $Tr(\mathbf{P}^a_k)$, thus giving the same MSE as the real quadrivariate Kalman filter.

Remark 15: It follows from (99) that the performance of the widely linear quaternion Kalman filter is identical to that of the real Kalman filter. The duality between $\mathbb{H}$ and $\mathbb{R}^4$ only exists for widely linear models, reflecting the fact that strictly linear models are suboptimal, and are thus suitable only for a very restrictive class of circular data.

Remark 16: When the underlying signal generating model is strictly linear, both the real Kalman filter and WL-QKF are over-parameterized and the quaternion Kalman filter provides better convergence performance.

X. Simulations

To illustrate the advantages offered by widely linear quaternion Kalman filters over their strictly linear counterparts, we considered the following scenarios: 1) filtering of a quaternion-valued signal from noisy measurements where the state model is known and the first- and second-order statistics of both state and measurement noise are also known; 2) multiple step ahead prediction of a nonstationary 3-D wind field; and 3) estimating the position of a target using bearings-only measurements.

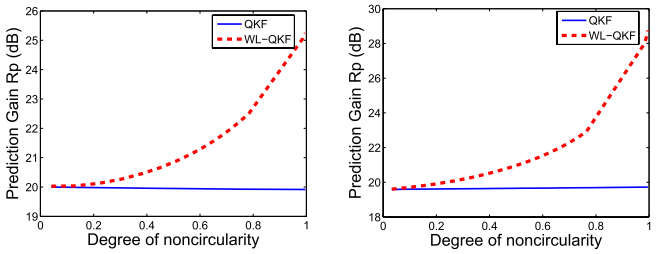


Fig. 1. Performance comparison between QKF and WL-QKF for the AR(4) process in (101), under varying degrees of noncircularity of the state and observation noise. (a) Noncircular observation noise. (b) Noncircular state noise.

A. Filtering of an Autoregressive Process

In the first experiment, the state model was described by a fourth-order quaternion autoregressive process, AR(4), given by

$$x_k = 1.4x_{k-1} - 0.7x_{k-1} + 0.04x_{k-1} - 0.05x_{k-1} + w_k \tag{100}$$

where the driving noise w_k was quadruply white zero mean Gaussian noise with the variance and pseudovariance defined as

$$\begin{aligned} E[w_k w_n^*] &= \alpha\delta_{k-n} & E[w_k w_n^{\imath *}] &= \beta\delta_{k-n} \\ E[w_k w_n^{\jmath *}] &= \gamma\,\delta_{k-n} & E[w_k w_n^{\kappa *}] &= \eta\delta_{k-n} \end{aligned}$$

where δ is the discrete Dirac delta function. In the first experiment, the state x_k was observed in the presence of additive quaternion-valued white noise v_n of varying degree of circularity,[4] that is

$$z_k = x_k + v_k. \tag{101}$$

In the second experiment, the state x_k was observed through a nonlinearity, given by

$$z_k = \arctan(x_k) + w_k. \tag{102}$$

The quantitative performance measure was the prediction gain $R_p = 10\log(\sigma_y^2/\sigma_e^2)$. where σ_y^2 and σ_e^2 are, respectively, the powers of the input signal and the output error. Fig. 1 compares the performances of the QKF and its widely linear counterpart, the WL-QKF, for the signal in (101). Fig. 1 shows the results for a circular state noise and an observation noise of varying degree of a circularity, while in Fig. 1 the observation noise was circular and the state noise noncircular. Conforming with the analysis, for both sets of experiments, when the noises were circular, the strictly linear and widely linear QKF had the same prediction gain. However, for noncircular noises, the QKF was inferior to the WL-QKF, as it did not cater for the full second-order statistics of the noises. This was also reflected by the performance of the QKF being independent of the noise noncircularity. Fig. 2 compares the performances of the QEKF and WL-QEKF for the nonlinear signal in (102). Similarly to the results for the linear observation model in the previous experiment, we observe that the WL-QEKF

[4]The noncircularity measure $r_s = \beta + \gamma + \eta/3\alpha$, where $r_s \in [0, 1]$ and the value $r_s = 0$ indicates a circular source while $r_s = 1$ indicates a highly noncircular source.

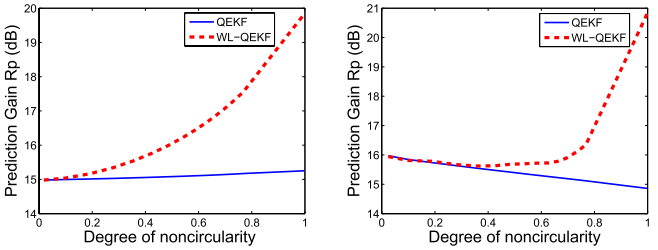


Fig. 2. Performance comparison between QEKF and WL-QEKF for the AR(4) process in (102), observed through output nonlinearity, for varying degrees of noncircularity of the state and observation noise. (a) Noncircular observation noise. (b) Noncircular state noise.

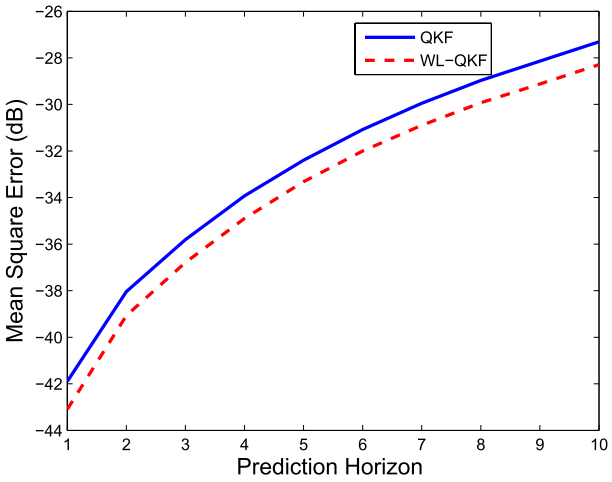


Fig. 3. Multistep ahead prediction of real-world 4-D wind data using QKF and WL-QKF.

outperformed the QEKF when either the observation noise or state noise were noncircular. However, unlike the previous experiment, for noncircular state noise, the performance of the strictly linear QEKF was dependent on the degree of noncircularity.

B. Multistep Ahead Prediction

The performances of the QKF and WL-QKF were next investigated for the multistep ahead prediction of a real-world noncircular and nonstationary 4-D wind field.[5] Fig. 3 shows the prediction performance in terms of the MSE for both the QKF and WL-QKF. Observe that, owing to its ability to utilize all the second-order information available, the WL-QKF was able to capture the underlying dynamics of the wind signal better than the QKF. This confirms that the wind signal is inherently noncircular (as shown in Fig. 4) and so should be processed using the widely linear Kalman filter. This is also true for most real-world scenarios, including orientation tracking in aeronautics.

C. Bearings-Only Tracking

In the bearings-only tracking problem, the aim is to estimate the position of an object based only on noisy bearings

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

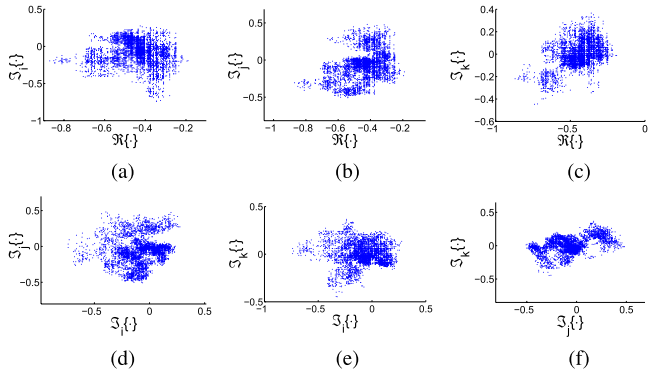


Fig. 4. Geometric view of the wind signal noncircularity via the six scatter plots of quaternion components. (a) q_r vs q_i. (b) q_r vs q_j. (c) q_r vs q_k. (d) q_i vs q_j. (e) q_i vs q_k. (f) q_j vs q_k

measurements of the object from one or more sensors. Bearings-only tracking is useful in military applications, for example, in submarine tracking by passive sonar. Since the range of the target must be estimated from bearings measurements, the problem is inherently nonlinear. To estimate the position $(x_k,\ y_k,\ z_k)$ and velocity $(\dot{x}_k,\ \dot{y}_k,\ \dot{z}_k)$ of a target using a system of L observers located at $(x^0_{i,k},\ y^0_{i,k},\ z^0_{i,k})$ for $i = 1, \ldots, L$, the WL-QEKF can be used with the state and observation model described as follows.

1) $\mathbf{x}_k = [\imath x_k + \jmath y_k + \kappa z_k, \imath \dot{x}_k + \jmath \dot{y}_k + \kappa \dot{z}_k]$.
2) Matrices $\mathbf{F}$ and $\mathbf{B}$ are defined as

$$\mathbf{F} = \begin{bmatrix} 1 & T \\ 0 & 1 \end{bmatrix} \qquad \mathbf{B} = \begin{bmatrix} \frac{T^2}{2} \\ T \end{bmatrix}$$

where T is the sampling interval.

3) $\mathbf{y}_k$ are the noise corrupted bearing measurements and $\mathbf{h}[\mathbf{x}_k]$ is a nonlinear defined as

$$\mathbf{h}[\mathbf{x}_k] = \begin{bmatrix} \theta_1, k + \imath\theta_{\frac{L}{2}+1,k} + \jmath\phi_{1,k} + \kappa\phi_{\frac{L}{2}+1,k} \\ \theta_{2,k} + \imath\theta_{\frac{L}{2}+2,k} + \jmath\phi_{2,k} + \kappa\phi_{\frac{L}{2}+2,k} \\ \vdots \\ \theta_{\frac{L}{2},k} + \imath\theta_{L,k} + \jmath\phi_{\frac{L}{2},k} + \kappa\phi_{L,k} \end{bmatrix}$$

where

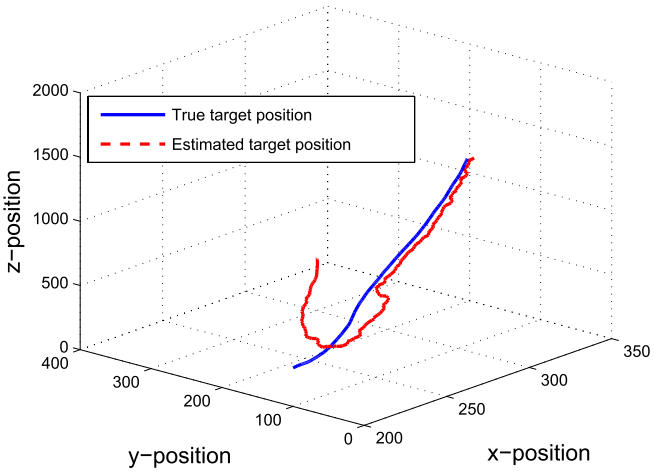
$$\theta_{i,k} = \arctan\frac{y_k - y^0_{i,k}}{x_k - x^0_{i,k}}$$

$$\phi_{i,k} = \arctan\frac{z_k - z^0_{i,k}}{((x_k - x^0_{i,k})^2 + (y_k - y^0_{i,k})^2)^{\frac{1}{2}}}$$

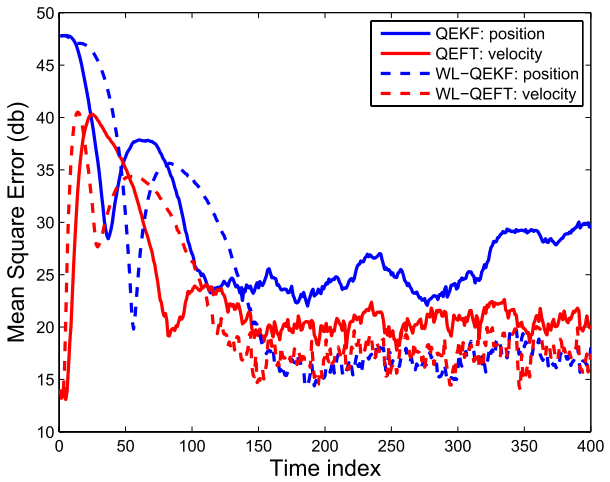
are, respectively, the azimuth and elevation angles.

4) $w_k = \ddot{x}_k + \jmath\ddot{x}_k + \kappa\ddot{x}_k$ is the zero mean noncircular state noise (accounting for the target accelerations) and $\mathbf{v}_k = [v_{1,k}\, v_{2,k} \ldots v_{L,k}]$ is the zero mean.

The function $\mathbf{h}[x_k]$ is nonholomorphic and therefore the first-order approximation using the quaternion Taylor series is widely linear. To illustrate the necessity of the widely linear model, we simulated a scenario where two static sensors are located at $(-1200, 1300, 0)$ and $(1000, 1500, 100)$. The sampling interval of $T = 0.1$ was used and the target was made to have an initial position of $(200, 100, 300)$ and velocity



(a)



(b)

Fig. 5. Bearings-only tracking using the WL-QEKF. (a) Trajectory of an object tracked in a 3-D space. (b) MSE performance.

$(2, 1, 0.5)$. The statistics of the state and observation noise were as follows:

$$w_k = \mathcal{N}(10, 10, 10, 10)$$
$$v_k = \mathcal{N}(0.001, 0.001, 0.001, 0.001).$$

Fig. 5(a) shows the trajectory of the target and the position estimated by the WL-QEKF. The MSE shown in Fig. 5 shows that the WL-QEKF achieved a better MSE performance than the QEKF, for both the position and velocity estimation of the target. This conforms with Remark #14 where we showed that due to quaternion nonlinear functions always being nonholomorphic, the extended Kalman filter must always be widely linear since the first-order quaternion Taylor series is also always widely linear.

XI. Conclusion

The quaternion Kalman (QKF) filter and its widely linear counterpart (WL-QKF) have been introduced for the processing of 3-D and 4-D signals in the quaternion domain. In this way, both second-order circular (proper) and noncircular (improper) are catered for, giving a natural solution in a division algebra where the data reside, as opposed to the existing approaches, which use the duality that between $\mathbb{H}$ and $\mathbb{R}^3$ to convert the quaternion quantities into real vectors. Both strictly linear and widely linear Kalman and extended Kalman filters have been developed, which was made possible by the recently introduced $\mathbb{HR}$-calculus. We have also illuminated that owing to the widely linear nature of the first-order quaternion Taylor series, the QEKF is in general widely linear, an issue that has so far prevented its development. The proposed class of quaternion Kalman filters fully operate in the quaternion domain where the data reside, and are a generic extension of their real- and complex-valued counterparts, simplifying into the existing widely and strictly linear complex solutions when operating on 2-D data [38]. The duality with the quadrivariate real Kalman filters and some computational issues have also been addressed. Simulations have shown that for a range of signals, for which either the state model is widely linear or the noises are noncircular, the WL-QKF filter, and the WL-QEKF offer better performance over their strictly linear counterparts.

Appendix A
Proof for the First-Order Quaternion Taylor Series

The Taylor series of a real quadrivariate function $u(q_a, q_b, q_c, q_d)$ is given by [39]

$$u(q_r, q_\imath, q_\jmath, q_\kappa) = u(q_{r0}, q_{\imath 0}, q_{\jmath 0}, q_{\kappa 0}) + \sum_{n=1}^{\infty} \frac{1}{n} \left[\Delta q_r \frac{\partial}{\partial q_r} + \Delta q_\imath \frac{\partial}{\partial q_\imath} + \Delta q_\jmath \frac{\partial}{\partial q_\jmath} + \Delta q_\kappa \frac{\partial}{\partial q_\kappa} \right]^n u(q_{r0}, q_{\imath 0}, q_{\jmath 0}, q_{\kappa 0}) \tag{103}$$

where $\Delta q_\eta = q_\eta - q_{\eta 0}$, for $\eta = \{r, \imath, \jmath, \kappa\}$. Using the identity

$$\begin{bmatrix} \frac{\partial f}{\partial q_r} \\ \frac{\partial f}{\partial q_\imath} \\ \frac{\partial f}{\partial q_\jmath} \\ \frac{\partial f}{\partial q_\kappa} \end{bmatrix}^T = \begin{bmatrix} \frac{\partial f}{\partial q} \\ \frac{\partial f}{\partial q^\imath} \\ \frac{\partial f}{\partial q^\jmath} \\ \frac{\partial f}{\partial q^\kappa} \end{bmatrix}^T \begin{bmatrix} 1 & \imath & \jmath & \kappa \\ 1 & \imath & -\jmath & -\kappa \\ 1 & -\imath & \jmath & -\kappa \\ 1 & -\imath & -\jmath & \kappa \end{bmatrix} \tag{104}$$

to related the real partial derivatives to quaternion partial derivatives, we next derive the quaternion Taylor series. For a real function of quaternion variables, we have $f(q, q^\imath, q^\jmath, q^\kappa) = u(q_r, q_\imath, q_\jmath, q_\kappa)$ and so substituting (104) into (103) and using the identities $\Delta q_r = 1/4(\Delta q + \Delta q^\imath + \Delta q^\jmath + \Delta q^\kappa)$, $\Delta q_\imath = 1/4\imath(\Delta q + \Delta q^\imath - \Delta q^\jmath - \Delta q^\kappa)$, $\Delta q_\jmath = 1/4\jmath(\Delta q - \Delta q^\imath + \Delta q^\jmath - \Delta q^k)$, and $\Delta q_\kappa = 1/4\kappa(\Delta q - \Delta q^\imath - \Delta q^\jmath + \Delta q^\kappa)$, we obtain the quaternion Taylor series

$$f(q, q^\imath, q^\jmath, q^\kappa) = f(q_0, q_0^\imath, q_0^\jmath, q_0^\kappa) + \sum_{n=1}^{\infty} \frac{1}{n} \left[\Delta q \frac{\partial}{\partial q} + \Delta q^\imath \frac{\partial}{\partial q^\imath} + \Delta q^\jmath \frac{\partial}{\partial q^\jmath} + \Delta q^\kappa \frac{\partial}{\partial q^\kappa} \right]^n \times f(q_0, q_0^\imath, q_0^\jmath, q_0^\kappa).$$

For a multivariate quaternion function, we thus have

$$f(\mathbf{q},\mathbf{q}^{\imath},\mathbf{q}^{\jmath},\mathbf{q}^{\kappa}) = f(\mathbf{q}_0,\mathbf{q}_0^{\imath},\mathbf{q}_0^{\jmath},\mathbf{q}_0^{\kappa}) + \sum_{n=1}^{\infty}\frac{1}{n}\left[\Delta\mathbf{q}^T\frac{\partial}{\partial\mathbf{q}} + \Delta\mathbf{q}^{\imath T}\frac{\partial}{\partial\mathbf{q}^{\imath}} + \Delta\mathbf{q}^{\jmath T}\frac{\partial}{\partial\mathbf{q}^{\jmath}} + \Delta\mathbf{q}^{\kappa T}\frac{\partial}{\partial\mathbf{q}^{\kappa}}\right]^n \times f(\mathbf{q}_0,\mathbf{q}_0^{\imath},\mathbf{q}_0^{\jmath},\mathbf{q}_0^{\kappa}).$$

Using the notation $\mathbf{v} = [\mathbf{q},\mathbf{q}^{\imath},\mathbf{q}^{\jmath},\mathbf{q}^{k}]^T$ and $\partial f/\partial\mathbf{q} = [\partial f/\partial\mathbf{q}, \partial f/\partial\mathbf{q}^{\imath}, \partial f/\partial\mathbf{q}^{\jmath}, \partial f/\partial\mathbf{q}^{\kappa}]^T$, the first-order quaternion Taylor series can be written as

$$f(\mathbf{v}) = f(\mathbf{v}_0) + \Delta\mathbf{v}^T\frac{\partial f}{\partial\mathbf{v}}. \tag{105}$$

REFERENCES

[1] M. St-Pierre and D. Gingras, "Comparison between the unscented Kalman filter and the extended Kalman filter for the position estimation module of an integrated navigation information system," in *Proc. Intell. Veh. Symp.*, Jun. 2004, pp. 831–835.

[2] F. Heimes, "Extended Kalman filter neural network training: Experimental results and algorithm improvements," in *Proc. IEEE Int. Conf. Syst., Man, Cybern.*, vol. 2. Oct. 1998, pp. 1639–1644.

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

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

[5] X. Wang and Y. Huang, "Convergence study in extended Kalman filter-based training of recurrent neural networks," *IEEE Trans. Neural Netw.*, vol. 22, no. 4, pp. 588–600, Apr. 2011.

[6] S. L. Goh and D. Mandic, "An augmented extended Kalman filter algorithm for complex-valued recurrent neural networks," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, vol. 5. May 2006.

[7] R. Sharaf and A. Noureldin, "Sensor integration for satellite-based vehicular navigation using neural networks," *IEEE Trans. Neural Netw.*, vol. 18, no. 2, pp. 589–594, Mar. 2007.

[8] D. Choukroun, I. Bar-Itzhack, and Y. Oshman, "Novel quaternion Kalman filter," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 42, no. 1, pp. 174–190, Jan. 2006.

[9] J. Marins, X. Yun, E. Bachmann, R. McGhee, and M. Zyda, "An extended Kalman filter for quaternion-based orientation estimation using MARG sensors," in *Proc. Int. Conf. Intell. Robots Syst.*, vol. 4. 2001, pp. 2003–2011.

[10] K. Shoemake, "Animating rotation with quaternion curves," *SIGGRAPH Comput. Graph.*, vol. 19, pp. 245–254, Jul. 1985.

[11] C. F. F. Karney, "Quaternions in molecular modeling," *J. Molecular Graph. Model.*, vol. 25, no. 5, pp. 595–604, 2007.

[12] C. Song, S. J. Kim, S. H. Kim, and H. Nam, "Robust control of the missile attitude based on quaternion feedback," *Control Eng. Pract.*, vol. 14, no. 7, pp. 811–818, 2006.

[13] A. Ude, "Filtering in a unit quaternion space for model-based object tracking," *Robot. Auto. Syst.*, vol. 28, nos. 2–3, pp. 163–172, 1999.

[14] V. Ntouskos, P. Papadakis, and F. Pirri, "A comprehensive analysis of human motion capture data for action recognition," in *Proc. Int. Conf. Comput. Vis. Theory Appl.*, Rome, Italy, 2012.

[15] A.-M. Zou, K. Kumar, and Z.-G. Hou, "Quaternion-based adaptive output feedback attitude control of spacecraft using Chebyshev neural networks," *IEEE Trans. Neural Netw.*, vol. 21, no. 9, pp. 1457–1471, Sep. 2010.

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

[17] 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.

[18] S. Javidi, C. Cheong Took, C. Jahanchahi, N. Le Bihan, and D. P. Mandic, "Blind extraction of improper quaternion sources," in *Proc. ICASSP*, May 2011, pp. 3708–3711.

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

[20] M. D. Shuster, "Approximate algorithms for fast optimal attitude computation," in *Proc. Guid. Control Conf.*, Mar. 1978, pp. 88–95.

[21] W. Grace, "Least squares estimate of spacecraft attitude," *SIAM Rev.*, vol. 7, no. 3, pp. 384–386, 1965.

[22] I. Bar-Itzhack and Y. Oshman, "Attitude determination from vector observations: Quaternion estimation," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-21, no. 1, pp. 128–136, Jan. 1985.

[23] D. Choukroun, I. Bar-Itzhack, and Y. Oshman, "Novel quaternion Kalman filter," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 42, no. 1, pp. 174–190, Jan. 2006.

[24] H. Rughooputh and S. Rughooputh, "Extended Kalman filter learning algorithm for hyper-complex multilayer neural networks," in *Proc. Int. Joint Conf. Neural Netw.*, vol. 3. 1999, pp. 1824–1828.

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

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

[27] J. Ward, *Quaternions and Cayley Numbers*. Norwell, MA, USA: Kluwer, 1997.

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

[29] B. K. P. Horn, "Closed-form solution of absolute orientation using unit quaternions," *J. Opt. Soc. Amer. A*, vol. 4, no. 4, pp. 629–642, 1987.

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

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

[32] W. Królikowski, "On 4-dimensional locally conformally flat almost Kahler manifolds," *Archivum Math.*, vol. 42, no. 3, pp. 215–223, 2006.

[33] K. Kreutz-Delgado, "The complex gradient operator and the $\mathbb{CR}$-calculus," Dept. Electr. Comput. Eng., Univ. California, San Diego, CA, USA, Tech. Rep., 2006.

[34] B. Picinbono and P. Bondon, "Second-order statistics of complex signals," *IEEE Trans. Signal Process.*, vol. 45, no. 2, pp. 411–420, Feb. 1997.

[35] B. C. Ujang, C. C. Took, and D. P. Mandic, "Quaternion-valued nonlinear adaptive filtering," *IEEE Trans. Neural Netw.*, vol. 22, no. 4, pp. 1193–1206, Aug. 2011.

[36] S. Haykin, *Adaptive Filter Theory*, 3rd ed. Englewood Cliffs, NJ, USA: Prentice-Hall, 1996.

[37] C. Jahanchahi, C. Took, and D. Mandic, "The widely linear quaternion recursive least squares filter," in *Proc. 2nd Int. Workshop CIP*, Jun. 2010, pp. 87–92.

[38] D. Dini and D. Mandic, "Class of widely linear complex Kalman filters," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 23, no. 5, pp. 775–786, May 2012.

[39] S. Bayin, *Mathematical Methods in Science and Engineering*. New York, NY, USA: Wiley, 2006.

Cyrus Jahanchahi received the M.Eng. degree in electrical and electronic engineering from Imperial College London, London, U.K., where he is currently pursuing the Ph.D. degree in signal processing.

His current research interests include linear and nonlinear adaptive filters and quaternion valued statistical analysis.

Danilo P. Mandic (M'99–SM'03–F'13) is a Professor of signal processing with Imperial College London, London, U.K. He was working on multivariate and nonlinear signal processing, data-driven signal analysis, and adaptive multidimensional learning systems. He was a Guest Professor with KU Leuven, Leuven, Belgium, and a Frontier Researcher with RIKEN, Japan. He has authored two research monographs, including *Recurrent Neural Networks for Prediction* in 2001, and *Complex Valued Nonlinear Adaptive Filters: Noncircularity, Widely Linear and Neural Models* in 2009.

Prof. Mandic was a member of the IEEE Technical Committee on Signal Processing Theory and Methods and was an Associate Editor of IEEE SIGNAL PROCESSING MAGAZINE, IEEE TRANSACTIONS ON SIGNAL PROCESSING, and IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS.