Brief Papers

Performance Bounds of Quaternion Estimators

Yili Xia, *Member, IEEE*, Cyrus Jahanchahi, Tohru Nitta, and Danilo P. Mandic, *Fellow, IEEE*

***Abstract*—The quaternion widely linear (WL) estimator has been recently introduced for optimal second-order modeling of the generality of quaternion data, both second-order circular (proper) and second-order noncircular (improper). Experimental evidence exists of its performance advantage over the conventional strictly linear (SL) as well as the semi-WL (SWL) estimators for improper data. However, rigorous theoretical and practical performance bounds are still missing in the literature, yet this is crucial for the development of quaternion valued learning systems for 3-D and 4-D data. To this end, based on the orthogonality principle, we introduce a rigorous closed-form solution to quantify the degree of performance benefits, in terms of the mean square error, obtained when using the WL models. The cases when the optimal WL estimation can simplify into the SWL or the SL estimation are also discussed.**

***Index Terms*—Augmented quaternion statistics, mean square error (MSE), quaternion widely linear (WL) model, semi-WL (SWL) model.**

I. Introduction

Standard techniques employed in multichannel statistical signal processing are typically not well equipped to fully exploit the coupled nature of the available information within the data channels. For example, univariate channel-wise processing is often inadequate and real-valued vectors are not a division algebra and suffer from the well-known mathematical deficiencies, e.g. the gimbal lock. Multivariate data typically come from 3-D and 4-D vector sensors, for example, from 3-D inertial sensors in body sensor networks, 3-D anemometers in wind energy, and three-axial seismometers in oil exploration [1], [2]. For this type of data, signal processing in the quaternion domain $\mathbb{H}$ has shown advantages over real vectors in $\mathbb{R}^3$ and $\mathbb{R}^4$, mostly owing to its division algebra and the accurate modeling of rotation and orientation. Quaternions account naturally for mutual information between the data channels, and offer a compact representation, free of mathematical deficiencies. In the context of learning systems, quaternions have been employed in Kalman filtering [3], [4], spectrum estimation [5], Fourier analysis [6], Taylor series expansion [7], least mean square estimation [8], and neural networks [9], [10].

As quaternions can be represented as a pair of complex numbers, it is natural to ask whether some recent breakthroughs in statistical signal processing of complex variable can be generalized to quaternions. One such breakthrough in complex-valued signal processing has been due to the widely linear (WL) model and augmented complex statistics [11]–[15]. This model, together with the corresponding augmented complex statistics, has been successfully used to design enhanced algorithms in communications and adaptive filters [14], [15]. A number of studies have shown that WL modeling offers theoretical and practical advantages over the standard strictly linear (SL) model, and is applicable to the generality of complex-valued random signals, both proper and improper.

The concept of $\mathbb{H}$ properness was introduced in [16] as the invariance of the probability density function (pdf) of a quaternion-valued variable q under some specific rotations around the angle of $\pi/2$, and this restriction was later relaxed to an arbitrary axis and angle of rotation φ. A variable q is said to be $\mathbb{H}$ proper, if $\mathrm{pdf}(q) = \mathrm{pdf}(e^{\upsilon\varphi} q)$ for any pure unit quaternion υ (whose real part vanishes) and any angle φ [17]. The Cayley–Dickson representation, whereby a quaternion variable is represented as a pair of complex variables, offers an insight into the use of complex-valued statistics for quaternion variables, leading to a less strict $\mathbb{C}$-properness condition. These statistical concepts have been explicitly or implicitly used in polarized seismic wave analysis [18], directionality detection in random fields for color image classification [19], and in degree reduction of Gaussian graphical models for covariance estimation [20], among others. These discoveries were followed by the quaternion WL model, together with the augmented quaternion statistics [21]–[23], to account for the generality of quaternion signals, both proper and improper. Such models have been shown to outperform the traditional SL quaternion processing, which relies only on the standard covariance matrix [3], [5], [6], [8], in diverse fields of learning systems including multivariate statistical analysis methods [7], [22], adaptive filtering [9], [10], [24], and independent component analysis [23], [25]. However, a rigorous quantitative analysis of statistical advantages of WL over SL processing in $\mathbb{H}$ is still missing, partly because different types of properness may occur based on the combinations of the four dimensions in $\mathbb{H}$.

To this end, we generalize our preliminary work in [26] and [27], which presented a closed-form solution to quantify the degree of performance benefits obtained when using the quasi-semi-WL (SWL) model over the SL model, in terms of the mean square error (MSE). For rigor, we here adopt a more general representation for quaternion imaginary units instead of the traditional canonical basis, and in this way, different properness conditions can be uniquely defined, regardless of the positions of imaginary units. Our analysis covers the general case of WL, SWL, and SL models. The so-derived theoretical and practical performance bounds are supported by case studies on the optimality of all the considered estimators in the context of learning systems.

II. Quaternion Estimators

We shall now briefly review the basic concepts in quaternion algebra, used in the design of quaternion estimators.

A quaternion variable $q \in \mathbb{H}$ is a skew field over $\mathbb{R}$, and comprises four real components (q_r, q_η, $q_{\eta'}$, and $q_{\eta''}$) and three imaginary units

Manuscript received April 18, 2014; revised October 22, 2014 and January 4, 2015; accepted January 4, 2015. Date of publication July 8, 2015; date of current version November 16, 2015. This work was supported in part by the Engineering and Physical Sciences Research Council under Grant EP/H026266/1, in part by the Natural Science Foundation of Jiangsu Province under Grant BK20140645, and in part by the National Natural Science Foundation of China under Grant 61401094.

Y. Xia is with the School of Information Science and Engineering, Southeast University, Nanjing 210018, China, and also with the Department of Electrical and Electronic Engineering, Imperial College London, London SW7 2AZ, U.K. (e-mail: yili.xia06@gmail.com).

C. Jahanchahi and D. P. Mandic are with the Department of Electrical and Electronic Engineering, Imperial College London, London SW7 2AZ, U.K. (e-mail: cyrus.jahanchahi05@imperial.ac.uk; d.mandic@imperial.ac.uk).

T. Nitta is with the Mathematical Neuroinformatics Group, National Institute of Advanced Industrial Science and Technology, Tsukuba 305-8568, Japan (e-mail: tohru-nitta@aist.go.jp).

Digital Object Identifier 10.1109/TNNLS.2015.2388782

(η, η', and η''), to give[1] [21]

$$q = q_r + \eta q_\eta + \eta' q_{\eta'} + \eta'' q_{\eta''} \tag{1}$$

where the imaginary units obey

$$\eta\eta' = \eta'', \eta'\eta'' = \eta, \eta''\eta = \eta'$$
$$\eta^2 = \eta'^2 = \eta''^2 = \eta\eta'\eta'' = -1.$$

Quaternion multiplication is not commutative, that is, for q_1 and $q_2 \in \mathbb{H}$, in general $q_1q_2 \neq q_2q_1$. The conjugate of a quaternion q is $q^* = q_r - \eta q_\eta - \eta' q_{\eta'} - \eta'' q_{\eta''}$, and the conjugate of a quaternion product obeys $(q_1q_2)^* = q_2^* q_1^*$. Another concept of particular interest to this brief is quaternion perpendicular involutions, defined as [28]

$$q^\eta = -\eta q \eta = q_r + \eta q_\eta - \eta' q_{\eta'} - \eta'' q_{\eta''} \tag{2}$$

for which the following properties hold:

$$(q^\eta)^\eta = q, \quad (q^\eta)^* = (q^*)^\eta, \quad (q_1q_2)^\eta = q_1^\eta q_2^\eta$$
$$(q_1+q_2)^\eta = q_1^\eta + q_2^\eta, \quad (q^\eta)^{\eta'} = (q^{\eta'})^\eta = q^{\eta''}.$$

Now, consider a real-valued MSE estimator that estimates a scalar random variable y, based on a random vector regressor $\mathbf{x}$, in the form

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

where $E[\cdot]$ is the statistical expectation operator. For jointly Gaussian, zero mean, $\mathbf{x}$ and y, the optimal solution is a linear estimator $\hat{y} = E[y|\mathbf{x}] = \mathbf{w}^T\mathbf{x}$, where $\mathbf{w}$ is the coefficient vector and $(\cdot)^T$ denotes the vector transpose operator.

The MSE estimation problem in the complex domain traditionally employs the so-called *SL model*, inherited from the real domain, which yields a linear estimator in the form

$$\hat{y}_{\mathrm{SL}} = E[y|\mathbf{x}] = \mathbf{w}^H\mathbf{x} \tag{3}$$

where $(\cdot)^H$ denotes the Hermitian transpose operator. However, this SL estimator is not the optimal solution for the generality of complex-valued Gaussian data, where the regression is linear both in $\mathbf{x}$ and $\mathbf{x}^*$, leading to the *WL estimator*, given by [12]

$$\hat{y}_{\mathrm{WL}} = E[y|\mathbf{x}] = \mathbf{h}^H\mathbf{x} + \mathbf{g}^H\mathbf{x}^* \tag{4}$$

which comprises both the SL part $\mathbf{h}^H\mathbf{x}$ and the conjugate part $\mathbf{g}^H\mathbf{x}^*$, where $\mathbf{h}$ and $\mathbf{g}$ are the coefficient vectors associated with $\mathbf{x}$ and $\mathbf{x}^*$, respectively. In the case of the jointly second-order circular or proper Gaussian processes, i.e., $E[y\mathbf{x}] = \mathbf{0}$ and $E[\mathbf{x}\mathbf{x}^T] = \mathbf{0}$, the WL estimator in (4) reduces to the SL estimator in (3), both giving identical MSE estimation performance, whereas in the other cases, the WL estimator offers significant performance advantages over the SL one, as justified in both theory and practice [12]–[15]. Although in this brief, we limit ourselves to Gaussian random variables, it is worth mentioning that the concept of circularity extends beyond second-order statistics, and there also exists a performance advantage offered by the WL model over the SL one for jointly noncircular non-Gaussian processes, for more detail, we refer to [29].

In a similar way, the MSE estimation problem in the quaternion domain $\mathbb{H}$ conventionally employs the SL estimation model, inherited from the complex domain, given by [3], [8]

$$\hat{y}_{\mathrm{SL}} = \mathbf{w}^H\mathbf{x}. \tag{5}$$

However, unlike the complex case, the most general regression for quaternion-valued Gaussian data is linear in both $\mathbf{x}$ and all its three involutions, referred to as *quaternion WL* model and given by [21], [22]

$$\hat{y}_{\mathrm{WL}} = \mathbf{w}^H\mathbf{x} + \mathbf{w}_\eta^H\mathbf{x}^\eta + \mathbf{w}_{\eta'}^H\mathbf{x}^{\eta'} + \mathbf{w}_{\eta''}^H\mathbf{x}^{\eta''} = \mathbf{w}^{aH}\mathbf{x}^a \tag{6}$$

or equivalently, its conjugate

$$\hat{y}^*_{\mathrm{WL}} = \mathbf{x}^H\mathbf{w} + \mathbf{x}^{\eta H}\mathbf{w}_\eta + \mathbf{x}^{\eta' H}\mathbf{w}_{\eta'} + \mathbf{x}^{\eta'' H}\mathbf{w}_{\eta''} = \mathbf{x}^{aH}\mathbf{w}^a \tag{7}$$

where $\mathbf{w}^a = [\mathbf{w}^T, \mathbf{w}_\eta^T, \mathbf{w}_{\eta'}^T, \mathbf{w}_{\eta''}^T]^T$ is termed the augmented coefficient vector and $\mathbf{x}^a = [\mathbf{x}^T, \mathbf{x}^{\eta T}, \mathbf{x}^{\eta' T}, \mathbf{x}^{\eta'' T}]^T$ is the augmented input vector.

Current statistical signal processing in $\mathbb{H}$ by and large employs the SL model, drawing upon the covariance matrix $\mathbf{R} = E[\mathbf{x}\mathbf{x}^H]$. However, based on (6), the modeling of both the second-order circular (proper) and noncircular (improper) signals is only possible using the augmented covariance matrix, given by [22]

$$\mathbf{R}^a = E[\mathbf{x}^a\mathbf{x}^{aH}] = \begin{bmatrix} \mathbf{R} & \mathbf{P} & \mathbf{S} & \mathbf{T} \\ \mathbf{P}^H & \mathbf{R}^\eta & \mathbf{T}^\eta & \mathbf{S}^\eta \\ \mathbf{S}^H & \mathbf{T}^{\eta H} & \mathbf{R}^{\eta'} & \mathbf{P}^{\eta'} \\ \mathbf{T}^H & \mathbf{S}^{\eta H} & \mathbf{P}^{\eta' H} & \mathbf{R}^{\eta''} \end{bmatrix} \tag{8}$$

which compromises the covariance matrix $\mathbf{R} = E[\mathbf{x}\mathbf{x}^H]$, the three pseudocovariance matrices $\mathbf{P} = E[\mathbf{x}\mathbf{x}^{\eta H}]$, $\mathbf{S} = E[\mathbf{x}\mathbf{x}^{\eta' H}]$, and $\mathbf{T} = E[\mathbf{x}\mathbf{x}^{\eta'' H}]$ as well as their involutions and Hermitian transposes.

We can also adopt the semiaugmented form to describe the WL estimation process in (6), given by [21]

$$\hat{y}_{\mathrm{WL}} = \mathbf{w}^{bH}\mathbf{x}^b + \mathbf{w}^{cH}\mathbf{x}^c \tag{9}$$

where $\mathbf{x}^b = [\mathbf{x}^T, \mathbf{x}^{\eta T}]^T$ and $\mathbf{x}^c = \mathbf{x}^{b\eta'} = [\mathbf{x}^{\eta' T}, \mathbf{x}^{\eta'' T}]^T$ are the two semiaugmented input vectors, for which the associated semiaugmented weight vectors are, respectively, given by $\mathbf{w}^{bH} = [\mathbf{w}^H, \mathbf{w}_\eta^H]$ and $\mathbf{w}^{cH} = [\mathbf{w}_{\eta'}^H, \mathbf{w}_{\eta''}^H]$. Also, the quaternion conjugate of the estimator in (9) can be expressed as

$$\hat{y}^*_{\mathrm{WL}} = \mathbf{x}^{bH}\mathbf{w}^b + \mathbf{x}^{cH}\mathbf{w}^c \tag{10}$$

to yield the augmented covariance matrix $\mathbf{R}^a$ in (8) in the semiaugmented form

$$\mathbf{R}^a = \begin{bmatrix} \mathbf{B} & \mathbf{C} \\ \mathbf{C}^H & \mathbf{B}^{\eta''} \end{bmatrix} \tag{11}$$

where

$$E[\mathbf{x}^b\mathbf{x}^{bH}] = \begin{bmatrix} \mathbf{R} & \mathbf{P} \\ \mathbf{P}^H & \mathbf{R}^\eta \end{bmatrix} = \mathbf{B} \tag{12}$$

$$E[\mathbf{x}^b\mathbf{x}^{cH}] = \begin{bmatrix} \mathbf{S} & \mathbf{T} \\ \mathbf{T}^\eta & \mathbf{S}^\eta \end{bmatrix} = \mathbf{C} \tag{13}$$

$$E[\mathbf{x}^c\mathbf{x}^{cH}] = \begin{bmatrix} \mathbf{R}^{\eta'} & \mathbf{P}^{\eta'} \\ \mathbf{P}^{\eta' H} & \mathbf{R}^{\eta''} \end{bmatrix} = \mathbf{B}^{\eta'}. \tag{14}$$

Quaternion proper ($\mathbb{H}$ proper) signals, have probability distributions that are rotation invariant with respect to all the six possible pairs of axes (combinations of η, η', and η''), and thus proper signals exhibit equal powers in all the quaternion components, so that in turn all the pseudocovariance matrices $\mathbf{P}$, $\mathbf{S}$, and $\mathbf{T}$ vanish.

Remark 1: For $\mathbb{H}$-proper signals, the augmented covariance matrix $\mathbf{R}^a$ in (8) is completely described by the standard covariance matrix $\mathbf{R}$ and its involutions, so that the SL estimation model in (5), based only on the covariance matrix $\mathbf{R}$, is second-order optimal for proper signals.

Another class of properness in $\mathbb{H}$ is the so-called $\mathbb{C}^\eta$ properness, whereby the quaternion random signal $\mathbf{x}$ is correlated with $\mathbf{x}^\eta$, but not with the remaining two perpendicular involutions $\mathbf{x}^{\eta'}$ and $\mathbf{x}^{\eta''}$, and hence the pseudocovariance matrices relating $\mathbf{x}$ and the remaining

[1] A most frequent choice of the imaginary axes is the canonical basis $\{\imath, \jmath, \kappa\}$, however, for compactness of our analysis of quaternion estimators, we adopt the more general representation $\{\eta, \eta', \eta''\}$, where η can be any element from the imaginary unit set $\{\imath, \jmath, \kappa\}$, η' is the one next to η cyclically in the imaginary unit set, and η'' is the one to the left.

two perpendicular involutions, $\mathbf{S}$ and $\mathbf{T}$ in (8), vanish. This implies that the matrix $\mathbf{C} = \mathbf{0}$ in (11), and that the augmented covariance matrix $\mathbf{R}^a$ simplifies into [21]

$$\mathbf{R}^a = \begin{bmatrix} \mathbf{B} & \mathbf{0} \\ \mathbf{0} & \mathbf{B}^{\eta'} \end{bmatrix}. \tag{15}$$

For the jointly $\mathbb{C}^\eta$-proper processes y and $\mathbf{x}$, the second-order optimal estimator in (9) then reduces to the so-called SWL estimator, given by [21]

$$\hat{y}_{\mathrm{SWL}} = \mathbf{w}^H\mathbf{x} + \mathbf{w}_\eta^H\mathbf{x}^\eta = \mathbf{w}^{bH}\mathbf{x}^b. \tag{16}$$

The conjugate of the SWL estimator in (16) is

$$\hat{y}^*_{\mathrm{SWL}} = \mathbf{x}^H\mathbf{w} + \mathbf{x}^{\eta H}\mathbf{w}_\eta = \mathbf{x}^{bH}\mathbf{w}^b. \tag{17}$$

Note that, inspired by the complex domain, a quasi-SWL model was proposed in [26], to simultaneously operate on both the quaternion regressor $\mathbf{x}$ and its conjugate $\mathbf{x}^*$. We here further generalize this approach, since unlike the complex case, there exist three kinds of quaternion SWL processing, as shown in (16).

III. MSE Analysis of Quaternion Estimators

MSE-based estimation in the quaternion domain $\mathbb{H}$ aims to find the optimal weight vectors so as to yield the minimum MSE $E[|y-\hat{y}|^2]$, where $\hat{y}$ can be obtained using any estimation model described above. The superior performance of the WL estimation processing over the SWL and SL estimation for second-order noncircular quaternion data has been illustrated experimentally [8], [22], [24], [25], however, the extent of the performance advantage offered by the WL processing for quaternion improper data has not yet been established theoretically. In this section, we use the orthogonality principle as a basis to introduce a closed-form solution for the optimal weight vectors for all the possible quaternion estimation models. We then provide the expressions for their respective minimum MSEs when operating on the generality of quaternion data (both proper and improper). This is followed by a performance comparison study, which quantifies the degree of performance benefits offered by the various WL processing schemes.

A. MSE Analysis of Widely Linear Estimator

Our aim is to find the optimal weight vectors $\mathbf{w}$, $\mathbf{w}_\eta$, $\mathbf{w}_{\eta'}$, and $\mathbf{w}_{\eta''}$ in (6) to yield the minimum MSE $E[|y-\hat{y}_{\mathrm{WL}}|^2]$. For this purpose, the first point to note is that the set of scalar quaternion variables $q(\omega)$ in the form $q(\omega) = \mathbf{a}^H\mathbf{x}(\omega) + \mathbf{b}^H\mathbf{x}^\eta(\omega) + \mathbf{c}^H\mathbf{x}^{\eta'}(\omega) + \mathbf{d}^H\mathbf{x}^{\eta''}(\omega)$, where $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$, and $\mathbf{d}$ belong to $\mathbb{H}^N$ constitutes a linear space. Upon introducing the scalar product $\langle q_1, q_2\rangle = E[q_1 q_2^*]$, this space then becomes a Hilbert subspace of the 1-D quaternion valued Hilbert space [26]. The WL estimate, $\hat{y}_{\mathrm{WL}}$, is the projection of y onto this subspace, and obeys the orthogonality principle

$$(y-\hat{y}_{\mathrm{WL}}) \perp \mathbf{x}, \quad (y-\hat{y}_{\mathrm{WL}}) \perp \mathbf{x}^\eta \tag{18}$$
$$(y-\hat{y}_{\mathrm{WL}}) \perp \mathbf{x}^{\eta'}, \quad (y-\hat{y}_{\mathrm{WL}}) \perp \mathbf{x}^{\eta''} \tag{19}$$

where the symbol $\perp$ indicates that all the components of $\mathbf{x}$ and its involutions are orthogonal to $(y-\hat{y}_{\mathrm{WL}})$ (their scalar product is zero). These orthogonality conditions can be written in terms of the expectations as

$$E[\mathbf{x}y^*] = E\big[\mathbf{x}\hat{y}^*_{\mathrm{WL}}\big], \quad E[\mathbf{x}^\eta y^*] = E\big[\mathbf{x}^\eta\hat{y}^*_{\mathrm{WL}}\big] \tag{20}$$
$$E[\mathbf{x}^{\eta'} y^*] = E\big[\mathbf{x}^{\eta'}\hat{y}^*_{\mathrm{WL}}\big], \quad E[\mathbf{x}^{\eta''} y^*] = E\big[\mathbf{x}^{\eta''}\hat{y}^*_{\mathrm{WL}}\big]. \tag{21}$$

Substituting (7) into (20) and (21), we obtain

$$\begin{bmatrix} \mathbf{R} & \mathbf{P} & \mathbf{S} & \mathbf{T} \\ \mathbf{P}^H & \mathbf{R}^\eta & \mathbf{T}^\eta & \mathbf{S}^\eta \\ \mathbf{S}^H & \mathbf{T}^{\eta H} & \mathbf{R}^{\eta'} & \mathbf{P}^{\eta'} \\ \mathbf{T}^H & \mathbf{S}^{\eta H} & \mathbf{P}^{\eta' H} & \mathbf{R}^{\eta''} \end{bmatrix} \begin{bmatrix} \mathbf{w} \\ \mathbf{w}_\eta \\ \mathbf{w}_{\eta'} \\ \mathbf{w}_{\eta''} \end{bmatrix} = \begin{bmatrix} \mathbf{r} \\ \mathbf{r}_\eta \\ \mathbf{r}_{\eta'} \\ \mathbf{r}_{\eta''} \end{bmatrix} \tag{22}$$

where the cross-correlation vectors $\mathbf{r} = E[\mathbf{x}y^*]$, $\mathbf{r}_\eta = E[\mathbf{x}^\eta y^*]$, $\mathbf{r}_{\eta'} = E[\mathbf{x}^{\eta'} y^*]$, and $\mathbf{r}_{\eta''} = E[\mathbf{x}^{\eta''} y^*]$. Observe that the square matrix on the left-hand side of (22) is the augmented covariance matrix $\mathbf{R}^a$ in (8).

To simplify the analysis, we adopt the semiaugmented form to represent the WL model, giving an equivalent to (22) in the form

$$\begin{bmatrix} \mathbf{B} & \mathbf{C} \\ \mathbf{C}^H & \mathbf{B}^{\eta'} \end{bmatrix} \begin{bmatrix} \mathbf{w}^b \\ \mathbf{w}^c \end{bmatrix} = \begin{bmatrix} \mathbf{r}^b \\ \mathbf{r}^c \end{bmatrix} \tag{23}$$

where $\mathbf{r}^b = [\mathbf{r}^T, \mathbf{r}_\eta^T]^T$ and $\mathbf{r}^c = [\mathbf{r}_{\eta'}^T, \mathbf{r}_{\eta''}^T]^T$. Assuming that $\mathbf{R}^a$ is invertible, from (23), we have

$$\begin{bmatrix} \mathbf{w}^b \\ \mathbf{w}^c \end{bmatrix} = \begin{bmatrix} \mathbf{A}_1 & \mathbf{A}_2 \\ \mathbf{A}_3 & \mathbf{A}_4 \end{bmatrix} \begin{bmatrix} \mathbf{r}^b \\ \mathbf{r}^c \end{bmatrix} \tag{24}$$

so that

$$\begin{bmatrix} \mathbf{B} & \mathbf{C} \\ \mathbf{C}^H & \mathbf{B}^{\eta'} \end{bmatrix} \begin{bmatrix} \mathbf{A}_1 & \mathbf{A}_2 \\ \mathbf{A}_3 & \mathbf{A}_4 \end{bmatrix} = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \tag{25}$$

where $\mathbf{I}$ is the identity matrix. Solving for $\mathbf{A}_1$, $\mathbf{A}_2$, $\mathbf{A}_3$, and $\mathbf{A}_4$ (see Appendix A for a detailed derivation), we have

$$\mathbf{A}_1 = (\mathbf{B} - \mathbf{C}\mathbf{B}^{-\eta'}\mathbf{C}^H)^{-1}, \quad \mathbf{A}_2 = -(\mathbf{B} - \mathbf{C}\mathbf{B}^{-\eta'}\mathbf{C}^H)^{-1}\mathbf{C}\mathbf{B}^{-\eta'}$$
$$\mathbf{A}_3 = -(\mathbf{B}^{\eta'} - \mathbf{C}^H\mathbf{B}^{-1}\mathbf{C})^{-1}\mathbf{C}^H\mathbf{B}^{-1}, \quad \mathbf{A}_4 = (\mathbf{B}^{\eta'} - \mathbf{C}^H\mathbf{B}^{-1}\mathbf{C})^{-1}$$

so that from (24), we arrive that

$$\mathbf{w}^b = \mathbf{A}_1\mathbf{r}^b + \mathbf{A}_2\mathbf{r}^c = (\mathbf{B} - \mathbf{C}\mathbf{B}^{-\eta'}\mathbf{C}^H)^{-1}(\mathbf{r}^b - \mathbf{C}\mathbf{B}^{-\eta'}\mathbf{r}^c) \tag{26}$$
$$\mathbf{w}^c = \mathbf{A}_3\mathbf{r}^b + \mathbf{A}_4\mathbf{r}^c = (\mathbf{B}^{\eta'} - \mathbf{C}^H\mathbf{B}^{-1}\mathbf{C})^{-1}(\mathbf{r}^c - \mathbf{C}^H\mathbf{B}^{-1}\mathbf{r}^b) \tag{27}$$

and upon using (9) and (10), the minimum MSE of the WL estimator, denoted as e^2_{WL}, can be expressed as

$$\begin{aligned} e^2_{\mathrm{WL}} &= E[|y-\hat{y}_{\mathrm{WL}}|^2] \\ &= E[|y|^2] + E\big[\hat{y}_{\mathrm{WL}}\hat{y}^*_{\mathrm{WL}}\big] - E\big[y\hat{y}^*_{\mathrm{WL}}\big] - E[\hat{y}_{\mathrm{WL}}y^*] \\ &= E[|y|^2] + E[(\mathbf{w}^{bH}\mathbf{x}^b + \mathbf{w}^{cH}\mathbf{x}^c)(\mathbf{x}^{bH}\mathbf{w}^b + \mathbf{x}^{cH}\mathbf{w}^c)] \\ &\quad - E[y(\mathbf{x}^{bH}\mathbf{w}^b + \mathbf{x}^{cH}\mathbf{w}^c)] - E[(\mathbf{w}^{bH}\mathbf{x}^b + \mathbf{w}^{cH}\mathbf{x}^c)y^*] \\ &= E[|y|^2] + \mathbf{w}^{bH}\mathbf{B}\mathbf{w}^b + \mathbf{w}^{cH}\mathbf{B}^{\eta'}\mathbf{w}^c + \mathbf{w}^{cH}\mathbf{C}^H\mathbf{w}^b \\ &\quad + \mathbf{w}^{bH}\mathbf{C}\mathbf{w}^c - \mathbf{r}^{bH}\mathbf{w}^b - \mathbf{r}^{cH}\mathbf{w}^c - \mathbf{w}^{bH}\mathbf{r}^b - \mathbf{w}^{cH}\mathbf{r}^c. \end{aligned} \tag{28}$$

Finally, from (23), we have $\mathbf{B}\mathbf{w}^b = \mathbf{r}^b - \mathbf{C}\mathbf{w}^c$ and $\mathbf{B}^{\eta'}\mathbf{w}^c = \mathbf{r}^c - \mathbf{C}^H\mathbf{w}^b$, so that upon substituting into (28), the MSE of the quaternion WL model, e^2_{WL}, can be expressed in a compact form as

$$e^2_{\mathrm{WL}} = E[|y|^2] - (\mathbf{r}^{bH}\mathbf{w}^b + \mathbf{r}^{cH}\mathbf{w}^c). \tag{29}$$

B. MSE Analysis of Semiwidely Linear (SWL) Estimator

Within SWL estimation, the estimate $\hat{y}_{\mathrm{SWL}}$, is the projection of y onto a linear Hilbert subspace consisting of $\mathbf{x}$ and one of its involutions $\mathbf{x}^\eta$ (16). The estimate $\hat{y}_{\mathrm{SWL}}$, which yields the minimum MSE, is then governed by the following orthogonality conditions:

$$(y-\hat{y}_{\mathrm{SWL}}) \perp \mathbf{x}, \quad (y-\hat{y}_{\mathrm{SWL}}) \perp \mathbf{x}^\eta \tag{30}$$

which can also be expressed in the semiaugmented form as

$$(y-\hat{y}_{\mathrm{SWL}}) \perp \mathbf{x}^b \tag{31}$$

so that

$$E[\mathbf{x}^b y^*] = E\big[\mathbf{x}^b \hat{y}^*_{\mathrm{SWL}}\big] = \mathbf{r}^b. \tag{32}$$

Substituting (17) into (32) yields $\mathbf{B}\mathbf{w}^b = \mathbf{r}^b$, so that

$$\mathbf{w}^b = \mathbf{B}^{-1}\mathbf{r}^b. \tag{33}$$

Finally, using (17), and (33), the minimum MSE based on the SWL processing, denoted by e^2_{SWL}, becomes

$$\begin{aligned} e^2_{\mathrm{SWL}} &= E\big[|y - \hat{y}_{\mathrm{SWL}}|^2\big] \\ &= E[|y|^2] + E\big[\hat{y}_{\mathrm{SWL}}\hat{y}^*_{\mathrm{SWL}}\big] - E\big[y\hat{y}^*_{\mathrm{SWL}}\big] - E\big[\hat{y}_{\mathrm{SWL}} y^*\big] \\ &= E[|y|^2] + \mathbf{w}^{bH}\mathbf{B}\mathbf{w}^b - \mathbf{r}^{bH}\mathbf{w}^b - \mathbf{w}^{bH}\mathbf{r}^b \\ &= E[|y|^2] - \mathbf{r}^{bH}\mathbf{w}^b = E[|y|^2] - \mathbf{r}^{bH}\mathbf{B}^{-1}\mathbf{r}^b. \end{aligned} \tag{34}$$

C. MSE Analysis of Strictly Linear Estimator

Out of all the models considered, the quaternion SL estimation model (5) is most restricted, whereby the estimate $\hat{y}_{\mathrm{SL}}$ represents a projection of y onto a linear Hilbert subspace defined by only $\mathbf{x}$. In other words, $\hat{y}_{\mathrm{SL}}$ obtained using the optimal weight vector $\mathbf{w}$ yields the minimum MSE according to the single orthogonality condition

$$(y - \hat{y}_{\mathrm{SL}}) \perp \mathbf{x}. \tag{35}$$

Similar to (32), we can now write

$$E[\mathbf{x}y^*] = E\big[\mathbf{x}\hat{y}^*_{\mathrm{SL}}\big] = \mathbf{r}. \tag{36}$$

Since the conjugate of the SL estimate in (36) is

$$\hat{y}^*_{\mathrm{SL}} = \mathbf{x}^H\mathbf{w} \tag{37}$$

then from (36), we have $\mathbf{R}\mathbf{w} = \mathbf{r}$ and by assuming the invertibility of the standard covariance matrix $\mathbf{R}$, the optimal weights assume the form

$$\mathbf{w} = \mathbf{R}^{-1}\mathbf{r}. \tag{38}$$

Finally, using (37) and (38), the minimum MSE based on the SL estimator, denoted by e^2_{SL}, for which the optimal weight coefficients are given in (38), can be obtained as

$$\begin{aligned} e^2_{\mathrm{SL}} &= E[|y - \hat{y}_{\mathrm{SL}}|^2] \\ &= E[|y|^2] + E\big[\hat{y}_{\mathrm{SL}}\hat{y}^*_{\mathrm{SL}}\big] - E\big[y\hat{y}^*_{\mathrm{SL}}\big] - E\big[\hat{y}_{\mathrm{SL}} y^*\big] \\ &= E[|y|^2] + \mathbf{w}^H\mathbf{R}\mathbf{w} - \mathbf{r}^H\mathbf{w} - \mathbf{w}^H\mathbf{r} \\ &= E[|y|^2] - \mathbf{r}^H\mathbf{R}^{-1}\mathbf{r}. \end{aligned} \tag{39}$$

D. Comparison Between WL and SWL Estimation

From (29) and (34), the performance advantage of the WL estimation over SWL estimation can be characterized by the difference of the respective error powers, given by

$$\begin{aligned} \triangle e^2_{\mathrm{WL}} &= e^2_{\mathrm{SWL}} - e^2_{\mathrm{WL}} \\ &= \mathbf{r}^{bH}\mathbf{w}^b + \mathbf{r}^{cH}\mathbf{w}^c - \mathbf{r}^{bH}\mathbf{B}^{-1}\mathbf{r}^b. \end{aligned} \tag{40}$$

To make this expression more physically meaningful, from (23), we have $\mathbf{B}\mathbf{w}^b + \mathbf{C}\mathbf{w}^c = \mathbf{r}^b$, which yields

$$\mathbf{w}^b = \mathbf{B}^{-1}(\mathbf{r}^b - \mathbf{C}\mathbf{w}^c). \tag{41}$$

Substituting (41) into (40), and using (27), we obtain the factorized form of (40), given by

$$\begin{aligned} \triangle e^2_{\mathrm{WL}} &= (\mathbf{r}^c - \mathbf{C}^H\mathbf{B}^{-1}\mathbf{r}^b)^H\mathbf{w}^c \\ &= (\mathbf{r}^c - \mathbf{C}^H\mathbf{B}^{-1}\mathbf{r}^b)^H(\mathbf{B}^{\eta'} - \mathbf{C}^H\mathbf{B}^{-1}\mathbf{C})^{-1}(\mathbf{r}^c - \mathbf{C}^H\mathbf{B}^{-1}\mathbf{r}^b). \end{aligned} \tag{42}$$

Remark 2: The term $\triangle e^2_{\mathrm{WL}}$ is always nonnegative, because the matrix $\mathbf{B}^{\eta'} - \mathbf{C}^H\mathbf{B}^{-1}\mathbf{C}$ is positive semidefinite (see Appendix B for the detailed proof).

Remark 3: The MSE of the SWL estimation is always larger or at most equal to that of the WL estimation. The equality between the MSEs of the SWL and WL estimators, that is $\triangle e^2_{\mathrm{WL}} = 0$, holds only when

$$\mathbf{r}^c - \mathbf{C}^H\mathbf{B}^{-1}\mathbf{r}^b = \mathbf{0} \tag{43}$$

which implies that $\mathbf{w}^c = \mathbf{0}$ in (27), that is, the quaternion WL estimation (9) reduces to the SWL estimation in (16).

We shall now consider several special cases of properness in some of the quaternion dimensions.

1) Jointly $\mathbb{C}^\eta$-Proper Case: Consider a special jointly $\mathbb{C}^\eta$-proper case between both the estimandum y and the regressor $\mathbf{x}^a = [\mathbf{x}^{bT}, \mathbf{x}^{cT}]^T$ within the quaternion WL processing (9), whereby

$$\mathbf{C} = E[\mathbf{x}^b\mathbf{x}^{cH}] = \mathbf{0}, \qquad \mathbf{r}^c = E[\mathbf{x}^c y^*] = \mathbf{0}. \tag{44}$$

It is clear that this jointly $\mathbb{C}^\eta$-proper assumption is sufficient for (43) to hold, and hence both the quaternion WL and the SWL estimation yield identical performance, with $\triangle e^2_{\mathrm{WL}} = 0$.

2) $\mathbb{C}^\eta$-Proper Regressor: When the $\mathbb{C}^\eta$-properness is only valid for the regressor $\mathbf{x}^a$, characterized by $\mathbf{C} = \mathbf{0}$, and no assumption is imposed on the estimandum y, the expressions for the optimal semiaugmented weight vectors $\mathbf{w}^b$ and $\mathbf{w}^c$ in (26) and (27) can be, respectively, simplified into

$$\mathbf{w}^b = \mathbf{B}^{-1}\mathbf{r}^b, \quad \mathbf{w}^c = \mathbf{B}^{-\eta'}\mathbf{r}^c. \tag{45}$$

Note that the term $\mathbf{w}^b = \mathbf{B}^{-1}\mathbf{r}^b$ remains the same to that in (33), obtained using the SWL estimation. This stems from the $\mathbb{C}^\eta$-properness assumption on the regressor $\mathbf{x}^a = [\mathbf{x}^{bT}, \mathbf{x}^{cT}]^T$, which implies that $\mathbf{x}^b$ and $\mathbf{x}^c$ are uncorrelated. Therefore, the Hilbert subspaces generated by $\mathbf{x}^b$ and $\mathbf{x}^c$ are orthogonal, since $\mathbf{x}^c$ does not affect the term coming from $\mathbf{x}^b$ only. Note that there still exists a performance advantage when using the WL estimation instead of the SWL processing, since $\triangle e^2_{\mathrm{WL}} = \mathbf{r}^{cH}\mathbf{B}^{-\eta'}\mathbf{r}^c$ is strictly positive.

E. Comparison Between SWL and SL Estimation

Now that we have rigorously established the advantage of the WL over various forms of SWL estimators, we shall next establish the theoretical performance advantage of SWL over SL estimation. From (34) and (39), the performance advantage of the SWL estimator over the SL one is characterized by

$$\triangle e^2_{\mathrm{SWL}} = e^2_{\mathrm{SL}} - e^2_{\mathrm{SWL}} = \mathbf{r}^{bH}\mathbf{w}^b - \mathbf{r}^H\mathbf{R}^{-1}\mathbf{r}. \tag{46}$$

For the quantitative analysis of $\triangle e^2_{\mathrm{SWL}}$, we next express (33) in a component-wise form, to give insight into the term $\mathbf{r}^{bH}\mathbf{w}^b$. Expanding the expression (33) gives

$$\begin{bmatrix} \mathbf{R} & \mathbf{P} \\ \mathbf{P}^H & \mathbf{R}^\eta \end{bmatrix}\begin{bmatrix} \mathbf{w} \\ \mathbf{w}_\eta \end{bmatrix} = \begin{bmatrix} \mathbf{r} \\ \mathbf{r}_\eta \end{bmatrix} \tag{47}$$

and consequently the solution for $\mathbf{w}^b = [\mathbf{w}^T, \mathbf{w}^T_\eta]^T$ becomes

$$\begin{bmatrix} \mathbf{w} \\ \mathbf{w}_\eta \end{bmatrix} = \begin{bmatrix} \tilde{\mathbf{A}}_1 & \tilde{\mathbf{A}}_2 \\ \tilde{\mathbf{A}}_3 & \tilde{\mathbf{A}}_4 \end{bmatrix}\begin{bmatrix} \mathbf{r} \\ \mathbf{r}_\eta \end{bmatrix}. \tag{48}$$

Now, similar to the analysis in (23)–(26), we have

$$\tilde{\mathbf{A}}_1 = (\mathbf{R} - \mathbf{P}\mathbf{R}^{-\eta}\mathbf{P}^H)^{-1}, \quad \tilde{\mathbf{A}}_2 = -(\mathbf{R} - \mathbf{P}\mathbf{R}^{-\eta}\mathbf{P}^H)^{-1}\mathbf{P}\mathbf{R}^{-\eta}$$

$$\tilde{\mathbf{A}}_3 = -(\mathbf{R}^\eta - \mathbf{P}^H\mathbf{R}^{-1}\mathbf{P})^{-1}\mathbf{P}^H\mathbf{R}^{-1}, \quad \tilde{\mathbf{A}}_4 = (\mathbf{R}^\eta - \mathbf{P}^H\mathbf{R}^{-1}\mathbf{P})^{-1}$$

TABLE I
THEORETICAL AND SIMULATED MSE MEASURES OF DIFFERENT QUATERNION ESTIMATORS, AS WELL AS THEIR PERFORMANCE DIFFERENCES

	$e^2_{\text{WL}}(29)$	$e^2_{\text{SWL}}(34)$	$e^2_{\text{SL}}(39)$	$\triangle e^2_{\text{WL}}(42)$	$\triangle e^2_{\text{SWL}}(52)$
simulated	0.003	11.0999	16.6430	11.0996	5.5431
theoretical	0	11.0936	16.6409	11.0936	5.5473

so that

$$\mathbf{w} = \tilde{\mathbf{A}}_1\mathbf{r} + \tilde{\mathbf{A}}_2\mathbf{r}_\eta = (\mathbf{R} - \mathbf{P}\mathbf{R}^{-\eta}\mathbf{P}^H)^{-1}(\mathbf{r} - \mathbf{P}\mathbf{R}^{-\eta}\mathbf{r}_\eta) \quad (49)$$

$$\mathbf{w}_\eta = \tilde{\mathbf{A}}_3\mathbf{r} + \tilde{\mathbf{A}}_4\mathbf{r}_\eta = (\mathbf{R}^\eta - \mathbf{P}^H\mathbf{R}^{-1}\mathbf{P})^{-1}(\mathbf{r}_\eta - \mathbf{P}^H\mathbf{R}^{-1}\mathbf{r}). \quad (50)$$

Therefore, the term $\mathbf{r}^{bH}\mathbf{w}^b$ can now be expanded as

$$\begin{aligned}\mathbf{r}^{bH}\mathbf{w}^b &= \mathbf{r}^H\mathbf{w} + \mathbf{r}_\eta^H\mathbf{w}_\eta \\ &= \mathbf{r}^H\mathbf{R}^{-1}(\mathbf{r} - \mathbf{P}\mathbf{w}_\eta) + \mathbf{r}_\eta^H\mathbf{w}_\eta \\ &= (\mathbf{r}_\eta - \mathbf{P}\mathbf{R}^{-1}\mathbf{r})^H\mathbf{w}_\eta + \mathbf{r}^H\mathbf{R}^{-1}\mathbf{r}. \end{aligned} \quad (51)$$

Finally, upon substituting (50) and (51) into (46), we obtain the expression for the difference in MSE between the SWL and SL estimation in the form

$$\begin{aligned}&\triangle e^2_{\text{SWL}} \\ &\quad = (\mathbf{r}_\eta - \mathbf{P}\mathbf{R}^{-1}\mathbf{r})^H(\mathbf{R}^\eta - \mathbf{P}^H\mathbf{R}^{-1}\mathbf{P})^{-1}(\mathbf{r}_\eta - \mathbf{P}^H\mathbf{R}^{-1}\mathbf{r}). \end{aligned} \quad (52)$$

Remark 4: The term $\triangle e^2_{\text{SWL}}$ in (52) is always nonnegative due to the positive semidefinite nature of $\mathbf{R}^\eta - \mathbf{P}^H\mathbf{R}^{-1}\mathbf{P}$ (see Appendix B for the detailed proof). The equality $\triangle e^2_{\text{SWL}} = 0$ holds only when

$$\mathbf{r}_\eta - \mathbf{P}^H\mathbf{R}^{-1}\mathbf{r} = \mathbf{0} \quad (53)$$

which implies $\mathbf{w}_\eta = \mathbf{0}$ in (50). In other words, for $\mathbf{w}_\eta = \mathbf{0}$, the quaternion SWL model in (16) reduces to the SL model in (5).

We shall now compare the SWL and SL estimators for several special cases of quaternion properness.

1) Jointly $\mathbb{R}^\eta$-Proper Case: Consider the so-called joint $\mathbb{R}^\eta$ properness of both the estimandum y and the regressor $\mathbf{x}^b = [\mathbf{x}^T, \mathbf{x}^{\eta T}]^T$ within the quaternion SWL processing (16), whereby

$$\mathbf{P} = E[\mathbf{x}\mathbf{x}^{\eta H}] = \mathbf{0}, \quad \mathbf{r}_\eta = [\mathbf{x}^\eta y^*] = \mathbf{0} \quad (54)$$

with no specific assumptions imposed on $\mathbf{C}$ and $\mathbf{r}^c$. This jointly $\mathbb{R}^\eta$-proper assumption is therefore sufficient for (53) to hold, and hence $\triangle e^2_{\text{SWL}} = 0$.

2) $\mathbb{R}^\eta$-Proper Regressor: In this case, the $\mathbb{R}^\eta$-proper assumption is valid only for the regressor $\mathbf{x}^b$, characterized by $\mathbf{P} = \mathbf{0}$, with no specific assumption imposed on the estimandum y. Then, the expressions for the optimal weight vectors $\mathbf{w}$ and $\mathbf{w}_\eta$ within SWL estimation in (49) and (50) can be greatly simplified into

$$\mathbf{w} = \mathbf{R}^{-1}\mathbf{r}, \quad \mathbf{w}_\eta = \mathbf{R}^{-\eta}\mathbf{r}_\eta. \quad (55)$$

Note that, $\mathbf{w} = \mathbf{R}^{-1}\mathbf{r}$ remains the same as that in (38), obtained using SL estimation. This is a direct consequence of the $\mathbb{R}^\eta$-properness assumption on the regressor $\mathbf{x}^b = [\mathbf{x}^T, \mathbf{x}^{\eta T}]^T$, which implies that $\mathbf{x}$ and $\mathbf{x}^\eta$ are uncorrelated. Therefore, the Hilbert subspaces generated by $\mathbf{x}$ and $\mathbf{x}^\eta$ are orthogonal, and $\mathbf{x}^\eta$ does not change the term coming from $\mathbf{x}$ only. Even in this case, there still exists a performance advantage of the SWL estimation over the SL estimation, since the MSE difference, $\triangle e^2_{\text{SWL}} = \mathbf{r}_\eta^H\mathbf{R}^{-\eta}\mathbf{r}_\eta$, is strictly positive.

F. Numerical Evaluation

We conducted an illustrative numerical experiment to evaluate our theoretical findings. The experiment was performed in a nonlinear system identification setting, where the system coefficient vector $\mathbf{w}$ to be estimated was a uniformly distributed quaternion vector random variable with length $L = 2$, while the system input vector $\mathbf{x}$ was generated using the quaternion-valued normal distribution, and was $\mathbb{H}$-proper, i.e., all the three pseudocovariance matrices $\mathbf{P}$, $\mathbf{S}$, and $\mathbf{T}$, as well as the semiaugmented pseudocovariance matrix $\mathbf{C}$, vanished. The desired system output was real valued, $y = 4\Re[\mathbf{w}^H\mathbf{x}]$. Although the input vector $\mathbf{x}$ in this case was $\mathbb{H}$ proper, all the cross-correlation vectors $\mathbf{r}$, $\mathbf{r}_\eta$, $\mathbf{r}^b$, and $\mathbf{r}^c$ did exist, therefore, both the conditions in (43) and (53) did not hold, indicating that the WL estimator was accordingly expected to offer performance advantage over the SWL and SL ones. The suitability of the WL estimator for this task was also straightforward to justify, since $y = 4\Re[\mathbf{w}^H\mathbf{x}] = \mathbf{w}^H\mathbf{x} + \mathbf{w}^{\eta H}\mathbf{x}^\eta + \mathbf{w}^{\eta' H}\mathbf{x}^{\eta'} + \mathbf{w}^{\eta'' H}\mathbf{x}^{\eta''}$. The theoretical and simulated MSE performances of all the quaternion estimators as well as their performance differences are given in Table I. The results were obtained by averaging 100 000 independent simulation trials and the close match between both the theoretical and experimental MSE measures for all estimators can be observed.

IV. CONCLUSION

Quaternion-valued learning systems are rapidly emerging; however, the extent to which the WL estimation outperforms various SWL and SL estimators has not been rigorously established. To fill this void in the literature, we have provided a closed-form solution to quantify the degree of the performance advantage, in terms of the MSE, offered by the second-order optimal WL estimation model over the SWL and conventional SL estimation. This has been achieved for the generality of quaternion-valued data (both proper and improper), and by employing the orthogonality conditions between the Hilbert spaces where the estimate and the regressor reside.

APPENDIX A
DETAILED PROOF OF (26)

From (25), we have

$$\mathbf{B}\mathbf{A}_1 + \mathbf{C}\mathbf{A}_3 = \mathbf{I} \quad (56)$$

$$\mathbf{C}^H\mathbf{A}_1 + \mathbf{B}^{\eta'}\mathbf{A}_3 = \mathbf{0} \quad (57)$$

and

$$\mathbf{B}\mathbf{A}_2 + \mathbf{C}\mathbf{A}_4 = \mathbf{0} \quad (58)$$

$$\mathbf{C}^H\mathbf{A}_2 + \mathbf{B}^{\eta'}\mathbf{A}_4 = \mathbf{I}. \quad (59)$$

From the invertibility of $\mathbf{B}$ in (56), we obtain

$$\mathbf{A}_1 = \mathbf{B}^{-1}(\mathbf{I} - \mathbf{C}\mathbf{A}_3) \quad (60)$$

and upon substituting (60) into (57), we have

$$(\mathbf{B}^{\eta'} - \mathbf{C}^H\mathbf{B}^{-1}\mathbf{C})\mathbf{A}_3 = -\mathbf{C}^H\mathbf{B}^{-1}\mathbf{I} \quad (61)$$

while the assumption of invertibility of $\mathbf{B}^{\eta'} - \mathbf{C}^H\mathbf{B}^{-1}\mathbf{C}$ gives

$$\mathbf{A}_3 = -(\mathbf{B}^{\eta'} - \mathbf{C}^H\mathbf{B}^{-1}\mathbf{C})^{-1}\mathbf{C}^H\mathbf{B}^{-1}. \quad (62)$$

Similarly, by assuming that $\mathbf{B}^{-\eta'}$ is invertible, from (57), we obtain

$$\mathbf{A}_3 = -\mathbf{B}^{-\eta'}\mathbf{C}^H\mathbf{A}_1 \quad (63)$$

while a substitution of (63) into (56) gives

$$(\mathbf{B}-\mathbf{C}\mathbf{B}^{-\eta'}\mathbf{C}^H)\mathbf{A}_1=\mathbf{I}. \tag{64}$$

From the invertibility of $\mathbf{B}-\mathbf{C}\mathbf{B}^{-\eta'}\mathbf{C}^H$, we arrive at

$$\mathbf{A}_1=(\mathbf{B}-\mathbf{C}\mathbf{B}^{-\eta'}\mathbf{C}^H)^{-1}. \tag{65}$$

In a similar way, from (58) and (59), we obtain

$$\mathbf{A}_2=-(\mathbf{B}-\mathbf{C}\mathbf{B}^{-\eta'}\mathbf{C}^H)^{-1}\mathbf{C}\mathbf{B}^{-\eta'}, \mathbf{A}_4=(\mathbf{B}^{\eta'}-\mathbf{C}^H\mathbf{B}^{-1}\mathbf{C})^{-1}.$$

Appendix B
Proof of Positive Semidefiniteness of $\mathbf{B}^{\eta'}-\mathbf{C}^H\mathbf{B}^{-1}\mathbf{C}$ and $\mathbf{R}^{\eta}-\mathbf{P}^H\mathbf{R}^{-1}\mathbf{P}$

Consider linear estimation of $\mathbf{y}$ in terms of $\mathbf{x}$ in such a way that $\hat{\mathbf{y}}=\mathbf{M}\mathbf{x}$, where $\mathbf{y}$ and $\mathbf{x}$ are two zero mean random vectors, and $\mathbf{M}$ is the weight matrix. The optimal weight matrix that yields the minimum MSE is governed by the orthogonality principle, so that $E[(\mathbf{y}-\hat{\mathbf{y}})\mathbf{x}^H]=\mathbf{0}$. By defining $\mathbf{\Gamma}_{xx}=E[\mathbf{x}\mathbf{x}^H]$, $\mathbf{\Gamma}_{xy}=E[\mathbf{x}\mathbf{y}^H]$, $\mathbf{\Gamma}_{yx}=E[\mathbf{y}\mathbf{x}^H]$, and $\mathbf{\Gamma}_{yy}=E[\mathbf{y}\mathbf{y}^H]$ and assuming that $\mathbf{\Gamma}_{xx}^{-1}$ exists, the solution is $\mathbf{M}=\mathbf{\Gamma}_{yx}\mathbf{\Gamma}_{xx}^{-1}$. Define $\mathbf{e}=\mathbf{y}-\hat{\mathbf{y}}$ as the estimation error vector, for which the covariance matrix can be written as

$$\mathbf{\Gamma}_{ee}=E[\mathbf{e}\mathbf{e}^H]=E[(\mathbf{y}-\hat{\mathbf{y}})(\mathbf{y}-\hat{\mathbf{y}})^H]. \tag{66}$$

By substituting $\hat{\mathbf{y}}=\mathbf{M}\mathbf{x}$ into (66), $\mathbf{\Gamma}_{ee}$ can be expanded as $\mathbf{\Gamma}_{ee}=\mathbf{\Gamma}_{yy}-\mathbf{\Gamma}_{yx}\mathbf{\Gamma}_{xx}^{-1}\mathbf{\Gamma}_{xy}$, which is positive semidefinite since for any $\mathbf{u}\neq\mathbf{0}$, $\mathbf{u}^H\mathbf{\Gamma}_{ee}\mathbf{u}=E[|\mathbf{e}^H\mathbf{u}|^2]\geq 0$.

Suppose now $\mathbf{y}=\mathbf{x}^{\eta}$. This gives $\mathbf{\Gamma}_{yy}=E[\mathbf{x}^{\eta}\mathbf{x}^{\eta H}]=\mathbf{R}^{\eta}$, $\mathbf{\Gamma}_{yx}=E[\mathbf{x}^{\eta}\mathbf{x}^H]=\mathbf{P}^H$, $\mathbf{\Gamma}_{xx}^{-1}=E[\mathbf{x}\mathbf{x}^H]^{-1}=\mathbf{R}^{-1}$, and $\mathbf{\Gamma}_{xy}=E[\mathbf{x}\mathbf{x}^{\eta H}]=\mathbf{P}$, and hence $\mathbf{\Gamma}_{ee}=\mathbf{R}^{\eta}-\mathbf{P}^H\mathbf{R}^{-1}\mathbf{P}$ is positive semidefinite.

Suppose now $\mathbf{x}=\mathbf{x}^b$ and $\mathbf{y}=\mathbf{x}^{b\eta'}$. This gives $\mathbf{\Gamma}_{yy}=E[\mathbf{x}^{b\eta'}\mathbf{x}^{b\eta' H}]=\mathbf{B}^{\eta'}$, $\mathbf{\Gamma}_{yx}=E[\mathbf{x}^{b\eta'}\mathbf{x}^{bH}]=E[\mathbf{x}^{c}\mathbf{x}^{bH}]=\mathbf{C}^H$, $\mathbf{\Gamma}_{xx}^{-1}=E[\mathbf{x}^b\mathbf{x}^{bH}]^{-1}=\mathbf{B}^{-1}$, and $\mathbf{\Gamma}_{xy}=E[\mathbf{x}^b\mathbf{x}^{b\eta' H}]=E[\mathbf{x}^b\mathbf{x}^{cH}]=\mathbf{C}$, and hence $\mathbf{\Gamma}_{ee}=\mathbf{B}^{\eta'}-\mathbf{C}^H\mathbf{B}^{-1}\mathbf{C}$ is positive semidefinite.

References

[1] P. Arena, L. Fortuna, G. Muscato, and M. G. Xibilia, *Neural Networks in Multidimensional Domains*. New York, NY, USA: Springer-Verlag, 1998.

[2] C. C. Took, G. Strbac, K. Aihara, and D. P. Mandic, "Quaternion-valued short-term joint forecasting of three-dimensional wind and atmospheric parameters," *Renew. Energy*, vol. 36, no. 6, pp. 1754–1760, Jun. 2010.

[3] A. M. Sabatini, "Quaternion-based extended Kalman filter for determining orientation by inertial and magnetic sensing," *IEEE Trans. Biomed. Eng.*, vol. 53, no. 7, pp. 1346–1356, Jul. 2006.

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

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

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

[7] J. Navarro-Moreno, R. M. Fernandez-Alcala, and J. C. Ruiz-Molina, "A quaternion widely linear series expansion and its applications," *IEEE Signal Process. Lett.*, vol. 19, no. 12, pp. 868–871, Dec. 2012.

[8] C. C. Took and D. P. Mandic, "The quaternion LMS algorithm for adaptive filtering of hypercomplex processes," *IEEE Trans. Signal Process.*, vol. 57, no. 4, pp. 1316–1327, Apr. 2009.

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

[10] Y. Xia, C. Jahanchahi, and D. P. Mandic, "Quaternion-valued echo state networks," *IEEE Trans. Neural Netw. Learn. Syst.*, 2014, doi: 10.1109/TNNLS.2014.2320715.

[11] F. D. Neeser and J. L. Massey, "Proper complex random processes with applications to information theory," *IEEE Trans. Inf. Theory*, vol. 39, no. 4, pp. 1293–1302, Jul. 1993.

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

[13] B. Picinbono, "Second-order complex random vectors and normal distributions," *IEEE Trans. Signal Process.*, vol. 44, no. 10, pp. 2637–2640, Oct. 1996.

[14] 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.

[15] P. J. Schreier and L. L. Scharf, *Statistical Signal Processing of Complex-Valued Data: The Theory of Improper and Noncircular Signals*. Cambridge, U.K.: Cambridge Univ. Press, 2010.

[16] N. N. Vakhania, "Random vectors with values in quaternion Hilbert spaces," *Theory Probab. Appl.*, vol. 43, no. 1, pp. 99–115, 1999.

[17] P. O. Amblard and N. L. Bihan, "On properness of quaternion valued random variables," in *Proc. IMA Conf. Math. Signal Process.*, Cirenceister, U.K., Dec. 2004, pp. 23–26.

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

[19] S. C. Olhede, D. Ramírez, and P. J. Schreier, "Detecting directionality in random fields using the monogenic signal," *IEEE Trans. Inf. Theory*, vol. 60, no. 10, pp. 6491–6510, Oct. 2014.

[20] A. Sloin and A. Wiesel, "Proper quaternion Gaussian graphical models," *IEEE Trans. Signal Process.*, vol. 62, no. 20, pp. 5487–5496, Oct. 2014.

[21] 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.

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

[23] J. Via, D. P. Palomar, L. Vielva, and I. Santamaria, "Quaternion ICA from second-order statistics," *IEEE Trans. Signal Process.*, vol. 59, no. 4, pp. 1586–1600, Apr. 2011.

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

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

[26] T. Nitta, "A theoretical foundation for the widely linear processing of quaternion-valued data," *Appl. Math.*, vol. 4, no. 12, pp. 1616–1620, Dec. 2013.

[27] C. Jahanchahi, S. Kanna, and D. Mandic, "A theoretical bound for the performance advantage of quaternion widely linear estimation," in *Proc. 19th Int. Conf. Digit. Signal Process.*, Hong Kong, Aug. 2014, pp. 254–258.

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

[29] B. Picinbono, "On circularity," *IEEE Trans. Signal Process.*, vol. 42, no. 12, pp. 3473–3482, Dec. 1994.