Proper Adaptive Filtering in Four-Dimensional Cayley-Dickson Algebras

Jesús Navarro-Moreno, Rosa María Fernández-Alcalá, José Domingo Jiménez-López and Juan Carlos Ruiz-Molina

Department of Statistics and Operations Research, University of Jaén, 23071 Jaén, Spain.
E-mails: {jnavarro,rmfernan,jdomingo,jcruiz}@ujaen.es.

Abstract

A family of hypercomplex algebras in four dimensions (4D) is proposed to devise adaptive filters. Such a family, called β-quaternions, has multiplication rules for the complex units that depend on a parameter β, and this family contains, as particular cases, both standard Hamilton quaternions and split quaternions. In this framework, two notions of properness for random vectors are introduced and their implications on the statistical processing involved are analyzed. Then, statistical tests to check properness in practice and a method to select the best algebra where the properness conditions could hold are provided. Also, proper adaptive filters are suggested and row and column updating problems are studied. The main advantage of the techniques proposed compared with the standard ones is that a notable reduction in the computational burden is achieved. Finally, simulation examples validate the proper adaptive filters and demonstrate that our scheme performs better than the traditional quaternion domain.

Keywords: Adaptive Filters, β-Quaternions, Cayley-Dickson Algebras, First and Second-Order Properness.

1. Introduction

Hypercomplex algebras based techniques play an increasingly prominent role in the development of solutions for signal processing problems [1, 2, 3, 4]. Hypercomplex algebras generalize the system of complex numbers by incorporating different amounts of imaginary units and different multiplication rules. Notably, the Cayley-Dickson (CD) algebras has gained special importance in signal processing. The CD construction is a recursive process that, by starting out from real numbers, defines higher-dimensional algebras by doubling the previous smaller ones. Complex numbers, standard (Hamilton) quaternions, and octonions are the most common CD algebras due to the fact that they are the only ones that provide normed division systems. However, these requirements are unnecessary in some applications, and the interest in other algebras is increasingly [5, 6, 7, 8, 9]. Among them, 4D hypercomplex algebras are generally sufficient in practical applications of signal processing methods [10].

The great advances in technology, biomedicine, robotics, and environmental sciences have meant that adaptive filtering techniques for multidimensional signals have become highly required and significant. Although the real or complex fields are the dominant frameworks for developing adaptive filtering algorithms [11, 12], hypercomplex algebras have also recently shown their potential usefulness [13, 14, 15, 16]. In this goal, (standard) quaternions have demonstrated their effectiveness in problems related to 3D or 4D signals and they constitute the most frequent hypercomplex algebra used in adaptive filtering practice (see, e.g., [17, 18, 19, 20, 21, 22, 23, 24]).

In the last decade, widely linear (WL) processing has turned out to be the most suitable type of processing in the quaternion domain [25]. However, for quaternion-valued signals with special second-order statistical properties, WL processing reduces to either semi-WL (SWL) or strictly linear (SL) processing, giving rise to algorithms with reduced computational cost [26]. Specifically, SWL and SL processing are used when the quaternion signal presents some properness properties. Quaternion adaptive filtering under properness conditions has been studied, e.g., in [17, 19, 20, 27].

Properness is actually an interesting statistical property that depends on the hypercomplex algebra considered. In this sense, a signal cannot be proper in the quaternion framework, but can be under a different hypercomplex environment. Moreover, multiple 4D CD algebras exist that could be considered as alternatives to quaternions, with each one of them own (different) properties [8]. For example, unlike the quaternion algebra, the split quaternions contain zero divisors, nilpotent elements, and nontrivial idempotents [28]. Thus, the choice of specific algebras that allow for efficient adaptive filtering becomes an important task [1, 2]. In fact, quaternions do not always get the best results possible in terms of performance, as we will see later.

Motivated by the above, a general family of 4D hypercomplex algebras is proposed as analysis framework for devising adaptive filtering algorithms. Such a family, called β-quaternions, has multiplication rules for the complex units depending on a real parameter β and contains, as particular cases, the quaternions ($\beta = -1$) as well as the split quaternions ($\beta = 1$), among others. In this setting, two concepts of properness for random vectors are introduced and the associated statistical processing is analyzed. The methodology proposed shows that properness properties constitute an effective approach to remarkably reduce the computational costs of the algorithms. Then, by taking advantage of the degree of freedom given by parameter β, estimations of β for which it is more probable to achieve the properness conditions are suggested. Thus, parameter β not only offers generality, but also flexibility. Additionally, statistical tests for checking the properness conditions in practice are also developed. Finally, adaptive filters under properness conditions are proposed. For these filters, recursive updating formula for when a new independent variable (row) or a new observation (column) become available are given. It is worth highlighting that the mathematical development of these algorithms has required the introduction of new metrics in β-quaternions and an analysis of their relationships. Simulation examples show the effectiveness of the approaches proposed and the inability of the standard quaternion system to operate accurately in some scenarios.

The main contributions in this paper can be summarized as follows:

(1) Two types of properness for β-quaternion random vectors are defined.

(2) The optimal (WL) estimator is introduced and calculated under the above two properness conditions, showing their computational advantages in these settings.

(3) Statistical tests are provided to check in practice whether data are proper, and a form to estimate β is proposed.

(4) Adaptive filters are studied under the two properness conditions above.

(5) Recursive updating methods when new information is available are suggested.

(6) Simulations of theoretical results confirm that the superior performance of the proposed techniques over the standard quaternion ones, under properness conditions.

The rest of the paper is arranged as follows. Section 2 briefly reviews the main aspects of the β-quaternion domain and introduces the mathematical tools that allow us to define different types of projections useful in adaptive filtering. Section 3 presents the concepts of properness, the associated processing and solutions to some classic estimation problems under properness conditions. Section 4 deals with the tasks of devising proper adaptive filters and their updating when new variables or observations are available. This section also addresses two problems of interest: estimating the best value of β to fulfil properness, if possible, and checking the properness conditions from sample information. Section 5 is devoted to experimental results and comparison with existing approaches. Finally, Section 6 concludes the work. In order to improve readability, all complementary results and technical proofs have been deferred to the Appendixes A–E.

1.1. Notation

Throughout the whole paper, we use boldface uppercase letters to represent matrices, boldface lowercase letters for column vectors, and lightface lowercase letters for scalar quantities. Superscript "T" denotes the transpose and the real part of a β-quaternion will be denoted by $\mathcal{R}\{\cdot\}$. Additionally, the notation $\mathbb{R}$ and $\mathbb{Q}_\beta$ is used to denote the real field and the β-quaternion algebra, respectively. $\mathbf{A} \in \mathbb{R}^{p\times q}$ (respectively, $\mathbf{A} \in \mathbb{Q}_\beta^{p\times q}$) means that $\mathbf{A}$ is a real (respectively, a β-quaternion) $p \times q$ matrix, and similarly $\mathbf{r} \in \mathbb{R}^p$ (respectively, $\mathbf{r} \in \mathbb{Q}_\beta^p$) means that $\mathbf{r}$ is a p-dimensional real (respectively, a β-quaternion) vector. $L_2^4(\Omega)$ denotes the space of second-order random vectors in $\mathbb{R}^4$. δ_{nl} is the Kronecker delta function, which is equal to one if $l = n$, and zero otherwise. $E[\cdot]$ is the statistical expectation operator, $\mathbf{0}_{p\times q}$ denotes the $p \times q$ zero matrix, $\mathbf{I}_p$ stands for the identity matrix of dimension p and "$\otimes$" is the Kronecker product. Furthermore, all the random variables are assumed to have zero-mean.

Finally, Table 1 summarizes the main symbols used in this paper.

2. Review of the β-Quaternion Domain

This section introduces the basic concepts in the β-quaternion domain and establishes some properties of interest.

Definition 1. *A random variable $x \in \mathbb{Q}_\beta$ is defined as $x = a + \mathrm{i}b + \mathrm{j}c + \mathrm{k}d$ where $a, b, c, d \in \mathbb{R}$ are random variables and the imaginary units* $(\mathrm{i}, \mathrm{j}, \mathrm{k})$ *satisfy the following multiplication rules:*

1	i	j	k
i	-*1*	k	−j
j	-k	β	-βi
k	j	βi	β

with $\beta \in \mathbb{R} - \{0\}$.

Two well-known 4D hypercomplex algebras are obtained as particular cases of β-quaternions: the standard (Hamilton) quaternion algebra when $\beta = -1$ and

Table 1: Summary of used symbols

Symbol	Meaning
$\mathbf{x}^{\nu^{\mathtt{T}}}$	Transpose of $\mathbf{x}^{\nu}$, $\nu = \mathrm{i}, \mathrm{j}, \mathrm{k}$
$\mathbf{x}^{\mathtt{H}_\alpha}$	α-Hermitian transpose, $\mathbf{x}^{(\alpha)^{\mathtt{T}}}$, of $\mathbf{x}$
$\mathbf{x}_r$	Real vector formed with the components of $\mathbf{x} \in \mathbb{Q}_\beta^p$
$\bar{\mathbf{x}}$	Augmented β-quaternion vector formed by $\mathbf{x} \in \mathbb{Q}_\beta^p$ and its auxiliary vectors $\mathbf{x}^{\mathrm{i}}$,$\mathbf{x}^{\mathrm{j}}$, $\mathbf{x}^{\mathrm{k}}$
$\mathbf{A}^{\mathtt{H}_\alpha}$	α-Hermitian transpose, $\mathbf{A}^{(\alpha)^{\mathtt{T}}}$, of a matrix $\mathbf{A} \in \mathbb{Q}_\beta^{p\times q}$
$\boldsymbol{\Upsilon}_{\mathbf{xy}}(\alpha,\beta)$	$E[\mathbf{x}\mathbf{y}^{\mathtt{H}_\alpha}]$, for $\mathbf{x} \in \mathbb{Q}_\beta^p$, $\mathbf{y} \in \mathbb{Q}_\beta^q$, $\alpha, \beta \in \mathbb{R} - \{0\}$
$\boldsymbol{\Upsilon}_{\mathbf{x}}(\alpha,\beta)$	$E[\mathbf{x}\mathbf{x}^{\mathtt{H}_\alpha}]$, for $\mathbf{x} \in \mathbb{Q}_\beta^p$, $\alpha, \beta \in \mathbb{R} - \{0\}$
$\boldsymbol{\Upsilon}^{\nu}_{\mathbf{xy}}(\beta)$	$E[\mathbf{x}\mathbf{y}^{\mathtt{H}_{3-2\nu}}]$, for $\mathbf{x} \in \mathbb{Q}_\beta^p$, $\mathbf{y} \in \mathbb{Q}_\beta^q$, $\beta \in \mathbb{R} - \{0\}$, $\nu = 1, 2$
$\boldsymbol{\Upsilon}^{\nu}_{\mathbf{x}}(\beta)$	$E[\mathbf{x}\mathbf{x}^{\mathtt{H}_{3-2\nu}}]$, for $\mathbf{x} \in \mathbb{Q}_\beta^p$, $\beta \in \mathbb{R} - \{0\}$, $\nu = 1, 2$
$\boldsymbol{\Upsilon}^{3}_{\mathbf{xy}}(\beta)$	$E[\mathbf{x}\mathbf{y}^{\mathtt{H}_\beta}]$, for $\mathbf{x} \in \mathbb{Q}_\beta^p$, $\mathbf{y} \in \mathbb{Q}_\beta^q$, $\beta \in \mathbb{R} - \{0\}$
$\boldsymbol{\Upsilon}^{3}_{\mathbf{x}}(\beta)$	$E[\mathbf{x}\mathbf{x}^{\mathtt{H}_\beta}]$, for $\mathbf{x} \in \mathbb{Q}_\beta^p$, $\beta \in \mathbb{R} - \{0\}$
$\hat{\boldsymbol{\Upsilon}}^{\nu}_{\mathbf{x}}(\beta)$	Sample autocorrelation matrix of $\mathbf{x} \in \mathbb{Q}_\beta^p$, for $\beta \in \mathbb{R} - \{0\}$, $\nu = 1, 2$
$\langle x, y\rangle_1$	$E[xy^{(1)}]$, for $x, y \in \mathbb{Q}_\beta$, $\beta > 0$
$\langle x, y\rangle_2$	$E[xy^{(-1)}]$, for $x, y \in \mathbb{Q}_\beta$, $\beta < 0$
$\langle x, y\rangle_3$	$E[xy^{(\beta)}]$, for $x, y \in \mathbb{Q}_\beta$, $\beta \neq 0$
$\prec \mathbf{x}, \mathbf{y} \succ_1$	$\mathbf{x}^{\mathtt{T}}\mathbf{y}^{(1)}$, for $\mathbf{x}, \mathbf{y} \in \mathbb{Q}_\beta^p$, $\beta > 0$
$\prec \mathbf{x}, \mathbf{y} \succ_2$	$\mathbf{x}^{\mathtt{T}}\mathbf{y}^{(-1)}$, for $\mathbf{x}, \mathbf{y} \in \mathbb{Q}_\beta^p$, $\beta < 0$
$\mathcal{G}_{\mathbf{x}}$	Closed linear subspace associated to $\mathbf{x} \in \mathbb{Q}_\beta^p$
$\mathcal{C}_{\mathcal{A}}$	Closed linear subspace associated to $\mathcal{A} = \{\mathbf{x}_1, \ldots, \mathbf{x}_m\}$ with $\mathbf{x}_i \in \mathbb{Q}_\beta^p$
$\mathcal{C}_{\mathcal{A}_p}$	Closed linear subspace associated to $\mathcal{A}_p = \{\boldsymbol{\chi}_1, \ldots, \boldsymbol{\chi}_p\}$ with $\boldsymbol{\chi}_i \in \mathbb{Q}_\beta^n$
$\mathcal{C}_{\dot{\mathcal{A}}_p}$	Closed linear subspace associated to $\dot{\mathcal{A}}_p = \{\boldsymbol{\chi}_1, \ldots, \boldsymbol{\chi}_p, \boldsymbol{\chi}_1^k, \ldots, \boldsymbol{\chi}_p^k\}$ with $\boldsymbol{\chi}_i \in \mathbb{Q}_\beta^n$
$\tilde{\boldsymbol{\zeta}}_\nu^{\mathtt{Q}_1}$	Projection of $\boldsymbol{\zeta} \in \mathbb{Q}_\beta^n$ onto the set $\mathcal{C}_{\mathcal{A}_p}$ under the metrics d_ν, $\nu = 1, 2$
$\tilde{\boldsymbol{\zeta}}_\nu^{\mathtt{Q}_2}$	Projection of $\boldsymbol{\zeta} \in \mathbb{Q}_\beta^n$ onto the set $\mathcal{C}_{\dot{\mathcal{A}}_p}$ under the metrics d_ν, $\nu = 1, 2$
$\tilde{\xi}_\nu^{\mathtt{Q}_i}$	Error associated to $\tilde{\boldsymbol{\zeta}}_\nu^{\mathtt{Q}_i}$, $i = 1, 2$, $\nu = 1, 2$
$\hat{y}^{\mathtt{Q}_i}$	$\mathtt{Q}_i$ estimator of $y \in \mathbb{Q}_\beta$ respect to $\mathbf{x} \in \mathbb{Q}_\beta^p$, $i = 1, 2$
$\epsilon^{\mathtt{Q}_i}$	Error associated to $\hat{y}^{\mathtt{Q}_i}$, $i = 1, 2$
$\tilde{\epsilon}^{\mathtt{Q}_i}_{\nu,\beta_0}$	Aproximate error associated to $\hat{y}^{\mathtt{Q}_i} \in \mathbb{Q}_{\beta_0}$, $i = 1, 2$, and $\nu = 1, 2$
$\hat{y}^{\mathtt{QWL}}$	QWL estimator of $y \in \mathbb{Q}_\beta$ respect to $\mathbf{x} \in \mathbb{Q}_\beta^p$
$\epsilon^{\mathtt{QWL}}$	Error associated to $\hat{y}^{\mathtt{QWL}}$

the split quaternions if $\beta = 1$ [8, 29]. Obviously, β-quaternion multiplication cannot satisfy the commutative law, although this product is associative and distributive over addition.

The following auxiliary β-quaternions will be necessary:

$$\begin{aligned} x^{(\alpha)} &= a - \mathrm{i}b + \mathrm{j}\frac{c}{\alpha} + \mathrm{k}\frac{d}{\alpha}, \qquad \alpha \in \mathbb{R} - \{0\} \\ x^{\mathrm{i}} &= a + \mathrm{i}b - \mathrm{j}c - \mathrm{k}d \\ x^{\mathrm{j}} &= a - \mathrm{i}b + \mathrm{j}c - \mathrm{k}d \\ x^{\mathrm{k}} &= a - \mathrm{i}b - \mathrm{j}c + \mathrm{k}d \end{aligned} \tag{1}$$

The inverse of $x \in \mathbb{Q}_\beta$ is $x^{-1} = \dfrac{x^{(-1)}}{xx^{(-1)}}$.

The auxiliary β-quaternions satisfies the next properties.

Property 1. *For $x, y \in \mathbb{Q}_\beta$ it follows that*

1. $(x^\nu)^\nu = x, \quad \nu = \mathrm{i}, \mathrm{j}, \mathrm{k}$
2. $(x^{\nu_1})^{\nu_2} = x^{\nu_3}, \quad \nu_1, \nu_2, \nu_3 = \mathrm{i}, \mathrm{j}, \mathrm{k}, \quad \nu_1 \neq \nu_2 \neq \nu_3$
3. $x^\nu y^\nu = (xy)^\nu, \quad \nu = \mathrm{i}, \mathrm{j}, \mathrm{k}$
4. $(x^\nu)^{-1} = (x^{-1})^\nu, \quad \nu = \mathrm{i}, \mathrm{j}, \mathrm{k}$
5. $(xy)^{(\alpha)} = y^{(\alpha)} x^{(\alpha)}$, *with* $\alpha = 1, -1$

Now, we extend some concepts to the vector setting. Consider a random vector $\mathbf{x} = [x_1, \ldots, x_p]^{\mathtt{T}} \in \mathbb{Q}_\beta^p$ where $x_i = a_i + \mathrm{i}b_i + \mathrm{j}c_i + \mathrm{k}d_i$, with a_i, b_i, c_i and d_i, $i = 1, \ldots, p$, real random variables. Then, the auxiliary vectors are: $\mathbf{x}^\nu = [x_1^\nu, \ldots, x_p^\nu]^{\mathtt{T}}$, for $\nu = \mathrm{i}, \mathrm{j}, \mathrm{k}$, and $\mathbf{x}^{(\alpha)} = [x_1^{(\alpha)}, \ldots, x_p^{(\alpha)}]^{\mathtt{T}}$, with $\alpha \neq 0$. We denote the α-Hermitian transpose of $\mathbf{x}$ as $\mathbf{x}^{\mathtt{H}_\alpha} \triangleq \mathbf{x}^{(\alpha)^{\mathtt{T}}}$. In a similar way, the α-Hermitian transpose of a matrix $\mathbf{A} \in \mathbb{Q}_\beta^{p \times q}$ can be defined as $\mathbf{A}^{\mathtt{H}_\alpha} = \mathbf{A}^{(\alpha)^{\mathtt{T}}}$. The inverse of β-quaternion matrices is briefly studied in Appendix B.

Consider two random vectors $\mathbf{x} \in \mathbb{Q}_\beta^p$ and $\mathbf{y} \in \mathbb{Q}_\beta^q$, we define the function:

$$\boldsymbol{\Upsilon}_{\mathbf{xy}}(\alpha, \beta) = E[\mathbf{x}\mathbf{y}^{\mathtt{H}_\alpha}], \qquad \alpha, \beta \in \mathbb{R} - \{0\} \tag{2}$$

and we denote $\boldsymbol{\Upsilon}_{\mathbf{x}}(\alpha,\beta) \triangleq \boldsymbol{\Upsilon}_{\mathbf{xx}}(\alpha,\beta)$. Three special cases are of interest:

$$\boldsymbol{\Upsilon}^1_{\mathbf{xy}}(\beta) \triangleq \boldsymbol{\Upsilon}_{\mathbf{xy}}(1,\beta) \qquad \boldsymbol{\Upsilon}^2_{\mathbf{xy}}(\beta) \triangleq \boldsymbol{\Upsilon}_{\mathbf{xy}}(-1,\beta) \qquad \boldsymbol{\Upsilon}^3_{\mathbf{xy}}(\beta) \triangleq \boldsymbol{\Upsilon}_{\mathbf{xy}}(\beta,\beta) \quad (3)$$

Similarly, when $\mathbf{x}$ and $\mathbf{y}$ coincide, we use the notation $\boldsymbol{\Upsilon}^\nu_{\mathbf{x}}(\beta) \triangleq \boldsymbol{\Upsilon}^\nu_{\mathbf{xx}}(\beta)$, $\nu = 1,2,3$.

The real vector associated to $\mathbf{x} \in \mathbb{Q}^p_\beta$ is

$$\mathbf{x}_r = [\mathbf{a}^{\mathrm{T}}, \mathbf{b}^{\mathrm{T}}, \mathbf{c}^{\mathrm{T}}, \mathbf{d}^{\mathrm{T}}]^{\mathrm{T}} \quad (4)$$

with $\mathbf{a} = [a_1,\ldots,a_p]^{\mathrm{T}}$, $\mathbf{b} = [b_1,\ldots,b_p]^{\mathrm{T}}$, $\mathbf{c} = [c_1,\ldots,c_p]^{\mathrm{T}}$, and $\mathbf{d} = [d_1,\ldots,d_p]^{\mathrm{T}}$. Furthermore, the augmented vector of $\mathbf{x}$ is

$$\bar{\mathbf{x}} = [\mathbf{x}^{\mathrm{T}}, \mathbf{x}^{\mathrm{k}\,\mathrm{T}}, \mathbf{x}^{\mathrm{i}\,\mathrm{T}}, \mathbf{x}^{\mathrm{j}\,\mathrm{T}}]^{\mathrm{T}} \quad (5)$$

The following relationship between the augmented vector and the real vector can be established

$$\bar{\mathbf{x}} = \mathcal{T}_p \mathbf{x}_r \quad (6)$$

where $\mathcal{T}_p = \mathcal{B} \otimes \mathbf{I}_p$ with

$$\mathcal{B} = \begin{pmatrix} 1 & \mathrm{i} & \mathrm{j} & \mathrm{k} \\ 1 & -\mathrm{i} & -\mathrm{j} & \mathrm{k} \\ 1 & \mathrm{i} & -\mathrm{j} & -\mathrm{k} \\ 1 & -\mathrm{i} & \mathrm{j} & -\mathrm{k} \end{pmatrix}$$

From Property 1, it is easy to prove the following result.

Property 2.

$$\boldsymbol{\Upsilon}_{\bar{\mathbf{x}}}(\alpha,\beta) = \begin{pmatrix} \boldsymbol{\Upsilon}_{\mathbf{x}}(\alpha,\beta) & \boldsymbol{\Upsilon}_{\mathbf{xx}^{\mathrm{k}}}(\alpha,\beta) & \boldsymbol{\Upsilon}_{\mathbf{xx}^{\mathrm{i}}}(\alpha,\beta) & \boldsymbol{\Upsilon}_{\mathbf{xx}^{\mathrm{j}}}(\alpha,\beta) \\ \boldsymbol{\Upsilon}^{\mathrm{k}}_{\mathbf{xx}^{\mathrm{k}}}(\alpha,\beta) & \boldsymbol{\Upsilon}^{\mathrm{k}}_{\mathbf{x}}(\alpha,\beta) & \boldsymbol{\Upsilon}^{\mathrm{k}}_{\mathbf{xx}^{\mathrm{j}}}(\alpha,\beta) & \boldsymbol{\Upsilon}^{\mathrm{k}}_{\mathbf{xx}^{\mathrm{i}}}(\alpha,\beta) \\ \boldsymbol{\Upsilon}^{\mathrm{i}}_{\mathbf{xx}^{\mathrm{i}}}(\alpha,\beta) & \boldsymbol{\Upsilon}^{\mathrm{i}}_{\mathbf{xx}^{\mathrm{j}}}(\alpha,\beta) & \boldsymbol{\Upsilon}^{\mathrm{i}}_{\mathbf{x}}(\alpha,\beta) & \boldsymbol{\Upsilon}^{\mathrm{i}}_{\mathbf{xx}^{\mathrm{k}}}(\alpha,\beta) \\ \boldsymbol{\Upsilon}^{\mathrm{j}}_{\mathbf{xx}^{\mathrm{j}}}(\alpha,\beta) & \boldsymbol{\Upsilon}^{\mathrm{j}}_{\mathbf{xx}^{\mathrm{i}}}(\alpha,\beta) & \boldsymbol{\Upsilon}^{\mathrm{j}}_{\mathbf{xx}^{\mathrm{k}}}(\alpha,\beta) & \boldsymbol{\Upsilon}^{\mathrm{j}}_{\mathbf{x}}(\alpha,\beta) \end{pmatrix} \quad (7)$$

The functions $\boldsymbol{\Upsilon}_{\mathbf{x}}(\alpha,\beta)$ and $\boldsymbol{\Upsilon}_{\mathbf{xx}^\nu}(\alpha,\beta)$, $\nu = \mathrm{i},\mathrm{j},\mathrm{k}$, can be obtained from Table A.7 (see Appendix A) in terms of the correlation functions (denoted by Υ) of the components of $\mathbf{x}_r$ (4).

Remark 1. *By setting* α *to 1, -1 or* β *in* (7)*, we get the corresponding matrices for* $\mathbf{\Upsilon}^i_{\bar{\mathbf{x}}}(\beta)$*,* $i = 1, 2, 3$*, respectively. Also, following the same procedure in Table A.7, we obtain the expressions for* $\mathbf{\Upsilon}^i_{\mathbf{x}}(\beta)$ *and* $\mathbf{\Upsilon}^i_{\mathbf{x}\mathbf{x}^\nu}(\beta)$*,* $i = 1, 2, 3$*,* $\nu = \mathrm{i}, \mathrm{j}, \mathrm{k}$*, in terms of the correlation functions of the components of* $\mathbf{x}_r$*.*

2.1. Some Concepts in Approximation Theory

A fundamental problem in signal processing is obtaining the projection of a desired random variable from a set of measurements. In order to successfully carry out this task, the choice of appropriate metrics is a crucial matter. Thus, we aim to introduce some metrics of great significance for our work. Two kinds of metrics are analyzed: stochastic and deterministic. We start with the first type.

Let $x, y \in \mathbb{Q}_\beta$ be random variables, we define the products

$$\begin{aligned} \langle x, y\rangle_1 &= E[xy^{(1)}], \quad \forall\beta > 0 \\ \langle x, y\rangle_2 &= E[xy^{(-1)}], \quad \forall\beta < 0 \\ \langle x, y\rangle_3 &= E[xy^{(\beta)}], \quad \forall\beta \neq 0 \end{aligned} \tag{8}$$

with $y^{(1)}$, $y^{(-1)}$ and $y^{(\beta)}$ given in (1).

Property 3. *For any random variables* $x, y, u, v \in \mathbb{Q}_\beta$ *and deterministic coefficients* $\lambda, \alpha, \rho, \theta \in \mathbb{Q}_\beta$*,* $\gamma, \delta \in \mathbb{R}$*, it follows that*

1. $\langle x, y\rangle_\nu^{(3-2\nu)} = \langle y, x\rangle_\nu, \qquad \nu = 1, 2$

2. $\langle \lambda x + \alpha y, \gamma u + \delta v\rangle_\nu = \lambda\gamma\langle x, u\rangle_\nu + \lambda\delta\langle x, v\rangle_\nu + \alpha\gamma\langle y, u\rangle_\nu + \alpha\delta\langle y, v\rangle_\nu, \quad \nu = 1, 2, 3$

3. $\langle \lambda x + \alpha y, \rho u + \theta v\rangle_\nu = \lambda\langle x, u\rangle_\nu\rho^{(3-2\nu)} + \lambda\langle x, v\rangle_\nu\theta^{(3-2\nu)} + \alpha\langle y, u\rangle_\nu\rho^{(3-2\nu)} + \alpha\langle y, v\rangle_\nu\theta^{(3-2\nu)}, \quad \nu = 1, 2$

4. $\langle x, x\rangle_\nu = 0 \Leftrightarrow x = 0, \quad \nu = 1, 2, 3$

5. $\langle x, y\rangle_\nu = 0 \Leftrightarrow \langle y, x\rangle_\nu = 0, \quad \nu = 1, 2, 3$

Definition 2. *Let* $x, y \in \mathbb{Q}_\beta$ *be random variables, then we define the following metrics:*

$$d_\nu(x, y) = \|x - y\|_\nu, \qquad \nu = 1, 2, 3 \tag{9}$$

where $\|x\|_\nu^2 = \mathcal{R}\{\langle x, x\rangle_\nu\}$, $\nu = 1, 2, 3$.

Given a random vector $\mathbf{x} \in \mathbb{Q}_\beta^p$, the closed linear subspace $\mathcal{G}_\mathbf{x}$ associated to $\mathbf{x}$ is the set of elements of the form $\sum_{i=1}^p \alpha_i x_i$, with x_i the ith component of $\mathbf{x}$ and $\alpha_i \in \mathbb{Q}_\beta$ deterministic coefficients. Following a reasoning similar to [7], the existence and uniqueness of the projection of a given random variable $y \in \mathbb{Q}_\beta$ onto the set $\mathcal{G}_\mathbf{x}$ under the metrics (9) is assured. This projection will be denoted by $\hat{y}_\nu$, $\nu = 1, 2, 3$. Moreover, $\hat{y}_\nu$ is the projection of y onto the set $\mathcal{G}_\mathbf{x}$ if, and only if, $\langle y - \hat{y}_\nu, x\rangle_\nu = 0$, $\forall x \in \mathcal{G}_\mathbf{x}$. As a consequence, if $\{\vartheta_1, \ldots, \vartheta_r\}$ is a basis of $\mathcal{G}_\mathbf{x}$ then,

$$\hat{y}_\nu = \sum_{i=1}^r h_i \vartheta_i \tag{10}$$

where the deterministic coefficients $h_i \in \mathbb{Q}_\beta$ are calculated from the system

$$\langle y, \vartheta_j\rangle_\nu = \sum_{i=1}^r h_i \langle \vartheta_i, \vartheta_j\rangle_\nu, \qquad j = 1, \ldots, r, \qquad \nu = 1, 2, 3 \tag{11}$$

The projections $\hat{y}_\nu$ are different for $\nu = 1, 2, 3$. Actually, $\hat{y}_3$ is the most interesting one since the metric space of β-quaternion random variables $(\mathbb{Q}_\beta, d_3)$ is isomorphic to the space of second-order random vectors in $\mathbb{R}^4$, $L_2^4(\Omega)$. Such an isomorphism is determined by the relation (6). However, the product $\langle\cdot,\cdot\rangle_3$ does not satisfy Properties 3.1 and 3.3 which could be a possible source of difficulties, especially in theoretical derivations. Taking into account these shortcomings, the products $\langle\cdot,\cdot\rangle_\nu$, $\nu = 1, 2$, are more convenient to use than $\langle\cdot,\cdot\rangle_3$. A detailed analysis and discussion will be given in Subsection 3.2.

The second type of metrics has a deterministic nature.

Definition 3. *Let* $\mathbf{x}, \mathbf{y} \in \mathbb{Q}_\beta^p$ *be two deterministic vectors. We define the following products:*

$$\begin{aligned} \prec \mathbf{x}, \mathbf{y} \succ_1 &= \mathbf{x}^T \mathbf{y}^{(1)}, \quad \forall \beta > 0 \\ \prec \mathbf{x}, \mathbf{y} \succ_2 &= \mathbf{x}^T \mathbf{y}^{(-1)}, \quad \forall \beta < 0 \end{aligned} \tag{12}$$

with $\mathbf{y}^{(\alpha)}$, $\alpha = 1, -1$, *defined in* (1).

Property 4. *For any deterministic vectors* $\mathbf{x}, \mathbf{y}, \mathbf{u}, \mathbf{v} \in \mathbb{Q}_\beta^p$ *and deterministic coefficients* $\lambda, \alpha, \rho, \theta \in \mathbb{Q}_\beta$, *it follows, for* $\nu = 1, 2$, *that*

1. $\prec \mathbf{x}, \mathbf{y} \succ_\nu^{(3-2\nu)} = \prec \mathbf{y}, \mathbf{x} \succ_\nu$
2. $\prec \lambda\mathbf{x} + \alpha\mathbf{y}, \rho\mathbf{u} + \theta\mathbf{v} \succ_\nu = \lambda \prec \mathbf{x}, \mathbf{u} \succ_\nu \rho^{(3-2\nu)} + \lambda \prec \mathbf{x}, \mathbf{v} \succ_\nu \theta^{(3-2\nu)}$ $+ \alpha \prec \mathbf{y}, \mathbf{u} \succ_\nu \rho^{(3-2\nu)} + \alpha \prec \mathbf{y}, \mathbf{v} \succ_\nu \theta^{(3-2\nu)}$
3. $\prec \mathbf{x}, \mathbf{x} \succ_\nu = 0 \Leftrightarrow \mathbf{x} = \mathbf{0}_{p\times 1}$

Definition 4. *Let* $\mathbf{x}, \mathbf{y} \in \mathbb{Q}_\beta^p$ *be two deterministic vectors. We define the metric*

$$d_\nu(\mathbf{x}, \mathbf{y}) = |\mathbf{x} - \mathbf{y}|_\nu, \qquad \nu = 1, 2 \tag{13}$$

where $|\mathbf{x}|_\nu^2 = \mathcal{R}\{\prec \mathbf{x}, \mathbf{x} \succ_\nu\}$, $\nu = 1, 2$.

Consider a set of deterministic vectors $\mathcal{A} = \{\mathbf{x}_1, \dots, \mathbf{x}_m\}$ with $\mathbf{x}_i \in \mathbb{Q}_\beta^p$. We define the closed linear subspace $\mathcal{C}_\mathcal{A}$ associated to $\mathcal{A}$ as the set whose elements are of the form $\sum_{i=1}^m \alpha_i \mathbf{x}_i$ with $\alpha_i \in \mathbb{Q}_\beta$ deterministic coefficients. Let $\mathbf{y} \in \mathbb{Q}_\beta^p$ be a deterministic vector then, the projection of $\mathbf{y}$ onto the set $\mathcal{C}_\mathcal{A}$ under the metrics (13) is denoted by $\tilde{\mathbf{y}}_\nu$, $\nu = 1, 2$. The existence and uniqueness of such projections is proved in a similar way as indicated above. Moreover, $\tilde{\mathbf{y}}_\nu$ is the projection of $\mathbf{y}$ onto the set $\mathcal{C}_\mathcal{A}$ if, and only if, $\prec \mathbf{y} - \tilde{\mathbf{y}}_\nu, \mathbf{x} \succ_\nu = 0$, $\forall \mathbf{x} \in \mathcal{C}_\mathcal{A}$. As a consequence, if $\{\boldsymbol{\vartheta}_1, \dots, \boldsymbol{\vartheta}_r\}$ is a basis of $\mathcal{C}_\mathcal{A}$ then,

$$\tilde{\mathbf{y}}_\nu = \sum_{i=1}^r h_i \boldsymbol{\vartheta}_i \tag{14}$$

where the deterministic coefficients $h_i \in \mathbb{Q}_\beta$ are obtained from the system

$$\prec \mathbf{y}, \boldsymbol{\vartheta}_j \succ_\nu = \sum_{i=1}^r h_i \prec \boldsymbol{\vartheta}_i, \boldsymbol{\vartheta}_j \succ_\nu, \qquad j = 1, \dots, r, \qquad \nu = 1, 2$$

3. Theoretical Proper Framework

In general, the aforementioned isomorphism between the spaces $(\mathbb{Q}_\beta, d_3)$ and $L_2^4(\Omega)$ implies the equivalence between the real processing and the WL β-quaternion processing. The WL processing requires to operate with augmented vectors $\bar{\mathbf{x}}$ and augmented second-order statistics $\boldsymbol{\Upsilon}_{\bar{\mathbf{x}}}(\alpha,\beta)$ given by (5) and (7), respectively. However, by imposing some restrictions on $\boldsymbol{\Upsilon}_{\bar{\mathbf{x}}}(\alpha,\beta)$ we can reduce the computational burden in the adaptive filtering algorithms as we will see later. This idea leads to the theoretical proper framework in which we develop the β-quaternion signal processing.

3.1. First and Second-Order Properness

In this section, two kind of properness are introduced: first and second-order properness which generalize the concepts of $\mathbb{Q}$-properness and $\mathbb{C}^{\mathrm{k}}$-properness, respectively, to the β-quaternion setting.

Definition 5. *A random vector $\mathbf{x} \in \mathbb{Q}_{\beta}^{p}$ is said to be $P_{\beta_0}^1$-proper (respectively, $N_{\beta_0}^1$-proper) if there exists a value $\beta_0 > 0$ (respectively, $\beta_0 < 0$) such that $\boldsymbol{\Upsilon}_{\mathbf{x}\mathbf{x}^{\nu}}^1(\beta_0)$ (respectively, $\boldsymbol{\Upsilon}_{\mathbf{x}\mathbf{x}^{\nu}}^2(\beta_0)$), $\nu = \mathrm{i}, \mathrm{j}, \mathrm{k}$, vanish.*

In like manner, two random vectors $\mathbf{x} \in \mathbb{Q}_{\beta}^{p_1}$ and $\mathbf{y} \in \mathbb{Q}_{\beta}^{p_2}$ are cross $P_{\beta_0}^1$-proper (respectively, cross $N_{\beta_0}^1$-proper) if there exists a value $\beta_0 > 0$ (respectively, $\beta_0 < 0$) such that $\boldsymbol{\Upsilon}_{\mathbf{x}\mathbf{y}^{\nu}}^1(\beta_0)$ (respectively, $\boldsymbol{\Upsilon}_{\mathbf{x}\mathbf{y}^{\nu}}^2(\beta_0)$), $\nu = \mathrm{i}, \mathrm{j}, \mathrm{k}$, vanish.

By applying the results in Table A.7, the first-order properness conditions can be characterized in terms of the components of the real vector (4) (see Table 2). Next, we introduce the notions associated with the second-order properness.

$P_{\beta_0}^1$-properness conditions	$N_{\beta_0}^1$-properness conditions
$\beta_0 > 0$	$\beta_0 < 0$
$\boldsymbol{\Upsilon}_{\mathbf{a}} = \boldsymbol{\Upsilon}_{\mathbf{b}} = \beta_0 \boldsymbol{\Upsilon}_{\mathbf{c}}, \quad \boldsymbol{\Upsilon}_{\mathbf{c}} = \boldsymbol{\Upsilon}_{\mathbf{d}}$	$\boldsymbol{\Upsilon}_{\mathbf{a}} = \boldsymbol{\Upsilon}_{\mathbf{b}} = -\beta_0 \boldsymbol{\Upsilon}_{\mathbf{c}}, \quad \boldsymbol{\Upsilon}_{\mathbf{c}} = \boldsymbol{\Upsilon}_{\mathbf{d}}$
$-\boldsymbol{\Upsilon}_{\mathbf{ab}}^{\mathsf{T}} = \boldsymbol{\Upsilon}_{\mathbf{ab}} = \beta_0 \boldsymbol{\Upsilon}_{\mathbf{cd}}, \quad -\boldsymbol{\Upsilon}_{\mathbf{cd}}^{\mathsf{T}} = \boldsymbol{\Upsilon}_{\mathbf{cd}}$	$\boldsymbol{\Upsilon}_{\mathbf{ab}}^{\mathsf{T}} = -\boldsymbol{\Upsilon}_{\mathbf{ab}} = \beta_0 \boldsymbol{\Upsilon}_{\mathbf{cd}}, \quad -\boldsymbol{\Upsilon}_{\mathbf{cd}}^{\mathsf{T}} = \boldsymbol{\Upsilon}_{\mathbf{cd}}$
$\boldsymbol{\Upsilon}_{\mathbf{ac}}^{\mathsf{T}} = \boldsymbol{\Upsilon}_{\mathbf{ac}} = -\boldsymbol{\Upsilon}_{\mathbf{bd}} = -\boldsymbol{\Upsilon}_{\mathbf{bd}}^{\mathsf{T}}$	$-\boldsymbol{\Upsilon}_{\mathbf{ac}}^{\mathsf{T}} = \boldsymbol{\Upsilon}_{\mathbf{ac}} = -\boldsymbol{\Upsilon}_{\mathbf{bd}} = \boldsymbol{\Upsilon}_{\mathbf{bd}}^{\mathsf{T}}$
$\boldsymbol{\Upsilon}_{\mathbf{ad}}^{\mathsf{T}} = \boldsymbol{\Upsilon}_{\mathbf{ad}} = \boldsymbol{\Upsilon}_{\mathbf{bc}} = \boldsymbol{\Upsilon}_{\mathbf{bc}}^{\mathsf{T}}$	$-\boldsymbol{\Upsilon}_{\mathbf{ad}}^{\mathsf{T}} = \boldsymbol{\Upsilon}_{\mathbf{ad}} = \boldsymbol{\Upsilon}_{\mathbf{bc}} = -\boldsymbol{\Upsilon}_{\mathbf{bc}}^{\mathsf{T}}$

Table 2: First-order properness conditions.

Definition 6. *A random vector* $\mathbf{x} \in \mathbb{Q}_\beta^p$ *is said to be* $P^2_{\beta_0}$*-proper (respectively,* $N^2_{\beta_0}$*-proper) if there exists a value* $\beta_0 > 0$ *(respectively,* $\beta_0 < 0$*) such that the* $\boldsymbol{\Upsilon}^1_{\mathbf{x}\mathbf{x}^\nu}(\beta_0)$ *(respectively,* $\boldsymbol{\Upsilon}^2_{\mathbf{x}\mathbf{x}^\nu}(\beta_0)$*),* $\nu = \mathrm{i}, \mathrm{j}$*, vanish.*

In like manner, two random vectors $\mathbf{x} \in \mathbb{Q}_\beta^{p_1}$ *and* $\mathbf{y} \in \mathbb{Q}_\beta^{p_2}$ *are cross* $P^2_{\beta_0}$*-proper (respectively, cross* $N^2_{\beta_0}$*-proper) if there exists a value* $\beta_0 > 0$ *(respectively,* $\beta_0 < 0$*) such that the* $\boldsymbol{\Upsilon}^1_{\mathbf{x}\mathbf{y}^\nu}(\beta_0)$ *(respectively,* $\boldsymbol{\Upsilon}^2_{\mathbf{x}\mathbf{y}^\nu}(\beta_0)$*),* $\nu = \mathrm{i}, \mathrm{j}$*, vanish.*

Similarly, in Table 3 we give the conditions on the real vector that characterizes the second-order properness.

$P^2_{\beta_0}$-properness conditions	$N^2_{\beta_0}$-properness conditions
$\beta_0 > 0$ $\boldsymbol{\Upsilon}_{\mathbf{a}} = \beta_0 \boldsymbol{\Upsilon}_{\mathbf{d}}, \quad \boldsymbol{\Upsilon}_{\mathbf{b}} = \beta_0 \boldsymbol{\Upsilon}_{\mathbf{c}}$ $\boldsymbol{\Upsilon}_{\mathbf{ab}} = -\beta_0 \boldsymbol{\Upsilon}^{\mathsf{T}}_{\mathbf{cd}}, \quad \boldsymbol{\Upsilon}_{\mathbf{ac}} = -\boldsymbol{\Upsilon}^{\mathsf{T}}_{\mathbf{bd}}$ $\boldsymbol{\Upsilon}_{\mathbf{ad}} = \boldsymbol{\Upsilon}^{\mathsf{T}}_{\mathbf{ad}}, \quad \boldsymbol{\Upsilon}_{\mathbf{bc}} = \boldsymbol{\Upsilon}^{\mathsf{T}}_{\mathbf{bc}}$	$\beta_0 < 0$ $\boldsymbol{\Upsilon}_{\mathbf{a}} = -\beta_0 \boldsymbol{\Upsilon}_{\mathbf{d}}, \quad \boldsymbol{\Upsilon}_{\mathbf{b}} = -\beta_0 \boldsymbol{\Upsilon}_{\mathbf{c}}$ $\boldsymbol{\Upsilon}_{\mathbf{ab}} = \beta_0 \boldsymbol{\Upsilon}^{\mathsf{T}}_{\mathbf{cd}}, \quad \boldsymbol{\Upsilon}_{\mathbf{ac}} = \boldsymbol{\Upsilon}^{\mathsf{T}}_{\mathbf{bd}}$ $\boldsymbol{\Upsilon}_{\mathbf{ad}} = -\boldsymbol{\Upsilon}^{\mathsf{T}}_{\mathbf{ad}}, \quad \boldsymbol{\Upsilon}_{\mathbf{bc}} = -\boldsymbol{\Upsilon}^{\mathsf{T}}_{\mathbf{bc}}$

Table 3: Second-order properness conditions.

Remark 2. *Notice that the* $\mathbb{Q}$*-properness concept introduced in [25] for the standard quaternion algebra is the particular case of* N^1_{-1}*-properness given here. Also, the* $\mathbb{C}^{\mathrm{k}}$*-properness notion suggested in [30] is the particular case of* N^2_{-1}*-properness.*

3.2. Proper Processing

We have previously highlighted the importance of selecting a suitable metric to find the projection. Now, this question is analyzed in a deeper way. Specifically, we aim to obtain the estimate of a desired random variable $y \in \mathbb{Q}_\beta$ from a set of random measurements $\mathbf{x} \in \mathbb{Q}_\beta^p$ by using a proper methodology. For that, we exploit the isomorphism between $(\mathbb{Q}_\beta, d_3)$ and $L_2^4(\Omega)$. Thus, and unless otherwise specified, we will use d_3 to find both the projection and its associated error. Afterwards, we determine the existing relationships between the three metrics in (9) under properness conditions.

Definition 7. *Consider a random variable* $y \in \mathbb{Q}_\beta$ *and a random vector* $\mathbf{x} \in \mathbb{Q}_\beta^p$*. The projection of* y *onto the closed linear subspace* $\mathcal{G}_{\mathbf{x}}$*, denoted by* $\hat{y}^{\mathtt{Q}_1}$*, is called the* $\mathtt{Q}_1$ *estimator of* y *respect to* $\mathbf{x}$*. Also, the* $\mathtt{Q}_2$ *estimator of* y *respect to* $\mathbf{x}$*, denoted by* $\hat{y}^{\mathtt{Q}_2}$*, is defined as the projection of* y *onto the set* $\mathcal{G}_{\dot{\mathbf{x}}}$ *with*

$$\dot{\mathbf{x}} = [\mathbf{x}^{\mathtt{T}}, \mathbf{x}^{\mathrm{k}^{\mathtt{T}}}]^{\mathtt{T}} \tag{15}$$

Finally, the projection of y *onto the set* $\mathcal{G}_{\bar{\mathbf{x}}}$*, where* $\bar{\mathbf{x}}$ *is the augmented vector given in* (5)*, is called QWL estimator of* y *respect to* $\mathbf{x}$ *and it will be denoted by* $\hat{y}^{\mathtt{QWL}}$*.*

These concepts are easily extended to the vectorial case. For example, the $\mathtt{Q}_1$ estimator of the random vector $\mathbf{y} \in \mathbb{Q}_\beta^q$ respect to $\mathbf{x}$ is $\hat{\mathbf{y}}^{\mathtt{Q}_1} = [\hat{y}_1^{\mathtt{Q}_1}, \ldots, \hat{y}_q^{\mathtt{Q}_1}]^{\mathtt{T}}$, where $\hat{y}_j^{\mathtt{Q}_1}$ is the projection of the jth component of $\mathbf{y}$ onto the set $\mathcal{G}_{\mathbf{x}}$.

The following result gives the expressions for the QWL, $\mathtt{Q}_1$ and $\mathtt{Q}_2$ estimators and establishes the relationships between them under properness conditions. It should be noted that, only when properness properties are assumed, all the derivations and calculations are restricted to a particular algebra associated to β_0. Otherwise, the results hold for any β-quaternion algebra.

Theorem 1. *Consider a random variable* $y \in \mathbb{Q}_\beta$ *and a random vector* $\mathbf{x} \in \mathbb{Q}_\beta^p$*. Then,*

1. *The QWL estimator of* y *respect to* $\mathbf{x}$ *is obtained as*

$$\hat{y}^{\mathtt{QWL}} = \mathbf{l}_1\mathbf{x} + \mathbf{l}_2\mathbf{x}^{\mathrm{k}} + \mathbf{l}_3\mathbf{x}^{\mathrm{i}} + \mathbf{l}_4\mathbf{x}^{\mathrm{j}}$$

where the deterministic vectors $\mathbf{l}_i^{\mathtt{T}} \in \mathbb{Q}_\beta^p$ *are computed through the equation*

$$[\mathbf{l}_1, \mathbf{l}_2, \mathbf{l}_3, \mathbf{l}_4] = \boldsymbol{\Upsilon}^3_{y\bar{\mathbf{x}}}(\beta)\boldsymbol{\Upsilon}^{3^{-1}}_{\bar{\mathbf{x}}}(\beta) \tag{16}$$

Furthermore, $\hat{y}^{\mathtt{QWL}}$ *is independent of the value of* β*, i.e., its real vector* $\hat{y}_r^{\mathtt{QWL}}$ *is identical in any* β*-quaternion algebra considered. Similarly, its associated error, given by*

$$\epsilon^{\mathtt{QWL}} = \|y - \hat{y}^{\mathtt{QWL}}\|_3^2, \tag{17}$$

is also independent of the value of β*.*

2. *The* $\mathbb{Q}_1$ *estimator of* y *respect to* $\mathbf{x}$ *is calculated as*

$$\hat{y}^{\mathbb{Q}_1} = \mathbf{f}\mathbf{x}$$

where the deterministic vector $\mathbf{f}^T \in \mathbb{Q}_\beta^p$ *is computed through the equation*

$$\mathbf{f} = \mathbf{\Upsilon}_{y\mathbf{x}}^3(\beta)\mathbf{\Upsilon}_{\mathbf{x}}^{3^{-1}}(\beta) \tag{18}$$

3. *The* $\mathbb{Q}_2$ *estimator of* y *respect to* $\mathbf{x}$ *is obtained as*

$$\hat{y}^{\mathbb{Q}_2} = \mathbf{g}\mathbf{x} + \mathbf{h}\mathbf{x}^{\mathrm{k}}$$

where the deterministic vectors $\mathbf{g}^T, \mathbf{h}^T \in \mathbb{Q}_\beta^p$ *are computed through the equation*

$$[\mathbf{g}, \mathbf{h}] = \mathbf{\Upsilon}_{y\dot{\mathbf{x}}}^3(\beta)\mathbf{\Upsilon}_{\dot{\mathbf{x}}}^{3^{-1}}(\beta) \tag{19}$$

4. $\epsilon^{\mathbb{QWL}} \leq \epsilon^{\mathbb{Q}_2} \leq \epsilon^{\mathbb{Q}_1}$ *with* $\epsilon^{\mathbb{Q}_i} = \|y - \hat{y}^{\mathbb{Q}_i}\|_3^2$, $i = 1, 2$.

5. *If* $\mathbf{x}$ *is* $P_{\beta_0}^1$*-proper (respectively,* $N_{\beta_0}^1$*-proper) then, the coefficients in* (16) *are computed as*

$$[\mathbf{l}_{1\beta_0}, \mathbf{l}_{2\beta_0}, \mathbf{l}_{3\beta_0}, \mathbf{l}_{4\beta_0}] = \Big[\mathbf{\Upsilon}_{y\mathbf{x}}^\nu(\beta_0)\mathbf{\Upsilon}_{\mathbf{x}}^{\nu^{-1}}(\beta_0), \mathbf{\Upsilon}_{y\mathbf{x}^{\mathrm{k}}}^\nu(\beta_0)\left(\mathbf{\Upsilon}_{\mathbf{x}}^{\nu^{-1}}(\beta_0)\right)^{\mathrm{k}}, \mathbf{\Upsilon}_{y\mathbf{x}^{\mathrm{i}}}^\nu(\beta_0)\left(\mathbf{\Upsilon}_{\mathbf{x}}^{\nu^{-1}}(\beta_0)\right)^{\mathrm{i}}, \mathbf{\Upsilon}_{y\mathbf{x}^{\mathrm{j}}}^\nu(\beta_0)\left(\mathbf{\Upsilon}_{\mathbf{x}}^{\nu^{-1}}(\beta_0)\right)^{\mathrm{j}}\Big], \quad \nu = 1, 2 \tag{20}$$

Moreover, if y *and* $\mathbf{x}$ *are cross* $P_{\beta_0}^1$*-proper (respectively,* $N_{\beta_0}^1$*-proper) then,* $\hat{y}^{\mathbb{QWL}} = \hat{y}^{\mathbb{Q}_1}$.

6. *If* $\mathbf{x}$ *is* $P_{\beta_0}^2$*-proper (respectively,* $N_{\beta_0}^2$*-proper) then, the coefficients in* (16) *are obtained as*

$$[\mathbf{l}_{1\beta_0}, \mathbf{l}_{2\beta_0}, \mathbf{l}_{3\beta_0}, \mathbf{l}_{4\beta_0}] = [\mathbf{\Upsilon}_{y\dot{\mathbf{x}}}^\nu(\beta_0)\mathbf{\Upsilon}_{\dot{\mathbf{x}}}^{\nu^{-1}}(\beta_0), \mathbf{\Upsilon}_{y\dot{\mathbf{x}}^{\mathrm{i}}}^\nu(\beta_0)\left(\mathbf{\Upsilon}_{\dot{\mathbf{x}}}^{\nu^{-1}}(\beta_0)\right)^{\mathrm{i}}], \quad \nu = 1, 2$$

Moreover, if y *and* $\mathbf{x}$ *are cross* $P_{\beta_0}^2$*-proper (respectively,* $N_{\beta_0}^2$*-proper) then,* $\hat{y}^{\mathbb{QWL}} = \hat{y}^{\mathbb{Q}_2}$.

Proof. See Appendix E.1.

Remark 3. *Theorem 1 shows that, unlike* $\hat{y}^{\mathtt{Q}_1}$ *and* $\hat{y}^{\mathtt{Q}_2}$ *whose real vectors depend on the* β*-quaternion algebra involved,* $\hat{y}^{\mathtt{QWL}}$ *always has the same real vector independently of* β*. Likewise, the computational complexity of the* QWL *estimator can be notably reduced whenever properness conditions are fulfilled (compare* (16) *with* (18) *and* (19)*). More importantly, this reduction in computational burden cannot be attained in the real domain.*

As noted above, the product $\langle\cdot,\cdot\rangle_3$ associated to the metric d_3 does not verify Properties 3.1 and 3.3, and this fact can cause difficulties to obtain $\hat{y}^{\mathtt{QWL}}$ and $\epsilon^{\mathtt{QWL}}$ in some situations. However, we establish a relationships between $\hat{y}^{\mathtt{QWL}}$ and the projections computed by means of the metrics d_ν, $\nu = 1, 2$, as well as the relationships between their associated errors under properness conditions.

Corollary 1.

1. *If* $\mathbf{x}$ *is* $P^i_{\beta_0}$*-proper and* y *and* $\mathbf{x}$ *are cross* $P^i_{\beta_0}$*-proper,* $i = 1, 2$*, then, the projection of* y *onto the set* $\mathcal{G}_{\mathbf{x}}$ *for* $i = 1$ *(equivalently, onto the set* $\mathcal{G}_{\dot{\mathbf{x}}}$ *for* $i = 2$*) under the metric* d_1 *coincides with* $\hat{y}^{\mathtt{QWL}}$ *and*

$$\epsilon^{\mathtt{QWL}} = \frac{\beta_0 + 1}{2\beta_0}\|y - \hat{y}^{\mathtt{QWL}}\|_1^2$$

2. *Similarly, if* $\mathbf{x}$ *is* $N^i_{\beta_0}$*-proper and* y *and* $\mathbf{x}$ *are cross* $N^i_{\beta_0}$*-proper,* $i = 1, 2$*, then, the projection of* y *onto the set* $\mathcal{G}_{\mathbf{x}}$ *for* $i = 1$*, (equivalently, onto the set* $\mathcal{G}_{\dot{\mathbf{x}}}$*, for* $i = 2$*) under the distance* d_2 *coincides with* $\hat{y}^{\mathtt{QWL}}$ *and*

$$\epsilon^{\mathtt{QWL}} = \frac{|\beta_0| + 1}{2|\beta_0|}\|y - \hat{y}^{\mathtt{QWL}}\|_2^2$$

As a consequence, the shortcomings that d_3 presents are overcome by computing both $\hat{y}^{\mathtt{QWL}}$ and $\epsilon^{\mathtt{QWL}}$ using either d_1 or d_2. This approach for the proper processing is shown schematically in Fig. 1.

Remark 4. *Although the proper processing offers an interesting reduction in computational burden, the issue of computing* $\boldsymbol{\Upsilon}_{\mathbf{x}}^{\nu^{-1}}(\beta_0)$ *or* $\boldsymbol{\Upsilon}_{\dot{\mathbf{x}}}^{\nu^{-1}}(\beta_0)$ *should be addressed, especially when these functions vary with time as occurs in dynamical*

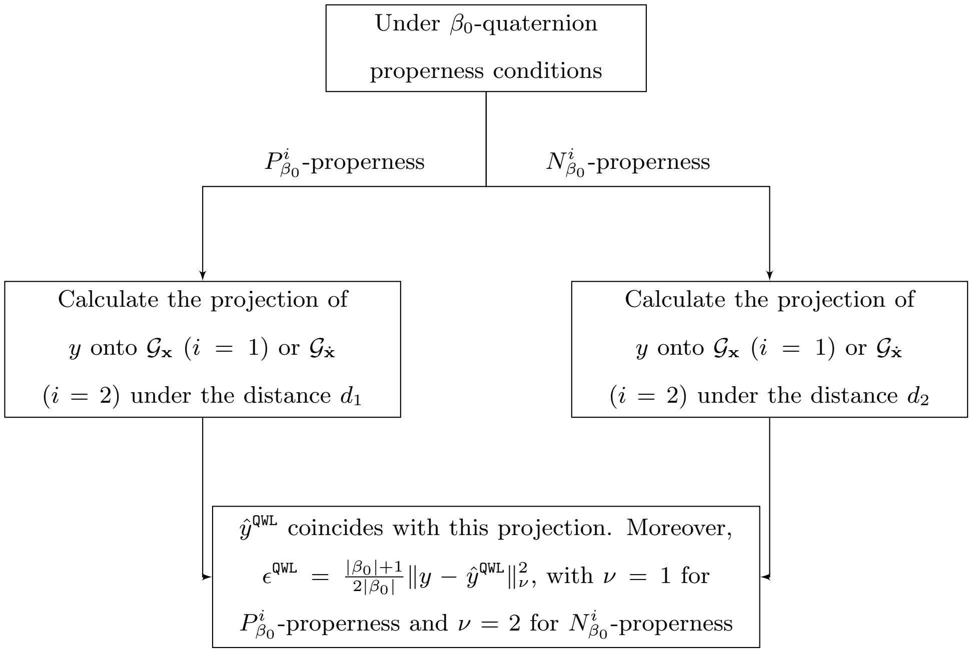

Figure 1: Flow chart of the approach proposed.

systems. In this regard, in Appendix C, filtering, one-stage prediction and fixed-interval smoothing algorithms are suggested under first and second-order properness conditions, in the case that the signal can be modeled via a state-space model.

4. Proper Adaptative Filters

In practice, the functions in (3) are usually unknown and need to be estimated from sample data. Furthermore, they can also vary with time when a random signal must be estimated. In this section, we address both shortcomings and tackle, from a sample perspective, the estimation problems analyzed above.

Notice that, in a first stage, it should be necessary to assess if the available data come from a proper random vector and then, a sample estimation algorithm should be applied. For this purpose, in Appendix D, statistical hypothesis tests to experimental check the properness properties are designed, and an estimation of parameter β_0 is provided. Next, adaptive sample algorithms to approach the proper estimators $\hat{y}^{\mathtt{Q}_i}$, $i = 1, 2$, suggested in Subsection 3.2 are analyzed in both first and second-order properness cases.

Suppose we take random samples from $y \in \mathbb{Q}_\beta$ and $\mathbf{x} \in \mathbb{Q}_\beta^p$, denoted by $y_{(1)}, \ldots, y_{(n)}$ and $\mathbf{x}_{(1)}, \ldots, \mathbf{x}_{(n)}$, respectively. Our goal is to solve the following adaptive filter problems:

1. First-order properness case.

   Find $\tilde{\mathbf{f}}_\nu^{\mathtt{T}} \in \mathbb{Q}_\beta^p$ such that minimizes the following cost function:

$$\min_{\tilde{\mathbf{f}}_\nu} \mathcal{R}\left\{\sum_{i=1}^{n}\left(y_{(i)} - \tilde{\mathbf{f}}_\nu \mathbf{x}_{(i)}\right)\left(y_{(i)} - \tilde{\mathbf{f}}_\nu \mathbf{x}_{(i)}\right)^{(3-2\nu)}\right\}, \qquad \nu = 1, 2 \tag{21}$$

2. Second-order properness case.

   Find $\tilde{\mathbf{g}}_\nu^{\mathtt{T}}, \tilde{\mathbf{h}}_\nu^{\mathtt{T}} \in \mathbb{Q}_\beta^p$ such that minimize the following cost function:

$$\min_{\tilde{\mathbf{g}}_\nu, \tilde{\mathbf{h}}_\nu} \mathcal{R}\left\{\sum_{i=1}^{n}\left(y_{(i)} - \tilde{\mathbf{g}}_\nu \mathbf{x}_{(i)} - \tilde{\mathbf{h}}_\nu \mathbf{x}_{(i)}^{\mathrm{k}}\right)\left(y_{(i)} - \tilde{\mathbf{g}}_\nu \mathbf{x}_{(i)} - \tilde{\mathbf{h}}_\nu \mathbf{x}_{(i)}^{\mathrm{k}}\right)^{(3-2\nu)}\right\} \tag{22}$$

   with $\nu = 1, 2$.

In order to solve these adaptive filter problems, we introduce two projection problems which are equivalent to (21) and (22). Denote $\mathbf{x}_{(ij)}$ the jth component of $\mathbf{x}_{(i)}$ and consider the vectors $\boldsymbol{\zeta}, \boldsymbol{\chi}_i \in \mathbb{Q}_\beta^n$ with jth components, $\boldsymbol{\zeta}_j = y_{(j)}$ and $\boldsymbol{\chi}_{ij} = \mathbf{x}_{(ji)}$, $i = 1, \ldots, p$, $j = 1, \ldots, n$. Consider also the matrices $\mathbf{X} = (\boldsymbol{\chi}_1, \ldots, \boldsymbol{\chi}_p)^{\mathtt{T}} \in \mathbb{Q}_\beta^{p\times n}$ and $\dot{\mathbf{X}} = \begin{pmatrix}\mathbf{X}\\ \mathbf{X}^{\mathrm{k}}\end{pmatrix} \in \mathbb{Q}_\beta^{2p\times n}$, and assume they have full column rank. Thus, the projections of $\boldsymbol{\zeta}$ onto the sets $\mathcal{C}_{\mathcal{A}_p}$ and $\mathcal{C}_{\dot{\mathcal{A}}_p}$ under the metrics d_ν in (13), with $\mathcal{A}_p = \{\boldsymbol{\chi}_1, \ldots, \boldsymbol{\chi}_p\}$ and $\dot{\mathcal{A}}_p = \{\boldsymbol{\chi}_1, \ldots, \boldsymbol{\chi}_p, \boldsymbol{\chi}_1^{\mathrm{k}}, \ldots, \boldsymbol{\chi}_p^{\mathrm{k}}\}$, are given by

$$\tilde{\boldsymbol{\zeta}}_\nu^{\mathtt{Q}_1} = \sum_{i=1}^{p} \tilde{f}_{\nu,i}\boldsymbol{\chi}_i \tag{23}$$

$$\tilde{\boldsymbol{\zeta}}_\nu^{\mathtt{Q}_2} = \sum_{i=1}^{p} \tilde{g}_{\nu,i}\boldsymbol{\chi}_i + \sum_{i=1}^{p} \tilde{h}_{\nu,i}\boldsymbol{\chi}_i^{\mathrm{k}} \tag{24}$$

respectively, and their associated errors are $\tilde{\xi}_\nu^{\mathtt{Q}_i} = \left|\boldsymbol{\zeta} - \tilde{\boldsymbol{\zeta}}_\nu^{\mathtt{Q}_i}\right|_\nu^2$, $i = 1, 2$, $\nu = 1, 2$.

Remark 5. *The quantities* $\tilde{\boldsymbol{\zeta}}_{\nu}^{\mathtt{Q}_i}$, $\tilde{f}_{\nu,i}$, $\tilde{g}_{\nu,i}$, $\tilde{h}_{\nu,i}$ *and* $\tilde{\xi}_{\nu}^{\mathtt{Q}_i}$ *depend also on the number of variables* p *and on the number of observations* n. *For simplicity of notation, this dependence will be included only when necessary. For example,* $\tilde{f}_{\nu,i}(n)$ *will specify that the coefficient* $\tilde{f}_{\nu,i}$ *is computed from* n *observations and* $\tilde{f}_{\nu,i}(p)$ *that is obtained from* p *variables.*

The following two results establish the closed relation between the adaptive filter problems (21) and (22), and the projection problems (23) and (24).

Theorem 2.

1. *The solution of* (21) *is unique and is given by*

$$\tilde{\mathbf{f}}_{\nu} = \boldsymbol{\zeta}^T \mathbf{X}^{H_{3-2\nu}} (\mathbf{X}\mathbf{X}^{H_{3-2\nu}})^{-1}, \qquad \nu = 1, 2 \tag{25}$$

Moreover, $\tilde{\mathbf{f}}_{\nu} = \left[\tilde{f}_{\nu,1}, \ldots, \tilde{f}_{\nu,p}\right]$ *where* $\tilde{f}_{\nu,i}$, $i = 1, \ldots, p$, *are the coefficients in* (23).

2. *The associated error of* $\tilde{\boldsymbol{\zeta}}_{\nu}^{\mathtt{Q}_1}$ *is given by*

$$\tilde{\xi}_{\nu}^{\mathtt{Q}_1} = \mathcal{R}\left\{\boldsymbol{\zeta}^T \boldsymbol{\zeta}^{(3-2\nu)} - \boldsymbol{\zeta}^T \big(\tilde{\mathbf{f}}_{\nu}\mathbf{X}\big)^{H_{3-2\nu}}\right\}, \qquad \nu = 1, 2 \tag{26}$$

Proof. See Appendix E.2.

Remark 6. *If* $\mathbf{x}$ *is* $P_{\beta_0}^1$*-proper and* y *and* $\mathbf{x}$ *are cross* $P_{\beta_0}^1$*-proper then, for sufficiently large* n *and under first and second-order ergodicity conditions,* $\tilde{\mathbf{f}}_{1,\beta_0}(n)$ *approaches* $\mathbf{f}_{\beta_0}$, *where* $\tilde{\mathbf{f}}_{1,\beta_0}(n)$ *and* $\mathbf{f}_{\beta_0}$ *denote the vectors in* (25) *and in* (18), *respectively, computed with the product associated to* β_0. *Also, by denoting*

$$\tilde{\epsilon}_{\nu,\beta_0}^{\mathtt{Q}_1}(n) = \frac{|\beta_0| + 1}{2|\beta_0| n} \tilde{\xi}_{\nu,\beta_0}^{\mathtt{Q}_1}(n), \qquad \nu = 1, 2 \tag{27}$$

it follows that, when n grows, $\tilde{\epsilon}_{1,\beta_0}^{\mathtt{Q}_1}(n)$ approaches $\epsilon^{\mathtt{QWL}}$ given in (17). Alternatively, $\tilde{\mathbf{f}}_{2,\beta_0}(n) \approx \mathbf{f}_{\beta_0}$ and $\tilde{\epsilon}_{2,\beta_0}^{\mathtt{Q}_1}(n) \approx \epsilon^{\mathtt{QWL}}$ in the $N_{\beta_0}^1$-properness case.

Theorem 3.

1. *The solution of* (22) *is unique and it is given by*

$$[\tilde{\mathbf{g}}_\nu, \tilde{\mathbf{h}}_\nu] = \boldsymbol{\zeta}^T \dot{\mathbf{X}}^{H_{3-2\nu}} (\dot{\mathbf{X}}\dot{\mathbf{X}}^{H_{3-2\nu}})^{-1}, \qquad \nu = 1, 2 \tag{28}$$

Moreover, $[\tilde{\mathbf{g}}_\nu, \tilde{\mathbf{h}}_\nu] = \left[\tilde{g}_{\nu,1}, \ldots, \tilde{g}_{\nu,p}, \tilde{h}_{\nu,1}, \ldots, \tilde{h}_{\nu,p}\right]$ *where* $\tilde{g}_{\nu,i}$, $\tilde{h}_{\nu,i}$, $i = 1, \ldots, p$, *are the coefficients in* (24).

2. *The associated error of* $\tilde{\boldsymbol{\zeta}}_\nu^{Q_2}$ *is given by*

$$\tilde{\xi}_\nu^{Q_2} = \mathcal{R}\left\{\boldsymbol{\zeta}^T \boldsymbol{\zeta}^{(3-2\nu)} - \boldsymbol{\zeta}^T \big([\tilde{\mathbf{g}}_\nu, \tilde{\mathbf{h}}_\nu]\dot{\mathbf{X}}\big)^{H_{3-2\nu}}\right\}, \qquad \nu = 1, 2 \tag{29}$$

3. $\tilde{\xi}_\nu^{Q_2} \leq \tilde{\xi}_\nu^{Q_1}$, $\nu = 1, 2$.

Proof. The proof is similar to that of Theorem 2.

Remark 7. *Similarly to Remark 6, under ergodicity conditions, if* $\mathbf{x}$ *is* $P_{\beta_0}^2$*-proper and* y *and* $\mathbf{x}$ *are cross* $P_{\beta_0}^2$*-proper then,* $[\tilde{\mathbf{g}}_{1,\beta_0}(n), \tilde{\mathbf{h}}_{1,\beta_0}(n)] \approx [\mathbf{g}_{\beta_0}, \mathbf{h}_{\beta_0}]$*, for sufficiently large* n*, where* $\tilde{\mathbf{g}}_{1,\beta_0}(n)$*,* $\tilde{\mathbf{h}}_{1,\beta_0}(n)$ *denote the vectors in* (28) *and* $\mathbf{g}_{\beta_0}$*,* $\mathbf{h}_{\beta_0}$ *the vectors in* (19) *all computed with the product associated to* β_0*. Also,* $\tilde{\epsilon}_{1,\beta_0}^{Q_2}(n) \approx \epsilon^{QWL}$ *given in* (17) *with*

$$\tilde{\epsilon}_{1,\beta_0}^{Q_2}(n) = \frac{|\beta_0| + 1}{2|\beta_0| n} \tilde{\xi}_{1,\beta_0}^{Q_2}(n) \tag{30}$$

Similar observations can be made in the $N_{\beta_0}^2$*-properness case.*

Remark 8. *The computational costs associated to* (25) *and* (28) *are of order* $O(p^3)$ *and* $O(8p^3)$*, respectively.*

Two problems associated with the updating of the matrix $\mathbf{X}$ usually appear in practice. On the one hand, a new row can be added to $\mathbf{X}$, i.e., a new independent random variable is available. Such a scenario appears, for example, in the filtering or one-stage prediction problems. On the other hand, a new column can be added to $\mathbf{X}$, i.e., new data from the independent random variables are incorporated to our study. In both cases, a method for updating the vectors (25) and (28) in a simple and efficient way is needed. Next, we provide recursive updating algorithms that alleviate the computational burden involved.

4.1. Row Updating

We study first the row updating problem for the adaptive filter associated to (21). The projection of $\boldsymbol{\zeta}$ onto $\mathcal{C}_{\mathcal{A}_{p+1}}$ with $\mathcal{A}_{p+1} = \{\boldsymbol{\chi}_1, \ldots, \boldsymbol{\chi}_{p+1}\}$ is now denoted by $\tilde{\boldsymbol{\zeta}}_\nu^{\mathtt{Q}_1}(p+1)$, its associated error by $\tilde{\xi}_\nu^{\mathtt{Q}_1}(p+1)$ and the vector (25) by $\tilde{\mathbf{f}}_\nu(p+1)$. Likewise, the projection of $\boldsymbol{\chi}_{i+1}$ onto the set $\mathcal{C}_{\mathcal{A}_i}$ with $\mathcal{A}_i = \{\boldsymbol{\chi}_1, \ldots, \boldsymbol{\chi}_i\}$ is denoted by $\tilde{\boldsymbol{\chi}}_\nu^{\mathtt{Q}_1}(i+1)$.

Following a similar reasoning to that used in the proof of Proposition 5.2.2 of [31] and taking into account Property 4 and (14), we get the following result.

Lemma 1. *The projections $\tilde{\boldsymbol{\chi}}_\nu^{Q_1}(i+1)$, $\nu = 1, 2$, can be recursively computed through the expression:*

$$\tilde{\boldsymbol{\chi}}_\nu^{Q_1}(i+1) = \sum_{j=1}^{i} l_{\nu,j}(i)\big(\boldsymbol{\chi}_{i+1-j} - \tilde{\boldsymbol{\chi}}_\nu^{Q_1}(i+1-j)\big), \qquad i \geq 1 \tag{31}$$

with $\tilde{\boldsymbol{\chi}}_\nu^{Q_1}(1) = \mathbf{0}_{n\times 1}$ and where

$$l_{\nu,i-k}(i) = \left(\prec \boldsymbol{\chi}_{i+1}, \boldsymbol{\chi}_{k+1} \succ_\nu - \sum_{j=0}^{k-1} l_{\nu,i-j}(i)\pi_{\nu,j} l_{\nu,k-j}^{(3-2\nu)}(k)\right)\pi_{\nu,k}^{-1}, \quad k = 1, \ldots, i-1 \tag{32}$$

with $l_{\nu,i}(i) = \prec \boldsymbol{\chi}_{i+1}, \boldsymbol{\chi}_1 \succ_\nu \pi_{\nu,0}^{-1}$ and

$$\pi_{\nu,i} = \prec \boldsymbol{\chi}_{i+1}, \boldsymbol{\chi}_{i+1} \succ_\nu - \sum_{j=0}^{i-1} l_{\nu,i-j}(i)\pi_{\nu,j} l_{\nu,i-j}^{(3-2\nu)}(i) \tag{33}$$

with the initial condition $\pi_{\nu,0} = \prec \boldsymbol{\chi}_1, \boldsymbol{\chi}_1 \succ_\nu$.

Theorem 4. *The projections $\tilde{\boldsymbol{\zeta}}_\nu^{Q_1}(p+1)$, $\nu = 1, 2$, can be recursively computed as:*

$$\tilde{\boldsymbol{\zeta}}_\nu^{Q_1}(p+1) = \sum_{j=1}^{p+1} \phi_{\nu,j}\big(\boldsymbol{\chi}_{p+2-j} - \tilde{\boldsymbol{\chi}}_\nu^{Q_1}(p+2-j)\big) \tag{34}$$

where

$$\phi_{\nu,p+1-k} = \left(\prec \boldsymbol{\zeta}, \boldsymbol{\chi}_{k+1} \succ_\nu - \sum_{j=0}^{k-1} \phi_{\nu,p+1-j}\pi_{\nu,j} l_{\nu,k-j}^{(3-2\nu)}(k)\right)\pi_{\nu,k}^{-1}, \quad k = 1, \ldots, p \tag{35}$$

with $\phi_{\nu,p+1} = \prec \boldsymbol{\zeta}, \boldsymbol{\chi}_1 \succ_\nu \pi_{\nu,0}^{-1}$, $l_{\nu,j}(i)$ and $\pi_{\nu,i}$ calculated in (32) *and* (33)*, respectively.*

Moreover, the associated error of $\tilde{\boldsymbol{\zeta}}_{\nu}^{\mathcal{Q}_1}(p+1)$ *is given by*

$$\tilde{\xi}_{\nu}^{\mathcal{Q}_1}(p+1) = \mathcal{R}\Big\{ \prec \boldsymbol{\zeta}, \boldsymbol{\zeta} \succ_{\nu} - \sum_{j=0}^{p} \phi_{\nu,p+1-j}\pi_{\nu,j}\phi_{\nu,p+1-j}^{(3-2\nu)} \Big\} \tag{36}$$

Proof. See Appendix E.3

Corollary 2. *The following assertions hold for* $\nu = 1, 2$*:*

1. $\tilde{\boldsymbol{\zeta}}_{\nu}^{\mathcal{Q}_1}(p+1) = \tilde{\boldsymbol{\zeta}}_{\nu}^{\mathcal{Q}_1}(p) + \phi_{\nu,1}\vec{\boldsymbol{\chi}}_{\nu}(p+1)$
2. $\tilde{\xi}_{\nu}^{\mathcal{Q}_1}(p+1) = \tilde{\xi}_{\nu}^{\mathcal{Q}_1}(p) - \phi_{\nu,1}\pi_{\nu,p}\phi_{\nu,1}^{(3-2\nu)}$
3. $\tilde{f}_{\nu,p+1-i}(p+1) = -\sum_{j=1}^{i} \tilde{f}_{\nu,p+1-i+j}(p+1) l_{\nu,p-i+j}(j) + \phi_{\nu,i+1}$

   for $i = 1, \ldots, p$*, with* $\tilde{f}_{\nu,p+1}(p+1) = \phi_{\nu,1}$ *and where* $\tilde{f}_{\nu,i}(p)$ *are given in* (25)*.*

Remark 9. *The algorithms provided in Theorem 4 and Corollary 2 can be adapted to the second-order properness case. For that, note that we have two new rows associated to* $p+1$*:* $\boldsymbol{\chi}_{p+1}^{T}$ *and* $\boldsymbol{\chi}_{p+1}^{\mathrm{k}T}$*. Thus, let* $\tilde{\boldsymbol{\zeta}}_{\nu}^{\mathcal{Q}_2}(2p+2)$ *be the projection of* $\boldsymbol{\zeta}$ *onto* $\mathcal{C}_{\dot{\mathcal{A}}_{2p+2}}$ *with* $\dot{\mathcal{A}}_{2p+2} = \{\boldsymbol{\chi}_1, \ldots, \boldsymbol{\chi}_{p+1}, \boldsymbol{\chi}_1^{\mathrm{k}}, \ldots, \boldsymbol{\chi}_{p+1}^{\mathrm{k}}\}$*, and denote by* $\tilde{\xi}_{\nu}^{\mathcal{Q}_2}(2p+2)$ *its associated error and by* $[\tilde{\mathbf{g}}_{\nu}(2p+2), \tilde{\mathbf{h}}_{\nu}(2p+2)]$ *the vector* (28) *for the two new rows added.*

The following steps are necessary for this adaptation:

Step 1: Define the set $\{\boldsymbol{\eta}_i\}_{i=1}^{2p+2}$ *as* $\boldsymbol{\eta}_i = \boldsymbol{\chi}_i$ *and* $\boldsymbol{\eta}_{p+1+i} = \boldsymbol{\chi}_i^{\mathrm{k}}$*,* $i = 1, \ldots, p+1$*.*

Step 2: Apply Lemma 1 to $\{\boldsymbol{\eta}_i\}_{i=1}^{2p+2}$*.*

Step 3: Obtain $\tilde{\boldsymbol{\zeta}}_{\nu}^{\mathcal{Q}_2}(2p+2)$*,* $\tilde{\xi}_{\nu}^{\mathcal{Q}_2}(2p+2)$ *and* $[\tilde{\mathbf{g}}_{\nu}(2p+2), \tilde{\mathbf{h}}_{\nu}(2p+2)]$ *by combining Step 2 and Corollary 2.*

Remark 10. *The computational cost of the algorithm given in Theorem 4 is of order* $O(p^2n^2)$*. Similarly, the computational cost of the algorithm in the second-order properness case is of order* $O(4p^2n^2)$*.*

4.2. Column Updating

Let us start by the first-order properness case. The new column is denoted by $\boldsymbol{\kappa}_{n+1}$. The vector $\boldsymbol{\zeta}$ and the matrix $\mathbf{X}$ are now denoted by $\boldsymbol{\zeta}(n)$ and $\mathbf{X}(n)$, respectively, when n observations are available. Then,

$$\mathbf{X}(n+1) = (\mathbf{X}(n), \boldsymbol{\kappa}_{n+1}) \quad \text{and} \quad \boldsymbol{\zeta}(n+1) = \begin{pmatrix} \boldsymbol{\zeta}(n) \\ y_{(n+1)} \end{pmatrix} \tag{37}$$

Also, the vector in (25) is denoted by $\tilde{\mathbf{f}}_\nu(n)$ and define the matrix $\mathbf{P}_\nu(n) = \left(\mathbf{X}(n)\mathbf{X}^{\mathtt{H}_{3-2\nu}}(n)\right)^{-1}$, $\nu = 1, 2$.

Theorem 5. *The vector $\tilde{\mathbf{f}}_\nu(n+1)$, $\nu = 1, 2$, is computed as:*

$$\tilde{\mathbf{f}}_\nu(n+1) = \tilde{\mathbf{f}}_\nu(n) + \left(y_{(n+1)} - \tilde{\mathbf{f}}_\nu(n)\boldsymbol{\kappa}_{n+1}\right)\lambda_\nu(n)\boldsymbol{\kappa}_{n+1}^{\mathtt{H}_{3-2\nu}}\mathbf{P}_\nu(n) \tag{38}$$

with $\lambda_\nu(n) = \left(1 + \boldsymbol{\kappa}_{n+1}^{\mathtt{H}_{3-2\nu}}\mathbf{P}_\nu(n)\boldsymbol{\kappa}_{n+1}\right)^{-1}$ *and*

$$\mathbf{P}_\nu(n+1) = \mathbf{P}_\nu(n) - \mathbf{P}_\nu(n)\boldsymbol{\kappa}_{n+1}\lambda_\nu(n)\boldsymbol{\kappa}_{n+1}^{\mathtt{H}_{3-2\nu}}\mathbf{P}_\nu(n) \tag{39}$$

Proof. See Appendix E.4.

Remark 11. *A similar algorithm can be suggested for the second-order properness case. For that, denote (28) by $[\tilde{\mathbf{g}}_\nu(n), \tilde{\mathbf{h}}_\nu(n)]$, the matrix $\dot{\mathbf{X}}$ by $\dot{\mathbf{X}}(n)$ and define the matrix $\dot{\mathbf{P}}_\nu(n) = \left(\dot{\mathbf{X}}(n)\dot{\mathbf{X}}^{\mathtt{H}_{3-2\nu}}(n)\right)^{-1}$, $\nu = 1, 2$. Then, by replacing $\tilde{\mathbf{f}}_\nu(n)$ and $\mathbf{P}_\nu(n)$ in Theorem 5 by $[\tilde{\mathbf{g}}_\nu(n), \tilde{\mathbf{h}}_\nu(n)]$ and $\dot{\mathbf{P}}_\nu(n)$, respectively, we get the desired result.*

Remark 12. *The computational cost of the algorithm given in Theorem 5 is of order $O(p^2n)$ and that of the algorithm for the second-order properness case is $O(4p^2n)$.*

In Fig. 2, we summarize the concepts studied in this subsection.

4.3. Summary of the Proposed Methodology

The methodology proposed, can be summarized by the following basic steps:

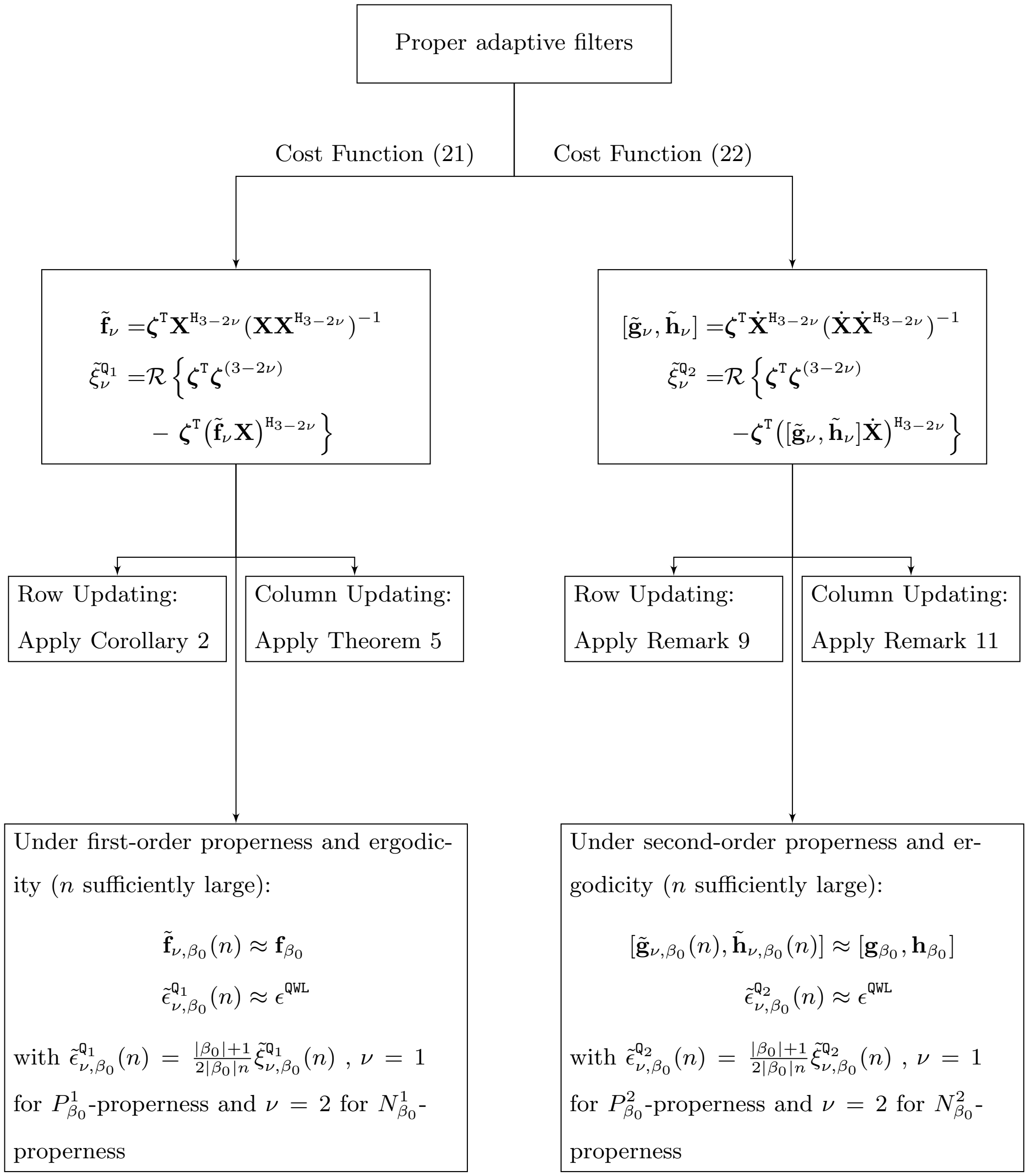

Figure 2: Scheme of the proper adaptive filters.

Step 1. Use the properness tests presented in Appendix D to check first-order and second-order properness properties and estimate the value of β, which includes the following sub-steps:

1.1. Fix a false-alarm probability α.

1.2. Check the first order-properness conditions given in (D.1) by means of the following decision rule:

$$\text{reject } H_0 \text{ if } \phi_\nu(\mathbf{x}_{(1)}, \dots, \mathbf{x}_{(n)}) \geq \chi^2_{1-\alpha,\eta_\nu}$$

with $\phi_\nu(\mathbf{x}_{(1)}, \dots, \mathbf{x}_{(n)})$, $\nu = 1, 2$, given in (D.2), $\eta_1 = 6p^2 + p - 1$, $\eta_2 = 6p^2 + 3p - 1$, and where $\chi^2_{1-\alpha,\eta_\nu}$ is the percentil of a chi-squared probability distribution.

1.3. If H_0 is rejected, check the second order-properness conditions given in (D.5) by means of the following decision rule:

$$\text{reject } H_0 \text{ if } \psi_\nu(\mathbf{x}_{(1)}, \dots, \mathbf{x}_{(n)}) \geq \chi^2_{1-\alpha,\eta_\nu}$$

with $\psi_\nu(\mathbf{x}_{(1)}, \dots, \mathbf{x}_{(n)})$, $\nu = 1, 2$, given in (D.6), $\eta_1 = 4p^2 - 1$, and $\eta_2 = 4p^2 + 2p - 1$.

1.4. The suitable value of β_0 is obtained from (D.4) (first-order) or (D.7) (second-order).

Step 2. In Fig. 2, choose the left branch (cost function (21)) in the first-order properness case, or the right branch (cost function (22)) in the second-order properness case.

Step 3. Update new information by applying Corollary 2 (row updating)/Theorem 5 (column updating) in the first-order properness case, or Remark 9 (row updating)/Remark 11 (column updating) in the second-order properness case.

5. Numerical Results

This section is devoted to the evaluation of the algorithms proposed using simulated data. The presentation is divided into three parts. Firstly, the properness tests devised in Appendix D are assessed. Second, the performance of the proper adaptive filters suggested in Subsection 4 is analyzed. Finally, a benchmark example is considered to compare our methodology with other methods existing in the literature.

5.1. Example 1: Properness Tests

Consider a random vector $\mathbf{x} = [x_1, x_2]^{\mathrm{T}} \in \mathbb{Q}_{\beta}^2$ such that the real vectors of x_1 and x_2 given by (4) have the following covariance matrices:

$$\boldsymbol{\Upsilon}_{\mathbf{x}_{ir}} = \begin{pmatrix} 8 & 0 & -\theta_1 & 0 \\ 0 & 8 & 0 & \theta_1 \\ -\theta_1 & 0 & 2 & 0 \\ 0 & \theta_1 & 0 & 2 \end{pmatrix}, \quad i = 1, 2 \tag{40}$$

and cross-covariance matrix:

$$\boldsymbol{\Upsilon}_{\mathbf{x}_{1r}\mathbf{x}_{2r}} = \begin{pmatrix} 0.8 & -0.4 & -0.5 & 0.6 \\ 0.4 & 0.8 & \theta_4 & 0.5 \\ -\theta_3 & \theta_5 & 0.2 & -0.1 \\ \theta_2 & \theta_3 & 0.1 & 0.2 \end{pmatrix}$$

Two β-quaternion algebras are considered: the algebras associated to the values $\beta_0 = 4$ and $\beta_0 = -4$, and four cases are studied:

1. Case 1: $\beta_0 = 4$, $\theta_1 = 1$, $\theta_2 = \theta_4 = \theta_5 = 0.6$ and $\theta_3 = 0.5$.

   $\mathbf{x}$ is P_4^1-proper (Table 2) and the GLRT statistic $\phi_1(\mathbf{x}_{(1)}, \ldots, \mathbf{x}_{(n)})$ converges to a $\chi^2(25)$ distribution (Theorem 8).

2. Case 2: $\beta_0 = -4$, $\theta_1 = 0$, $\theta_2 = \theta_5 = -0.6$, $\theta_4 = 0.6$ and $\theta_3 = -0.5$.

   $\mathbf{x}$ is N_{-4}^1-proper (Table 2) and the GLRT statistic $\phi_2(\mathbf{x}_{(1)}, \ldots, \mathbf{x}_{(n)})$ converges to a $\chi^2(29)$ distribution (Theorem 8).

3. Case 3: $\beta_0 = 4$, $\theta_1 = 1$, $\theta_2 = 0.6$, $\theta_3 = 0.5$ and $\theta_4 = \theta_5 = -1$.

   $\mathbf{x}$ is P_4^2-proper (Table 3) and the GLRT statistic $\psi_1(\mathbf{x}_{(1)}, \ldots, \mathbf{x}_{(n)})$ converges to a $\chi^2(15)$ distribution (Theorem 9).

4. Case 4: $\beta_0 = -4$, $\theta_1 = 0$, $\theta_2 = -0.6$, $\theta_3 = -0.5$, $\theta_4 = -1$ and $\theta_5 = 1$.

   $\mathbf{x}$ is N^2_{-4}-proper (Table 3) and the GLRT statistic $\psi_2(\mathbf{x}_{(1)}, \ldots, \mathbf{x}_{(n)})$ converges to a $\chi^2(19)$ distribution (Theorem 9).

On the one hand, the behavior of the cumulative distribution functions (CDFs) of $\phi_\nu(\mathbf{x}_{(1)}, \ldots, \mathbf{x}_{(n)})$ and $\psi_\nu(\mathbf{x}_{(1)}, \ldots, \mathbf{x}_{(n)})$, $\nu = 1, 2$, is studied in terms of the sample size n. For that, and through Monte Carlo simulations, 2000 values of the statistics $\phi_\nu(\mathbf{x}_{(1)}, \ldots, \mathbf{x}_{(n)})$ and $\psi_\nu(\mathbf{x}_{(1)}, \ldots, \mathbf{x}_{(n)})$, $\nu = 1, 2$, have been generated for $n = 50, 100, 250$. The corresponding empirical CDFs are displayed in Fig. 3. This figure shows, for all the properness cases considered, the rapid convergence of the empirical CDFs to the target chi-squared distributions.

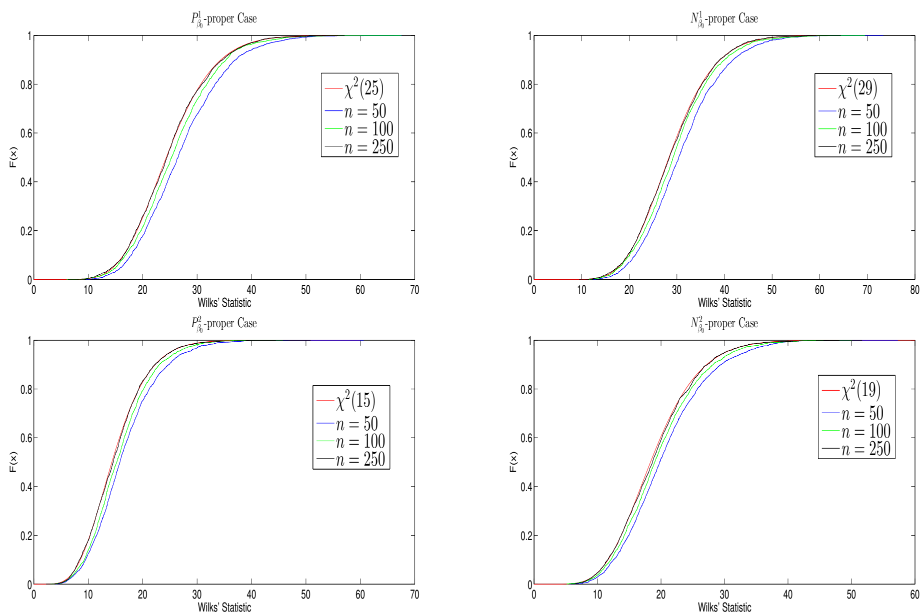

Figure 3: Convergence of the empirical CDFs to the target chi-squared distributions.

On the other hand, we assess the accuracy of the estimations of β_0 for several values of n. Table 4 depicts the 95% confidence intervals (CIs) for β_0 obtained from 2000 simulated runs, where the maximizations in (D.4) and (D.7) have been made using the optimization toolbox in Matlab. As can be expected, the amplitudes of the CIs become narrower as n grows.

n	50	100	250
Case 1	[3.0378, 5.295]	[3.2929, 4.8229]	[3.5755, 4.5407]
Case 2	[-5.3537, -3.0251]	[-4.8633, -3.266]	[-4.5621, -3.5617]
Case 3	[3.0274, 5.2808]	[3.305, 4.7978]	[3.5799, 4.525]
Case 4	[-5.3252, -3.0012]	[-4.8502, -3.2906]	[-4.5502, -3.5673]

Table 4: 95% CIs for β_0.

5.2. Example 2: Proper Adaptive Filters

In order to investigate the performance of the framework proposed, we carry out a comparative analysis between the theoretic proper processing addressed in Subsection 3.2 and the proper adaptive filters. To this end, we conduct a simulation example in which the QWL errors and their corresponding approximated errors provided by the adaptive methodology are compared. Likewise, we illustrate the adaptation of the algorithms suggested to the row and column updating cases. Finally, we show that the standard quaternion domain is not always the best hypercomplex algebra for solving adaptive filter problems since, as we will see, it fails to give accurate estimations under general properness conditions.

We consider the following state-space model:

$$\begin{aligned} x_{t+1} &= f x_t + u_t \\ y_t &= x_t + v_t \end{aligned} \tag{41}$$

with $f = 0.16 + \text{i}0.4 - \text{j}0.24 + \text{k}0.32 \in \mathbb{Q}_\beta$, and where x_0, u_t and v_t are independent β-quaternions random variables whose real vectors (4) have covariance matrices $\boldsymbol{\Upsilon}_{\mathbf{x}_{0r}}$ given by (40) and

$$\boldsymbol{\Upsilon}_{\mathbf{u}_{tr}\mathbf{u}_{sr}} = \begin{pmatrix} \frac{\theta_2}{4} & 0 & 0 & 0 \\ 0 & \frac{1}{4} & 0 & 0 \\ 0 & 0 & \frac{1}{16} & 0 \\ 0 & 0 & 0 & \frac{\theta_2}{16} \end{pmatrix} \delta_{ts}, \quad \boldsymbol{\Upsilon}_{\mathbf{v}_{tr}\mathbf{v}_{sr}} = \begin{pmatrix} 0.4 & 0 & 0 & 0 \\ 0 & 0.4 & 0 & 0 \\ 0 & 0 & 0.1 & 0 \\ 0 & 0 & 0 & 0.1 \end{pmatrix} \delta_{ts},$$

The state-space model (41) is analyzed in two different frameworks: the algebras associated to the values $\beta_0 = 4$ and $\beta_0 = -4$, and where we also assume first and second-order properness conditions. From this model and for each of the above scenarios, we compute the filtering errors, $\epsilon^{\texttt{QWL}}(t)$, and the fixed-interval smoothing errors for $m = 9$, $\epsilon^{\texttt{QWL}}(t|9)$, which are compared with the approximate errors (27) and (30), denoted by $\tilde{\epsilon}^{\texttt{Q}_i}_{\nu,\beta_0}(t)$ for the filtering case and by $\tilde{\epsilon}^{\texttt{Q}_i}_{\nu,\beta_0}(t|9)$ for the smoothing case. Such approximated errors are obtained from 10000 trajectories generated by Monte Carlo simulation. The standard quaternion algebra is also considered and their approximated errors $\tilde{\epsilon}^{\texttt{Q}_i}_{2,-1}(t)$ and $\tilde{\epsilon}^{\texttt{Q}_i}_{2,-1}(t|9)$ are calculated.

The results corresponding to the filtering problem (row updating) are depicted in Fig. 4 and those associated to the smoothing problem (column updating) in Fig. 5. Both figures study four cases of properness: two of first-order, P_4^1 and N_{-4}^1, and two of second-order, P_4^2 and N_{-4}^2. Tables 4 and 5 explain the contents of such figures. For instance, the first row in Table 5 indicates that the subfigure (a) of Fig. 4 is obtained by setting $\theta_1 = \theta_2 = 1$ in the covariance matrices associated to model (41) and thus, from Table 2, x_0, u_t and v_t are P_4^1-proper. Hence, Theorem 6 (T6) in Appendix C can be used to obtain the optimal error $\epsilon^{\texttt{QWL}}(t)$. Additionally, following the left branch of the scheme in Fig. 2, we have calculated $\tilde{\epsilon}^{\texttt{Q}_1}_{1,4}(t)$ via the recursion in Corollary 2 (C2). The approximate error $\tilde{\epsilon}^{\texttt{Q}_1}_{2,-1}(t)$ is obtained through SL standard quaternion techniques. Note that we compare the error $\tilde{\epsilon}^{\texttt{Q}_1}_{2,-1}(t)$ associated to the standard quaternion estimate $\tilde{\boldsymbol{\zeta}}^{\texttt{Q}_1}_{2,-1}(t)$ (and not $\tilde{\epsilon}^{\texttt{Q}_2}_{2,-1}(t)$) with the error $\tilde{\epsilon}^{\texttt{Q}_1}_{1,4}(t)$ associated with $\tilde{\boldsymbol{\zeta}}^{\texttt{Q}_1}_{1,4}(t)$ since both estimates have identical computational cost. On the other hand, the third row in Table 5 indicates that the subfigure (c) of Fig. 4 is obtained by taking $\theta_1 = -1$ and $\theta_2 = 8$ in the system (41) and this fact entails a P_4^2-properness case. Now, four errors are computed: $\epsilon^{\texttt{QWL}}(t)$ from Remark 13 (R13), $\tilde{\epsilon}^{\texttt{Q}_2}_{1,4}(t)$ by using Remark 9, $\tilde{\epsilon}^{\texttt{Q}_2}_{2,-1}(t)$ via the SWL standard quaternion methodology, and $\tilde{\epsilon}^{\texttt{Q}_1}_{1,4}(t)$ from C2. Two comments are in order. First, error $\tilde{\epsilon}^{\texttt{Q}_1}_{1,4}(t)$ allows us to evaluate the loss in performance due to a P_4^1-properness treatment of the problem. Second, similarly to the above case, we consider $\tilde{\epsilon}^{\texttt{Q}_2}_{2,-1}(t)$ in the analysis since $\tilde{\boldsymbol{\zeta}}^{\mathtt{Q}_2}_{1,4}(t)$ and $\tilde{\boldsymbol{\zeta}}^{\mathtt{Q}_2}_{2,-1}(t)$ have identical computational complexity. Similar interpretations can be made for the rows of Table 6 corresponding to Fig. 5.

Figure 4						
Subfig.	Case	Properness	$\epsilon^{\mathtt{QWL}}(t)$	$\tilde{\epsilon}^{\mathtt{Q}_i}_{\nu,\beta_0}(t)$	$\tilde{\epsilon}^{\mathtt{Q}_i}_{2,-1}(t)$	$\tilde{\epsilon}^{\mathtt{Q}_1}_{\nu,\beta_0}(t)$
(a)	$\theta_1 = \theta_2 = 1$	P^1_4	T6	C2	SL stand.	
(b)	$\theta_1 = 0,\ \theta_2 = 1$	N^1_{-4}	T6	C2	SL stand.	
(c)	$\theta_1 = -1,\ \theta_2 = 8$	P^2_4	R13	R9	SWL stand.	C2
(d)	$\theta_1 = 0,\ \theta_2 = 8$	N^2_{-4}	R13	R9	SWL stand.	C2

Table 5: Properness properties analyzed in the subfigures of Fig. 4 and the form in which the errors are calculated for the filtering problem.

Figure 5						
Subfig.	Case	Properness	$\epsilon^{\mathtt{QWL}}(t\|9)$	$\tilde{\epsilon}^{\mathtt{Q}_i}_{\nu,\beta_0}(t\|9)$	$\tilde{\epsilon}^{\mathtt{Q}_i}_{2,-1}(t\|9)$	$\tilde{\epsilon}^{\mathtt{Q}_1}_{\nu,\beta_0}(t\|9)$
(a)	$\theta_1 = \theta_2 = 1$	P^1_4	T7	T5	SL stand.	
(b)	$\theta_1 = 0,\ \theta_2 = 1$	N^1_{-4}	T7	T5	SL stand.	
(c)	$\theta_1 = -1,\ \theta_2 = 8$	P^2_4	R13	R11	SWL stand.	T5
(d)	$\theta_1 = 0,\ \theta_2 = 8$	N^2_{-4}	R13	R11	SWL stand.	T5

Table 6: Properness properties analyzed in the subfigures of Fig. 5 and the form in which the errors are calculated for the smoothing problem.

Fig. 4 shows, in all the cases studied, that $\tilde{\epsilon}^{\mathtt{Q}_i}_{\nu,\beta_0}(t)$ is very close to $\epsilon^{\mathtt{QWL}}(t)$, i.e., the adaptive filters suggested, $\tilde{\boldsymbol{\zeta}}^{\mathtt{Q}_i}_{\nu,\beta_0}(t)$, are able to provide an accurate estimation performance very close to the optimal estimator. Also, since for all t, $\tilde{\epsilon}^{\mathtt{Q}_i}_{\nu,\beta_0}(t) < \tilde{\epsilon}^{\mathtt{Q}_i}_{2,-1}(t)$ then, $\tilde{\boldsymbol{\zeta}}^{\mathtt{Q}_i}_{\nu,\beta_0}(t)$ outperform $\tilde{\boldsymbol{\zeta}}^{\mathtt{Q}_i}_{\nu,-1}(t)$ for $\beta_0 = -4$ or 4, $i = 1, 2$ and $\nu = 1, 2$. Therefore, we have shown that, if the properness conditions hold in a β-quaternion algebra with $\beta_0 \neq -1$ then, the standard quaternion algebra is not the best setting to address the adaptive filtering problem. Also, under second-order properness conditions (subfigures 4(c) and 4(d)), we observe, as expected, that $\tilde{\epsilon}^{\mathtt{Q}_2}_{\nu,\beta_0}(t) < \tilde{\epsilon}^{\mathtt{Q}_1}_{\nu,\beta_0}(t)$, for all t, and thus, $\tilde{\boldsymbol{\zeta}}^{\mathtt{Q}_2}_{\nu,\beta_0}(t)$ is the preferable

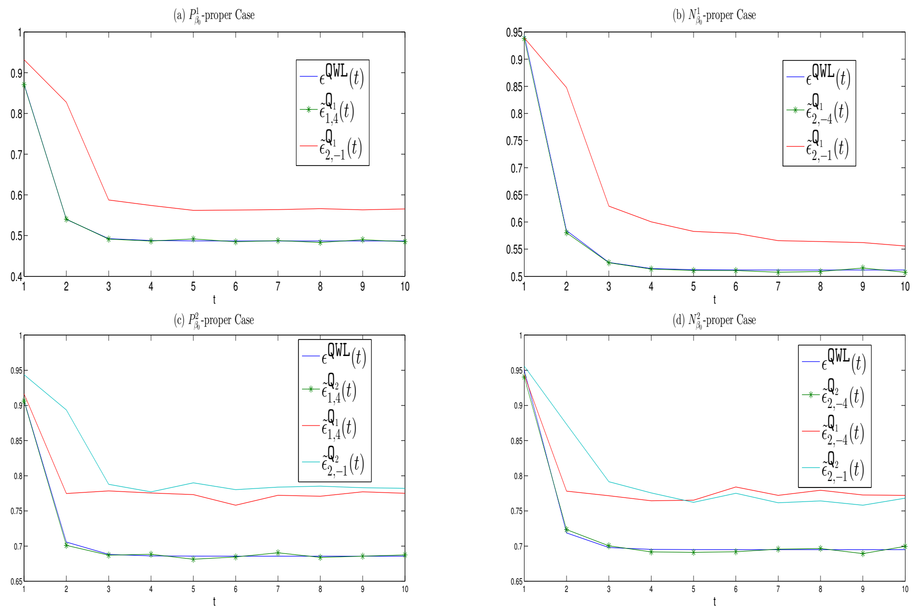

Figure 4: Comparative study between the theoretic proper processing and the proper adaptive filters for the filtering case.

estimator. Similar interpretations apply to Fig. 5.

5.3. Example 3: Proper Adaptive Filters vs. Standard Quaternion Filter

In this example, our aim is to contrast the results obtained using the proposed method and those existing in the literature. In particular, the proper adaptive filtering approach suggested in Section 4 is compared with the quaternion standard filtering methodology [17], for different β-quaternion algebras and properness scenarios.

For this purpose, we consider the following versatile equation of motion looked at [32], which can be used in a variety of situations including bearings-only and rotation tracking:

$$\frac{\partial \varphi}{\partial t} = \phi, \qquad \frac{\partial \phi}{\partial t} = u, \tag{42}$$

where u represents the input of the system, and φ is the variable of interest with rate of change ϕ. Equation (42) enables the following discrete-time state-space

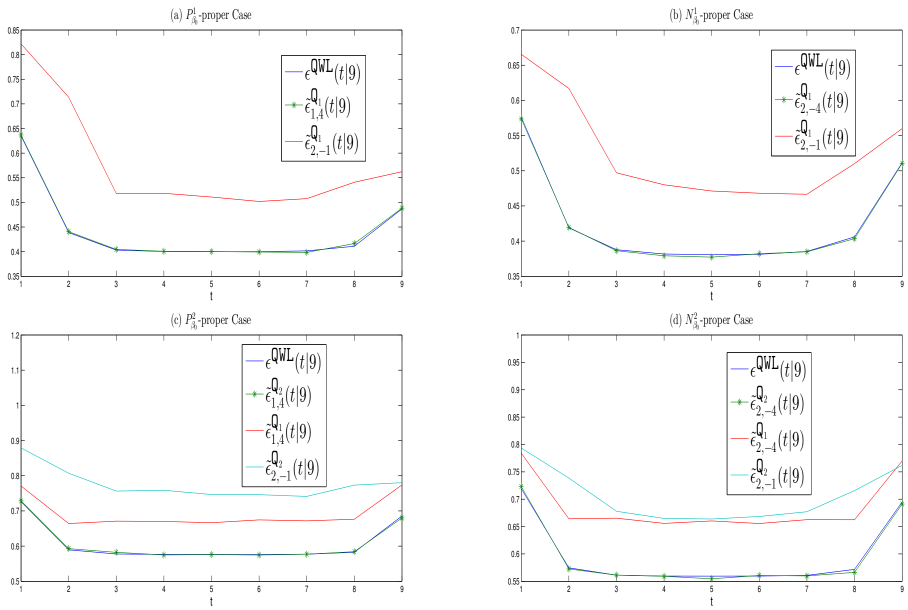

Figure 5: Comparative study between the theoretic proper processing and the proper adaptive filters for the fixed-interval smoothing case.

model:

$$\mathbf{x}_{t+1} = \mathbf{F}\mathbf{x}_t + \mathbf{G}u_t, \quad t = 1, \ldots, 100 \tag{43}$$

with $\mathbf{x}_t = [\varphi_t, \phi_t]^{\mathtt{T}}$, and initial condition $\mathbf{x}_0 = \mathbf{0}_{2\times 1}$. Moreover, $\mathbf{F} = \begin{pmatrix} 1 & \Delta T \\ 0 & 1 \end{pmatrix}$, and $\mathbf{G} = \begin{bmatrix} \Delta T^2/2 \\ \Delta T \end{bmatrix}$, where $\Delta T = 0.1$ represents the sampling interval, and u_t is a β-quaternion white noise whose real vector has this covariance matrix:

$$\mathbf{\Upsilon}_{u_{tr}u_{sr}} = \begin{pmatrix} 3|\beta_0| & 0 & 0 & 0 \\ 0 & 3|\beta_0| & 0 & 0 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 3 \end{pmatrix} \delta_{ts}$$

By way of illustration, we examine two first-order properness scenarios: $P^1_{\beta_0}$ and $N^1_{\beta_0}$, and three different β-quaternion algebras in each scenario: $\beta_0 = 5$, $\beta_0 = 20$, and $\beta_0 = 10000$ in the $P^1_{\beta_0}$-proper setting, and $\beta_0 = -5$, $\beta_0 = -20$, and $\beta_0 = -10000$ in the $N^1_{\beta_0}$-proper setting.

Then, we assume that the observations are obtained from the equation

$$\mathbf{y}_t = \mathbf{x}_t + \mathbf{v}_t \tag{44}$$

where $\mathbf{v}_t = [v_{1t}, v_{2t}]^{\mathtt{T}}$ is a β-quaternion white noise vector, such that the real covariance matrices associated to v_{1t} and v_{2t} are of the form

$$\mathbf{\Upsilon}_{v_{itr}v_{isr}} = \begin{pmatrix} |\beta_0| & 0 & 0 & \lambda \\ 0 & |\beta_0| & \lambda & 0 \\ 0 & \lambda & 1 & 0 \\ \lambda & 0 & 0 & 1 \end{pmatrix} \delta_{ts}, \quad i = 1, 2$$

In this setting, the following two cases are analyzed:

- **Case 1.** $P^1_{\beta_0}$**-proper scenario:** $\lambda = 2$, and $\mathbf{\Upsilon}_{v_{1tr}v_{2sr}} = \mathbf{0}_{4\times 4}$
- **Case 2.** $N^1_{\beta_0}$**-proper scenario:** $\lambda = 0$, and

$$\mathbf{\Upsilon}_{v_{1tr}v_{2sr}} = \begin{pmatrix} -0.5|\beta_0| & 0.7|\beta_0| & 0 & -0.9 \\ -0.7|\beta_0| & -0.5|\beta_0| & -0.9 & 0 \\ 0 & 0.9 & 0.5 & -0.7 \\ 0.9 & 0 & 0.7 & 0.5 \end{pmatrix} \delta_{ts}$$

For the above-mentioned cases, the filtering errors $\tilde{\epsilon}^{\mathtt{Q}_1}_{\nu,\beta_0}(t)$, $\nu = 1, 2$, given in (27) are compared with their counterparts in the quaternion domain, i.e., the quaternion strictly linear (QSL) filtering errors, denoted by $\epsilon^{\mathtt{QSL}}(t)$, and computed from the quaternion standard filter [17]. As a measure of comparison, the differences between both errors, $DE_{\nu,\beta_0}(t) = \epsilon^{\mathtt{QSL}}(t) - \tilde{\epsilon}^{\mathtt{Q}_1}_{\nu,\beta_0}(t)$, are calculated for the different values of β_0. These differences are displayed in Fig. 6 for $\beta_0 = 5, 20, 10000$ in a $P^1_{\beta_0}$-proper scenario (subfigure (a)), and $\beta_0 = -5, -20, -10000$ in a $N^1_{\beta_0}$-proper scenario (subfigure (b)). Note that the behavior of these differences stabilizes for values of $|\beta_0| > 10000$.

As expected, under $P^1_{\beta_0}$, as well as $N^1_{\beta_0}$,-properness conditions, the proposed methodology produces better estimators than the quaternion standard filters ($DE_{\nu,\beta_0}(t) > 0$). In addition, in the $P^1_{\beta_0}$-proper scenario, the maximum difference is reached when $\beta_0 = 5$, whereas in the $N^1_{\beta_0}$-proper scenario, the greatest advantage of our methodology is achieved for $\beta_0 = -10000$.

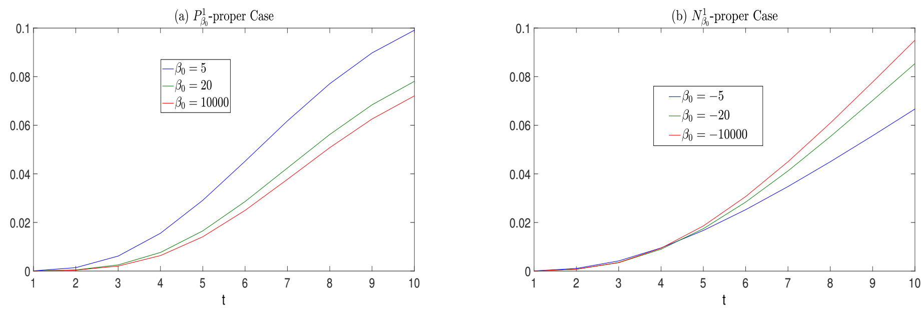

Figure 6: Difference between the standard QSL and proper adaptive filtering errors.

6. Conclusions

The β-quaternion domain has been suggested as a framework for developing new recursive adaptive filters. This family of 4D hypercomplex algebras contains a wide variety of number systems, including quaternions and split quaternions, where the desired properness properties could be attained. The reduction in computational complexity is the major advantage of the proper algorithms. Also, simulation examples have shown the superiority of some 4D hypercomplex algebras over the standard quaternions. As a conclusion, the choice of the best algebra for processing a hypercomplex signal depends on their resultant statistical properties, and ignoring this issue would mean missing out on the opportunity of enhancing the effectiveness of the algorithms.

It is worth mentioning that, as indicated in Section 4, the application of the proposed adaptive filters requires the existence and computation of inverses in (25) and (28). To address this drawback, ill-conditioned matrices should be dealt with biased regression methods, which is one of the ongoing research challenges.

7. Acknowledgments

This work has been supported in part by the Project PID2021-124486NB-I00 of the "Plan Estatal de I+D+i", Ministerio de Educación y Ciencia, Spain, the I+D+i Project with reference number 1256911 of "Programa Operativo FEDER

Andalucía 2014-2020", Junta de Andalucía, and Project EI-FQM2-2023 of "Plan de Apoyo a la Investigación 2023-2024" from the University of Jaén.

Appendix A. Expressions of $\boldsymbol{\Upsilon}_{\boldsymbol{x}}(\boldsymbol{\alpha}, \boldsymbol{\beta})$ and $\boldsymbol{\Upsilon}_{\boldsymbol{x}\boldsymbol{x}^{\nu}}(\boldsymbol{\alpha}, \boldsymbol{\beta})$

$\boldsymbol{\Upsilon}_{\mathbf{x}}(\alpha,\beta) = \boldsymbol{\Upsilon}_{\mathbf{a}} + \boldsymbol{\Upsilon}_{\mathbf{b}} + \frac{\beta}{\alpha}\boldsymbol{\Upsilon}_{\mathbf{c}} + \frac{\beta}{\alpha}\boldsymbol{\Upsilon}_{\mathbf{d}} + \mathrm{i}\big(-\boldsymbol{\Upsilon}_{\mathbf{ab}} + \boldsymbol{\Upsilon}^{\mathsf{T}}_{\mathbf{ab}} - \frac{\beta}{\alpha}\boldsymbol{\Upsilon}_{\mathbf{cd}} + \frac{\beta}{\alpha}\boldsymbol{\Upsilon}^{\mathsf{T}}_{\mathbf{cd}}\big) + \mathrm{j}\big(\frac{1}{\alpha}\boldsymbol{\Upsilon}_{\mathbf{ac}} + \boldsymbol{\Upsilon}^{\mathsf{T}}_{\mathbf{ac}} - \frac{1}{\alpha}\boldsymbol{\Upsilon}_{\mathbf{bd}} - \boldsymbol{\Upsilon}^{\mathsf{T}}_{\mathbf{bd}}\big) + \mathrm{k}\big(\frac{1}{\alpha}\boldsymbol{\Upsilon}_{\mathbf{ad}} + \boldsymbol{\Upsilon}^{\mathsf{T}}_{\mathbf{ad}} + \frac{1}{\alpha}\boldsymbol{\Upsilon}_{\mathbf{bc}} + \boldsymbol{\Upsilon}^{\mathsf{T}}_{\mathbf{bc}}\big)$	$\boldsymbol{\Upsilon}_{\mathbf{xx}^{\mathrm{i}}}(\alpha,\beta) = \boldsymbol{\Upsilon}_{\mathbf{a}} + \boldsymbol{\Upsilon}_{\mathbf{b}} - \frac{\beta}{\alpha}\boldsymbol{\Upsilon}_{\mathbf{c}} - \frac{\beta}{\alpha}\boldsymbol{\Upsilon}_{\mathbf{d}} + \mathrm{i}\big(-\boldsymbol{\Upsilon}_{\mathbf{ab}} + \boldsymbol{\Upsilon}^{\mathsf{T}}_{\mathbf{ab}} + \frac{\beta}{\alpha}\boldsymbol{\Upsilon}_{\mathbf{cd}} - \frac{\beta}{\alpha}\boldsymbol{\Upsilon}^{\mathsf{T}}_{\mathbf{cd}}\big) + \mathrm{j}\big(-\frac{1}{\alpha}\boldsymbol{\Upsilon}_{\mathbf{ac}} + \boldsymbol{\Upsilon}^{\mathsf{T}}_{\mathbf{ac}} + \frac{1}{\alpha}\boldsymbol{\Upsilon}_{\mathbf{bd}} - \boldsymbol{\Upsilon}^{\mathsf{T}}_{\mathbf{bd}}\big) + \mathrm{k}\big(-\frac{1}{\alpha}\boldsymbol{\Upsilon}_{\mathbf{ad}} + \boldsymbol{\Upsilon}^{\mathsf{T}}_{\mathbf{ad}} - \frac{1}{\alpha}\boldsymbol{\Upsilon}_{\mathbf{bc}} + \boldsymbol{\Upsilon}^{\mathsf{T}}_{\mathbf{bc}}\big)$
$\boldsymbol{\Upsilon}_{\mathbf{xx}^{\mathrm{j}}}(\alpha,\beta) = \boldsymbol{\Upsilon}_{\mathbf{a}} - \boldsymbol{\Upsilon}_{\mathbf{b}} + \frac{\beta}{\alpha}\boldsymbol{\Upsilon}_{\mathbf{c}} - \frac{\beta}{\alpha}\boldsymbol{\Upsilon}_{\mathbf{d}} + \mathrm{i}\big(\boldsymbol{\Upsilon}_{\mathbf{ab}} + \boldsymbol{\Upsilon}^{\mathsf{T}}_{\mathbf{ab}} + \frac{\beta}{\alpha}\boldsymbol{\Upsilon}_{\mathbf{cd}} + \frac{\beta}{\alpha}\boldsymbol{\Upsilon}^{\mathsf{T}}_{\mathbf{cd}}\big) + \mathrm{j}\big(\frac{1}{\alpha}\boldsymbol{\Upsilon}_{\mathbf{ac}} + \boldsymbol{\Upsilon}^{\mathsf{T}}_{\mathbf{ac}} + \frac{1}{\alpha}\boldsymbol{\Upsilon}_{\mathbf{bd}} + \boldsymbol{\Upsilon}^{\mathsf{T}}_{\mathbf{bd}}\big) + \mathrm{k}\big(-\frac{1}{\alpha}\boldsymbol{\Upsilon}_{\mathbf{ad}} + \boldsymbol{\Upsilon}^{\mathsf{T}}_{\mathbf{ad}} + \frac{1}{\alpha}\boldsymbol{\Upsilon}_{\mathbf{bc}} - \boldsymbol{\Upsilon}^{\mathsf{T}}_{\mathbf{bc}}\big)$	$\boldsymbol{\Upsilon}_{\mathbf{xx}^{\mathrm{k}}}(\alpha,\beta) = \boldsymbol{\Upsilon}_{\mathbf{a}} - \boldsymbol{\Upsilon}_{\mathbf{b}} - \frac{\beta}{\alpha}\boldsymbol{\Upsilon}_{\mathbf{c}} + \frac{\beta}{\alpha}\boldsymbol{\Upsilon}_{\mathbf{d}} + \mathrm{i}\big(\boldsymbol{\Upsilon}_{\mathbf{ab}} + \boldsymbol{\Upsilon}^{\mathsf{T}}_{\mathbf{ab}} - \frac{\beta}{\alpha}\boldsymbol{\Upsilon}_{\mathbf{cd}} - \frac{\beta}{\alpha}\boldsymbol{\Upsilon}^{\mathsf{T}}_{\mathbf{cd}}\big) + \mathrm{j}\big(-\frac{1}{\alpha}\boldsymbol{\Upsilon}_{\mathbf{ac}} + \boldsymbol{\Upsilon}^{\mathsf{T}}_{\mathbf{ac}} - \frac{1}{\alpha}\boldsymbol{\Upsilon}_{\mathbf{bd}} + \boldsymbol{\Upsilon}^{\mathsf{T}}_{\mathbf{bd}}\big) + \mathrm{k}\big(\frac{1}{\alpha}\boldsymbol{\Upsilon}_{\mathbf{ad}} + \boldsymbol{\Upsilon}^{\mathsf{T}}_{\mathbf{ad}} - \frac{1}{\alpha}\boldsymbol{\Upsilon}_{\mathbf{bc}} - \boldsymbol{\Upsilon}^{\mathsf{T}}_{\mathbf{bc}}\big)$

Table A.7: Expressions for $\boldsymbol{\Upsilon}_{\mathbf{x}}(\alpha,\beta)$ and $\boldsymbol{\Upsilon}_{\mathbf{xx}^{\nu}}(\alpha,\beta)$, $\nu = \mathrm{i},\mathrm{j},\mathrm{k}$, in terms of the correlation functions of the components of $\mathbf{x}_r$.

Appendix B. Inverse of β-quaternion matrices

We give a procedure to obtain the inverse of a matrix by using the Schur complement. Let $\mathbf{M} \in \mathbb{Q}_{\beta}^{p+q\times p+q}$ be the matrix

$$\mathbf{M} = \begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{pmatrix}$$

with $\mathbf{A} \in \mathbb{Q}_{\beta}^{p\times p}$, $\mathbf{B} \in \mathbb{Q}_{\beta}^{p\times q}$, $\mathbf{C} \in \mathbb{Q}_{\beta}^{q\times p}$ and $\mathbf{D} \in \mathbb{Q}_{\beta}^{q\times q}$. If $\mathbf{D}$ is invertible then,

$$\mathbf{M}^{-1} = \begin{pmatrix} \left(\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C}\right)^{-1} & -\left(\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C}\right)^{-1}\mathbf{B}\mathbf{D}^{-1} \\ -\mathbf{D}^{-1}\mathbf{C}\left(\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C}\right)^{-1} & \mathbf{D}^{-1} + \mathbf{D}^{-1}\mathbf{C}\left(\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C}\right)^{-1}\mathbf{B}\mathbf{D}^{-1} \end{pmatrix}$$

Property 5. *Consider the matrices* $\mathbf{A} \in \mathbb{Q}_\beta^{p\times p}$, $\mathbf{B} \in \mathbb{Q}_\beta^{p\times q}$, $\mathbf{C} \in \mathbb{Q}_\beta^{q\times q}$ *and* $\mathbf{D} \in \mathbb{Q}_\beta^{q\times p}$. *Suppose that* $\mathbf{A}$ *and* $\mathbf{C}$ *are invertible then,*

$$(\mathbf{A}+\mathbf{BCD})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}\left(\mathbf{C}^{-1}+\mathbf{DA}^{-1}\mathbf{B}\right)^{-1}\mathbf{DA}^{-1}$$

The next result can be proved from Property 1.

Property 6.

1. *If* $\mathbf{A} \in \mathbb{Q}_\beta^{p\times q}, \mathbf{B} \in \mathbb{Q}_\beta^{q\times r}$ *then,* $(\mathbf{AB})^{H_\alpha} = \mathbf{B}^{H_\alpha}\mathbf{A}^{H_\alpha}$, *with* $\alpha = 1, -1$.
2. *If* $\mathbf{A} \in \mathbb{Q}_\beta^{p\times p}$ *then,* $(\mathbf{A}^\nu)^{-1} = (\mathbf{A}^{-1})^\nu, \quad \nu = \mathrm{i}, \mathrm{j}, \mathrm{k}$

Appendix C. Proper State-Space Models: Estimation

In this section, we consider two state-space models from which efficient algorithms for filtering, one-stage prediction and fixed-interval smoothing are suggested. Firstly, we study the first-order properness case and subsequently, we analyze the second-order case.

Consider a random signal vector $\mathbf{x}_t \in \mathbb{Q}_\beta^p$, $t \in \mathbb{N}$. Assume that $\mathbf{x}_t$ cannot be observed directly but it is observed through the equation

$$\mathbf{y}_t = \mathbf{x}_t + \mathbf{v}_t, \qquad t \geq 1$$

with $\mathbf{v}_t \in \mathbb{Q}_\beta^p$ a white noise independent of $\mathbf{x}_t$. Suppose that the signal verifies the state model

$$\mathbf{x}_{t+1} = \mathbf{F}_t\mathbf{x}_t + \mathbf{u}_t, \qquad t > 0$$

where $\mathbf{F}_t \in \mathbb{Q}_\beta^{p\times p}$ is deterministic and $\mathbf{u}_t \in \mathbb{Q}_\beta^p$ is a white noise independent of $\mathbf{x}_0$ and $\mathbf{v}_t$.

Filtering and one-stage predictor

Our first aim is to devise the filter and the one-stage predictor of $\mathbf{x}_t$ on the basis of the information provided by the observations $\{\mathbf{y}_1, \dots, \mathbf{y}_t\}$ under properness conditions. These estimates are denoted by $\hat{\mathbf{x}}^{\mathtt{QWL}}(t|t)$ and $\hat{\mathbf{x}}^{\mathtt{QWL}}(t+1|t)$, respectively. From Properties 3.1 and 3.3, Corollary 1 and the classical Kalman filter we can state the following result.

Theorem 6. *Suppose that* $\mathbf{x}_0$*,* $\mathbf{u}_t$ *and* $\mathbf{v}_t$ *are* $P^1_{\beta_0}$*-proper (respectively,* $N^1_{\beta_0}$*-proper), such that* $\boldsymbol{\Upsilon}^{\nu}_{\mathbf{u}_t\mathbf{u}_s}(\beta_0) = \mathbf{Q}_{\nu,t}\delta_{ts}$ *and* $\boldsymbol{\Upsilon}^{\nu}_{\mathbf{v}_t\mathbf{v}_s}(\beta_0) = \mathbf{R}_{\nu,t}\delta_{ts}$*,* $\nu = 1, 2$*. Taking* $\nu = 1$ *for the* $P^1_{\beta_0}$*-proper case and* $\nu = 2$ *for the* $N^1_{\beta_0}$*-proper case then, the filter and the one-stage predictor can be obtained, under the product associated to* β_0*, as*

$$\begin{aligned}
\hat{\mathbf{x}}^{QWL}(t+1|t) &= \mathbf{F}_t\hat{\mathbf{x}}^{QWL}(t|t) \\
\mathbf{P}_{t+1} &= \mathbf{F}_t\mathbf{L}_t\mathbf{F}_t^{H_{3-2\nu}} + \mathbf{Q}_{\nu,t} \\
\mathbf{K}_{t+1} &= \mathbf{P}_{t+1}\left(\mathbf{P}_{t+1} + \mathbf{R}_{\nu,t+1}\right)^{-1} \\
\hat{\mathbf{x}}^{QWL}(t+1|t+1) &= \hat{\mathbf{x}}^{QWL}(t+1|t) + \mathbf{K}_{t+1}\left[\mathbf{y}_{t+1} - \hat{\mathbf{x}}^{QWL}(t+1|t)\right] \\
\mathbf{L}_{t+1} &= \mathbf{P}_{t+1} - \mathbf{K}_{t+1}\mathbf{P}_{t+1}
\end{aligned} \tag{C.1}$$

for $t \geq 1$*, and with* $\hat{\mathbf{x}}^{QWL}(0|0) = \mathbf{0}_{p\times 1}$ *and* $\mathbf{L}_0 = \boldsymbol{\Upsilon}^{\nu}_{\mathbf{x}_0}(\beta_0)$*.*

Moreover, the error associated to the ith component of $\mathbf{x}_{t+1}$*,* $i = 1, \ldots, p$*, is given by*

$$\begin{aligned}
\epsilon_i^{QWL}(t+1) &= \|x_{t+1i} - \hat{x}_i^{QWL}(t+1|t+1)\|_3^2 = \frac{|\beta_0|+1}{2|\beta_0|}\mathcal{R}\{l_{t+1}(i,i)\} \\
\epsilon_i^{QWL}(t+1|t) &= \|x_{t+1i} - \hat{x}_i^{QWL}(t+1|t)\|_3^2 = \frac{|\beta_0|+1}{2|\beta_0|}\mathcal{R}\{p_{t+1}(i,i)\}
\end{aligned}$$

where $l_{t+1}(i,i)$ *and* $p_{t+1}(i,i)$ *are the* (i,i)*-th entries in the matrices* $\mathbf{L}_{t+1}$ *and* $\mathbf{P}_{t+1}$*, respectively.*

Fixed-interval Smoothing

Our second objective is to compute the fixed-interval smoother of $\mathbf{x}_t$ based on future data. Specifically, we estimate $\mathbf{x}_t$, $t \leq m$, from the information contained in the set $\{\mathbf{y}_1, \ldots, \mathbf{y}_m\}$ under properness conditions. This estimator will be denoted by $\hat{\mathbf{x}}^{\mathtt{QWL}}(t|m)$.

Taking into account Properties 3.1 and 3.3, Corollary 1 and the Rauch-Tung-Striebel formulas [11] we have the following result.

Theorem 7. *Under the conditions of Theorem 6, taking* $\nu = 1$ *for the* $P^1_{\beta_0}$*-proper case and* $\nu = 2$ *for the* $N^1_{\beta_0}$*-proper case, then the fixed-interval smoother* *is obtained for* $t \leq m$ *through the following recursions*

$$\hat{\mathbf{x}}^{QWL}(t|m) = \hat{\mathbf{x}}^{QWL}(t|t) + \mathbf{D}_t \left(\hat{\mathbf{x}}^{QWL}(t+1|m) - \hat{\mathbf{x}}^{QWL}(t+1|t)\right)$$
$$\mathbf{J}_t = \mathbf{L}_t + \mathbf{D}_t \left(\mathbf{J}_{t+1} - \mathbf{P}_{t+1}\right) \mathbf{D}_t^{H_{3-2\nu}}$$

where the computations are all performed whith the product associated to β_0*. Also,* $\mathbf{D}_t = \mathbf{L}_t \mathbf{F}_t^{H_{3-2\nu}} \mathbf{P}_{t+1}^{-1}$ *and* $\hat{\mathbf{x}}^{QWL}(t|t)$*,* $\hat{\mathbf{x}}^{QWL}(t+1|t)$*,* $\mathbf{L}_t$ *and* $\mathbf{P}_{t+1}$ *given in* (C.1).

The error associated to the components x_{ti}*,* $i = 1, \dots, p$*, is given by*

$$\epsilon_i^{QWL}(t|m) = \|x_{ti} - \hat{x}_i^{QWL}(t|m)\|_3^2 = \frac{|\beta_0| + 1}{2|\beta_0|} \mathcal{R}\{j_t(i,i)\}$$

where $j_t(i,i)$ *is the* (i,i)*-th entry in the matrix* $\mathbf{J}_t$*.*

Remark 13. *Similar algorithms to those suggested in Theorems 6 and 7 can be devised for the second-order properness case. Suppose* $\mathbf{x}_t$ *can be modelled via the state model*

$$\dot{\mathbf{x}}_{t+1} = \mathbf{G}_t \dot{\mathbf{x}}_t + \dot{\mathbf{u}}_t, \qquad t > 0$$

with $\dot{\mathbf{x}}_t$ *given in* (15)*,* $\mathbf{G}_t \in \mathbb{Q}_\beta^{2p \times 2p}$ *and* $\mathbf{u}_t$ *a white noise independent of* $\dot{\mathbf{x}}_0$ *and* $\mathbf{v}_t$*. Assume that* $\mathbf{x}_0$*,* $\mathbf{u}_t$ *and* $\mathbf{v}_t$ *are either* $P_{\beta_0}^2$*-proper or* $N_{\beta_0}^2$*-proper, such that* $\mathbf{\Upsilon}_{\dot{\mathbf{u}}_t \dot{\mathbf{u}}_s}^{\nu}(\beta_0) = \mathbf{E}_{\nu,t}\delta_{ts}$ *and* $\mathbf{\Upsilon}_{\dot{\mathbf{v}}_t \dot{\mathbf{v}}_s}^{\nu}(\beta_0) = \mathbf{W}_{\nu,t}\delta_{ts}$*. Then,* $\hat{\mathbf{x}}^{QWL}(t+1|t)$ *,* $\hat{\mathbf{x}}^{QWL}(t|t)$ *and* $\hat{\mathbf{x}}^{QWL}(t|m)$ *can be obtained by interchanging the vector* $\mathbf{y}_t$ *by* $\dot{\mathbf{y}}_t$ *and the matrices* $\mathbf{F}_t$*,* $\mathbf{Q}_{\nu,t}$ *and* $\mathbf{R}_{\nu,t}$ *by* $\mathbf{G}_t$*,* $\mathbf{E}_{\nu,t}$ *and* $\mathbf{W}_{\nu,t}$ *in Theorems 6 and 7.*

Remark 14. *The computational complexities of the algorithms proposed for the first and second-order properness cases are of orders* $O(p^3 n)$ *and* $O(8p^3 n)$*, respectively, with* n *the number of observations. However, the cost is of order* $O(64p^3 n)$ *when a QWL methodology is used.*

Appendix D. Properness Tests

Tools to check the properness properties in practice are needed. More concretely, we aim to test whether a random vector $\mathbf{x} \in \mathbb{Q}_\beta^p$ satisfies any of the properness conditions from the information supplied by a random sample. We start studying the first-order proper case. We have two objectives: 1) test if $\mathbf{x}$ is $P^1_{\beta_0}$-proper (or $N^1_{\beta_0}$-proper) and 2) estimate the suitable value of β_0 under which such properness properties could be satisfied by $\mathbf{x}$, if possible. To this end, we consider the following statistical hypothesis test:

$$\begin{aligned} &H_0: \ \exists\ \beta_0 \text{ such that } \mathbf{x} \text{ is } P^1_{\beta_0}\text{-proper (respectively, } \mathbf{x} \text{ is } N^1_{\beta_0}\text{-proper)} \\ &H_1: \nexists\ \beta_0 \text{ such that } \mathbf{x} \text{ is } P^1_{\beta_0}\text{-proper (respectively, } \mathbf{x} \text{ is } N^1_{\beta_0}\text{-proper)} \end{aligned} \tag{D.1}$$

Following a similar proof to that of Theorem 1 in [7], the next result can be proved.

Theorem 8. *Given n independent and identically distributed random samples $\mathbf{x}_{(1)}, \ldots, \mathbf{x}_{(n)}$ of a random vector $\mathbf{x} \in \mathbb{Q}^p_\beta$ such that $\mathbf{x}_r$ follows a Gaussian distribution, then the generalized likelihood ratio test (GLRT) for* (D.1) *($\nu = 1$ for the case of $P^1_{\beta_0}$-properness and $\nu = 2$ for the case of $N^1_{\beta_0}$-properness) is given by*

$$\phi_\nu(\mathbf{x}_{(1)}, \ldots, \mathbf{x}_{(n)}) = -n\left[\ln\left(\det(\hat{\mathbf{\Upsilon}}^\nu_{\bar{\mathbf{x}}}(\beta_0))\right) - \ln\left(\det(\mathbf{\Upsilon}^\nu_{\mathbb{Q}_1}(\beta_0))\right)\right], \qquad \nu = 1, 2 \tag{D.2}$$

where the determinant is computed under the product associated to β_0,

$$\hat{\mathbf{\Upsilon}}^\nu_{\bar{\mathbf{x}}}(\beta_0) = \mathcal{T}_p \hat{\mathbf{\Upsilon}}_{\mathbf{x}_r} \mathcal{T}_p^{H_{3-2\nu}}, \qquad \nu = 1, 2 \tag{D.3}$$

with $\hat{\mathbf{\Upsilon}}_{\mathbf{x}_r}$ the sample autocorrelation matrix and

$$\mathbf{\Upsilon}^\nu_{\mathbb{Q}_1}(\beta_0) = \operatorname{diag}\left(\hat{\mathbf{\Upsilon}}^\nu_{\mathbf{x}}(\beta_0), \hat{\mathbf{\Upsilon}}^{\nu\mathrm{k}}_{\mathbf{x}}(\beta_0), \hat{\mathbf{\Upsilon}}^{\nu\mathrm{i}}_{\mathbf{x}}(\beta_0), \hat{\mathbf{\Upsilon}}^{\nu\mathrm{j}}_{\mathbf{x}}(\beta_0)\right), \qquad \nu = 1, 2$$

The parameter β_0 can be estimated as

$$\begin{aligned} \beta_0 &= \max_{\beta>0}\left[c_1 - \frac{n}{2}\left[\ln\left(\det(\mathbf{\Upsilon}^1_{\mathbb{Q}_1}(\beta))\right) + 4p\right]\right], \qquad \nu = 1 \\ \beta_0 &= \max_{\beta<0}\left[c_1 - \frac{n}{2}\left[\ln\left(\det(\mathbf{\Upsilon}^2_{\mathbb{Q}_1}(\beta))\right) + 4p\right]\right], \qquad \nu = 2 \end{aligned} \tag{D.4}$$

with $c_1 = -2np\ln(2\pi) + np\ln(16|\beta|)$.

Also, assuming that H_0 is true, the distribution of the statistics $\phi_1(\mathbf{x}_{(1)}, \ldots, \mathbf{x}_{(n)})$ and $\phi_2(\mathbf{x}_{(1)}, \ldots, \mathbf{x}_{(n)})$ tend to a chi-squared distribution with degrees of freedom equal to $6p^2 + p - 1$ and $6p^2 + 3p - 1$, respectively, as the sample size tends to infinity.

We consider now the case of second-order properness. Similarly to the above case, we pursue the twofold objective of: 1) testing if $\mathbf{x}$ is $P^2_{\beta_0}$-proper (or $N^2_{\beta_0}$-proper) and 2) estimating the suitable value of β_0 under which $P^2_{\beta_0}$- or $N^2_{\beta_0}$-properness properties could be satisfied by $\mathbf{x}$, if possible. For that, consider the hypothesis test problem as follows:

$$\begin{aligned} &H_0: \ \exists\ \beta_0 \text{ such that } \mathbf{x} \text{ is } P^2_{\beta_0}\text{-proper (respectively, } \mathbf{x} \text{ is } N^2_{\beta_0}\text{-proper)} \\ &H_1: \nexists\ \beta_0 \text{ such that } \mathbf{x} \text{ is } P^2_{\beta_0}\text{-proper (respectively, } \mathbf{x} \text{ is } N^2_{\beta_0}\text{-proper)} \end{aligned} \tag{D.5}$$

Then, both objectives are achieved in the next result.

Theorem 9. *Given n independent and identically distributed random samples $\mathbf{x}_{(1)}, \dots, \mathbf{x}_{(n)}$ of a random vector $\mathbf{x} \in \mathbb{Q}^p_\beta$ such that $\mathbf{x}_r$ follows a Gaussian distribution, then the GLRT statistic for* (D.5) *($\nu = 1$ for the case of $P^2_{\beta_0}$-properness and $\nu = 2$ for the case of $N^2_{\beta_0}$-properness) is given by*

$$\psi_\nu(\mathbf{x}_{(1)}, \dots, \mathbf{x}_{(n)}) = -n\left[\ln\left(\det(\hat{\mathbf{\Upsilon}}^\nu_{\bar{\mathbf{x}}}(\beta_0))\right) - \ln\left(\det(\mathbf{\Upsilon}^\nu_{\mathcal{Q}_2}(\beta_0))\right)\right], \qquad \nu = 1, 2 \tag{D.6}$$

where the determinant is computed under the product associated to β_0, $\hat{\mathbf{\Upsilon}}^\nu_{\bar{\mathbf{x}}}(\beta_0)$ is given in (D.3) *and*

$$\mathbf{\Upsilon}^\nu_{\mathcal{Q}_2}(\beta_0) = \operatorname{diag}\left(\hat{\mathbf{\Upsilon}}^\nu_{\dot{\mathbf{x}}}(\beta_0), \hat{\mathbf{\Upsilon}}^{\nu\mathrm{i}}_{\dot{\mathbf{x}}}(\beta_0)\right), \qquad \nu = 1, 2$$

with $\dot{\mathbf{x}}$ given in (15).

The parameter β_0 can be estimated as

$$\begin{aligned} \beta_0 &= \max_{\beta>0}\left[c_1 - \frac{n}{2}\left[\ln\left(\det(\mathbf{\Upsilon}^1_{\mathcal{Q}_2}(\beta))\right) + 4p\right]\right], \qquad \nu = 1 \\ \beta_0 &= \max_{\beta<0}\left[c_1 - \frac{n}{2}\left[\ln\left(\det(\mathbf{\Upsilon}^2_{\mathcal{Q}_2}(\beta))\right) + 4p\right]\right], \qquad \nu = 2 \end{aligned} \tag{D.7}$$

with $c_1 = -2np\ln(2\pi) + np\ln(16|\beta|)$.

Also, assuming that H_0 is true, the distribution of the statistics $\psi_1(\mathbf{x}_{(1)}, \dots, \mathbf{x}_{(n)})$ and $\psi_2(\mathbf{x}_{(1)}, \dots, \mathbf{x}_{(n)})$ tend to a chi-squared distribution with degrees of freedom equal to $4p^2-1$ and $4p^2+2p-1$, respectively, as the sample size tends to infinity.

Appendix E. Proofs of the Theorems

Appendix E.1. Proof of Theorem 1

Firstly, from (10) and (11), (16) follows. Also, $\hat{y}^{\texttt{QWL}}$ and $\epsilon^{\texttt{QWL}}$ are independent of the value of β due to the fact that the space of random β-quaternions $(\mathbb{Q}_\beta, d_3)$ and $L_2^4(\Omega)$ are isomorphic and then, the QWL and the real processing are equivalent.

The expressions (18) and (19) are obtained in a similar way as (16).

The proof of point 4 is immediate since $\mathcal{G}_{\mathbf{x}} \subseteq \mathcal{G}_{\dot{\mathbf{x}}} \subseteq \mathcal{G}_{\bar{\mathbf{x}}}$.

Next, point 5 is proved under the $P^1_{\beta_0}$-properness condition. The proof of the $N^1_{\beta_0}$-proper case is similar. From (3) and (6) we get

$$\begin{aligned} \boldsymbol{\Upsilon}^3_{y\bar{\mathbf{x}}}(\beta_0) &= [1,\mathrm{i},\mathrm{j},\mathrm{k}]\boldsymbol{\Upsilon}_{\mathbf{y}_r\mathbf{x}_r}\mathcal{T}_p^{\texttt{H}_{\beta_0}}, \qquad \boldsymbol{\Upsilon}^3_{\bar{\mathbf{x}}}(\beta_0) = \mathcal{T}_p\boldsymbol{\Upsilon}_{\mathbf{x}_r}\mathcal{T}_p^{\texttt{H}_{\beta_0}} \\ \boldsymbol{\Upsilon}^1_{y\bar{\mathbf{x}}}(\beta_0) &= [1,\mathrm{i},\mathrm{j},\mathrm{k}]\boldsymbol{\Upsilon}_{\mathbf{y}_r\mathbf{x}_r}\mathcal{T}_p^{\texttt{H}_1}, \qquad \boldsymbol{\Upsilon}^1_{\bar{\mathbf{x}}}(\beta_0) = \mathcal{T}_p\boldsymbol{\Upsilon}_{\mathbf{x}_r}\mathcal{T}_p^{\texttt{H}_1} \end{aligned} \tag{E.1}$$

Assume that $\mathbf{x}$ is only $P^1_{\beta_0}$-proper. Then, from (16), (E.1), Definition 5 and (7), we have that

$$\begin{aligned} [\mathbf{l}_{1\beta_0},\mathbf{l}_{2\beta_0},\mathbf{l}_{3\beta_0},\mathbf{l}_{4\beta_0}] &= \boldsymbol{\Upsilon}^3_{y\bar{\mathbf{x}}}(\beta_0)\boldsymbol{\Upsilon}^{3^{-1}}_{\bar{\mathbf{x}}}(\beta_0) = \boldsymbol{\Upsilon}^1_{y\bar{\mathbf{x}}}(\beta_0)\boldsymbol{\Upsilon}^{1^{-1}}_{\bar{\mathbf{x}}}(\beta_0) \\ &= [\boldsymbol{\Upsilon}^1_{y\mathbf{x}}(\beta_0), \boldsymbol{\Upsilon}^1_{y\mathbf{x}^{\mathrm{k}}}(\beta_0), \boldsymbol{\Upsilon}^1_{y\mathbf{x}^{\mathrm{i}}}(\beta_0), \boldsymbol{\Upsilon}^1_{y\mathbf{x}^{\mathrm{j}}}(\beta_0)] \\ &\quad \operatorname{diag}\left(\boldsymbol{\Upsilon}^{1^{-1}}_{\mathbf{x}}(\beta_0), \boldsymbol{\Upsilon}^{1\mathrm{k}^{-1}}_{\mathbf{x}}(\beta_0), \boldsymbol{\Upsilon}^{1\mathrm{i}^{-1}}_{\mathbf{x}}(\beta_0), \boldsymbol{\Upsilon}^{1\mathrm{j}^{-1}}_{\mathbf{x}}(\beta_0)\right) \end{aligned} \tag{E.2}$$

and taking Property 6 in Appendix B into account we get (20). Finally, if y and $\mathbf{x}$ are also cross $P^1_{\beta_0}$-proper, by applying Definition 5 to (E.2), we have that $\hat{y}^{\texttt{QWL}} = \hat{y}^{\texttt{Q}_1}$.

The proof of point 6 is similar to that of the previous point.

Appendix E.2. Proof of Theorem 2

Denote the components of $\tilde{\mathbf{f}}_\nu$ in (21) by $\tilde{f}_{\nu,i}$, $i = 1,\ldots,p$. Thus, solving the adaptive filter problem (21) is equivalent to solve the optimization problem: $\min_{\tilde{f}_{\nu,i}}\{\tilde{\xi}^{\texttt{Q}_1}_\nu\}$, and hence, it is equivalent to find the projection $\tilde{\boldsymbol{\zeta}}^{\texttt{Q}_1}_\nu$ given in (23). Then, by using (14) and Property 6.1 in Appendix B, we get (25) and (26).

Appendix E.3. Proof of Theorem 4

Define

$$\vec{\boldsymbol{\chi}}_\nu(i) = \boldsymbol{\chi}_i - \tilde{\boldsymbol{\chi}}_\nu^{\mathtt{Q}_1}(i)$$

then, $\{\vec{\boldsymbol{\chi}}_\nu(i)\}_{i=1}^{p+1}$ is a orthogonal basis of $\mathcal{C}_{\mathcal{A}_{p+1}}$. Hence, (34) follows from (14). Also,

$$\prec \boldsymbol{\zeta}, \vec{\boldsymbol{\chi}}_\nu(k+1) \succ_\nu = \prec \tilde{\boldsymbol{\zeta}}_\nu^{\mathtt{Q}_1}(p+1), \vec{\boldsymbol{\chi}}_\nu(k+1) \succ_\nu = \phi_{\nu,p+1-k}\pi_{\nu,k}, \qquad 0 \leq k \leq p \tag{E.3}$$

with $\pi_{\nu,k} = \prec \vec{\boldsymbol{\chi}}_\nu(k+1), \vec{\boldsymbol{\chi}}_\nu(k+1) \succ_\nu$ given in (33).

By using (31) and Property 4, (E.3) can be rewritten as

$$\phi_{\nu,p+1-k}\pi_{\nu,k} = \prec \boldsymbol{\zeta}, \boldsymbol{\chi}_{k+1} \succ_\nu - \sum_{j=0}^{k-1} \prec \boldsymbol{\zeta}, \vec{\boldsymbol{\chi}}_\nu(j+1) \succ_\nu l_{\nu,k-j}^{(3-2\nu)}(k)$$

and thus, (35) is obtained from (E.3).

Finally, from (34) and the ortogonality of the set $\{\vec{\boldsymbol{\chi}}_\nu(i)\}_{i=1}^{p+1}$, we have (36).

Appendix E.4. Proof of Theorem 5

From (37), we get $\mathbf{P}_\nu^{-1}(n+1) = \mathbf{P}_\nu^{-1}(n) + \boldsymbol{\kappa}_{n+1}\boldsymbol{\kappa}_{n+1}^{\mathtt{H}_{3-2\nu}}$. Then, by setting $\mathbf{A} = \mathbf{P}_\nu^{-1}(n)$, $\mathbf{B} = \boldsymbol{\kappa}_{n+1}$, $\mathbf{C} = 1$ and $\mathbf{D} = \boldsymbol{\kappa}_{n+1}^{\mathtt{H}_{3-2\nu}}$ in Property 5 of Appendix B, we obtain (39).

Now, from (37) and (39), we have

$$\begin{aligned}
\tilde{\mathbf{f}}_\nu(n+1) &= \boldsymbol{\zeta}^{\mathtt{T}}(n+1)\mathbf{X}^{\mathtt{H}_{3-2\nu}}(n+1)\mathbf{P}_\nu(n+1) \\
&= \left(\boldsymbol{\zeta}^{\mathtt{T}}(n)\mathbf{X}^{\mathtt{H}_{3-2\nu}}(n) + y_{(n+1)}\boldsymbol{\kappa}_{n+1}^{\mathtt{H}_{3-2\nu}}\right)\mathbf{P}_\nu(n+1) \\
&= \tilde{\mathbf{f}}_\nu(n) - \tilde{\mathbf{f}}_\nu(n)\boldsymbol{\kappa}_{n+1}\lambda_\nu(n)\boldsymbol{\kappa}_{n+1}^{\mathtt{H}_{3-2\nu}}\mathbf{P}_\nu(n) \\
&\quad + y_{(n+1)}\left(1 - \boldsymbol{\kappa}_{n+1}^{\mathtt{H}_{3-2\nu}}\mathbf{P}_\nu(n)\boldsymbol{\kappa}_{n+1}\lambda_\nu(n)\right)\boldsymbol{\kappa}_{n+1}^{\mathtt{H}_{3-2\nu}}\mathbf{P}_\nu(n) \\
&= \tilde{\mathbf{f}}_\nu(n) - \tilde{\mathbf{f}}_\nu(n)\boldsymbol{\kappa}_{n+1}\lambda_\nu(n)\boldsymbol{\kappa}_{n+1}^{\mathtt{H}_{3-2\nu}}\mathbf{P}_\nu(n) + y_{(n+1)}\lambda_\nu(n)\boldsymbol{\kappa}_{n+1}^{\mathtt{H}_{3-2\nu}}\mathbf{P}_\nu(n)
\end{aligned}$$

and (38) follows.

References

[1] D. Alfsmann, On families of 2N-dimensional hypercomplex algebras suitable for digital signal processing, in: Proc. 14th European Signal Process. Conf., 2006, pp. 1-4.

[2] D. Alfsmann, H.G. Göckler, S.J. Sangwine, and T.A. Ell, Hypercomplex algebras in digital signal processing: benefits and drawbacks, in: Proc. 15th European Signal Process. Conf., 2007, pp. 1322-1326.

[3] S.L. Hahn and K.M. Snopek, Complex and Hypercomplex Analytic Signals: Theory and Applications, Artech House Publishers, 2016.

[4] E. Guizzo, T. Weyde, S. Scardapane, and D. Comminiello, Learning speech emotion representations in the quaternion domain, IEEE/ACM Trans. Audio, Speech, and Language Process., 31 (2023) 1200-1212.

[5] S.H. Pei, J.H. Chang, and J.J. Ding, Commutative reduced biquaternions and their Fourier transform for signal and image processing applications, IEEE Trans. Signal Process. 52 (7) (2004) 2012-2031.

[6] D. Schulz, J. Seitz, and J. P. C. L. da Costa, Widely linear SIMO filtering for hypercomplex numbers, 2011 IEEE Inform. Theory Workshop, 2011, pp. 390-395.

[7] J. Navarro-Moreno, R.M. Fernández-Alcalá J.D. López-Jiménez, and J.C. Ruiz-Molina, Tessarine signal processing under the $\mathbb{T}$-properness condition, J. Franklin Inst. 357 (14) (2020) 10100-10126.

[8] G. Vieira and M.E. Valle, A general framework for hypercomplex-valued extreme learning machines, J. Comp. Math. Data Science 3 (2022) 100032.

[9] E. Grassucci, G. Mancini, C. Brignone, A. Uncini, and D. Comminiello, Dual quaternion ambisonics array for six-degree-of-freedom acoustic representation, Pattern Recognit. Lett. 166 (2023) 24-30.

[10] F. Catoni, D. Boccaletti, R. Cannata, V. Catoni, E. Nichelatti, and P. Zampetti, The Mathematics of Minkowski Space-Time. With an Introduction to Commutative Hypercomplex Numbers, Birkhäuser Verlag, Berlin, 2008.

[11] T. Kailath, A.H. Sayed, and B. Hassibi, Linear Estimation, Pretince Hall, New Jersey, 2000.

[12] A.H. Sayed, Fundamentals of Adaptive Filtering, Wiley, New Jersey, 2003.

[13] T. Mizoguchi and I. Yamada, An algebraic translation of Cayley-Dickson linear systems and its applications to online learning, IEEE Trans. Signal Process. 62 (6) (2014) 1438–1453.

[14] F. Ortolani, D. Comminiello, M. Scarpiniti, and A. Uncini, Frequency domain adaptive filtering in hypercomplex systems, in: Proc. Int. Workshop Neural Netw., 2015, pp. 47–56.

[15] F. Ortolani, M. Scarpiniti, D. Comminiello, and A. Uncini, On 4-dimensional hypercomplex algebras in adaptive signal processing, in: Neural Advances Process. Nonlinear Dynamic Signals, 2017, pp. 131-140.

[16] A. Li, B. Yang, H. Huo, and F. Hussain, Hypercomplex graph collaborative filtering, in: Proc. ACM Web Conference, 2022, pp. 1914–1922.

[17] C. Cheong Took and D.P. Mandic, A quaternion widely linear adaptive filter, IEEE Trans. Signal Process. 58 (8) (2010) 4427-4431.

[18] W. Liu, Channel equalization and beamforming for quaternion-valued wireless communication systems, J. Franklin Inst. 354 (18) (2017) 8721-8733.

[19] M. Xiang, C. Cheong Took, and D. P. Mandic. Cost-effective quaternion minimum mean square error estimation: From widely linear to four-channel processing, Signal Process. 136 (2017) 81-91.

[20] M. Xiang, S. Kanna, and D. P. Mandic, Performance analysis of quaternion-valued adaptive filters in non-stationary environments, IEEE Trans. Signal Process. 66 (6) (2018) 1566-1579.

[21] G. Wang and R. Xue, Quaternion filtering based on quaternion involutions and its application in signal processing, IEEE Access 7 (2019) 149068-149079.

[22] T. Ogunfunmiand and C. Safarian, The quaternion stochastic information gradient algorithm for nonlinear adaptive systems, IEEE Trans. Signal Process. 67 (23) (2019) 5909-5921.

[23] D. Lin, Q. Zhang, S. Chen, and S. Wang, Maximum total quaternion correntropy for adaptive filtering, IEEE Trans. Signal Process. 70 (2022) 4967-4980.

[24] E. C. Mengüç, N. Acir, and D. P. Mandic, A Class of Online Censoring Based Quaternion-Valued Least Mean Square Algorithms, IEEE Signal Process. Lett. 30 (2023) 244-248.

[25] C. Cheong Took and D.P. Mandic, Augmented second-order statistics of quaternion random signals, Signal Process. 91 (2011) 214-224.

[26] T. Nitta, M. Kobayashi, and D.P. Mandic, Hypercomplex widely linear estimation through the lens of underpinning geometry, IEEE Trans. Signal Process. 67 (15) (2019) 3985-3994.

[27] Y. Xia, C. Jahanchahi, T. Nitta, and D. P. Mandic, Performance bounds of quaternion estimators, IEEE Trans. Neural Netw. Learn. Syst. 26 (12) (2015) 3287-3292.

[28] M. Jafari and Y. Yaili, Matrix theory over the split quaternions, Int. J. Geometry 3 (2) (2014) 57-69.

[29] I. Ali, A note on quaternion matrices and split quaternion matrix pencils, J. Appl. Math. Comput. 58 (2018) 323–334.

[30] P. O. Amblard and N. L. Bihan, On Properness of Quaternion Valued Random Variables, in: Proc. IMA Conf. Math. Signal Process., 2004, pp. 23–26.

[31] P.J. Brockwell and R.A. Davis, Time Series: Theory and Methods, Springer-Verlag, New York, 2nd edition, 1991.

[32] S.P. Talebi, S. Werner, and D.P. Mandic, Quaternion-valued distributed filtering and control, IEEE Trans. Autom. Control 65 (10) (2020) 4246-4256.