Available online at www.sciencedirect.com

ScienceDirect

Journal of the Franklin Institute 357 (2020) 10100–10126

www.elsevier.com/locate/jfranklin

Tessarine signal processing under the $\mathbb{T}$-properness condition

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

Department of Statistics and Operations Research, University of Jaén, 23071 Jaén, Spain

Received 9 March 2020; received in revised form 22 June 2020; accepted 4 August 2020

Available online 8 August 2020

Abstract

The paper analyzes the processing of 4D commutative hypercomplex or tessarine signals under properness conditions. Firstly, the concept of $\mathbb{T}$-properness is introduced and a procedure to test experimentally whether a tessarine random signal is proper or not is proposed. Then, for the class of $\mathbb{T}$-proper signals, the linear minimum mean square error estimation problem is addressed. In this regard, it should be highlighted that although the tessarine algebra is not a Hilbert space, a metric which guarantees the existence and unicity of the optimal estimator is defined. Moreover, the equivalence, under $\mathbb{T}$-properness conditions, between the optimal estimator based on a tessarine widely linear processing and the one based on a tessarine strictly linear (TSL) processing is also shown, attaining thus a notable reduction in computational burden. Finally, two $\mathbb{T}$-proper models, a TSL state-space model and a TSL stationary model, from which the optimal estimator can be recursively obtained are considered. In both cases, simulated examples are developed where the superiority of TSL processing over the counterparts in the quaternion domain is exhibited.

* Corresponding author.

E-mail addresses: jnavarro@ujaen.es (J. Navarro-Moreno), rmfernan@ujaen.es (R.M. Fernández-Alcalá), jdomingo@ujaen.es (J.D. Jiménez-López), jcruiz@ujaen.es (J.C. Ruiz-Molina).

https://doi.org/10.1016/j.jfranklin.2020.08.002

0016-0032/

1. Introduction

Over the last few years, the use of hypercomplex algebras has gained a growing reputation in the statistical processing of multidimensional signals, as evidenced by recent research work (see, e.g., [1,2], and references therein).

Hypercomplex signals arise as n-dimensional generalizations of complex signals in relation to Clifford and Cayley-Dickson algebras, which are formed by a real part and a vector part of imaginary variables. Essentially, families of hypercomplex algebras of order 2^n are defined on the real field, giving rise to structures that allow the modeling of a wide variety of multidimensional physical phenomena in a simple entity that naturally collects the information shared between its components and geometric characteristics thereof [3]. Also, hypercomplex algebras have shown their superiority over the real field in some applied problems (see, e.g., [4,5]).

The natural extension of real and complex numbers is the 4D hypercomplex numbers. However, this generalization cannot be done without losing some peculiarities of algebraic operations. In particular, one must either renounce the commutative property of the product or acknowledge that the product among some non-zero numbers is zero [6]. Following the first possibility, Hamilton [7] introduced the non-commutative quaternions which have become the most famous and studied system of hypercomplex numbers. The quaternion algebra plays an important role in quantum physics, kinematic, differential geometry, image and signal processing, etc. (e.g., [8–13]). Nevertheless, the lack of commutativity restricts the applications of quaternions since all results about complex numbers cannot be generalized in quaternions [14].

Alternatively, commutative algebras have been constructed with structures suitable for signal processing. The first family of commutative 4D hypercomplex numbers, the tessarines or reduced biquaternions, was introduced by Segre [15]. The first attempt to apply tessarine algebra to signal processing was made by Schütte in 1990 [16] and since then, tessarines have been satisfactorily applied in digital signal and image processing [14,17–22], neural networks [5,23,24], or electromagnetism [25].

Both 4D hypercomplex algebras, quaternions and tessarines, have some advantages compared to each other and none of them is the best in all circumstances and situations [1,26–28]. In favor of the set of tessarines is the commutative property of the multiplication which can reduce the complexity of many problems [19,21]. Moreover, the existence of zero divisors can, on the one hand, cause difficulties and, on the other, be an advantage in terms of, for example, the potential benefits for reduced computational burden or to admit an orthogonal decomposition [1,3,19]. Additional advantages of tessarines over quaternions in signal processing can be found in [14,19,21].

Once chosen a specific 4D hypercomplex algebra as a study framework, the next step is to decide what class of signal processing has to be used. Actually, the kind of signal properness determines the type of treatment that one has to follow in its processing. The concepts of signal properness are well established in the quaternion domain [29]. The most general processing is the so-called widely linear (WL), which is suitable for improper signals, and is also equivalent to the linear processing in $\mathbb{R}^4$ [30]. Other types of processing are appropriate when the signal presents properness properties. $\mathbb{Q}$-properness or $\mathbb{C}^i$-properness constitute the two principal types of quaternion properness and the optimal processing is then reduced to the strictly linear (SL) or semi-widely linear (SWL) processing, respectively. The main practical repercussion of both SL and SWL processing is the significative reduction in the computational burden of the associated algorithms in comparison with the WL ones [31–33]. Notably, this saving in computational complexity cannot be attained when the algorithms are formulated in the real domain.

In contrast with the quaternion domain, the concept of properness and the use of augmented statistics has not yet been exploited in the tessarine signal processing. In this way, the purpose of this paper is to introduce the concept of properness for tessarine random signals and analyze its implications on the processing. Thus, we first extend the definition of $\mathbb{Q}$-properness for quaternion random signals to the tessarine domain and propose a methodology to test whether a tessarine random vector is proper or not. In particular, a generalized likelihood ratio test (GLRT) is provided for testing the $\mathbb{T}$-properness of a set of data and the performance of such a test is illustrated by means of a simulation example. As in the quaternion domain, the more general scenario is the improper case and thus, the optimal processing is the tessarine widely linear (TWL) processing. However, under $\mathbb{T}$-properness conditions, a tessarine strictly linear (TSL) processing is then the more appropriate one since it entails a reduction in the computational burden of the problem.

To complete our study, the estimation problem is addressed for $\mathbb{T}$-proper tessarine random signal vectors. The usual approach to tackle the linear minimum mean square error (MMSE) estimation problem in the real, complex or quaternion domains has been by exploiting their Hilbert space structures. For these algebras, it is well-known that the optimal linear estimator is the orthogonal projection of the signal of interest onto the set of observations. However, the set of tessarine random variables is not a Hilbert space and thus, the projection theorem cannot be applied. Hence, we initially prove that it is a metric space satisfying the necessary properties to guarantee the existence and uniqueness of this projection. Then, the general forms of the optimal estimators by using both a TWL and a TSL processing are provided. Interestingly, under $\mathbb{T}$-properness conditions, the benefits of the TWL processing dissipate because in such a case the TSL processing leads to the optimal estimator whereas it entails a lower computational cost.

At this point, although the solution to the estimation problem is completely determined, it may imply computational difficulties because of the inversion of a matrix whose dimension is closely related to the number of available observations [34–36]. In order to overcome this handicap, we propose two models, a TSL state-space model and a TSL stationary model, where the TSL estimator is the optimal one, and devise efficient algorithms for the computation of such TSL estimates. Finally, two simulation examples where the TSL estimation algorithms proposed here present a better behavior than the counterparts in the quaternion domain are provided.

In essence, the contributions of this paper can be summarized in the following points: *(i)* the definition and characterization of the $\mathbb{T}$-properness property in the tessarine domain are provided and a $\mathbb{T}$-properness test is proposed; *(ii)* a new analytic framework is defined to ensure the existence and unicity of the optimal estimator in the tessarine domain; *(iii)* the concepts of TWL and TSL processing are formulated and the advantages of TSL processing under $\mathbb{T}$-proper conditions are exploited in linear estimation problems; and, *(iv)* two $\mathbb{T}$-proper models, the TSL state-space model and the TSL stationary model, are devised. In the case of a TSL state-space model, a Kalman filter like algorithm is devised and for the TSL stationary model the Durbin-Levison recursion is generalized for the purpose of estimation.

The rest of the paper is structured as follows. Section 2 introduces the notation and briefly reviews the main aspects of the quaternion signal processing. Section 3 outlines some elements of tessarine algebra which set the bases of our developments. Then, in Section 4, our definition of properness for tessarine random variables is introduced and a $\mathbb{T}$-properness test is provided. In Section 5, a metric on the set of tessarine random variables, which guarantees the existence and uniqueness of the projection of any tessarine random signal onto the closed linear space spanned by the observation set, is defined and a method for obtaining such a projection under a TWL as well as a TSL processing is proposed. Next, Section 6 provides efficient algorithms for the estimation of $\mathbb{T}$- proper tessarine random signals under a TSL state-space model and a TSL stationary model. Finally, two simulation examples illustrating the superiority of TSL estimators over the quaternion SL (QSL) ones have been included. Note that, to improve readability, all technical proofs have been deferred to an Appendix A.

2. Preliminaries

2.1. Notation

Throughout this paper, all the random variables are assumed to have zero-mean. Next, we introduce the basic notation. We use boldface uppercase letters to denote matrices, boldface lowercase letters for column vectors, and lightface lowercase letters for scalar quantities. Superscripts "$*$", "$\mathtt{T}$" and "$\mathtt{H}$" represent the tessarine (or quaternion) conjugate, transpose and Hermitian transpose, respectively. The real part of a tessarine (or quaternion) will be denoted by $\mathcal{R}\{\cdot\}$.

Moreover, the notation $\mathbb{Z}$, $\mathbb{R}$, $\mathbb{T}$ and $\mathbb{Q}$ is used to denote the set of integer, real, tessarine and quaternion numbers, respectively. In this sense, $\mathbf{A} \in \mathbb{R}^{p\times q}$ (respectively, $\mathbf{A} \in \mathbb{T}^{p\times q}$ or $\mathbf{A} \in \mathbb{Q}^{p\times q}$) means that $\mathbf{A}$ is a real (respectively, tessarine or quaternion) $p \times q$ matrix, and similarly $\mathbf{r} \in \mathbb{R}^p$ (respectively, $\mathbf{r} \in \mathbb{T}^p$ or $\mathbf{r} \in \mathbb{Q}^p$) means that $\mathbf{r}$ is a p-dimensional real (respectively, tessarine or quaternion) vector. δ_{nl} is the Kronecker delta function, which is equal to one if $l = n$, and zero otherwise. Finally, $E[\cdot]$ is the expectation operator, $\mathbf{0}_{p\times q}$ denotes the $p \times q$ zero matrix, $\mathbf{I}_p$ represents the identity matrix of dimension p and "$\otimes$" stands for the Kronecker product.

2.2. Review of the quaternion processing

Definition 1. A random variable $x \in \mathbb{Q}$ is defined by

$$x = a + \mathrm{i}b + \mathrm{j}c + \mathrm{k}d$$

where the random variables $a, b, c, d \in \mathbb{R}$ and the imaginary units (i, j, k) satisfy the identities

$$\mathrm{ij} = \mathrm{k}, \quad \mathrm{jk} = \mathrm{i}, \quad \mathrm{ki} = \mathrm{j}, \quad \mathrm{i}^2 = \mathrm{j}^2 = \mathrm{k}^2 = -1 \tag{1}$$

The conjugate of $x \in \mathbb{Q}$ is defined by $x^* = a - \mathrm{i}b - \mathrm{j}c - \mathrm{k}d$ and its involution over a pure unit quaternion ν is $x^\nu = -\nu x \nu$, with $\nu = \mathrm{i}, \mathrm{j}, \mathrm{k}$.

The autocorrelation and cross-correlation of the random variables $x, y \in \mathbb{Q}$ are denoted by $\lambda_x = E[xx^*]$ and $\lambda_{xy} = E[xy^*]$, respectively. Similarly, the autocorrelation and cross-correlation matrices of the random vectors $\mathbf{x} \in \mathbb{Q}^{p_1}$ and $\mathbf{y} \in \mathbb{Q}^{p_2}$ are denoted by $\mathbf{\Lambda_x} = E[\mathbf{x}\mathbf{x}^{\mathtt{H}}]$ and $\mathbf{\Lambda_{xy}} = E[\mathbf{x}\mathbf{y}^{\mathtt{H}}]$, respectively.

Definition 2. A random vector $\mathbf{x} \in \mathbb{Q}^p$ is said to be $\mathbb{Q}$-proper if, and only if, the matrices $\mathbf{\Lambda}_{\mathbf{x}\mathbf{x}^\nu}$, $\nu = \mathrm{i}, *, \mathrm{k}$, vanish. Similarly, two random vectors $\mathbf{x} \in \mathbb{Q}^{p_1}$ and $\mathbf{y} \in \mathbb{Q}^{p_2}$ are cross $\mathbb{Q}$-proper if, and only if, the matrices $\mathbf{\Lambda}_{\mathbf{x}\mathbf{y}^\nu}$, $\nu = \mathrm{i}, *, \mathrm{k}$, vanish.

There exist several statistical tests to experimentally check if a quaternion random vector is proper or improper (see, e.g., [37,38]).

The closed linear subspace $\mathcal{H}_{\mathcal{A}}$ associated to a set of random vectors $\mathcal{A} = \{\mathbf{x}_1, \ldots, \mathbf{x}_m\}$, with $\mathbf{x}_i \in \mathbb{Q}^p$, $i = 1, \ldots, m$, is the set of elements of the form $\sum_{i=1}^m \sum_{j=1}^p \alpha_{ij} x_{ij}$, with x_{ij} the jth-component of $\mathbf{x}_i$ and $\alpha_{ij} \in \mathbb{Q}$ deterministic coefficients.

The quaternion algebra (1) and the definition of the following inner product $\langle x, y\rangle_{\mathtt{Q}} = \lambda_{xy}$ endow $\mathbb{Q}$ with the structure of a Hilbert space [36]. Then, using the Best Approximation Theorem [39], it is shown that for any $y \in \mathbb{Q}$, under the norm

$$\|x\|_{\mathtt{Q}} = \lambda_x^{\frac{1}{2}} \tag{2}$$

there exists a unique projection onto the set $\mathcal{H}_{\mathcal{A}}$ of the form

$$\hat{y}^{\mathtt{Q}} = \sum_{i=1}^{m} \mathbf{h}_i^{\mathrm{T}} \mathbf{x}_i \tag{3}$$

where the deterministic vectors $\mathbf{h}_i \in \mathbb{Q}^p$ are computed through the equation

$$[\mathbf{h}_1^{\mathrm{T}}, \ldots, \mathbf{h}_m^{\mathrm{T}}] = \mathbf{\Lambda}_{y\mathbf{z}} \mathbf{\Lambda}_{\mathbf{z}}^{-1} \tag{4}$$

with $\mathbf{z} = [\mathbf{x}_1^{\mathrm{T}}, \ldots, \mathbf{x}_m^{\mathrm{T}}]^{\mathrm{T}}$. Moreover, the error associated to $\hat{y}^{\mathtt{Q}}$ is given by

$$\epsilon^{\mathtt{Q}} = \|y - \hat{y}^{\mathtt{Q}}\|_{\mathtt{Q}}^2 \tag{5}$$

The projection $\hat{y}^{\mathtt{Q}}$ is the linear MMSE estimator of y given the information supplied by the set $\mathcal{A}$ and it is called the QSL estimator. From now on, we denote such an estimator as $\hat{y}^{\mathtt{QSL}}$.

In the quaternion domain, the complete description of the second-order statistical properties of the random variable $x \in \mathbb{Q}$ is given by the augmented vector $\bar{\mathbf{x}} = [x, x^{\mathrm{i}}, x^{\mathrm{j}}, x^{\mathrm{k}}]^{\mathrm{T}}$ (see [30] for further details). Thus, a better approximation of y is attained if we consider the information set $\bar{\mathcal{A}} = \{\bar{\mathbf{x}}_1, \ldots, \bar{\mathbf{x}}_m\}$, with $\bar{\mathbf{x}}_i \in \mathbb{Q}^{4p}$, $i = 1, \ldots, m$. In such a case, the projection of y onto the set $\mathcal{H}_{\bar{\mathcal{A}}}$ is called QWL estimator and it is given by

$$\hat{y}^{\mathtt{QWL}} = \sum_{i=1}^{m} \left(\mathbf{h}_{1i}^{\mathrm{T}} \mathbf{x}_i + \mathbf{h}_{2i}^{\mathrm{T}} \mathbf{x}_i^{\mathrm{i}} + \mathbf{h}_{3i}^{\mathrm{T}} \mathbf{x}_i^{\mathrm{j}} + \mathbf{h}_{4i}^{\mathrm{T}} \mathbf{x}_i^{\mathrm{k}} \right)$$

where the deterministic vectors $\mathbf{h}_{ji} \in \mathbb{Q}^p$ are computed through the equation

$$\left[\mathbf{h}_{11}^{\mathrm{T}}, \ldots, \mathbf{h}_{1m}^{\mathrm{T}}, \mathbf{h}_{21}^{\mathrm{T}}, \ldots, \mathbf{h}_{2m}^{\mathrm{T}}, \mathbf{h}_{31}^{\mathrm{T}}, \ldots, \mathbf{h}_{3m}^{\mathrm{T}}, \mathbf{h}_{41}^{\mathrm{T}}, \ldots, \mathbf{h}_{4m}^{\mathrm{T}}\right] = \mathbf{\Lambda}_{y\bar{\mathbf{z}}} \mathbf{\Lambda}_{\bar{\mathbf{z}}}^{-1}$$

with $\bar{\mathbf{z}} = [\mathbf{x}_1^{\mathrm{T}}, \ldots, \mathbf{x}_m^{\mathrm{T}}, \mathbf{x}_1^{\mathrm{iT}}, \ldots, \mathbf{x}_m^{\mathrm{iT}}, \mathbf{x}_1^{\mathrm{jT}}, \ldots, \mathbf{x}_m^{\mathrm{jT}}, \mathbf{x}_1^{\mathrm{kT}}, \ldots, \mathbf{x}_m^{\mathrm{kT}}]^{\mathrm{T}}$.

The QWL processing is the most suitable one in terms of performance whenever the signal is improper. In fact, since $\mathcal{H}_{\mathcal{A}} \subseteq \mathcal{H}_{\bar{\mathcal{A}}}$ then $\epsilon^{\mathtt{QWL}} \leq \epsilon^{\mathtt{QSL}}$, where $\epsilon^{\mathtt{QWL}} = \|y - \hat{y}^{\mathtt{QWL}}\|_{\mathtt{Q}}^2$ and $\epsilon^{\mathtt{QSL}} = \|y - \hat{y}^{\mathtt{QSL}}\|_{\mathtt{Q}}^2$. Hence, in general, the QWL estimation error is lower than the QSL estimation error. However, the computation of the QSL estimator entails a lower computational cost than the one associated with the QWL estimator [31]. Finally, under $\mathbb{Q}$-properness conditions, both QWL and QSL estimators coincide [30] and hence, in this situation the QSL estimator shows both advantages: minimum error and lower computational cost.

3. Tessarine domain

In this section, the basic concepts in the tessarine domain as well as some properties of interest are established.

3.1. Random variables

Definition 3. A random variable $x \in \mathbb{T}$ is defined by [27]

$$x = a + \mathrm{i}b + \mathrm{j}c + \mathrm{k}d$$

where the random variables $a, b, c, d \in \mathbb{R}$ and the imaginary units (i, j, k) satisfy the identities

$$\mathrm{ij} = \mathrm{k}, \quad \mathrm{jk} = \mathrm{i}, \quad \mathrm{ki} = -\mathrm{j}, \quad \mathrm{i}^2 = \mathrm{k}^2 = -1, \quad \mathrm{j}^2 = 1$$

Notice that we do not distinguish between quaternion and tessarine product (the type will be clear from the context). The main advantage of tessarine algebra over quaternion algebra is the commutative property of multiplication. The tessarines also have their limitations. The set of tessarines is not a division algebra, and their geometric meaning is unfamiliar to most engineers. However, they have almost no influence on signal and image processing applications [19].

Given a random variable $x \in \mathbb{T}$, the real vector formed by its components is denoted by $\mathbf{x}_r$, i.e., $\mathbf{x}_r = [a, b, c, d]^{\mathrm{T}}$. Moreover, the conjugate of x is

$$x^* = a - \mathrm{i}b + \mathrm{j}c - \mathrm{k}d$$

and, similar to the quaternion processing, we define the following auxiliary tessarines:

$$x^{\mathrm{i}} = a + \mathrm{i}b - \mathrm{j}c - \mathrm{k}d$$

$$x^{\mathrm{k}} = a - \mathrm{i}b - \mathrm{j}c + \mathrm{k}d$$

The triplet $(x^{\mathrm{i}}, x^*, x^{\mathrm{k}})$ constitutes the principal conjugations [6]. Unlike the quaternion domain, the principal conjugations are not involutions of x, but they are auto-involutive [23]. The principal conjugations have the same properties as the conjugate of complex and hyperbolic numbers. In particular, the following property will be useful in our derivations.

Property 1. *Let be $x, y \in \mathbb{T}$, then*

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

Additionally, the pseudo autocorrelation and pseudo cross-correlation of the random variables $x, y \in \mathbb{T}$ are denoted by $\gamma_x = E[xx^*]$ and $\gamma_{xy} = E[xy^*]$, respectively. Observe that γ_x and γ_{xy} are not true correlations because, for any random variable $x \in \mathbb{T}$, $E[xx^*]$ is a hyperbolic number (a complex number with imaginary part j). Similarly, the pseudo autocorrelation and pseudo cross-correlation matrices of the random vectors $\mathbf{x} \in \mathbb{T}^{p_1}$ and $\mathbf{y} \in \mathbb{T}^{p_2}$ are denoted by $\mathbf{\Gamma_x} = E[\mathbf{xx}^{\mathrm{H}}]$ and $\mathbf{\Gamma_{xy}} = E[\mathbf{xy}^{\mathrm{H}}]$, respectively.

3.2. Random signals

Definition 4. A tessarine random signal is a stochastic process of the form

$$x(t) = a(t) + \mathrm{i}b(t) + \mathrm{j}c(t) + \mathrm{k}d(t), \quad t \in \mathbb{Z}$$

where $a(t)$, $b(t)$, $c(t)$, and $d(t)$ are real random signals.

Now, we extend the above concepts to the vectorial case.

Definition 5. A tessarine random signal vector is of the form

$$\mathbf{x}(t) = [x_1(t), \ldots, x_p(t)]^{\mathrm{T}}, \quad t \in \mathbb{Z}$$

where $x_i(t) = a_i(t) + \mathrm{i}b_i(t) + \mathrm{j}c_i(t) + \mathrm{k}d_i(t) \in \mathbb{T}$, $i = 1, \ldots, p$, with $a_i(t)$, $b_i(t)$, $c_i(t)$, and $d_i(t)$ real random signals.

The real vectors associated to $a_i(t)$, $b_i(t)$, $c_i(t)$, and $d_i(t)$ are denoted by

$$\mathbf{a}(t) = [a_1(t), \ldots, a_p(t)]^{\mathrm{T}}, \qquad \mathbf{b}(t) = [b_1(t), \ldots, b_p(t)]^{\mathrm{T}}$$
$$\mathbf{c}(t) = [c_1(t), \ldots, c_p(t)]^{\mathrm{T}}, \qquad \mathbf{d}(t) = [d_1(t), \ldots, d_p(t)]^{\mathrm{T}}$$

and the real vector formed with the components of $\mathbf{x}(t)$ is denoted by $\mathbf{x}_r(t)$, i.e.,

$$\mathbf{x}_r(t) = [\mathbf{a}^{\mathrm{T}}(t), \mathbf{b}^{\mathrm{T}}(t), \mathbf{c}^{\mathrm{T}}(t), \mathbf{d}^{\mathrm{T}}(t)]^{\mathrm{T}}, \quad t \in \mathbb{Z} \tag{6}$$

Moreover, the following auxiliary tessarine vectors are defined:

$$\mathbf{x}^{\nu}(t) = [x_1^{\nu}(t), \ldots, x_p^{\nu}(t)]^{\mathrm{T}}, \quad \nu = \mathrm{i}, *, \mathrm{k}, \quad t \in \mathbb{Z}$$

Definition 6. The pseudo autocorrelation function of the random signal vector $\mathbf{x}(t) \in \mathbb{T}^p$ is defined as $\boldsymbol{\Gamma}_{\mathbf{x}}(t, s) = E[\mathbf{x}(t)\mathbf{x}^{\mathrm{H}}(s)]$, $\forall t, s \in \mathbb{Z}$, and the pseudo cross-correlation function of the random signal vectors $\mathbf{x}(t) \in \mathbb{T}^{p_1}$ and $\mathbf{y}(t) \in \mathbb{T}^{p_2}$ is defined as $\boldsymbol{\Gamma}_{\mathbf{xy}}(t, s) = E[\mathbf{x}(t)\mathbf{y}^{\mathrm{H}}(s)]$, $\forall t, s \in \mathbb{Z}$.

By using a similar notation for the correlation functions of real signal vectors, the proof of the following property is straightforward.

Property 2. *For all $t, s \in \mathbb{Z}$:*

$$\begin{aligned}\boldsymbol{\Gamma}_{\mathbf{x}}(t, s) =& \boldsymbol{\Gamma}_{\mathbf{a}}(t, s) + \boldsymbol{\Gamma}_{\mathbf{b}}(t, s) + \boldsymbol{\Gamma}_{\mathbf{c}}(t, s) + \boldsymbol{\Gamma}_{\mathbf{d}}(t, s)\\
&+ \mathrm{i}\big(-\boldsymbol{\Gamma}_{\mathbf{ab}}(t, s) + \boldsymbol{\Gamma}_{\mathbf{ba}}(t, s) - \boldsymbol{\Gamma}_{\mathbf{cd}}(t, s) + \boldsymbol{\Gamma}_{\mathbf{dc}}(t, s)\big)\\
&+ \mathrm{j}\big(\boldsymbol{\Gamma}_{\mathbf{ac}}(t, s) + \boldsymbol{\Gamma}_{\mathbf{bd}}(t, s) + \boldsymbol{\Gamma}_{\mathbf{ca}}(t, s) + \boldsymbol{\Gamma}_{\mathbf{db}}(t, s)\big)\\
&+ \mathrm{k}\big(-\boldsymbol{\Gamma}_{\mathbf{ad}}(t, s) + \boldsymbol{\Gamma}_{\mathbf{bc}}(t, s) - \boldsymbol{\Gamma}_{\mathbf{cb}}(t, s) + \boldsymbol{\Gamma}_{\mathbf{da}}(t, s)\big)\end{aligned}$$

$$\begin{aligned}\boldsymbol{\Gamma}_{\mathbf{xx}^{\mathrm{i}}}(t, s) =& \boldsymbol{\Gamma}_{\mathbf{a}}(t, s) + \boldsymbol{\Gamma}_{\mathbf{b}}(t, s) - \boldsymbol{\Gamma}_{\mathbf{c}}(t, s) - \boldsymbol{\Gamma}_{\mathbf{d}}(t, s)\\
&+ \mathrm{i}\big(-\boldsymbol{\Gamma}_{\mathbf{ab}}(t, s) + \boldsymbol{\Gamma}_{\mathbf{ba}}(t, s) + \boldsymbol{\Gamma}_{\mathbf{cd}}(t, s) - \boldsymbol{\Gamma}_{\mathbf{dc}}(t, s)\big)\\
&+ \mathrm{j}\big(-\boldsymbol{\Gamma}_{\mathbf{ac}}(t, s) - \boldsymbol{\Gamma}_{\mathbf{bd}}(t, s) + \boldsymbol{\Gamma}_{\mathbf{ca}}(t, s) + \boldsymbol{\Gamma}_{\mathbf{db}}(t, s)\big)\\
&+ \mathrm{k}\big(\boldsymbol{\Gamma}_{\mathbf{ad}}(t, s) - \boldsymbol{\Gamma}_{\mathbf{bc}}(t, s) - \boldsymbol{\Gamma}_{\mathbf{cb}}(t, s) + \boldsymbol{\Gamma}_{\mathbf{da}}(t, s)\big)\end{aligned}$$

$$\begin{aligned}\boldsymbol{\Gamma}_{\mathbf{xx}^{*}}(t, s) =& \boldsymbol{\Gamma}_{\mathbf{a}}(t, s) - \boldsymbol{\Gamma}_{\mathbf{b}}(t, s) + \boldsymbol{\Gamma}_{\mathbf{c}}(t, s) - \boldsymbol{\Gamma}_{\mathbf{d}}(t, s)\\
&+ \mathrm{i}\big(\boldsymbol{\Gamma}_{\mathbf{ab}}(t, s) + \boldsymbol{\Gamma}_{\mathbf{ba}}(t, s) + \boldsymbol{\Gamma}_{\mathbf{cd}}(t, s) + \boldsymbol{\Gamma}_{\mathbf{dc}}(t, s)\big)\\
&+ \mathrm{j}\big(\boldsymbol{\Gamma}_{\mathbf{ac}}(t, s) - \boldsymbol{\Gamma}_{\mathbf{bd}}(t, s) + \boldsymbol{\Gamma}_{\mathbf{ca}}(t, s) - \boldsymbol{\Gamma}_{\mathbf{db}}(t, s)\big)\\
&+ \mathrm{k}\big(\boldsymbol{\Gamma}_{\mathbf{ad}}(t, s) + \boldsymbol{\Gamma}_{\mathbf{bc}}(t, s) + \boldsymbol{\Gamma}_{\mathbf{cb}}(t, s) + \boldsymbol{\Gamma}_{\mathbf{da}}(t, s)\big)\end{aligned}$$

$$\begin{aligned}\mathbf{\Gamma}_{\mathbf{x}\mathbf{x}^{\mathrm{k}}}(t,s) =&\mathbf{\Gamma}_{\mathbf{a}}(t,s)-\mathbf{\Gamma}_{\mathbf{b}}(t,s)-\mathbf{\Gamma}_{\mathbf{c}}(t,s)+\mathbf{\Gamma}_{\mathbf{d}}(t,s)\\ &+\mathrm{i}\big(\mathbf{\Gamma}_{\mathbf{ab}}(t,s)+\mathbf{\Gamma}_{\mathbf{ba}}(t,s)-\mathbf{\Gamma}_{\mathbf{cd}}(t,s)-\mathbf{\Gamma}_{\mathbf{dc}}(t,s)\big)\\ &+\mathrm{j}\big(-\mathbf{\Gamma}_{\mathbf{ac}}(t,s)+\mathbf{\Gamma}_{\mathbf{bd}}(t,s)+\mathbf{\Gamma}_{\mathbf{ca}}(t,s)-\mathbf{\Gamma}_{\mathbf{db}}(t,s)\big)\\ &+\mathrm{k}\big(-\mathbf{\Gamma}_{\mathbf{ad}}(t,s)-\mathbf{\Gamma}_{\mathbf{bc}}(t,s)+\mathbf{\Gamma}_{\mathbf{cb}}(t,s)+\mathbf{\Gamma}_{\mathbf{da}}(t,s)\big)\end{aligned}$$

As in the QWL processing, we define the augmented vector

$$\bar{\mathbf{x}}(t)=[\mathbf{x}^{\mathrm{T}}(t),\mathbf{x}^{\mathrm{i}^{\mathrm{T}}}(t),\mathbf{x}^{*^{\mathrm{T}}}(t),\mathbf{x}^{\mathrm{k}^{\mathrm{T}}}(t)]^{\mathrm{T}},\quad t\in\mathbb{Z} \tag{7}$$

The following relationship between the augmented vector and the real vector $\mathbf{x}_r(t)$ given in (6) can be established:

$$\bar{\mathbf{x}}(t)=2\mathcal{T}_p\mathbf{x}_r(t) \tag{8}$$

where $\mathcal{T}_p=\frac{1}{2}\mathcal{B}\otimes\mathbf{I}_p$ with

$$\mathcal{B}=\begin{pmatrix}1 & \mathrm{i} & \mathrm{j} & \mathrm{k}\\ 1 & \mathrm{i} & -\mathrm{j} & -\mathrm{k}\\ 1 & -\mathrm{i} & \mathrm{j} & -\mathrm{k}\\ 1 & -\mathrm{i} & -\mathrm{j} & \mathrm{k}\end{pmatrix} \tag{9}$$

and $\mathcal{T}_p^{\mathrm{H}}\mathcal{T}_p=\mathbf{I}_{4p}$.

Based on Property 1, the pseudo autocorrelation function of the augmented vector $\bar{\mathbf{x}}(t)$ given in (7) is provided in the following property.

Property 3.

$$\mathbf{\Gamma}_{\bar{\mathbf{x}}}(t,s)=\begin{pmatrix}\mathbf{\Gamma}_{\mathbf{x}}(t,s) & \mathbf{\Gamma}_{\mathbf{x}\mathbf{x}^{\mathrm{i}}}(t,s) & \mathbf{\Gamma}_{\mathbf{x}\mathbf{x}^{*}}(t,s) & \mathbf{\Gamma}_{\mathbf{x}\mathbf{x}^{\mathrm{k}}}(t,s)\\ \mathbf{\Gamma}^{\mathrm{i}}_{\mathbf{x}\mathbf{x}^{\mathrm{i}}}(t,s) & \mathbf{\Gamma}^{\mathrm{i}}_{\mathbf{x}}(t,s) & \mathbf{\Gamma}^{\mathrm{i}}_{\mathbf{x}\mathbf{x}^{\mathrm{k}}}(t,s) & \mathbf{\Gamma}^{\mathrm{i}}_{\mathbf{x}\mathbf{x}^{*}}(t,s)\\ \mathbf{\Gamma}^{*}_{\mathbf{x}\mathbf{x}^{*}}(t,s) & \mathbf{\Gamma}^{*}_{\mathbf{x}\mathbf{x}^{\mathrm{k}}}(t,s) & \mathbf{\Gamma}^{*}_{\mathbf{x}}(t,s) & \mathbf{\Gamma}^{*}_{\mathbf{x}\mathbf{x}^{\mathrm{i}}}(t,s)\\ \mathbf{\Gamma}^{\mathrm{k}}_{\mathbf{x}\mathbf{x}^{\mathrm{k}}}(t,s) & \mathbf{\Gamma}^{\mathrm{k}}_{\mathbf{x}\mathbf{x}^{*}}(t,s) & \mathbf{\Gamma}^{\mathrm{k}}_{\mathbf{x}\mathbf{x}^{\mathrm{i}}}(t,s) & \mathbf{\Gamma}^{\mathrm{k}}_{\mathbf{x}}(t,s)\end{pmatrix},\quad t,s\in\mathbb{Z} \tag{10}$$

To conclude this section, the concept of stationarity is introduced.

Definition 7. A random signal vector $\mathbf{x}(t)\in\mathbb{T}^p$ is said to be stationary if the pseudo autocorrelation function of the augmented vector $\bar{\mathbf{x}}(t)$ verifies that $\mathbf{\Gamma}_{\bar{\mathbf{x}}}(t+l,t)=\mathbf{\Gamma}_{\bar{\mathbf{x}}}(s+l,s)$, $\forall t,s,l\in\mathbb{Z}$, i.e., it only depends on the lag l.

4. $\mathbb{T}$-properness

Next, the concept of $\mathbb{Q}$-properness introduced in [29] for quaternions is adapted to the tessarine domain and a $\mathbb{T}$-properness statistical test is proposed.

Definition 8. A random signal vector $\mathbf{x}(t)\in\mathbb{T}^p$ is said to be $\mathbb{T}$-proper if, and only if, the functions $\mathbf{\Gamma}_{\mathbf{x}\mathbf{x}^{\nu}}(t,s)$, $\nu=\mathrm{i},*,\mathrm{k}$, vanish $\forall t,s\in\mathbb{Z}$. In like manner, two random signal vectors $\mathbf{x}(t)\in\mathbb{T}^{p_1}$ and $\mathbf{y}(t)\in\mathbb{T}^{p_2}$ are cross $\mathbb{T}$-proper if, and only if, the functions $\mathbf{\Gamma}_{\mathbf{x}\mathbf{y}^{\nu}}(t,s)$, $\nu=\mathrm{i},*,\mathrm{k}$, vanish $\forall t,s\in\mathbb{Z}$. Finally, $\mathbf{x}(t)$ and $\mathbf{y}(t)$ are jointly $\mathbb{T}$-proper if, and only if, are $\mathbb{T}$-proper and cross $\mathbb{T}$-proper.

From Property 2 we have the following result.

Proposition 1. *A random signal vector* $\mathbf{x}(t) \in \mathbb{T}^p$ *is* $\mathbb{T}$*-proper if, and only if, the following equalities are verified* $\forall t, s \in \mathbb{Z}$*:*

$$\begin{aligned} \mathbf{\Gamma_a}(t,s) = \mathbf{\Gamma_b}(t,s) = \mathbf{\Gamma_c}(t,s) = \mathbf{\Gamma_d}(t,s) \\ -\mathbf{\Gamma_{ab}}(t,s) = \mathbf{\Gamma_{ba}}(t,s) = -\mathbf{\Gamma_{cd}}(t,s) = \mathbf{\Gamma_{dc}}(t,s) \\ \mathbf{\Gamma_{ac}}(t,s) = \mathbf{\Gamma_{bd}}(t,s) = \mathbf{\Gamma_{ca}}(t,s) = \mathbf{\Gamma_{db}}(t,s) \\ -\mathbf{\Gamma_{ad}}(t,s) = \mathbf{\Gamma_{bc}}(t,s) = -\mathbf{\Gamma_{cb}}(t,s) = \mathbf{\Gamma_{da}}(t,s) \end{aligned} \tag{11}$$

Remark 1. Similarly to Definition 8, a random vector $\mathbf{x} \in \mathbb{T}^p$ is said to be $\mathbb{T}$-proper if, and only if, the matrices $\mathbf{\Gamma}_{\mathbf{x}\mathbf{x}^\nu}$, $\nu = \mathrm{i}, *, \mathrm{k}$, vanish. Also, two random vectors $\mathbf{x} \in \mathbb{T}^{p_1}$ and $\mathbf{y} \in \mathbb{T}^{p_2}$ are cross $\mathbb{T}$-proper if, and only if, the matrices $\mathbf{\Gamma}_{\mathbf{x}\mathbf{y}^\nu}$, $\nu = \mathrm{i}, *, \mathrm{k}$, vanish. Then, the conditions (11) become

$$\begin{aligned} &\mathbf{\Gamma_a} = \mathbf{\Gamma_b} = \mathbf{\Gamma_c} = \mathbf{\Gamma_d} \\ &\mathbf{\Gamma}_{\mathbf{ab}}^{\mathrm{T}} = -\mathbf{\Gamma_{ab}} = -\mathbf{\Gamma_{cd}} = \mathbf{\Gamma}_{\mathbf{cd}}^{\mathrm{T}} \\ &\mathbf{\Gamma}_{\mathbf{ac}}^{\mathrm{T}} = \mathbf{\Gamma_{ac}} = \mathbf{\Gamma_{bd}} = \mathbf{\Gamma}_{\mathbf{bd}}^{\mathrm{T}} \\ &\mathbf{\Gamma}_{\mathbf{ad}}^{\mathrm{T}} = -\mathbf{\Gamma_{ad}} = \mathbf{\Gamma_{bc}} = -\mathbf{\Gamma}_{\mathbf{bc}}^{\mathrm{T}} \end{aligned} \tag{12}$$

It should be highlighted that although the definition of $\mathbb{T}$-properness is similar to that of $\mathbb{Q}$-properness given in Definition 2, the former implies different relations between the correlation matrices of the real components from those obtained under $\mathbb{Q}$-properness (compare (12) with [29, Lemma 8]).

The objective now is to test whether a random vector $\mathbf{x} \in \mathbb{T}^p$ is $\mathbb{T}$-proper from the information supplied by a random sample. More specifically, we aim to test the following hypotheses

$$\begin{aligned} &H_0 : \mathbf{x} \text{ is } \mathbb{T}\text{-proper (i.e., } \mathbf{\Gamma}_{\mathbf{x}\mathbf{x}^{\mathrm{i}}} = \mathbf{\Gamma}_{\mathbf{x}\mathbf{x}^{*}} = \mathbf{\Gamma}_{\mathbf{x}\mathbf{x}^{\mathrm{k}}} = \mathbf{0}_{p\times p}) \\ &H_1 : \mathbf{x} \text{ is not } \mathbb{T}\text{-proper} \end{aligned} \tag{13}$$

Theorem 1 (. $\mathbb{T}$-Properness GLRT) *Given n independent and identically distributed random samples* $\mathbf{x}_1, \dots, \mathbf{x}_n$ *of a random vector* $\mathbf{x} \in \mathbb{T}^p$ *such that* $\mathbf{x}_r$ *follows a Gaussian distribution, then the GLRT statistic is given by*

$$\phi(\mathbf{x}_1, \dots, \mathbf{x}_n) = -n[\ln|\hat{\mathbf{\Gamma}}_{\bar{\mathbf{x}}}| - \ln|\mathbf{\Gamma}_{\mathbb{T}}|] \tag{14}$$

where $\hat{\mathbf{\Gamma}}_{\bar{\mathbf{x}}} = 4\mathcal{T}_p\hat{\mathbf{\Gamma}}_{\mathbf{x}_r}\mathcal{T}_p^{\mathrm{H}}$, *with* $\hat{\mathbf{\Gamma}}_{\mathbf{x}_r}$ *the sample autocorrelation matrix, and*

$$\mathbf{\Gamma}_{\mathbb{T}} = \begin{pmatrix} \hat{\mathbf{\Gamma}}_{\mathbf{x}} & \mathbf{0}_{p\times p} & \mathbf{0}_{p\times p} & \mathbf{0}_{p\times p} \\ \mathbf{0}_{p\times p} & \hat{\mathbf{\Gamma}}_{\mathbf{x}}^{\mathrm{i}} & \mathbf{0}_{p\times p} & \mathbf{0}_{p\times p} \\ \mathbf{0}_{p\times p} & \mathbf{0}_{p\times p} & \hat{\mathbf{\Gamma}}_{\mathbf{x}}^{*} & \mathbf{0}_{p\times p} \\ \mathbf{0}_{p\times p} & \mathbf{0}_{p\times p} & \mathbf{0}_{p\times p} & \hat{\mathbf{\Gamma}}_{\mathbf{x}}^{\mathrm{k}} \end{pmatrix} \tag{15}$$

Also, assuming that H_0 *is true, the distribution of the statistic* $\phi(\mathbf{x}_1, \dots, \mathbf{x}_n)$ *tends to a chi-squared distribution with degrees of freedom equal to* $2p(3p+1)$ *as the sample size tends to infinity. Thus,* H_0 *should be rejected if* $\phi(\mathbf{x}_1, \dots, \mathbf{x}_n)$ *is too large in terms of the tabulated* $\chi^2(2p(3p+1))$ *distribution.*

4.1. Example

Consider a random vector $\mathbf{x} \in \mathbb{T}^2$ whose real vector is Gaussian with autocorrelation matrix

$$\boldsymbol{\Gamma}_{\mathbf{x}_r} = \begin{pmatrix} 2+a_1 & 0.4 & 0 & -0.1 & 1 & 0.3 & 0 & -0.2 \\ 0.4 & 2 & 0.1 & 0 & 0.3 & 1 & 0.2 & 0 \\ 0 & 0.1 & 2 & 0.4 & 0 & 0.2 & 1 & 0.3 \\ -0.1 & 0 & 0.4 & 2 & -0.2 & 0 & 0.3 & 1 \\ 1 & 0.3 & 0 & -0.2 & 2 & 0.4 & a_2 & -0.1 \\ 0.3 & 1 & 0.2 & 0 & 0.4 & 2 & 0.1 & 0 \\ 0 & 0.2 & 1 & 0.3 & a_2 & 0.1 & 2 & 0.4 \\ -0.2 & 0 & 0.3 & 1 & -0.1 & 0 & 0.4 & 2 \end{pmatrix}$$

with a_1 taking the values 0 and 1 and a_2 the values 0 and 0.7. Observe that, from Proposition 1, when $a_1 = a_2 = 0$ the vector $\mathbf{x}$ is $\mathbb{T}$-proper.

Firstly, the behavior of the cumulative distribution function (CDF) of the Wilks' statistic $\phi(\mathbf{x}_1, \ldots, \mathbf{x}_n)$ is analyzed in terms of the sample size. In this particular example, under H_0 ($a_1 = a_2 = 0$), the Wilks' statistic should converge to a $\chi^2(28)$ distribution. From this idea, through Monte Carlo simulation, 2000 values of the statistic $\phi(\mathbf{x}_1, \ldots, \mathbf{x}_n)$ have been generated, varying the sample size n. By way of illustration, the results obtained for $n = 25$, 100 and 250 are presented in Fig. 1, where the corresponding CDFs are displayed. At a glance, this figure shows the convergence of the simulated CDFs towards the CDF of the $\chi^2(28)$.

Secondly, the receiver operating characteristic (ROC) curve is depicted in Fig. 2 for the cases: $a_1 = 1, a_2 = 0$ (case 1), $a_1 = 0, a_2 = 0.7$ (case 2), and $a_1 = 1, a_2 = 0.7$ (case 3), and for the sample sizes $n = 25, 50, 75$ and 100. As could be expected, the detection probability increases as the sample size increases. Moreover, it is noteworthy that the ROC curves improve (i.e., the curve is closer to the upper left-hand corner of the plot) for higher values a_1 and a_2 (compare case 1 with case 3, and case 2 with case 3). As a consequence, the proposed $\mathbb{T}$-properness test will have a better performance in terms of both sensitivity and specificity as we deviate from the $\mathbb{T}$-property condition.

5. Tessarine processing

As was pointed out in Section 2.2, the existence and uniqueness of the optimal quaternion estimator relies on the structure of Hilbert space. However, the tessarine algebra is not a Hilbert space and thus, this strategy to address the estimation problem is not applicable. The problem lies with the appropriate definition of a norm in the tessarine domain. In other words, the existence of zero divisors implies that the Euclidean norm is not multiplicative (note that the product xx^* is not a real number and thus, a norm cannot be defined as in Eq. (2)).

Next, we establish the necessary bases to define the TWL and TSL estimators similar to the quaternion ones. First, we prove that $\mathbb{T}$ is a metric space. Also, this metric space fulfills a series of properties that guarantees the existence and uniqueness of the projection of a tessarine random variable. Finally, a method to obtain this projection is provided.

Given the random variables $x, y \in \mathbb{T}$, define the product

$$\langle x, y \rangle = \gamma_{xy} = E[xy^*] \tag{16}$$

The properties below follow from Property 1.

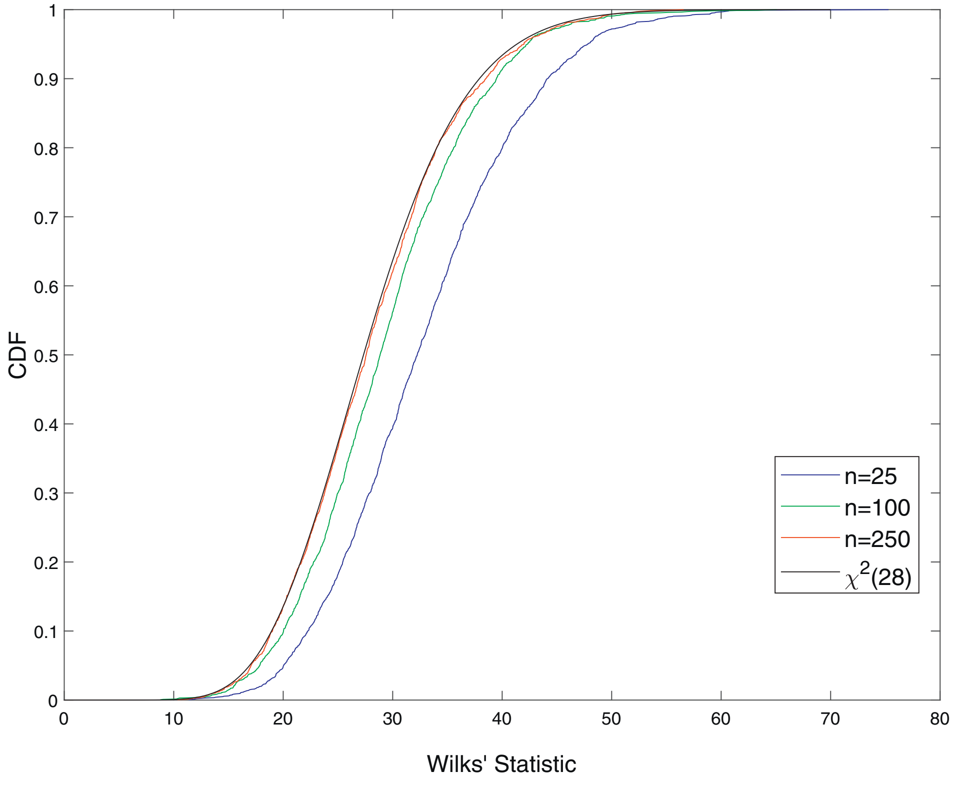

Fig. 1. CDF of a $\chi^2(28)$ and CDFs of the Wilks' statistic $\phi(\mathbf{x}_1, \ldots, \mathbf{x}_n)$ for several values of n and under H_0.

Property 4. *Consider four random variables* $x, y, u, v \in \mathbb{T}$ *and four deterministic coefficients* $\alpha, \beta, \gamma, \delta \in \mathbb{T}$. *Then,*

1. $\langle x, y\rangle = \langle y, x\rangle^*$
2. $\langle \alpha x + \beta y, \gamma u + \delta v\rangle = \alpha\gamma^*\langle x, u\rangle + \alpha\delta^*\langle x, v\rangle + \beta\gamma^*\langle y, u\rangle + \beta\delta^*\langle y, v\rangle$
3. $\langle x, x\rangle = 0 \Leftrightarrow x = 0$

From the product (16), the following function is defined:

$$\|x\| = (\mathcal{R}\{\gamma_x\})^{\frac{1}{2}} \tag{17}$$

Note that the value of this function coincides with Eq. (2), nevertheless, it is not a norm because it is not absolutely scalable, i.e., $\|\alpha x\| \neq |\alpha| \|x\|$ with $\alpha \in \mathbb{T}$ a deterministic coefficient and $|\alpha| = \mathcal{R}\{\alpha\alpha^*\}^{\frac{1}{2}}$.

Now, this function is going to endow $\mathbb{T}$ with a metric. To do so, the following properties, obtained by its coincidence with Eq. (2), are required.

Property 5. *Consider two random variables* $x, y \in \mathbb{T}$. *Then,*

1. $\|x\| \geq 0$
2. $\|x + y\| \leq \|x\| + \|y\|$

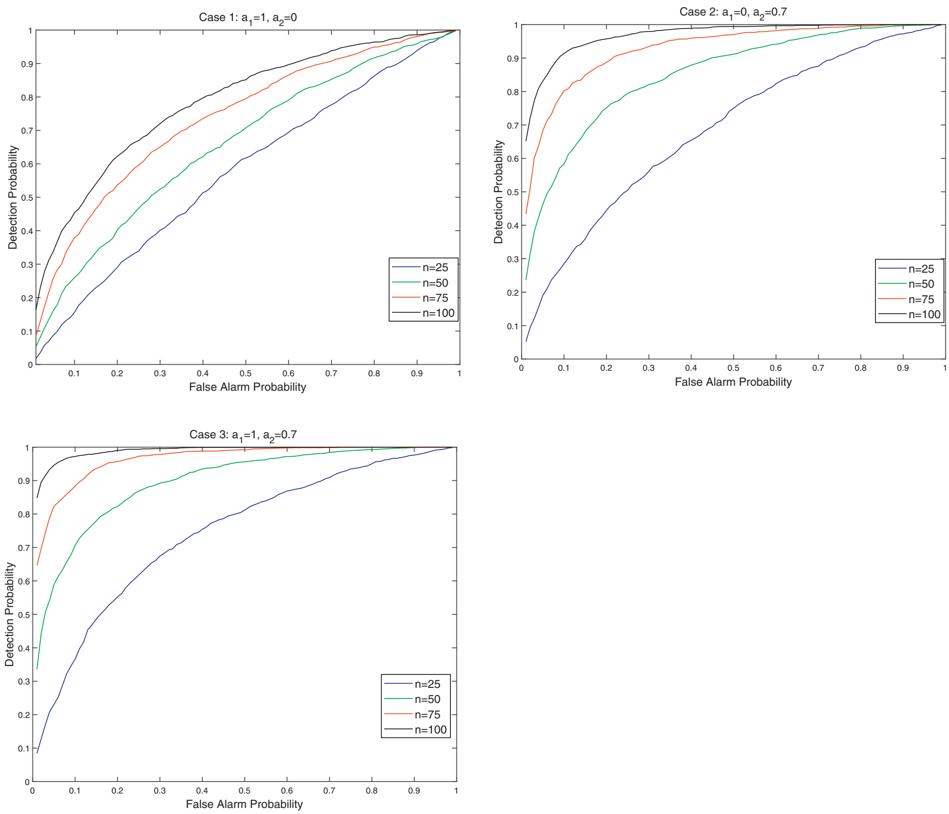

Fig. 2. ROC curves for cases 1, 2 and 3.

3. $\|x\| = 0 \Rightarrow x = 0$
4. $\|x + y\|^2 + \|x - y\|^2 = 2(\|x\|^2 + \|y\|^2)$
5. $\|y - x\| \geq \|x\| - \|y\|$
6. $\|\alpha x\| = |\alpha| \|x\|, \quad \alpha \in \mathbb{R}$

Now, from Eq. (17) the following function is defined:

$$d(x, y) = \|x - y\| \tag{18}$$

Notice that, from Property 5, it is easy to check that Eq. (18) is a metric on $\mathbb{T}$.

Definition 9. The closed linear subspace $\mathcal{G}_{\mathcal{C}}$ associated to the set of random vectors $\mathcal{C} = \{\mathbf{x}_1, \ldots, \mathbf{x}_m\}$, with $\mathbf{x}_i \in \mathbb{T}^p$, $i = 1, \ldots, m$, is the set of elements of the form $\sum_{i=1}^{m} \sum_{j=1}^{p} \alpha_{ij} x_{ij}$, with x_{ij} the jth-component of $\mathbf{x}_i$ and $\alpha_{ij} \in \mathbb{T}$ deterministic coefficients.

In general, in a metric space, neither the existence nor the uniqueness of the projection of an element on a set is assured. However, the properties of Eq. (18) will guarantee the existence and uniqueness of the projection of a given random variable $y \in \mathbb{T}$ onto the set $\mathcal{G}_{\mathcal{C}}$. This projection will be denoted by $\hat{y}$.

Theorem 2. *Consider a random variable $y \in \mathbb{T}$ and the set of random vectors $\mathcal{C} = \{\mathbf{x}_1, \dots, \mathbf{x}_m\}$, with $\mathbf{x}_i \in \mathbb{T}^p$, $i = 1, \dots, m$. Then, there exists a unique projection of y, $\hat{y}$, onto the set $\mathcal{G}_\mathcal{C}$.*

The error associated to $\hat{y}$ is denoted by $\epsilon = \|y - \hat{y}\|^2$. Our aim now is to provide a method to compute $\hat{y}$.

Theorem 3. *Consider a random variable $y \in \mathbb{T}$ and the set of random vectors $\mathcal{C} = \{\mathbf{x}_1, \dots, \mathbf{x}_m\}$, with $\mathbf{x}_i \in \mathbb{T}^p$, $i = 1, \dots, m$. Then, $\hat{y}$ is the projection of y onto the set $\mathcal{G}_\mathcal{C}$ if, and only if, $\langle y - \hat{y}, x\rangle = 0$, $\forall x \in \mathcal{G}_\mathcal{C}$.*

The following result is easily proven from Theorem 3.

Corollary 1. *Let $\{\vartheta_1, \dots, \vartheta_r\}$ be a basis of $\mathcal{G}_\mathcal{C}$. Then, the projection of the random variable $y \in \mathbb{T}$ onto the set $\mathcal{G}_\mathcal{C}$ can be written as $\hat{y} = \sum_{i=1}^r h_i \vartheta_i$ where the deterministic coefficients $h_i \in \mathbb{T}$ are obtained from the system*

$$\langle y, \vartheta_j\rangle = \sum_{i=1}^r h_i \langle \vartheta_i, \vartheta_j\rangle, \qquad j = 1, \dots, r$$

Similarly to the quaternion domain, the following definition is established.

Definition 10. Consider a random variable $y \in \mathbb{T}$ and the set of random vectors $\mathcal{C} = \{\mathbf{x}_1, \dots, \mathbf{x}_m\}$, with $\mathbf{x}_i \in \mathbb{T}^p$, $i = 1, \dots, m$. The projection of y onto the set $\mathcal{G}_\mathcal{C}$, $\hat{y}$, is called the TSL estimator of y respect to $\mathcal{C}$ and, from now on, it will be denoted by $\hat{y}^{\mathrm{TSL}}$. If the information set is formed by augmented vectors defined in (7), i.e., if we consider $\bar{\mathcal{C}} = \{\bar{\mathbf{x}}_1, \dots, \bar{\mathbf{x}}_m\}$ with $\bar{\mathbf{x}}_i \in \mathbb{T}^{4p}$, $i = 1, \dots, m$, then, the projection of y onto the set $\mathcal{G}_{\bar{\mathcal{C}}}$ is called TWL estimator of y respect to $\mathcal{C}$ and it will be denoted by $\hat{y}^{\mathrm{TWL}}$.

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

Theorem 4. *Consider a random variable $y \in \mathbb{T}$ and the set of random vectors $\mathcal{C}$ given in the above definition. Then,*

1. *The TWL estimator of y respect to $\mathcal{C}$ is obtained as*
$$\hat{y}^{\mathrm{TWL}} = \sum_{i=1}^m \left(\mathbf{h}_{1i}^{\mathrm{T}}\mathbf{x}_i + \mathbf{h}_{2i}^{\mathrm{T}}\mathbf{x}_i^{\mathrm{i}} + \mathbf{h}_{3i}^{\mathrm{T}}\mathbf{x}_i^{*} + \mathbf{h}_{4i}^{\mathrm{T}}\mathbf{x}_i^{\mathrm{k}}\right)$$
where the deterministic vectors $\mathbf{h}_{ji} \in \mathbb{T}^p$ are computed through the equation
$$\left[\mathbf{h}_{11}^{\mathrm{T}}, \dots, \mathbf{h}_{1m}^{\mathrm{T}}, \mathbf{h}_{21}^{\mathrm{T}}, \dots, \mathbf{h}_{2m}^{\mathrm{T}}, \mathbf{h}_{31}^{\mathrm{T}}, \dots, \mathbf{h}_{3m}^{\mathrm{T}}, \mathbf{h}_{41}^{\mathrm{T}}, \dots, \mathbf{h}_{4m}^{\mathrm{T}}\right] = \mathbf{\Gamma}_{y\bar{\mathbf{z}}}\mathbf{\Gamma}_{\bar{\mathbf{z}}}^{-1} \tag{19}$$
with $\bar{\mathbf{z}} = [\mathbf{x}_1^{\mathrm{T}}, \dots, \mathbf{x}_m^{\mathrm{T}}, \mathbf{x}_1^{\mathrm{iT}}, \dots, \mathbf{x}_m^{\mathrm{iT}}, \mathbf{x}_1^{*\mathrm{T}}, \dots, \mathbf{x}_m^{*\mathrm{T}}, \mathbf{x}_1^{\mathrm{kT}}, \dots, \mathbf{x}_m^{\mathrm{kT}}]^{\mathrm{T}}$.
2. *The TSL estimator of y respect to $\mathcal{C}$ is calculated as*
$$\hat{y}^{\mathrm{TSL}} = \sum_{i=1}^m \mathbf{g}_i^{\mathrm{T}}\mathbf{x}_i \tag{20}$$
where the deterministic vectors $\mathbf{g}_i \in \mathbb{T}^p$ are computed through the equation
$$\left[\mathbf{g}_1^{\mathrm{T}}, \dots, \mathbf{g}_m^{\mathrm{T}}\right] = \mathbf{\Gamma}_{y\mathbf{z}}\mathbf{\Gamma}_{\mathbf{z}}^{-1} \tag{21}$$
with $\mathbf{z} = [\mathbf{x}_1^{\mathrm{T}}, \dots, \mathbf{x}_m^{\mathrm{T}}]^{\mathrm{T}}$.

3. *The error associated to the TWL estimator is lower than or equal to the TSL estimation error, i.e.,* $\epsilon^{\text{TWL}} \leq \epsilon^{\text{TSL}}$ *where* $\epsilon^{\text{TWL}} = \|y - \hat{y}^{\text{TWL}}\|^2$ *and* $\epsilon^{\text{TSL}} = \|y - \hat{y}^{\text{TSL}}\|^2$.
4. *If* $\mathbf{x}_1, \ldots, \mathbf{x}_m$ *are jointly* $\mathbb{T}$*-proper and* y *is cross* $\mathbb{T}$*-proper with each element of* $\{\mathbf{x}_1, \ldots, \mathbf{x}_m\}$ *then,* $\hat{y}^{\text{TWL}} = \hat{y}^{\text{TSL}}$.

Since a complete description of the second-order statistical properties of $\mathbf{x}$ is given by any combination of the four components of the augmented vector $\bar{\mathbf{x}}$, the TWL estimator is the optimal estimator[1] of y on the basis of the information supplied by $\mathcal{C}$. The main advantage of the TSL estimator over the above one is that its calculation entails a lower computational burden than that associated with the TWL estimator. However, under $\mathbb{T}$-properness properties both estimators coincide and thus, the TSL processing is the appropriate one.

6. $\mathbb{T}$-proper models: state-space and stationary models

This section presents two models where the TSL estimator is the optimal one. Moreover, these models have the additional advantage of avoiding the computation of the inverse matrix $\Gamma_{\mathbf{z}}^{-1}$ in Eq. (21), and this advantage is more marked when the number of observations is higher. To do that, two recursive algorithms are proposed, the first one is devised from the assumption that the signal is described by a state-space model, and the second one is obtained under stationarity conditions.

6.1. TSL state-space models

Consider a random signal vector $\mathbf{x}(t) \in \mathbb{T}^p$ verifying the following state model:

$$\mathbf{x}(t+1) = \mathbf{F}_1(t)\mathbf{x}(t) + \mathbf{F}_2(t)\mathbf{x}^{\text{i}}(t) + \mathbf{F}_3(t)\mathbf{x}^*(t) + \mathbf{F}_4(t)\mathbf{x}^{\text{k}}(t) + \mathbf{u}(t) \tag{22}$$

where $\mathbf{F}_i(t) \in \mathbb{T}^{p\times p}$, $i = 1, \ldots, 4$, are deterministic matrices and $\mathbf{u}(t) \in \mathbb{T}^p$ is a random noise vector with $\mathbf{\Gamma_u}(t, s) = \mathbf{Q}_1(t)\delta_{ts}$.

Assume that the state cannot be observed directly but it is obtained from the following observation equation:

$$\mathbf{z}(t) = \mathbf{x}(t) + \mathbf{v}(t) \tag{23}$$

where $\mathbf{v}(t) \in \mathbb{T}^p$ is a random noise vector with $\mathbf{\Gamma_v}(t, s) = \mathbf{R}_1(t)\delta_{ts}$.

By considering the augmented vectors $\bar{\mathbf{x}}(t)$ and $\bar{\mathbf{z}}(t)$ of $\mathbf{x}(t)$ and $\mathbf{z}(t)$, respectively, defined as in Eq. (7), and using Property 1, Eqs. (22) and (23), the following TWL state-space model can be introduced:

$$\begin{aligned}\bar{\mathbf{x}}(t+1) &= \mathbf{\Phi}(t)\bar{\mathbf{x}}(t) + \bar{\mathbf{u}}(t)\\ \bar{\mathbf{z}}(t) &= \bar{\mathbf{x}}(t) + \bar{\mathbf{v}}(t)\end{aligned} \tag{24}$$

where

$$\mathbf{\Phi}(t) = \begin{pmatrix} \mathbf{F}_1(t) & \mathbf{F}_2(t) & \mathbf{F}_3(t) & \mathbf{F}_4(t) \\ \mathbf{F}_2^{\text{i}}(t) & \mathbf{F}_1^{\text{i}}(t) & \mathbf{F}_4^{\text{i}}(t) & \mathbf{F}_3^{\text{i}}(t) \\ \mathbf{F}_3^*(t) & \mathbf{F}_4^*(t) & \mathbf{F}_1^*(t) & \mathbf{F}_2^*(t) \\ \mathbf{F}_4^{\text{k}}(t) & \mathbf{F}_3^{\text{k}}(t) & \mathbf{F}_2^{\text{k}}(t) & \mathbf{F}_1^{\text{k}}(t) \end{pmatrix}$$

[1] Similarly to QWL processing, TWL processing is isomorphic to the real processing and thus, any WL algorithm could be equivalently expressed through a real formalism. In other words, the three approaches (QWL, TWL and $\mathbb{R}^4$) are completely equivalent. However, this equivalence vanishes under properness conditions.

Additionally, assume that $\bar{\mathbf{x}}(0)$ is uncorrelated with the augmented noises $\bar{\mathbf{u}}(t)$ and $\bar{\mathbf{v}}(t)$, and they verify the following conditions:

$$\boldsymbol{\Gamma}_{\bar{\mathbf{u}}}(t,s)=\mathbf{Q}(t)\delta_{ts},\quad \boldsymbol{\Gamma}_{\bar{\mathbf{v}}}(t,s)=\mathbf{R}(t)\delta_{ts},\quad \boldsymbol{\Gamma}_{\bar{\mathbf{u}}\bar{\mathbf{v}}}(t,s)=\mathbf{0}_{4p\times 4p} \tag{25}$$

With the purpose of a TSL processing, we state the following proposition.

Proposition 2. *If* $\mathbf{x}(0)$ *and* $\mathbf{u}(t)$ *are* $\mathbb{T}$*-proper and* $\boldsymbol{\Phi}(t)$ *is of the form*

$$\boldsymbol{\Phi}(t)=\begin{pmatrix}\mathbf{F}_1(t) & \mathbf{0}_{p\times p} & \mathbf{0}_{p\times p} & \mathbf{0}_{p\times p}\\ \mathbf{0}_{p\times p} & \mathbf{F}_1^{\mathrm{i}}(t) & \mathbf{0}_{p\times p} & \mathbf{0}_{p\times p}\\ \mathbf{0}_{p\times p} & \mathbf{0}_{p\times p} & \mathbf{F}_1^{*}(t) & \mathbf{0}_{p\times p}\\ \mathbf{0}_{p\times p} & \mathbf{0}_{p\times p} & \mathbf{0}_{p\times p} & \mathbf{F}_1^{\mathrm{k}}(t)\end{pmatrix} \tag{26}$$

then, $\mathbf{x}(t)$ *in (22) is* $\mathbb{T}$*-proper. If additionally* $\mathbf{v}(t)$ *is* $\mathbb{T}$*-proper then,* $\mathbf{x}(t)$ *and* $\mathbf{z}(t)$ *are jointly* $\mathbb{T}$*-proper.*

Now, our aim is to devise, under the metric (18), the TSL filter of $\mathbf{x}(t)$ and the TSL one-stage predictor of $\mathbf{x}(t+1)$ on the basis of the set of observations $\mathcal{C}=\{\mathbf{z}(1),\dots,\mathbf{z}(t)\}$. This filter is denoted by $\hat{\mathbf{x}}^{\mathrm{TSL}}(t|t)=[\hat{x}_1^{\mathrm{TSL}}(t|t),\dots,\hat{x}_p^{\mathrm{TSL}}(t|t)]^{\mathrm{T}}$, where $\hat{x}_j^{\mathrm{TSL}}(t|t)$ is the TSL estimator of $x_j(t)$ obtained from the information supplied by $\mathcal{C}$. Similarly, $\hat{\mathbf{x}}^{\mathrm{TSL}}(t+1|t)$ denotes the TSL one-stage predictor of $\mathbf{x}(t+1)$ obtained from $\mathcal{C}$, $\hat{\mathbf{x}}^{\mathrm{TWL}}(t|t)$ the TWL filter and $\hat{\mathbf{x}}^{\mathrm{TWL}}(t+1|t)$ the TWL one-stage predictor.

Theorem 5. *If* $\mathbf{x}(0)$, $\mathbf{u}(t)$ *and* $\mathbf{v}(t)$ *are* $\mathbb{T}$*-proper and* $\boldsymbol{\Phi}(t)$ *is as in Eq. (26) then,* $\hat{\mathbf{x}}^{\mathrm{TWL}}(t|t)=\hat{\mathbf{x}}^{\mathrm{TSL}}(t|t)$ *and* $\hat{\mathbf{x}}^{\mathrm{TWL}}(t+1|t)=\hat{\mathbf{x}}^{\mathrm{TSL}}(t+1|t)$ *which can be recursively obtained for* $t\geq 1$ *as*

$$\begin{aligned}\hat{\mathbf{x}}^{\mathrm{TSL}}(t+1|t)&=\mathbf{F}_1(t)\hat{\mathbf{x}}^{\mathrm{TSL}}(t|t)\\ \mathbf{P}_1(t+1)&=\mathbf{F}_1(t)\mathbf{L}_1(t)\mathbf{F}_1^{\mathrm{H}}(t)+\mathbf{Q}_1(t)\\ \mathbf{K}_1(t+1)&=\mathbf{P}_1(t+1)(\mathbf{P}_1(t+1)+\mathbf{R}_1(t+1))^{-1}\\ \hat{\mathbf{x}}^{\mathrm{TSL}}(t+1|t+1)&=\hat{\mathbf{x}}^{\mathrm{TSL}}(t+1|t)+\mathbf{K}_1(t+1)\left[\mathbf{z}(t+1)-\hat{\mathbf{x}}^{\mathrm{TSL}}(t+1|t)\right]\\ \mathbf{L}_1(t+1)&=\mathbf{P}_1(t+1)-\mathbf{K}_1(t+1)\mathbf{P}_1(t+1)\end{aligned} \tag{27}$$

with $\hat{\mathbf{x}}^{\mathrm{TSL}}(0|0)=\mathbf{0}_{p\times 1}$ *and* $\mathbf{L}_1(0)=\boldsymbol{\Gamma}_{\mathbf{x}}(0,0)$.

Moreover, the errors associated to the component $x_i(t)$, $i=1,\dots,p$, *are given by*

$$\begin{aligned}\epsilon_i^{\mathrm{TSL}}(t)&=\|x_i(t)-\hat{x}_i^{\mathrm{TSL}}(t|t)\|^2=\mathcal{R}\{l_{ii}(t)\}\\ \epsilon_i^{\mathrm{TSL}}(t+1|t)&=\|x_i(t+1)-\hat{x}_i^{\mathrm{TSL}}(t+1|t)\|^2=\mathcal{R}\{p_{ii}(t)\}\end{aligned} \tag{28}$$

where $l_{ii}(t)$ *and* $p_{ii}(t)$ *are the* (i,i)*th entries in the matrices* $\mathbf{L}_1(t)$ *and* $\mathbf{P}_1(t)$*, respectively.*

Remark 2. Under the conditions of Theorem 5, the TSL estimators (filter and predictor) have two advantages: they coincide with the TWL estimators and also, the computational burden involved is reduced, since the dimension of the first ones is p while the dimension of the second ones is $4p$. In relation to this second advantage, notice that the same property is shared by the quaternion techniques (see, e.g., [31]). As a consequence, it should be highlighted that the computational complexity of TSL estimators is equal to that of their counterparts in the quaternion domain; i.e., both are of order $O(p^3)$. However, TWL estimators and QWL estimators are of order $O(64p^3)$.

6.2. Example

We consider an example where QSL processing is not the best technique to address the estimation problem, but TSL processing is a superior methodology in terms of performance. For this purpose, the TSL filter and the TSL one-stage predictor, $\hat{x}^{\mathrm{TSL}}(t|t)$ and $\hat{x}^{\mathrm{TSL}}(t+1|t)$, are compared with their counterparts in the quaternion domain $\hat{x}^{\mathrm{QSL}}(t|t)$ and $\hat{x}^{\mathrm{QSL}}(t+1|t)$, in the following case.

Consider the tessarine state-space model:

$$\begin{aligned} x(t+1) &= f x(t) + u(t), \qquad 1 \le t \le 100 \\ z(t) &= x(t) + v(t), \qquad 1 \le t \le 100 \end{aligned} \tag{29}$$

with $f = 0.1 + \mathrm{i}0.2 + \mathrm{j}0.3 + \mathrm{k}0.4 \in \mathbb{T}$, $x(0)$, $u(t)$ and $v(t)$ are tessarine random signals whose real vectors (6) verify

$$\begin{aligned} \mathbf{\Gamma}_{\mathbf{u}_r}(t,s) &= \begin{pmatrix} 0.3 & 0 & \alpha & 0 \\ 0 & 0.3 & 0 & \alpha \\ \alpha & 0 & 0.3 & 0 \\ 0 & \alpha & 0 & 0.3 \end{pmatrix} \delta_{ts} \quad \mathbf{\Gamma}_{\mathbf{v}_r}(t,s) = \begin{pmatrix} 2 & 0 & \beta & 0 \\ 0 & 2 & 0 & \beta \\ \beta & 0 & 2 & 0 \\ 0 & \beta & 0 & 2 \end{pmatrix} \delta_{ts}, \\ \mathbf{\Gamma}_{\mathbf{x}_r}(0,0) &= \begin{pmatrix} 0.04 & 0 & -0.025 & 0 \\ 0 & 0.04 & 0 & -0.025 \\ -0.025 & 0 & 0.04 & 0 \\ 0 & -0.025 & 0 & 0.04 \end{pmatrix} \end{aligned} \tag{30}$$

We conduct a performance comparison study of the TSL and QSL approaches for the following values of the parameters: $\alpha = 0, 0.1, 0.2, 0.29$ and $\beta = 0, -0.7, -1.5, -1.9$.

From Proposition 1 it is easy to check that $x(0)$, $u(t)$ and $v(t)$ are $\mathbb{T}$-proper for any values of α and β. Then, by using Theorem 5 we have that $\hat{x}^{\mathrm{TWL}}(t|t) = \hat{x}^{\mathrm{TSL}}(t|t)$ and $\hat{x}^{\mathrm{TWL}}(t+1|t) = \hat{x}^{\mathrm{TSL}}(t+1|t)$. Moreover, these estimates as well as their associated errors can be obtained via (27) and (28), respectively. These errors depending on the values of α and β are now denoted by $\epsilon^{\mathrm{TSL}}_{(\alpha,\beta)}(t)$ and $\epsilon^{\mathrm{TSL}}_{(\alpha,\beta)}(t+1|t)$, respectively.

Next, the errors of the QSL estimators are computed. Similarly to the tessarine case, they are denoted by $\epsilon^{\mathrm{QSL}}_{(\alpha,\beta)}(t)$ and $\epsilon^{\mathrm{QSL}}_{(\alpha,\beta)}(t+1|t)$. To do so, the autocorrelation function of $x(t)$ in the quaternion domain, $\lambda_x(t,s)$, should be previously determined. From Eq. (8) we have

$$\lambda_x(t,s) = [1,0,0,0]\mathcal{B}\Gamma_{\mathbf{x}_r}(t,s)\mathcal{B}^{\mathrm{H}}[1,0,0,0]^{\mathrm{T}}, \quad t \ge s \tag{31}$$

with $\mathbf{\Gamma}_{\mathbf{x}_r}(0,0)$ and $\mathcal{B}$ given in Eqs. (30) and (9), respectively, and

$$\begin{aligned} \Gamma_{\mathbf{x}_r}(t,t) &= \mathbf{F}_r \Gamma_{\mathbf{x}_r}(t-1,t-1)\mathbf{F}_r^{\mathrm{T}} + \mathbf{\Gamma}_{\mathbf{u}_r}(t-1,t-1), \quad t > 1 \\ \Gamma_{\mathbf{x}_r}(t,s) &= \mathbf{F}_r^{t-s}\Gamma_{\mathbf{x}_r}(s,s), \qquad t > s \end{aligned}$$

being

$$\mathbf{F}_r = \begin{pmatrix} 0.1 & -0.2 & 0.3 & -0.4 \\ 0.2 & 0.1 & 0.4 & 0.3 \\ 0.3 & -0.4 & 0.1 & -0.2 \\ 0.4 & 0.3 & 0.2 & 0.1 \end{pmatrix}$$

Then, $\epsilon^{\mathrm{QSL}}_{(\alpha,\beta)}(t)$ and $\epsilon^{\mathrm{QSL}}_{(\alpha,\beta)}(t+1|t)$ can be obtained via Eqs. (3)–(5).

Table 1
Temporal averages $\overline{DF}(\alpha, \beta)$ and $\overline{DP}(\alpha, \beta)$ in terms of α and β.

	$\overline{DF}(\alpha, \beta)$				$\overline{DP}(\alpha, \beta)$			
	β				β			
α	-1.9	-1.5	-0.7	0	-1.9	-1.5	-0.7	0
0.29	1.42	0.99	0.56	0.34	0.74	0.52	0.29	0.18
0.2	1.19	0.8	0.42	0.23	0.66	0.46	0.26	0.16
0.1	0.94	0.59	0.28	0.15	0.57	0.38	0.22	0.14
0	0.69	0.4	0.17	0.08	0.45	0.29	0.16	0.11

To compare TSL estimators with QSL estimators, we compute the following differences:

$$DF_{(\alpha,\beta)}(t) = \epsilon^{\mathtt{QSL}}_{(\alpha,\beta)}(t) - \epsilon^{\mathtt{TSL}}_{(\alpha,\beta)}(t)$$

$$DP_{(\alpha,\beta)}(t) = \epsilon^{\mathtt{QSL}}_{(\alpha,\beta)}(t+1|t) - \epsilon^{\mathtt{TSL}}_{(\alpha,\beta)}(t+1|t)$$

as well as their temporal averages given by

$$\overline{DF}(\alpha, \beta) = \frac{1}{100}\sum_{t=1}^{100} DF_{(\alpha,\beta)}(t), \qquad \overline{DP}(\alpha, \beta) = \frac{1}{99}\sum_{t=1}^{99} DP_{(\alpha,\beta)}(t) \tag{32}$$

Fig. 3 depicts $DF_{(\alpha,\beta)}(t)$ and $DP_{(\alpha,\beta)}(t)$ for the values of the parameters α and β indicated above. Also, Table 1 shows the temporal averages given by Eq. (32). As we can see, TSL processing outperforms QSL processing for all the values of α and β considered in this example. More specifically, the difference in performance decreases when both α and β tend toward 0 since in this limit case ($\alpha = \beta = 0$) the noises $u(t)$ and $v(t)$ are also $\mathbb{Q}$-proper. In the opposite case, when the values of α and β become more different from 0, the noises $u(t)$ and $v(t)$ move away from the $\mathbb{Q}$-properness condition and thus, the differences are increasingly important. As a conclusion, QSL processing performs worse than TSL processing, with this inferior performance being more evident as the parameters α and β grow in absolute value.

Finally, the computational burden of the different approaches is analyzed. From footnote 1 and [31], any WL hypercomplex filter or predictor (tessarine or quaternion) that could be used in this example has a computational complexity of order $O(64)$. However, from Remark 2 it follows that their SL counterparts have a computational complexity of order $O(1)$. From this analysis and the above performance results, we can conclude that TSL processing is the more adequate in this example.

6.3. TSL stationary models

In this section, a TSL model is proposed for the one-stage prediction problem of a stationary $\mathbb{T}$-proper tessarine signal $\mathbf{x}(t)$. Our aim is to obtain the TSL one-stage predictor, under the metric (18), of $\mathbf{x}(t+1)$ on the basis of the observation set $\mathcal{C} = \{\mathbf{x}(1), \dots, \mathbf{x}(t)\}$. This predictor is denoted by $\hat{\mathbf{x}}^{\mathtt{TSL}}(t+1|t) = [\hat{x}_1^{\mathtt{TSL}}(t+1|t), \dots, \hat{x}_p^{\mathtt{TSL}}(t+1|t)]^{\mathrm{T}}$. Similarly, the TWL one-stage predictor is denoted by $\hat{\mathbf{x}}^{\mathtt{TWL}}(t+1|t)$.

From Eq. (20), the TSL one-stage predictor takes the form

$$\hat{\mathbf{x}}^{\mathtt{TSL}}(t+1|t) = \sum_{i=1}^{t} \mathbf{G}(t,i)\mathbf{x}(t+1-i) \tag{33}$$

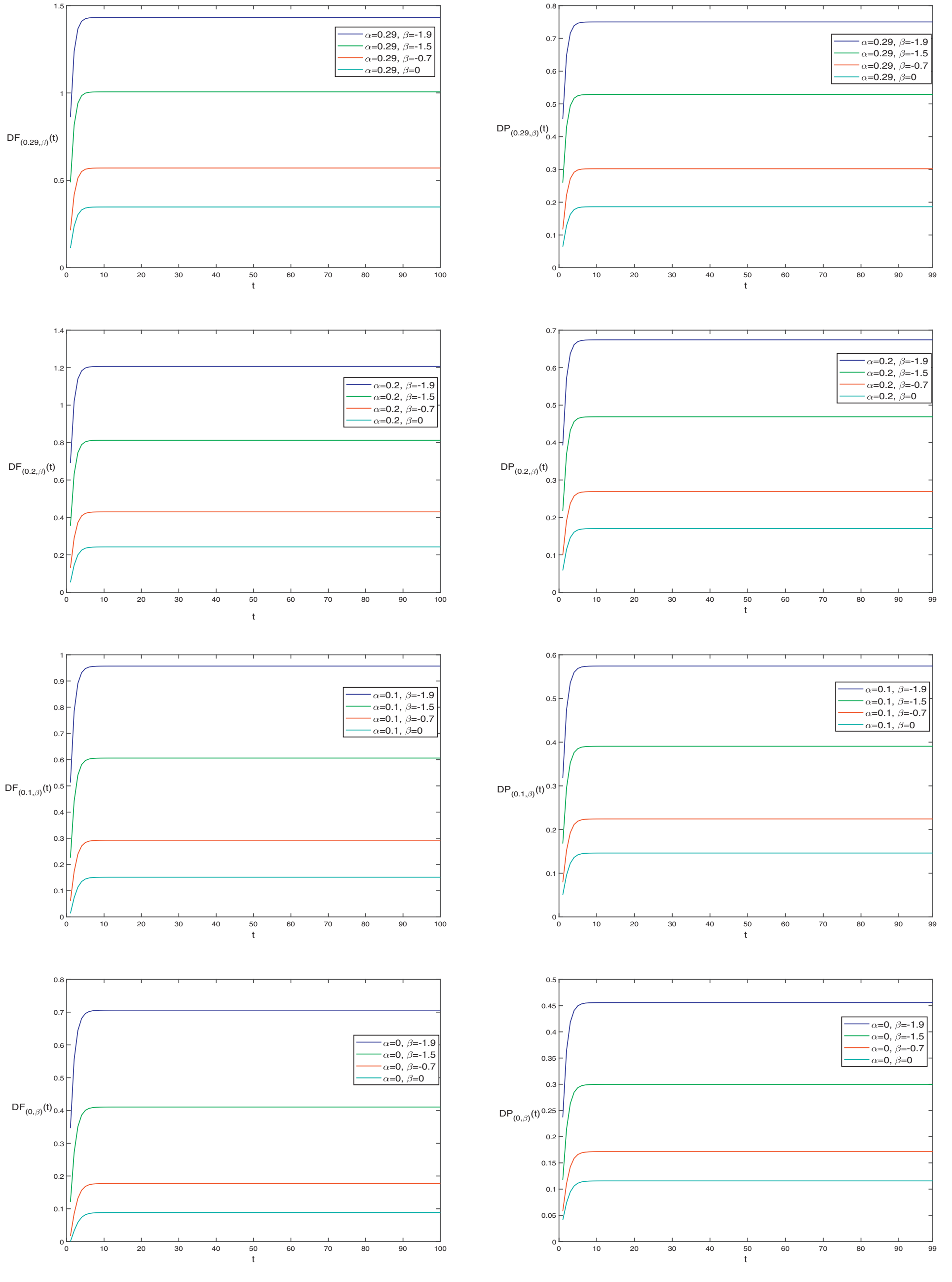

Fig. 3. Differences of the errors for the SL filters $DF_{(\alpha,\beta)}(t)$ (left column) and the SL one-stage predictors $DP_{(\alpha,\beta)}(t)$ (right column).

where the deterministic matrices $\mathbf{G}(t, i) \in \mathbb{T}^{p\times p}$ are obtained via (21).

Although the prediction problem is simplified due to the reduction in the dimensions implied by the $\mathbb{T}$-properness property, we have the inconvenience of computing (33) as the number of observations t grows. However, the stationary condition allows us to obtain the matrices $\mathbf{G}(t, i)$ in an efficient form.

Theorem 6. *Let $\mathbf{x}(t)$ be a stationary $\mathbb{T}$-proper random signal. Then, the one-stage predictors verify $\hat{\mathbf{x}}^{\mathtt{TWL}}(t+1|t) = \hat{\mathbf{x}}^{\mathtt{TSL}}(t+1|t)$ and the matrices $\mathbf{G}(t, i)$ in Eq. (33) can be calculated of the form*

$$\begin{aligned}
\mathbf{G}(t,t) &= \mathbf{H}(t-1)\boldsymbol{\Omega}^{-1}(t-1)\\
\mathbf{J}(t,t) &= \mathbf{H}^{\mathtt{H}}(t-1)\boldsymbol{\Pi}^{-1}(t-1)\\
\mathbf{G}(t,i) &= \mathbf{G}(t-1,i) - \mathbf{G}(t,t)\mathbf{J}(t-1,t-i), \qquad i=1,\dots,t-1\\
\mathbf{J}(t,i) &= \mathbf{J}(t-1,i) - \mathbf{J}(t,t)\mathbf{G}(t-1,t-i), \qquad i=1,\dots,t-1
\end{aligned} \tag{34}$$

where

$$\begin{aligned}
\boldsymbol{\Pi}(t) &= \boldsymbol{\Pi}(t-1) - \mathbf{G}(t,t)\mathbf{H}^{\mathtt{H}}(t-1)\\
\boldsymbol{\Omega}(t) &= \boldsymbol{\Omega}(t-1) - \mathbf{J}(t,t)\mathbf{H}(t-1)\\
\mathbf{H}(t) &= \boldsymbol{\Gamma}_{\mathbf{x}}(t+1) - \sum_{i=1}^{t}\mathbf{G}(t,j)\boldsymbol{\Gamma}_{\mathbf{x}}(t+1-i)
\end{aligned}$$

and with initial conditions

$$\begin{aligned}
\boldsymbol{\Pi}(0) &= \boldsymbol{\Omega}(0) = \boldsymbol{\Gamma}_{\mathbf{x}}(0)\\
\mathbf{H}(0) &= \boldsymbol{\Gamma}_{\mathbf{x}}(1)
\end{aligned}$$

with $\boldsymbol{\Gamma}_{\mathbf{x}}(\omega) = \boldsymbol{\Gamma}_{\mathbf{x}}(t+\omega, t)$. Moreover, the error associated to the component $x_i(t+1)$ is given by

$$\epsilon_i^{\mathtt{TSL}}(t+1|t) = \|x_i(t+1) - \hat{x}_i^{\mathtt{TSL}}(t+1|t)\|^2 = \mathcal{R}\{\pi_{ii}(t)\}, \quad i=1,\dots,p \tag{35}$$

where $\pi_{ii}(t)$ is the element (i, i) of the matrix $\boldsymbol{\Pi}(t)$.

Remark 3. In a similar way to the discussion in Remark 2, the computational complexity of the TSL algorithm given in Theorem 6 is equal to that of its counterpart in the quaternion domain (both procedures are of order $O(n^2)$, with n the number of observations). On the other hand, the associated WL algorithms for stationary signals have a computational burden of order $O(16n^2)$.

6.4. Example

In this section, another simulation example where TSL processing outperforms QSL processing is developed. For this purpose, the TSL predictor $\hat{x}^{\mathtt{TSL}}(t+1|t)$ is compared with its counterpart in the quaternion domain $\hat{x}^{\mathtt{QSL}}(t+1|t)$.

Consider the stationary $\mathbb{T}$-proper tessarine signal $\{x(t), t=1,\dots,100\}$ with pseudo autocorrelation function

$$\begin{aligned}
\gamma_x(0) &= 4(\alpha^2+5) + \mathrm{j}16\alpha\\
\gamma_x(1) &= -8 - \mathrm{j}4\alpha\\
\gamma_x(\omega) &= 0, \quad \omega \neq -1, 0, 1
\end{aligned} \tag{36}$$

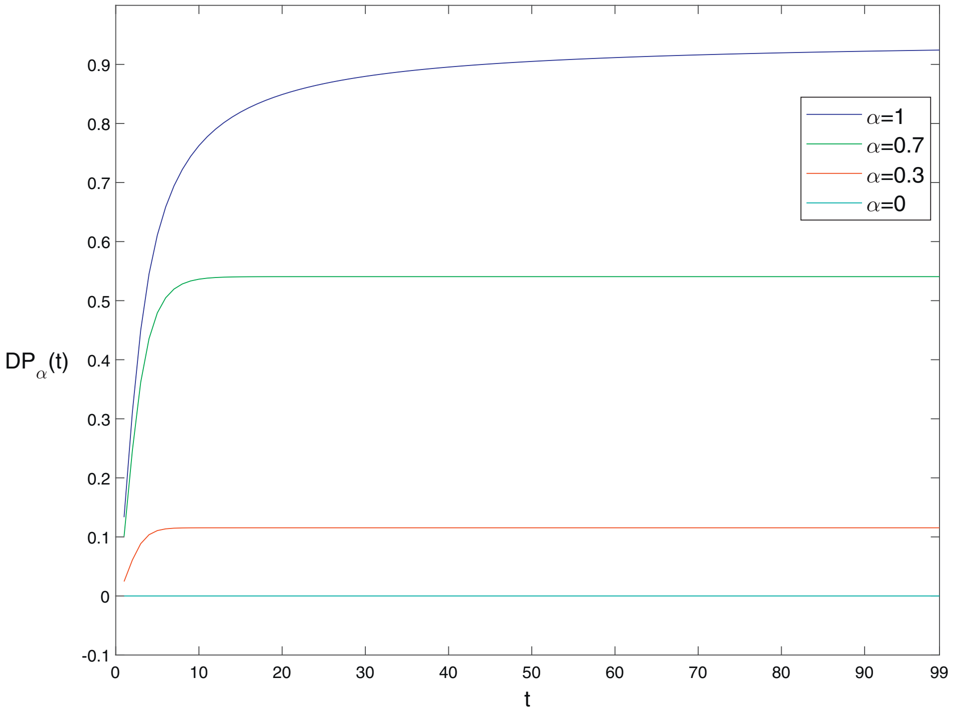

Fig. 4. Differences of the errors for the SL one-stage predictors.

with $\gamma_x(\omega) = \gamma_x(t+\omega, t)$. Next, we compare the performances of the TSL and QSL approaches for the values $\alpha = 0, 0.3, 0.7, 1$.

From Theorem 6, it follows that $\hat{\mathbf{x}}^{\mathtt{TWL}}(t+1|t) = \hat{\mathbf{x}}^{\mathtt{TSL}}(t+1|t)$ for any value of α and this estimator as well as its associated error can be calculated via (34) and (35), respectively. This error is denoted by $\epsilon_\alpha^{\mathtt{TSL}}(t+1|t)$.

Now, we compute the error of the QSL one-stage predictor, denoted by $\epsilon_\alpha^{\mathtt{QSL}}(t+1|t)$. From Eqs. (8) and (36) we obtain that

$$\Gamma_{\mathbf{x}_r}(0) = \begin{pmatrix} \alpha^2+5 & 0 & 4\alpha & 0 \\ 0 & \alpha^2+5 & 0 & 4\alpha \\ 4\alpha & 0 & \alpha^2+5 & 0 \\ 0 & 4\alpha & 0 & \alpha^2+5 \end{pmatrix}, \quad \Gamma_{\mathbf{x}_r}(1) = \begin{pmatrix} -2 & 0 & -\alpha & 0 \\ 0 & -2 & 0 & -\alpha \\ -\alpha & 0 & -2 & 0 \\ 0 & -\alpha & 0 & -2 \end{pmatrix}$$

with $\Gamma_{\mathbf{x}_r}(\omega) = \mathbf{0}_{4\times 4}$, $\omega \neq -1, 0, 1$, and $\Gamma_{\mathbf{x}_r}(\omega) = \Gamma_{\mathbf{x}_r}(t+\omega, t)$. Then, following a similar reasoning to Eq. (31), we obtain the autocorrelation function of $x(t)$ in the quaternion domain. Finally, from Eqs. (3)–(5) we compute $\epsilon_\alpha^{\mathtt{QSL}}(t+1|t)$.

Similarly to the previous example, we calculate the following error performance measures:

$$DP_\alpha(t) = \epsilon_\alpha^{\mathtt{QSL}}(t+1|t) - \epsilon_\alpha^{\mathtt{TSL}}(t+1|t)$$

$$\overline{DP}(\alpha) = \frac{1}{99}\sum_{t=1}^{99} DP_\alpha(t)$$

Table 2
Temporal averages $\overline{DP}(\alpha)$.

α	0	0.3	0.7	1
$\overline{DP}(\alpha)$	0	0.11	0.52	0.85

$DP_\alpha(t)$ is displayed in Fig. 4 and $\overline{DP}(\alpha)$ is calculated for different values of α in Table 2. These results confirm the better performance of the proposed TSL one-stage predictor in relation to the QSL one for all values of α except for $\alpha = 0$. In this case, the signal $x(t)$ is $\mathbb{T}$-proper and $\mathbb{Q}$-proper simultaneously and hence, both methodologies have the same performance.

Finally, we study the computational complexity of WL and SL algorithms for this particular example. Taking Remark 3 into account, the computational order of both TSL and QSL methodologies is $O(99^2)$ which is notably inferior compared to WL approaches (they are of order $O(396^2)$).

To summarize, this example illustrates that the proposed TSL algorithm is the more suitable method for all the values of α except for $\alpha = 0$, value for which both TSL and QSL methodologies have an equal performance and computational burden and thus, any of them can be chosen to solve the problem.

7. Concluding remarks

Hypercomplex algebras have undoubtedly become one of the most attractive topics due to their wide applications in image and signal processing. Although quaternions play a predominant role in these research fields, significant additional benefits are introduced by tessarines. Commutativity, reduced computational complexity, the possibility of extending some important results from the complex-valued signal processing and a better performance in some scenarios make tessarines particularly well suited to applications. Actually, the handicap of not being a normed division algebra does not have important consequences on the processing of the tessarine signal. In this sense, this paper has aimed to contribute to the comparison of both methodologies by focusing on some particular processing where their performance can be different. With this purpose, $\mathbb{T}$-properness has been introduced and $\mathbb{T}$-proper models and estimation algorithms have been devised.

It should be emphasized that neither, tessarine or quaternion processing, is better than the other under all circumstances. Which is to be applied more fruitfully may depend on the specific second order conditions. Then, under $\mathbb{T}$-properness conditions it is expected that our proposed solutions outperform its quaternion counterparts, and vice versa under $\mathbb{Q}$-properness conditions.

On the whole, some questions still remain unanswered or need to be analyzed in depth. For instance, the developing of new tessarine methods or additional examples where tessarines are able to solve real problems in a superior way to other conventional methods.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRediT authorship contribution statement

Jesús Navarro-Moreno: Conceptualization, Methodology, Investigation, Writing - original draft. **Rosa María Fernández-Alcalá:** Formal analysis, Writing - review & editing. **José Domingo Jiménez-López:** Visualization, Software, Validation. **Juan Carlos Ruiz-Molina:** Formal analysis, Supervision, Validation, Writing - review & editing.

Acknowledgments

This work has been supported in part by the I+D+i Project with reference number 1256911, under "Programa Operativo FEDER Andalucía 2014–2020", Junta de Andalucía, and the Project EI_FQM2_2019 of "Plan de Apoyo a la Investigación", University of Jaén, Spain.

Appendix A

A.1. Proof of Theorem 1

Firstly, we prove that the GLRT statistic is given by (14). Let $\mathcal{M}_{\mathbb{R}}$ denote the subset of real autocorrelation matrices of dimension $4p \times 4p$ of the form

$$\mathcal{M}_{\mathbb{R}} = \left\{ \mathbf{\Gamma} : \mathbf{\Gamma} = \begin{pmatrix} \mathbf{F}_1 & -\mathbf{F}_2 & \mathbf{F}_3 & -\mathbf{F}_4 \\ \mathbf{F}_2 & \mathbf{F}_1 & \mathbf{F}_4 & \mathbf{F}_3 \\ \mathbf{F}_3 & -\mathbf{F}_4 & \mathbf{F}_1 & -\mathbf{F}_2 \\ \mathbf{F}_4 & \mathbf{F}_3 & \mathbf{F}_2 & \mathbf{F}_1 \end{pmatrix}, \quad \mathbf{F}_i \in \mathbb{R}^{p\times p}, \quad i = 1, \dots, 4 \right\}$$

Likewise, consider the subset of tessarine pseudo autocorrelation matrices of dimensions $4p \times 4p$, $\mathcal{M}_{\mathbb{T}} = \{\mathbf{\Gamma}_{\bar{\mathbf{x}}} : \mathbf{\Gamma}_{\mathbf{x}\mathbf{x}^{\mathrm{i}}} = \mathbf{\Gamma}_{\mathbf{x}\mathbf{x}^{*}} = \mathbf{\Gamma}_{\mathbf{x}\mathbf{x}^{\mathrm{k}}} = \mathbf{0}_{p\times p}\}$. From Eqs. (8) and (12) we have the following relations between $\mathcal{M}_{\mathbb{R}}$ and $\mathcal{M}_{\mathbb{T}}$:

$$\begin{aligned} &\forall \mathbf{\Gamma}_{\bar{\mathbf{x}}} \in \mathcal{M}_{\mathbb{T}}, \quad \exists \mathbf{\Gamma} \in \mathcal{M}_{\mathbb{R}} \text{ such that } \mathbf{\Gamma}_{\bar{\mathbf{x}}} = 4\mathcal{T}_p \mathbf{\Gamma} \mathcal{T}_p^{\mathrm{H}} \\ &\forall \mathbf{\Gamma} \in \mathcal{M}_{\mathbb{R}}, \quad \exists \mathbf{\Gamma}_{\bar{\mathbf{x}}} \in \mathcal{M}_{\mathbb{T}} \text{ such that } \mathbf{\Gamma} = \tfrac{1}{4}\mathcal{T}_p^{\mathrm{H}} \mathbf{\Gamma}_{\bar{\mathbf{x}}} \mathcal{T}_p \end{aligned} \tag{37}$$

Hence, Eq. (13) is equivalent to test

$$\begin{aligned} &H_0 : \mathbf{\Gamma}_{\mathbf{x}_r} \in \mathcal{M}_{\mathbb{R}} \\ &H_1 : \mathbf{\Gamma}_{\mathbf{x}_r} \notin \mathcal{M}_{\mathbb{R}} \end{aligned}$$

on the basis of the information of the real vectors $\mathbf{x}_{1r}, \dots, \mathbf{x}_{nr}$.

It is well-known that the Gaussian distribution satisfies the conditions of the Wilks' Theorem [40]. Moreover, the log-likelihood is

$$\ln L(\mathbf{x}_{1r}, \dots, \mathbf{x}_{nr}/\mathbf{\Gamma}_{\mathbf{x}_r}) = c - \frac{n}{2}[\ln|\mathbf{\Gamma}_{\mathbf{x}_r}| + \operatorname{trace}(\mathbf{\Gamma}_{\mathbf{x}_r}^{-1}\hat{\mathbf{\Gamma}}_{\mathbf{x}_r})] \tag{38}$$

with $c = -2np\ln(2\pi)$ and under H_1 it is obtained

$$\max_{H_1} \ln L(\mathbf{x}_{1r}, \dots, \mathbf{x}_{nr}/\mathbf{\Gamma}_{\mathbf{x}_r}) = c - \frac{n}{2}[\ln|\hat{\mathbf{\Gamma}}_{\mathbf{x}_r}| + 4p] \tag{39}$$

From (37) and taking into account that $\operatorname{trace}(\mathbf{AB}) = \operatorname{trace}(\mathbf{BA})$, to maximize (38) under H_0 is equivalent to

$$\max_{\mathbf{\Gamma}_{\bar{\mathbf{x}}} \in \mathcal{M}_{\mathbb{T}}} \left[c_1 - \frac{n}{2}[\ln|\mathbf{\Gamma}_{\bar{\mathbf{x}}}| + \operatorname{trace}(\mathbf{\Gamma}_{\bar{\mathbf{x}}}^{-1}\hat{\mathbf{\Gamma}}_{\bar{\mathbf{x}}})] \right] \tag{40}$$

with $c_1 = c + np\ln(16)$ and $\hat{\mathbf{\Gamma}}_{\bar{\mathbf{x}}} = 4\mathcal{T}_p \hat{\mathbf{\Gamma}}_{\mathbf{x}_r} \mathcal{T}_p^{\mathrm{H}}$.

Reasoning as in Lemma 3 of [37], we have that the optimal solution of Eqs. (40) is (15) and then (40) is equal to $c_1 - \frac{n}{2}[\ln|\boldsymbol{\Gamma}_{\mathbb{T}}| + 4p]$. Now, it is straightforward that Eq. (39) is equal to $c_1 - \frac{n}{2}[\ln|\hat{\boldsymbol{\Gamma}}_{\bar{\mathbf{x}}}| + 4p]$. Hence, we have that the Wilks' statistic is given by Eq. (14).

Now, in order to find the asymptotic distribution of Eq. (14), we consider that a real autocorrelation matrix of dimensions $4p \times 4p$ is completely specified by $m_1 = \sum_{i=1}^{4p} i$ real numbers. In the same way, the dimension of the set $\mathcal{M}_{\mathbb{R}}$ is $m_0 = 2p + 4\sum_{i=1}^{p-1} i$, then the degrees of freedom of Eq. (14) under H_0 for a large enough sample size n is $m_1 - m_0 = 2p(3p+1)$, i.e., the Wilks' statistic tends in distribution to $\chi^2(2p(3p+1))$.

A.2. Proof of Theorem 2

The proof is divided in two parts. In the first one, the existence of at least one projection is established. The second part is devoted to prove the unicity.

To start with, and in order to demonstrate the existence of at least one projection, we assume that $\|y\| = \delta$. Applying point 5 of Property 5 for any $x \in \mathcal{G}_{\mathcal{C}}$ such that $\|x\| > 2\delta$ we have that $d(y, x) = \|y - x\| \geq \|x\| - \|y\| > \delta$. Then, since $0 \in \mathcal{G}_{\mathcal{C}}$, if the projection $\hat{y}$ exists, then $\hat{y} \in \{x \in \mathcal{G}_{\mathcal{C}} : \|x\| \leq 2\delta\}$. However, this set is closed and bounded, and hence compact[2]. Thus, applying the classical result of the existence of the projection onto a compact set in a metric space [41], the existence of at least one projection is clear.

Next, we prove the unicity of this projection. Suppose that there are two projections $\hat{y}_1$ and $\hat{y}_2$ of y and $\|y - \hat{y}_i\| = l$, $i = 1, 2$. First, we prove that $(\hat{y}_1 + \hat{y}_2)/2$ is also a projection of y. Applying points 2 and 6 of Property 5 we have

$$d(y, \frac{\hat{y}_1 + \hat{y}_2}{2}) = \|\frac{1}{2}(y - \hat{y}_1) + \frac{1}{2}(y - \hat{y}_2)\| \leq \frac{1}{2}\|y - \hat{y}_1\| + \frac{1}{2}\|y - \hat{y}_2\| = l$$

Now, consider the new random variables $(y - \hat{y}_1)/2, (y - \hat{y}_2)/2 \in \mathbb{T}$. Applying point 4 of Property 5 to these two variables we have

$$\|\frac{\hat{y}_2 - \hat{y}_1}{2}\|^2 = -\|y - \frac{\hat{y}_2 + \hat{y}_1}{2}\|^2 + 2\Big(\|\frac{y - \hat{y}_1}{2}\|^2 + \|\frac{y - \hat{y}_2}{2}\|^2\Big)$$

Then, using that $(\hat{y}_1 + \hat{y}_2)/2$ is a projection and point 6 of Property 5, we obtain that $\|\hat{y}_2 - \hat{y}_1\|^2 = 0$ and hence $\hat{y}_1 = \hat{y}_2$.

A.3. Proof of Theorem 3

We start by proving the necessary condition. To do that, consider $x \in \mathcal{G}_{\mathcal{C}}$. From point 2 of Property 4 we get

$$\begin{aligned}\langle y - x, y - x\rangle &= \langle y - \hat{y} + \hat{y} - x, y - \hat{y} + \hat{y} - x\rangle \\ &= \langle y - \hat{y}, y - \hat{y}\rangle + \langle \hat{y} - x, \hat{y} - x\rangle + \langle y - \hat{y}, \hat{y} - x\rangle + \langle \hat{y} - x, y - \hat{y}\rangle \end{aligned} \tag{41}$$

Since $\hat{y} - x \in \mathcal{G}_{\mathcal{C}}$, it follows that $\langle y - \hat{y}, \hat{y} - x\rangle = 0$ and then, from (41), we have that $\langle y - x, y - x\rangle = \langle y - \hat{y}, y - \hat{y}\rangle + \langle \hat{y} - x, \hat{y} - x\rangle$. As a consequence, $\|y - \hat{y}\|^2 \leq \|y - x\|^2$, $\forall x \in \mathcal{G}_{\mathcal{C}}$, which means that $\hat{y}$ is the projection of y.

[2] Note that, as quaternions, the set $\mathbb{T}$ of all tessarines inherits the topology from $\mathbb{R}^4$. For instance, we say that $S \subset \mathbb{T}$ is compact if, and only if, the set $\{(a, b, c, d) : a + \mathrm{i}b + \mathrm{j}c + \mathrm{k}d \in S\} \subset \mathbb{R}^4$ is compact [5].

Next, we demonstrate the sufficient condition. Suppose there is a random variable $w \in \mathcal{G}_C$ such that $\langle y - \hat{y}, w \rangle = z \neq 0$. We distinguish two cases depending on whether $\mathcal{R}\{z\}$ is equal or different to 0. Suppose that $\mathcal{R}\{z\} \neq 0$. Then, for $\alpha \in \mathbb{R}$ and from point 2 of Property 4, we have that

$$\langle y - \hat{y} - \alpha w, y - \hat{y} - \alpha w \rangle = \langle y - \hat{y}, y - \hat{y} \rangle + \alpha^2 \langle w, w \rangle - \alpha \langle y - \hat{y}, w \rangle - \alpha \langle w, y - \hat{y} \rangle$$

Next, taking $\alpha = \frac{\mathcal{R}\{z\}}{\|w\|^2}$ in the above equation it follows that

$$\|y - \hat{y} - \alpha w\|^2 = \|y - \hat{y}\|^2 - \frac{\mathcal{R}^2\{z\}}{\|w\|^2}$$

which implies that $\|y - \hat{y} - \alpha w\|^2 < \|y - \hat{y}\|^2$, and this is a contradiction to the fact that $\hat{y}$ is the projection of y.

Now, suppose that $\mathcal{R}\{z\} = 0$. Since $z \neq 0$ then, there is at least an imaginary unit different to 0. Assume that the component "i" is different to 0 (similar reasoning can be followed with the other imaginary units). Consider the new tessarine random variable $w_1 = \mathrm{i}w$. From point 2 of Property 4 we have that $\langle y - \hat{y}, w_1 \rangle = z_1 \neq 0$ and $\mathcal{R}\{z_1\} \neq 0$. Then, repeating the previous reasoning for the new variable w_1, the result holds.

A.4. Proof of Theorem 4

Points 1 and 2 are obtained from Corollary 1. The proof of point 3 is immediate due to the fact that $\mathcal{G}_C \subseteq \mathcal{G}_{\bar{C}}$. Finally, from Eqs. (10), (19) and Definition 8 we have that

$$[\mathbf{h}_{11}^{\mathrm{T}}, \ldots, \mathbf{h}_{1m}^{\mathrm{T}}, \mathbf{h}_{21}^{\mathrm{T}}, \ldots, \mathbf{h}_{2m}^{\mathrm{T}}, \mathbf{h}_{31}^{\mathrm{T}}, \ldots, \mathbf{h}_{3m}^{\mathrm{T}}, \mathbf{h}_{41}^{\mathrm{T}}, \ldots, \mathbf{h}_{4m}^{\mathrm{T}}]$$
$$= [\boldsymbol{\Gamma}_{y\mathbf{z}}, \mathbf{0}_{1\times pm}, \mathbf{0}_{1\times pm}, \mathbf{0}_{1\times pm}] \begin{pmatrix} \boldsymbol{\Gamma}_{\mathbf{z}}^{-1} & \mathbf{0}_{pm\times pm} & \mathbf{0}_{pm\times pm} & \mathbf{0}_{pm\times pm} \\ \mathbf{0}_{pm\times pm} & \boldsymbol{\Gamma}_{\mathbf{z}}^{\mathrm{i}^{-1}} & \mathbf{0}_{pm\times pm} & \mathbf{0}_{pm\times pm} \\ \mathbf{0}_{pm\times pm} & \mathbf{0}_{pm\times pm} & \boldsymbol{\Gamma}_{\mathbf{z}}^{*^{-1}} & \mathbf{0}_{pm\times pm} \\ \mathbf{0}_{pm\times pm} & \mathbf{0}_{pm\times pm} & \mathbf{0}_{pm\times pm} & \boldsymbol{\Gamma}_{\mathbf{z}}^{\mathrm{k}^{-1}} \end{pmatrix}$$

and thus, $\hat{y}^{\mathtt{TWL}} = \hat{y}^{\mathtt{TSL}}$.

A.5. Proof of Proposition 2

By applying the principle of recursion in Eq. (24) we get

$$\bar{\mathbf{x}}(t) = \left(\prod_{i=0}^{t-1} \boldsymbol{\Phi}(i) \right) \bar{\mathbf{x}}(0) + \sum_{i=0}^{t-2} \left(\prod_{j=i+1}^{t-1} \boldsymbol{\Phi}(j) \right) \bar{\mathbf{u}}(i) + \bar{\mathbf{u}}(t-1)$$

from which the $\mathbb{T}$-properness of $\mathbf{x}(t)$ follows immediately. Finally, since $\mathbf{x}(t)$ and $\mathbf{z}(t)$ are uncorrelated, they are proven to be cross $\mathbb{T}$-proper. Also, the $\mathbb{T}$-properness of $\mathbf{v}(t)$ implies that $\mathbf{z}(t)$ is $\mathbb{T}$-proper.

A.6. Proof of Theorem 5

From point 4 of Theorem 4 it follows that $\hat{\mathbf{x}}^{\mathtt{TWL}}(t|t) = \hat{\mathbf{x}}^{\mathtt{TSL}}(t|t)$ and $\hat{\mathbf{x}}^{\mathtt{TWL}}(t+1|t) = \hat{\mathbf{x}}^{\mathtt{TSL}}(t+1|t)$. The proof of Eq. (27) consists of two parts. In the first one, a recursive algorithm for the optimal filter and the one-stage predictor of the augmented state $\bar{\mathbf{x}}(t)$ is derived. In the second part, the result is proven by using $\mathbb{T}$-properness conditions on that algorithm.

Denote by $\hat{\bar{\mathbf{x}}}(t|t)$ and $\hat{\bar{\mathbf{x}}}(t+1|t)$ the optimal filter and the one-stage predictor of the augmented state $\bar{\mathbf{x}}(t)$ which obeys (24). It is clear that,

$$\begin{aligned} \hat{\mathbf{x}}^{\mathtt{TWL}}(t|t) &= [\mathbf{I}_p, \mathbf{0}_{p\times 3p}]\hat{\bar{\mathbf{x}}}(t|t) \\ \hat{\mathbf{x}}^{\mathtt{TWL}}(t+1|t) &= [\mathbf{I}_p, \mathbf{0}_{p\times 3p}]\hat{\bar{\mathbf{x}}}(t+1|t) \end{aligned} \tag{42}$$

Now, following a similar reasoning to that given in [35], and under the tessarine product, we obtain that $\hat{\bar{\mathbf{x}}}(t|t)$ and $\hat{\bar{\mathbf{x}}}(t+1|t)$ are given by the tessarine Kalman filter for $t \geq 1$,

$$\begin{aligned} \hat{\bar{\mathbf{x}}}(t+1|t) &= \mathbf{\Phi}(t)\hat{\bar{\mathbf{x}}}(t|t) \\ \mathbf{P}(t+1) &= \mathbf{\Phi}(t)\mathbf{L}(t)\mathbf{\Phi}^{\mathtt{H}}(t) + \mathbf{Q}(t) \\ \mathbf{K}(t+1) &= \mathbf{P}(t+1)(\mathbf{P}(t+1) + \mathbf{R}(t+1))^{-1} \\ \hat{\bar{\mathbf{x}}}(t+1|t+1) &= \hat{\bar{\mathbf{x}}}(t+1|t) + \mathbf{K}(t+1)\left[\bar{\mathbf{z}}(t+1) - \hat{\bar{\mathbf{x}}}(t+1|t)\right] \\ \mathbf{L}(t+1) &= \mathbf{P}(t+1) - \mathbf{K}(t+1)\mathbf{P}(t+1) \end{aligned} \tag{43}$$

with $\hat{\bar{\mathbf{x}}}(0|0) = \mathbf{0}_{4p\times 1}$, $\mathbf{L}(0) = \mathbf{\Gamma}_{\bar{\mathbf{x}}}(0,0)$, and $\mathbf{Q}(t)$ and $\mathbf{R}(t)$ given in Eq. (25).

Moreover,

$$\begin{aligned} \mathbf{L}(t) &= E[(\bar{\mathbf{x}}(t) - \hat{\bar{\mathbf{x}}}(t|t))(\bar{\mathbf{x}}(t) - \hat{\bar{\mathbf{x}}}(t|t))^{\mathtt{H}}] \\ \mathbf{P}(t) &= E[(\bar{\mathbf{x}}(t) - \hat{\bar{\mathbf{x}}}(t+1|t))(\bar{\mathbf{x}}(t) - \hat{\bar{\mathbf{x}}}(t+1|t))^{\mathtt{H}}] \end{aligned} \tag{44}$$

Then, applying the $\mathbb{T}$-properness conditions and Eq. (26) on Eqs. (42) and (43), we obtain (27) with $\mathbf{F}_1(t)$ defined in Eq. (22) and $\mathbf{K}_1(t)$, $\mathbf{P}_1(t)$ and $\mathbf{L}_1(t)$ the matrices of dimensions $p\times p$ formed by the first p rows and p columns of the matrices $\mathbf{K}(t)$, $\mathbf{P}(t)$ and $\mathbf{L}(t)$, respectively. Finally, Eq. (28) is a consequence of Eq. (44).

A.7. Proof of Theorem 6

The proof follows similar steps to that of Theorem 5. Firstly, by using the methodology developed in [36], we prove that $\hat{\bar{\mathbf{x}}}(t+1|t)$ can be calculated via an algorithm similar to the one given in Theorem 6. Finally, from this fact and the $\mathbb{T}$-properness condition, Eqs. (34) and (35) are devised.

References

[1] D. Alfsmann, H.G. Göckler, S.J. Sangwine, T.A. Ell, Hypercomplex algebras in digital signal processing: benefits and drawbacks, in: Proceedings of the 15th European Signal Processing Conference, 2007, pp. 1322–1326.

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

[3] D. Alfsmann, On families of 2n-dimensional hypercomplex algebras suitable for digital signal processing, in: Proceedings of the 14th European Signal Processing Conference, 2006, pp. 1–4.

[4] J.B. Kuipers, Quaternions and Rotation Sequences: A Primer with Applications to Orbits, Aerospace and Virtual Reality, Princeton University Press, New Jersey, 2002.

[5] F.Z. de Castro, M.E. Valle, A broad class of discrete-time hypercomplex-valued hopfield neural networks, Neural Netw. 122 (2020) 54–67.

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

[7] W.R. Hamilton, On quaternions, or on a new system of imaginaries in algebra, Philosophical Mag. 25 (3) (1844) 92–101.

[8] A.J. Hanson, Visualizing Quaternions, Morgan Kaufmann, San Francisco, 2005.

[9] H.J. Luinge, P.H. Veltink, Measuring orientation of human body segments using miniature gyroscopes and accelerometers, Med. Biol. Eng. Comput. 43 (2) (2005) 273–282.

[10] E. Chen, Y. Xu, X. Yang, W. Zhang, Quaternion based optical flow estimation for robust object tracking, Dig. Signal Process. 23 (1) (2013) 118–125.

[11] F. Tobar, D.P. Mandic, Design of positive-definite quaternion kernels, IEEE Signal Process. Letters 22 (11) (2015) 2117–2121.

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

[13] C.C. Took, K. Strbac, K. Aihara, D.P. Mandic, Quaternion-valued short term forecasting of three-dimensional wind and atmospheric parameters, Renew. Energy 36 (2) (2011) 1754–1760.

[14] L. Guo, M. Zhu, X. Ge, Reduced biquaternion canonical transform, convolution and correlation, Signal Process. 91 (2011) 2147–2153.

[15] C. Segre, The real representations of complex elements and extension to bicomplex systems, Math. Ann. 40 (1892) 413–467.

[16] H.D. Schütte, J. Wenzel, Hypercomplex numbers in digital signal processing, in: Proceedings of the International Symposium on Circuits and Systems, 1990, pp. 1557–1560.

[17] V. Sercov, A. Petrovsky, D. Lushtyk, Digital hypercomplex all-pass filters: a novel filters bank building block, in: Proceedings of the 6th International Workshop on Systems, Signals and Image Processing, 1999, pp. 181–184.

[18] G. Sommer, Geometric Computing with Clifford Algebras, Springer-Verlag, Berlin, 2001.

[19] S.H. Pei, J.H. Chang, 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.

[20] S. Miron, N.L. Bihan, J. Mars, High resolution vector sensor array processing based on biquaternions, in: Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, 2006, pp. 1077–1080.

[21] S.H. Pei, J.H. Chang, J.J. Ding, M.Y. Chen, Eigenvalues and singular value decompositions of reduced biquaternion matrices, IEEE Trans. Circuits Syst. I 525 (2008) 2673–2685.

[22] A.T. Kamal, M.T. El-Melegy, Color image processing using reduced biquaternions with applications to face recognition in a PCA framework, in: Proceedings of the IEEE International Conference on Computer Vision Workshops, 2017, pp. 3039–3046.

[23] A. Esposito, M. Faundez-Zanuy, F.C. Morabito, E. Pasero, Neural Advances in Processing Nonlinear Dynamic Signals, Springer, Switzerland, 2018.

[24] M. Kobayashi, Twin-multistate commutative quaternion hopfield neural networks, Neurocomputing 320 (2018) 150–156.

[25] G. Livadiotis, Complex symmetric formulation of maxwell equations for fields and potentials, Mathematics 6 (114) (2018).

[26] F. Ortolani, M. Scarpiniti, D. Comminiello, A. Uncini, On the influence of microphone array geometry on the behavior of hypercomplex adaptive filters, in: Proceedings of the 5th IEEE Microwaves, Radar and Remote Sensing Symposium, 2017, pp. 37–42.

[27] F. Ortolani, D. Comminiello, M. Scarpiniti, A. Uncini, On 4-dimensional hypercomplex algebras in adaptive signal processing, Neural Adv. Process. Nonlinear Dyn. Signals (2017) 131–140.

[28] F. Ortolani, Hypercomplex Adaptive Filtering, University di Roma, 2017 Ph.d thesis.

[29] J. Vía, D. Ramírez, I. Santamaría, Properness and widely linear processing of quaternion random vectors, IEEE Trans. Inform. Theory 56 (7) (2010) 3502–3515.

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

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

[32] J. Navarro-Moreno, R.M. Fernández-Alcalá, J.C. Ruiz-Molina, Semi-widely simulation and estimation of continuous-time $\mathbb{C}^\eta$-proper quaternion random signals, IEEE Trans. Signal Process. 63 (18) (2015) 4999–5012.

[33] J. Navarro-Moreno, J.C. Ruiz-Molina, Semi-widely linear estimation of $\mathbb{C}^\eta$-proper quaternion random signal vectors under gaussian and stationary conditions, Signal Process. 119 (2016) 56–66.

[34] D.H. Dini, C. Jahanchachi, D.P. Mandic, Kalman filtering for widely linear complex and quaternion valued bearings only tracking, IET Signal Process. 6 (5) (2012) 435–445.

[35] C. Jahanchachi, D.P. Mandic, A class of quaternion kalman filters, IEEE Trans. Neural Netw. Learn. Syst. 25 (3) (2014) 533–544.

[36] J. Navarro-Moreno, R.M. Fernández-Alcalá, C.C. Took, D.P. Mandic, Prediction of wide-sense stationary quaternion random signals, Signal Process. 93 (2013) 2573–2580.

[37] J. Vía, D.P. Palomar, L. Vielva, Generalized likelihood ratios for testing the properness of quaternion gaussian vectors, IEEE Trans. Signal Process. 59 (4) (2011) 1356–1370.
[38] P. Ginzberg, A.T. Walden, Testing for quaternion propriety, IEEE Trans. Signal Process. 59 (7) (2011) 3025–3034.
[39] P.J. Brockwell, R.A. Davis, Time Series: Theory and Methods, 2nd, Springer-Verlag, New York, 1991.
[40] H. Chernoff, On the distribution of the likelihood ratio, Ann. Math. Statist. 25 (1954) 573–578.
[41] E.W. Cheney, Introduction to Approximation Theory, 2nd ed., AMS Chelsea Publishing, Providence, 1982.