BLIND EXTRACTION OF IMPROPER QUATERNION SOURCES

[1]S. Javidi, [1]C. Cheong Took, [1]C. Jahanchahi, [2]N. Le Bihan and [1]D. P. Mandic

[1]Imperial College London,UK, [2]GIPSA-Lab, Grenoble, France
{soroush.javidi, c.cheong-took, cyrus.jahanchahi05, d.mandic}@imperial.ac.uk,
nicolas.le-bihan@gipsa-lab.grenoble-inp.fr

ABSTRACT

Blind extraction of quaternion-valued latent sources is addressed based on their local temporal properties. The extraction criterion is based on the minimum mean square widely linear prediction error, thus allowing for the extraction of both proper and improper quaternion sources. The use of the widely linear adaptive predictor is justified by the relationship between the mean square prediction error and the crosscorrelation and cross-pseudocorrelations of the source signals. Simulations on benchmark improper quaternion sources together with a real-world example of EEG artifact removal illustrate the usefulness of the proposed methodology.

Index Terms— Quaternion blind source extraction, improper quaternion signals, quaternion noncircularity, quaternion widely linear model (QWL), widely linear quaternion LMS (WL-QLMS), augmented quaternion statistics.

1. INTRODUCTION

Blind Source Separation (BSS) in the real domain is well understood, and a number of solutions have been proposed for applications in communications, radar and brain imaging. Traditional extensions of BSS methodologies to the complex domain $\mathbb{C}$ assume source signals with circular (rotation invariant) distributions, resulting in straightforward extensions of the corresponding algorithms in $\mathbb{R}$. This way, the distributions are considered independent of the signal phase, however, this assumption is too stringent for most real-world complex signals, leading to under-modelling for noncircular sources.

Recent results in augmented complex statistics have removed any assumption regarding the signal distribution, and instead provide general and accurate models by considering the information in both the covariance $E\{\mathbf{z}\mathbf{z}^H\}$ and the pseudo-covariance $E\{\mathbf{z}\mathbf{z}^T\}$ matrices of $\mathbf{z} \in \mathbb{C}^N$ [1]. A signal is termed second-order circular (proper) if the pseudo-covariance vanishes, and is otherwise called second-order noncircular (improper). This has led to the widely linear (WL) stochastic model [1], where the mean square estimate of a complex-valued signal is obtained from the weighted sum of the linear estimator and its conjugate counterpart [2].

Extensions of the widely linear model and augmented statistics to the four dimensional quaternion domain $\mathbb{H}$ have recently received considerable attention due to their enhanced accuracy in modelling rotation and the coupling between signal components. In [3], the concept of proper quaternion random variables (also known as $\mathbb{Q}$-proper) was discussed as invariance of the probability distribution to rotations by angle $\frac{\pi}{2}$, and was generalised to any arbitrary angle in [4]. A unifying framework has recently been proposed in [5] which defines a set of four bases from which to construct augmented quaternion statistics, with a similar approach given in [6]. These bases can be seen as the quaternion analogue to the complex bases $\{\mathbf{z}, \mathbf{z}^*\}$ in augmented complex statistics, and allow for the exploitation of the complete second-order information present in quaternion signals. The quaternion widely linear model uses those bases to allow for the optimal minimum mean square error modelling of both $\mathbb{Q}$-proper and $\mathbb{Q}$-improper quaternion signals [5, 6, 7].

Existing blind source separation methodologies for the quaternion domain include a semi-blind block-based algorithm in [8], based on the calculation of rotation angle of whitened quaternion data, and the maximum likelihood approach in [9] where the choice of nonlinearities for the score function was discussed. On the other hand, quaternion-valued blind source extraction (BSE) algorithms, designed so that only a few sources of interest from large-scale mixtures are recovered, are still in their infancy but have huge potential due to their ability to extract vector sources. Their introduction would offer both a reduced computational cost and will relax the need for further post-processing for the selection of the desired sources. This is especially important in real-world applications, such as EEG conditioning for brain computer interfacing (BCI), where we may only be interested in removing artifacts from an observed mixture coming from as many as 64 recording channels.

To this end, we introduce a class of BSE algorithms based on the local temporal structure of quaternion source signals. A quaternion widely linear predictor is used to extract both $\mathbb{Q}$-proper and $\mathbb{Q}$-improper sources, based on the smallest normalised prediction error, making such BSE independent of source powers. This is a generalisation of the complex widely linear prediction based BSE algorithm in [10], and is supported by simulations on both $\mathbb{Q}$-proper and $\mathbb{Q}$-improper signals, coming from synthetic and real-world scenarios.

2. QUATERNION WIDELY LINEAR MODEL

Consider the quaternion signal $y(k) = y_a(k) + \imath y_b(k) + \jmath y_c(k) + \kappa y_d(k)$, where $y_a(k), y_b(k), y_c(k)$ and $y_d(k)$ are real-valued scalars, and $\imath, \jmath$ and κ are orthogonal unit vectors, where $\imath^2 = \jmath^2 = \kappa^2 = -1$. Its optimal linear mean square estimate in terms of the observation $\mathbf{x}(k) \in \mathbb{H}^N$ is given by the widely linear model [5]. To show this, we can express the mean square error (MSE) estimator for a quaternion-valued signal[1] $y \in \mathbb{H}$ in terms of the MSE estimators of its respective components, that is

$$\hat{y}_\alpha = E\{y_\alpha | x_a, x_b, x_c, x_d\}, \qquad \alpha = \{a, b, c\} \quad (1)$$

such that $\hat{y} = \hat{y}_a + \imath\hat{y}_b + \jmath\hat{y}_c + \kappa\hat{y}_d$. By employing the perpendicular involutions (self-inverse mappings) [11]

$$y^\beta = -\beta y \beta, \quad \beta = \{\imath, \jmath, \kappa\},$$

the MSE estimator in (1) can be written as[2]

$$\begin{aligned}\hat{y} &= E\{y|x, x^\imath, x^\jmath, x^\kappa\} + \imath E\{y^\imath|x, x^\imath, x^\jmath, x^\kappa\} \\ &\quad + \jmath E\{y^\jmath|x, x^\imath, x^\jmath, x^\kappa\} + \kappa E\{y^\kappa|x, q^\imath, x^\jmath, x^\kappa\},\end{aligned}$$

This results in the so called widely linear estimator

$$\begin{aligned}y(k) &= \mathbf{h}^H(k)\mathbf{x}(k) + \mathbf{g}^H(k)\mathbf{x}^\imath(k) \\ &\qquad + \mathbf{u}^H(k)\mathbf{x}^\jmath(k) + \mathbf{v}^H(k)\mathbf{x}^\kappa(k) \quad (2)\end{aligned}$$

where $\mathbf{h}, \mathbf{g}, \mathbf{u}$ and $\mathbf{v}$ are coefficient vectors and the symbol $(\cdot)^H$ denotes the Hermitian transpose operator. Thus, the complete second-order information in the observation $\mathbf{x}(k)$ is contained in the augmented covariance matrix

$$\begin{aligned}\mathcal{C}^a_{\mathbf{x}} &= E\{\mathbf{x}^a\mathbf{x}^{aH}\} \\ &= \begin{bmatrix} \mathcal{C}_{\mathbf{x}\mathbf{x}} & \mathcal{C}_{\mathbf{x}^\imath} & \mathcal{C}_{\mathbf{x}^\jmath} & \mathcal{C}_{\mathbf{x}^\kappa} \\ \mathcal{C}^H_{\mathbf{x}^\imath} & \mathcal{C}_{\mathbf{x}^\imath\mathbf{x}^\imath} & \mathcal{C}_{\mathbf{x}^\imath\mathbf{x}^\jmath} & \mathcal{C}_{\mathbf{x}^\imath\mathbf{x}^\kappa} \\ \mathcal{C}^H_{\mathbf{x}^\jmath} & \mathcal{C}_{\mathbf{x}^\jmath\mathbf{x}^\imath} & \mathcal{C}_{\mathbf{x}^\jmath\mathbf{x}^\jmath} & \mathcal{C}_{\mathbf{x}^\jmath\mathbf{x}^\kappa} \\ \mathcal{C}^H_{\mathbf{x}^\kappa} & \mathcal{C}_{\mathbf{x}^\kappa\mathbf{x}^\imath} & \mathcal{C}_{\mathbf{x}^\kappa\mathbf{x}^\jmath} & \mathcal{C}_{\mathbf{x}^\kappa\mathbf{x}^\kappa} \end{bmatrix} \in \mathbb{H}^{4N\times 4N} \quad (3)\end{aligned}$$

where $\mathbf{x}^a = [\mathbf{x}^T, \mathbf{x}^{\imath T}, \mathbf{x}^{\jmath T}, \mathbf{x}^{\kappa T}]^T$ is the augmented input vector. The matrices $\mathcal{C}_{\mathbf{x}^\imath}, \mathcal{C}_{\mathbf{x}^\jmath}, \mathcal{C}_{\mathbf{x}^\kappa}$ are called respectively the $\imath$-, $\jmath$- and κ-covariance matrices (or the pseudo-covariance matrices $\mathcal{C}_{\mathbf{x}^\beta} = E\{\mathbf{x}\mathbf{x}^{\beta H}\}$), while $\mathcal{C}_{\mathbf{x}\mathbf{x}} = E\{\mathbf{x}\mathbf{x}^H\}$ is the standard covariance matrix. It is important to note that a $\mathbb{Q}$-proper random vector, $\mathbf{x}(k)$ is not correlated with its involutions; in this case the pseudo-covariance matrices vanish, and the augmented covariance matrix in (3) becomes real-valued diagonal. A detailed account of the quaternion augmented statistics and WL model can be found in [5, 6, 7].

[1] The index k is omitted for brevity.

[2] Since $y_a = \frac{1}{4}(y + y^\imath + y^\jmath + y^\kappa)$, $y_b = \frac{1}{4}(y + y^\imath - y^\jmath - y^\kappa)$, $y_c = \frac{1}{4}(y - y^\imath + y^\jmath - y^\kappa)$, and $y_d = \frac{1}{4}(y - y^\imath - y^\jmath + y^\kappa)$ [5].

3. TEMPORAL BSE OF QUATERNION SIGNALS

Consider the observation vector $\mathbf{x} \in \mathbb{H}^N$, a linear mixture of the latent sources $\mathbf{s} = [s_1, \ldots, s_N]^T \in \mathbb{H}^{N_s}$, given by

$$\mathbf{x}(k) = \mathbf{A}\mathbf{s}(k) \quad (4)$$

where $\mathbf{A} \in \mathbb{H}^{N\times N_s}$ is the matrix of mixing coefficients. The sources are considered independent, with no assumptions made regarding their $\mathbb{Q}$-properness. The mixing matrix is assumed full rank and invertible, and is for simplicity considered to be square. Ideally, the recovered source $y(k) = \mathbf{w}^H\mathbf{x}(k)$, where $\mathbf{w} \in \mathbb{H}^N$ is a demixing vector, such that $\mathbf{b}^H = \mathbf{w}^H\mathbf{A}$ has a single non-zero element b_n, corresponding to the nth source. If $\mathbf{x}(k)$ is whitened, then b_n is of unit magnitude and an arbitrary rotation angle.

The proposed algorithm calculates the demixing vector $\mathbf{w}(k)$ by discriminating between the sources based on their degree of widely linear predictability, measured by the normalised mean square prediction error (MSPE); the extraction architecture is shown in Figure 1. The error $e(k)$ at the output of the widely linear predictor is given by

$$e(k) = y(k) - y_{WL}(k) \quad (5)$$

where $y_{WL}(k)$ is the widely linear predictor output, given in (2). The MSPE $E\{|e(k)|^2\}$ is normalised so that the relative temporal structure, and hence predictability, of the sources is unaffected by differences in the magnitude of the observed mixtures (scaling ambiguity), and the cost function is given by

$$\mathcal{J}(\mathbf{w}, \mathbf{h}, \mathbf{g}, \mathbf{u}, \mathbf{v}) = \frac{E\{|e(k)|^2\}}{E\{|y(k)|^2\}}. \quad (6)$$

Minimising this cost function with respect to the predictor coefficients results in differences between the prediction errors for various sources, and serves as a basis for the proposed BSE. After some simplification, the MSPE can be expressed as

$$\begin{aligned}E\{|e(k)|^2\} &= \xi_0 \\ &- 2\sum_{m=1}^{M} \Re\Big\{\xi_m h_m(k) + \xi_{\imath|m} g_m(k) + \xi_{\jmath|m} u_m(k) + \xi_{\kappa|m} v_m(k)\Big\} \\ &+ 2\sum_{m,\ell=1}^{M} \Re\Big\{h^*_m(k)\xi_{\imath|\ell-m} g_\ell(k) + h^*_m(k)\xi_{\jmath|\ell-m} u_\ell(k) \\ &\qquad + h^*_m(k)\xi_{\kappa|\ell-m} v_\ell(k) + g^*_m(k)\xi^\imath_{\kappa|\ell-m} u_\ell(k) \\ &\qquad + g^*_m(k)\xi^\imath_{\jmath|\ell-m} v_\ell(k) + u^*_m(k)\xi^\jmath_{\imath|\ell-m} v_\ell(k)\Big\} \\ &+ \sum_{m,\ell=1}^{M} \Re\Big\{h^*_m(k)\xi_{\ell-m} h_\ell(k) + g^*_m(k)\xi^\imath_{\ell-m} g_\ell(k) \\ &\qquad + u^*_m(k)\xi^\jmath_{\ell-m} u_\ell(k) + v^*_m(k)\xi^\kappa_{\ell-m} v_\ell(k)\Big\} \quad (7)\end{aligned}$$

where $\xi_{\alpha|\ell-m} \triangleq \mathbf{w}^H \mathbf{A} \mathcal{C}_{\mathbf{s}^\alpha}(\ell-m)\mathbf{A}^{\alpha H}\mathbf{w}^\alpha$ and $\xi_{\ell-m} \triangleq \mathbf{w}^H \mathbf{A}\mathcal{C}_{\mathbf{ss}}(\ell-m)\mathbf{A}^H\mathbf{w}$ and $\Re\{\cdot\}$ denotes the real or scalar part of a quaternion variable. The real-valued MSPE is related to the cross-correlation and cross-pseudo-correlation of the source components; as the sources are assumed orthogonal, these matrices are diagonal. For $\mathbb{Q}$-proper sources, the pseudo-covariances and thus the terms $\xi_{\alpha|\ell-m}$ vanish, simplifying the expression for the MSPE in (7).

A gradient based weight update based on the widely linear predictor is derived using the conjugate gradient within $\mathbb{HR}$ calculus [12], yielding

$$\nabla_{\mathbf{w}^*}\mathcal{J} = \frac{1}{\sigma_y^2(k)}\Big(\breve{\mathbf{x}}_1(k)e^*(k) - \frac{1}{2}e(k)\breve{\mathbf{x}}_2(k) - \frac{\sigma_e^2(k)}{\sigma_y^2(k)}\big(\mathbf{x}(k)y^*(k) - \frac{1}{2}y(k)\mathbf{x}^*(k)\big)\Big) \qquad (8)$$

with

$$\breve{\mathbf{x}}_1(k) = \mathbf{x}(k) - \sum_{m=1}^{M} h_m^*(k)\mathbf{x}(k-m)$$
$$\breve{\mathbf{x}}_2(k) = \mathbf{x}^*(k) - \sum_{m=1}^{M}\big(\mathbf{x}^*(k-m)h_m(k) - \mathbf{x}^{\imath *}(k-m)g_m(k) - \mathbf{x}^{\jmath *}(k-m)u_m(k) - \mathbf{x}^{\kappa *}(k-m)v_m(k)\big). \qquad (9)$$

The demixing vector $\mathbf{w}$ is then normalised to avoid spurious solutions. The moving average estimates σ_y^2 and σ_e^2 of the variance of $y(k)$ and $e(k)$ are given by

$$\sigma_e^2(k) = \gamma_e\sigma_e^2(k-1) + (1-\gamma_e)|e(k)|^2$$
$$\sigma_y^2(k) = \gamma_y\sigma_y^2(k-1) + (1-\gamma_y)|y(k)|^2 \qquad (10)$$

where γ_e and γ_y are the respective forgetting factors[3].

Finally, the gradient for the update of the widely linear predictor coefficients in Figure 1 is given by

$$\nabla_{\breve{\mathbf{w}}^{a*}} = \frac{1}{\sigma_y^2(k)}\Big(-\mathbf{y}^a(k)e^*(k) + \frac{1}{2}e(k)\mathbf{y}^{a*}(k)\Big) \qquad (11)$$

where the vectors $\breve{\mathbf{w}}^a = [\mathbf{h}^T, \mathbf{g}^T, \mathbf{u}^T, \mathbf{v}^T]^T$, $\mathbf{y}(k) = [y(k-1), \ldots, y(k-L)]^T$, $\mathbf{y}^a(k) = [\mathbf{y}^T(k), \mathbf{y}^{\imath T}(k), \mathbf{y}^{\jmath T}(k), \mathbf{y}^{\kappa T}(k)]^T$ and L is the predictor filter length. The algorithm in (11) is therefore a normalised variant of the WL-QLMS algorithm [7]. Note that in the derivation of the updates, non-commutativity of the quaternion multiplication should be taken into account. As desired, in the extraction of $\mathbb{Q}$-proper sources, the elements of $\breve{\mathbf{w}}^a$ become $\mathbf{h} \neq \mathbf{0}, \mathbf{g} = \mathbf{u} = \mathbf{v} = \mathbf{0}$.

4. SIMULATIONS

To illustrate the performance of the proposed BSE algorithm two experimental settings were considered: synthetic benchmark data and real-world EEG data. In the first experiment, two $\mathbb{Q}$-improper benchmark sources of length $N_s = 1000$ were mixed using a random quaternion-valued square mixing matrix. Following [9], source s_1 was chosen as a pure phase-modulated two-point cyclic polytope with improperness measure[4] $r_{s_1} = 1$, and source s_2 was an AR(4) signal generated using noncircular quaternion Gaussian noise, where $r_{s_2} = 0.4387$. The sources were recovered using the proposed extraction algorithms in (8) and (11); the step-size was empirically chosen as $\mu_{\mathbf{w}} = 0.9$, predictor length $L = 10$, step-sizes for the WL predictor coefficient updates $\mu_{\breve{\mathbf{w}}^a} = 0.01$, and forgetting factors in (10) as $\gamma_e = \gamma_y = 0.975$. For these parameters, the MSPE of s_1 and s_2 were respectively 5.7931 and 1.1137. The performances were assessed using the Performance Index (PI) given by

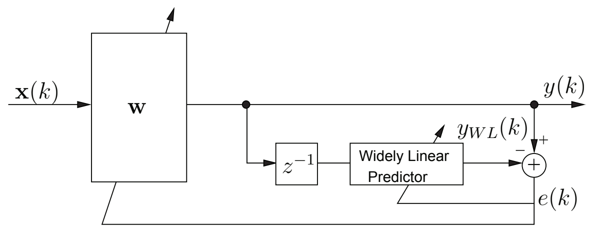

Fig. 1. The prediction-based extraction architecture

$$PI = 10\log_{10}\left(\frac{1}{N}\Big(\sum_{i=1}^{N}\frac{|b_i|^2}{\max\{|b_1|^2, \ldots, |b_N|^2\}} - 1\Big)\right)$$

where $\mathbf{b}^H = \mathbf{w}^H\mathbf{A} = [b_1, \ldots, b_N]^H$; the PI indicates the proximity of $\mathbf{b}$ to a vector with a single non-zero element. As desired, based on (11) the source s_2 with the smallest MSPE was first extracted, taking around 100 samples to converge to the PI of -43.24 dB, as shown in Figure 2. When the same sources were extracted using the standard linear predictor the algorithm diverged, since due to the $\mathbb{Q}$-improperness of the sources the standard linear model was inadequate.

In the next experiment, the line noise and electroencephalogram (EOG) artifacts were extracted from an EEG mixture, recorded from 12 electrodes positioned according to the 10-20 system at AF8, AF4, AF7, AF3, C3, C4, PO7, PO3, PO4, PO8 and the left and right mastoids. In addition, 4 electrodes were placed around both eye sockets to directly record the reference EOG signals[5]. The frontal, central and occipital electrodes were combined into three 4-tuple quaternion-valued EEG signals. The widely linear predictor had $L = 10$ coefficients, step-sizes $\mu_{\mathbf{w}} = 0.9$ and $\mu_{\breve{\mathbf{w}}^a} = 9 \times 10^{-3}$, forgetting factors $\gamma_e = \gamma_y = 0.975$. Deflation was utilised to remove consecutive artifacts from the mixture; the real and imaginary components of the first and second extracted

[3]If $\mathbf{x}(k)$ is whitened, the source estimate power $\sigma_y^2(k) = 1$.

[4]The $\mathbb{Q}$-improperness index $r_s = \frac{|E\{ss^{\imath *}\}| + |E\{ss^{\jmath *}\}| + |E\{ss^{\kappa *}\}|}{3E\{ss^*\}}$ where $r_s \in [0,1]$ and the value $r_s = 0$ indicates a $\mathbb{Q}$-proper source, while for a highly $\mathbb{Q}$-improper source $r_s = 1$.

[5]The EOG measurements were not known to the BSE process, they only served as a reference for performance assessment.

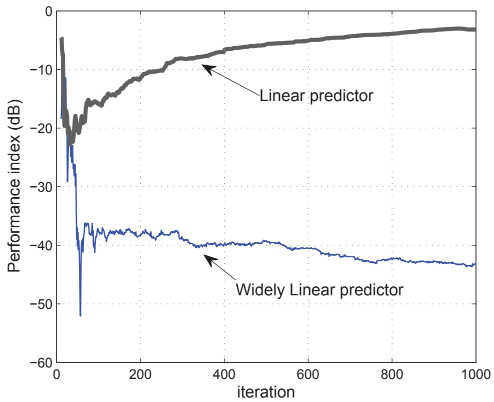

Fig. 2. Learning curves for the quaternion BSE.

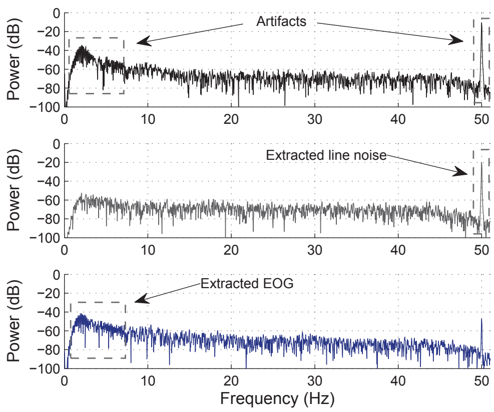

Fig. 3. Power spectra of the reference EOG artifact (*top*), extracted line noise (*middle*) and extracted EOG (*bottom*) using the widely linear predictor.

quaternion-valued signal contained respectively the line noise and EOG artifacts. The power spectra of the EOG artifact, extracted line noise and extracted EOG signal are shown in Figure 3, with the boxed segments highlighting the extracted undesired components. The first extracted signal was the 50Hz line noise, whereas the second extracted signal contains the EOG artifacts corresponding to the 1-8Hz activity. Figure 4 presents the corresponding results for the strictly linear QLMS predictor; the bottom panel shows a 30 dB worse performance for the suppression of the power line noise.

5. CONCLUSIONS

A blind source extraction algorithm capable of extracting both $\mathbb{Q}$-proper and $\mathbb{Q}$-improper quaternion-valued signals based on their temporal profile has been introduced. This has been achieved by minimising the normalised mean square prediction error at the output of a quaternion widely linear predictor. The performance of the algorithm has been demonstrated on the extraction of quaternion synthetic sources, and in EEG artifact extraction, conforming with the analysis.

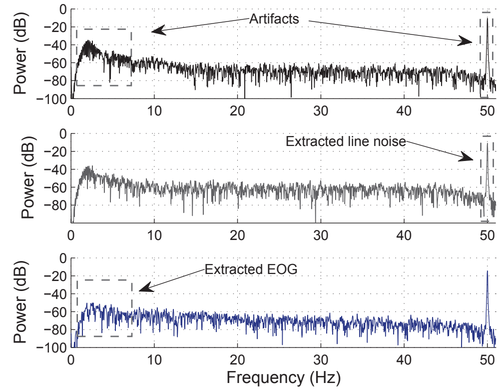

Fig. 4. Power spectra of the reference EOG artifact (*top*), extracted line noise (*middle*) and extracted EOG (*bottom*) using the strictly linear predictor.

References

[1] B. Picinbono and P. Chevalier. Widely linear estimation with complex data. *IEEE Transactions on Signal Processing*, 43(8):2030–2033, 1995.

[2] D. P. Mandic and S. L. Goh. *Complex Valued Nonlinear Adaptive Filters: Noncircularity, Widely Linear and Neural Models*. Wiley, 2009.

[3] N. N. Vakhania. Random vectors with values in quaternion Hilbert spaces. *Theory of Probability and its Applications*, 43(1):99–115, January 1999.

[4] P. O. Amblard and N. Le Bihan. On properness of quaternion valued random variables. In *IMA Conference on Mathematics in Signal Processing*, 2004.

[5] C. Cheong Took and D. P. Mandic. Augmented second order statistics of quaternion random signals. *Signal Processing*, 91(2):214–224, February 2011.

[6] J. Vía, D. Ramírez, and I. Santamaría. Properness and widely linear processing of quaternion random vectors. *IEEE Transactions on Information Theory*, 56(7):3502–3515, 2010.

[7] C. Cheong Took and D. P. Mandic. A quaternion widely linear adaptive filter. *IEEE Transactions on Signal Processing*, 58(8):4427–4431, August 2010.

[8] V. Zarzoso and A. K. Nandi. Closed-form semi-blind separation of three sources from three real-valued instantaneous linear mixtures via quaternions. In *Sixth Int. Symp. on Sig. Proc. and its App.*, volume 1, pages 1–4, 2001.

[9] N. Le Bihan and S. Buchholz. Quaternionic independent component analysis using hypercomplex nonlinearities. In *7th IMA Conference on Mathematics in Signal Processing*, 2006.

[10] S. Javidi, D. P. Mandic, and A. Cichocki. Complex blind source extraction from noisy mixtures using second order statistics. *IEEE Transactions on Circuits and Systems I: Regular Papers*, 57(7):1404–1416, July 2010.

[11] T. A. Ell and S. J. Sangwine. Quaternion involutions and anti-involutions. *Computers & Mathematics with Applications*, 53(1):137–143, January 2007.

[12] D. P. Mandic, C. Jahanchahi, and C. Cheong Took. A quaternion gradient operator and its applications. *IEEE Signal Processing Letters*, 18(1):47–50, 2011.