Fast Independent Component Analysis Algorithm for Quaternion Valued Signals

Soroush Javidi, *Student Member, IEEE*, Clive Cheong Took, *Member, IEEE*,
and Danilo P. Mandic, *Senior Member, IEEE*

Abstract—An extension of the fast independent component analysis algorithm is proposed for the blind separation of both $\mathbb{Q}$-proper and $\mathbb{Q}$-improper quaternion-valued signals. This is achieved by maximizing a negentropy-based cost function, and is derived rigorously using the recently developed $\mathbb{HR}$ calculus in order to implement Newton optimization in the augmented quaternion statistics framework. It is shown that the use of augmented statistics and the associated widely linear modeling provides theoretical and practical advantages when dealing with general quaternion signals with noncircular (rotation-dependent) distributions. Simulations using both benchmark and real-world quaternion-valued signals support the approach.

Index Terms—Augmented quaternion statistics, independent component analysis, quaternion blind source separation, quaternion noncircularity, quaternion widely linear modeling.

I. INTRODUCTION

THE independent component analysis (ICA) methodology is a popular framework for the separation of latent sources from an observed mixture, based on the assumption of statistical independence of the sources [1]. In its standard form, the latent sources are assumed to be linearly mixed, and subject to additive noise. For this scenario, algorithms proposed in the past two decades include those based on the second- and higher-order statistics (second order blind identification and joint approximate diagonalization of eigenmatrices algorithms [2], [3]), and those based on information theoretic criteria such as the maximization of likelihood and minimization of mutual information [4].

The fast independent component analysis (FastICA) algorithm [5], a fast converging algorithm based on the maximization of non-Gaussianity and implemented using an approximative Newton optimization method, has become a standard for the separation of both sub- and super-Gaussian sources. The algorithm has shown to exhibit cubic convergence when in the deflationary mode (using the kurtosis cost function) and local quadratic convergence in the symmetric orthogonalization approach [6]. The FastICA algorithm was subsequently extended to allow for the separation of complex

The authors are with the Department of Electrical and Electronic Engineering, Imperial College London, London SW7 2AZ, U.K. (e-mail: soroush.javidi02@imperial.ac.uk; c.cheong-took@imperial.ac.uk; d.mandic@imperial.ac.uk).

circular sources in [7], and more recently it was generalized to enable separation of complex noncircular sources [8], this was achieved based on the fourth-order moments and through the strong uncorrelating transform [9]. A FastICA algorithm for the separation of both circular and noncircular sources based on a negentropy-based criterion was addressed in [10].

The recent progress in supervised and unsupervised adaptive signal processing algorithms in the complex domain [11] has been made possible through to the advances in the so-called augmented statistics [12] and the framework for analysis of non-analytic functions called the $\mathbb{CR}$ calculus (also known as Wirtinger calculus) [13].

On the other hand, the advances in multidimensional sensor technology (3-D anemometers, 3-D inertial body sensors, robotics) have highlighted the need for adaptive signal processing algorithms in the quaternion domain, which is a natural domain for the processing of 3-D and 4-D signals. The literature on quaternion-valued signal processing includes the algebraic [14], [15] as well as statistical approaches [16], [17]. More recent developments include the analysis of quaternion-valued random variables via augmented quaternion statistics [18], the unitary diagonalization of quaternion matrices [19], and the so-called $\mathbb{HR}$ calculus, which is a unified framework for the analysis of non-analytic quaternion functions [20].

These advances have been exploited through widely linear modeling of quaternion signals [18], [21], allowing the incorporation of the full second-order information and leading to the class of widely linear quaternion least mean square algorithms [22]. For nonlinear signal models, both split and fully quaternion nonlinear models have been successfully implemented [23], [24], whereas for unsupervised adaptive algorithms a quaternion ICA algorithm based on likelihood maximization and the concept of Infomax was proposed by Le Bihan and Buchholz in [25]. In their study, it was concluded that a fully quaternion nonlinearity results in better separation performance. Moreover, real-valued and complex-valued ICA models are shown to be special instances of the quaternion-valued ICA model (all being division algebras), which highlights the generality of quaternions. In other words, quaternion-valued ICA can solve both the real-valued and complex-valued blind source separation paradigm, but the opposite does not necessarily hold. For instance, inter-channel correlation of a quadrivariate independent component is accounted for in a quaternion model, whereas complex-valued ICA models do not allow for the modelling of

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

2   IEEE TRANSACTIONS ON NEURAL NETWORKS

correlation beyond two dimensions. Therefore, the complex-valued ICA model is not likely to reconstruct the quadrivariate sources as shown in Section IV-D.

In this paper, we propose a FastICA algorithm suitable for the separation of both $\mathbb{Q}$-proper and $\mathbb{Q}$-improper (second-order noncircular) quaternion-valued signals from an observed linear mixture. This is achieved by means of the recently introduced augmented quaternion statistics [18], [21], widely linear modeling [18], [21], [22], and $\mathbb{HR}$ calculus [20]. An augmented Newton method is proposed, whereby at a cost of additional complexity we are able to capture the complete statistical properties of the mixing signals and thus ensure successful separation of latent sources. The performance is validated using synthetic $\mathbb{Q}$-proper and $\mathbb{Q}$-improper signals in both deflationary and simultaneous separation scenarios, together with real-world case studies of electroencephalogram (EEG) artifact extraction and in the context of Alamouti coding in communications.

II. PRELIMINARIES ON QUATERNION SIGNALS

A. Quaternion Algebra

Consider the quaternion variable $q = q_a + \iota q_b + J q_c + \kappa q_d \in \mathbb{H}$, where q_a, q_b, q_c, and q_d are real-valued scalars, and $\iota, J,$ and κ are both orthogonal unit vectors and imaginary units, such that

$$\iota J = -J\iota = \kappa \quad J\kappa = -\kappa J = \iota \quad \kappa \iota = -\iota \kappa = J$$
$$\iota J \kappa = \iota^2 = J^2 = \kappa^2 = -1. \tag{1}$$

These identities illustrate the noncommutative property of quaternion products, whereby $q_1 q_2 \neq q_2 q_1$. The variable q can also be written in terms of its real (scalar) part $\Re\{q\} = q_a$ and its vector part $\Im\{q\} = \iota\Im_\iota\{q\} + J\Im_J\{q\} + \kappa\Im_\kappa\{q\}$, such that $q = \Re\{q\} + \Im\{q\}$. Alternatively, by adopting the Cayley–Dickson notation, q can be constructed from a pair of complex quantities $z_1 = q_a + \iota q_b$ and $z_2 = q_c + \iota q_d$, such that $q = z_1 + z_2 J$, however, in this paper direct quaternionic notation will be used.

We next consider three self-inverse mappings[1] or *involutions* [26] about the $\iota, J,$ and κ axes, given by

$$q^\iota = -\iota q \iota = q_a + \iota q_b - J q_c - \kappa q_d$$
$$q^J = -J q J = q_a - \iota q_b + J q_c - \kappa q_d$$
$$q^\kappa = -\kappa q \kappa = q_a - \iota q_b - J q_c + \kappa q_d \tag{2}$$

which form the bases for augmented quaternion statistics [18], [21]. Intuitively, an involution represents a rotation along a single imaginary axis, while the conjugate operator $(\cdot)^*$, also an involution, rotates along all the three imaginary directions, that is

$$q^* = q_a - \iota q_b - J q_c - \kappa q_d. \tag{3}$$

The involutions have the property that $(q_1 q_2)^\alpha = q_1^\alpha q_2^\alpha$, $\alpha = \{\iota, J, \kappa\}$, while $(q_1 q_2)^* = q_2^* q_1^*$. Finally, the norm (modulus) of a quaternion variable q is defined by

$$\|q\|_2 = \sqrt{qq^*} = \sqrt{q^*q} = \sqrt{q_a^2 + q_b^2 + q_c^2 + q_d^2} \tag{4}$$

[1] A self-inverse mapping operator $\text{sinv}(\cdot)$ is such that $\text{sinv}(\text{sinv}(q)) = q$.

while for a vector $\mathbf{q}$ in a quaternion Hilbert space [16], the 2-norm is defined as $\|\mathbf{q}\|_2 = \sqrt{\mathbf{q}^H \mathbf{q}}$.

B. Augmented Statistics and Widely Linear Modeling

For a random vector $\mathbf{q} = \mathbf{q}_a + \iota \mathbf{q}_b + J \mathbf{q}_c + \kappa \mathbf{q}_d \in \mathbb{H}^N$, the probability density function (pdf) is defined in terms of the joint pdf of its components, such that $p_Q(\mathbf{q}) \triangleq p_{Q_a, Q_b, Q_c, Q_d}(\mathbf{q}_a, \mathbf{q}_b, \mathbf{q}_c, \mathbf{q}_d)$. Its mean is then calculated in terms of each respective component as $E\{\mathbf{q}\} = E\{\mathbf{q}_a\} + \iota E\{\mathbf{q}_b\} + J E\{\mathbf{q}_c\} + \kappa E\{\mathbf{q}_d\}$. The covariance matrix of quadrivariate real-valued component vectors $\mathcal{C}_{\mathbf{qq}}^r = E\{\mathbf{q}^r \mathbf{q}^{rT}\} \in \mathbb{R}^{4N \times 4N}$ describes the second-order relationship between the respective components of $\mathbf{q}$, where $\mathbf{q}^r = [\mathbf{q}_a^T, \mathbf{q}_b^T, \mathbf{q}_c^T, \mathbf{q}_d^T]^T$. Representing the components of $\mathcal{C}_{\mathbf{qq}}^r$ by their equivalent quaternion-valued counterparts allows for the complete second-order statistical information to be captured directly in $\mathbb{H}$ [18]. This is achieved by expressing the components of the quaternion random vector $\mathbf{q}$ through its involutions (2), such that

$$\mathbf{q}_a = \frac{1}{4}\left(\mathbf{q} + \mathbf{q}^\iota + \mathbf{q}^J + \mathbf{q}^\kappa\right), \quad \mathbf{q}_b = \frac{1}{4}\left(\mathbf{q} + \mathbf{q}^\iota - \mathbf{q}^J - \mathbf{q}^\kappa\right)$$
$$\mathbf{q}_c = \frac{1}{4}\left(\mathbf{q} - \mathbf{q}^\iota + \mathbf{q}^J - \mathbf{q}^\kappa\right), \quad \mathbf{q}_d = \frac{1}{4}\left(\mathbf{q} - \mathbf{q}^\iota - \mathbf{q}^J + \mathbf{q}^\kappa\right). \tag{5}$$

In analogy to the complex domain[2] where both $\mathbf{z}$ and $\mathbf{z}^*$ are used to define the augmented statistics [27], [28], it can be shown that the bases $\mathbf{q}, \mathbf{q}^\iota, \mathbf{q}^J,$ and $\mathbf{q}^\kappa$ provide a suitable means to define the quaternion augmented statistics [18]. This way, the augmented random vector $\mathbf{q}^a = [\mathbf{q}^T, \mathbf{q}^{\iota T}, \mathbf{q}^{J T}, \mathbf{q}^{\kappa T}]^T$ is used to define the augmented covariance matrix

$$\mathcal{C}_{\mathbf{q}}^a = E\left\{\mathbf{q}^a \mathbf{q}^{aH}\right\}$$
$$= \begin{bmatrix} \mathcal{C}_{\mathbf{qq}} & \mathcal{C}_{\mathbf{q}^\iota} & \mathcal{C}_{\mathbf{q}^J} & \mathcal{C}_{\mathbf{q}^\kappa} \\ \mathcal{C}_{\mathbf{q}^\iota}^H & \mathcal{C}_{\mathbf{q}^\iota \mathbf{q}^\iota} & \mathcal{C}_{\mathbf{q}^\iota \mathbf{q}^J} & \mathcal{C}_{\mathbf{q}^\iota \mathbf{q}^\kappa} \\ \mathcal{C}_{\mathbf{q}^J}^H & \mathcal{C}_{\mathbf{q}^J \mathbf{q}^\iota} & \mathcal{C}_{\mathbf{q}^J \mathbf{q}^J} & \mathcal{C}_{\mathbf{q}^J \mathbf{q}^\kappa} \\ \mathcal{C}_{\mathbf{q}^\kappa}^H & \mathcal{C}_{\mathbf{q}^\kappa \mathbf{q}^\iota} & \mathcal{C}_{\mathbf{q}^\kappa \mathbf{q}^J} & \mathcal{C}_{\mathbf{q}^\kappa \mathbf{q}^\kappa} \end{bmatrix} \in \mathbb{H}^{4N \times 4N} \tag{6}$$

which describes the complete second-order information available. The block matrices in (6), $\mathcal{C}_{\mathbf{q}^\iota}, \mathcal{C}_{\mathbf{q}^J}, \mathcal{C}_{\mathbf{q}^\kappa}$, are respectively termed the ι-, J-, and κ-covariance matrices $E\{\mathbf{qq}^{\alpha H}\}$, $\alpha = \{\iota, J, \kappa\}$, while $\mathcal{C}_{\mathbf{qq}} = E\{\mathbf{qq}^H\}$ is the standard covariance matrix. The ι-, J-, and κ-covariance matrices are also called the complementary or pseudo-covariance matrices [28].

In the domain of second-order statistics, the concept of properness[3] has been extended from the complex to the quaternion domain and has been discussed in [16] and [17]. Following on the involution-based augmented bases in [18] and [21], a random vector is considered $\mathbb{Q}$-proper if it is not correlated with its involutions, or equivalently, $\mathcal{C}_{\mathbf{q}^\iota} = \mathcal{C}_{\mathbf{q}^J} = \mathcal{C}_{\mathbf{q}^\kappa} = \mathbf{0}$, that is, all the cross-covariance matrices vanish, it is otherwise termed $\mathbb{Q}$-improper [18]. Note that,

[2] In the complex domain, the real and imaginary components can be represented in terms of the conjugate coordinates $\mathbf{z}$ and $\mathbf{z}^*$ as $\mathbf{z}_r = (1/2)(\mathbf{z} + \mathbf{z}^*)$ and $\mathbf{z}_i = (1/2J)(\mathbf{z} - \mathbf{z}^*)$.

[3] Properness refers to second-order circularity (rotation invariant pdfs), that is, equal power in all the components, $q_a, q_b, q_c,$ and q_d, of a quaternion random variable $q = q_a + \iota q_b + J q_c + \kappa q_d$.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

JAVIDI *et al.*: FastICA ALGORITHM FOR QUATERNION VALUED SIGNALS

3

for a $\mathbb{Q}$-proper random vector, the augmented covariance matrix in (6) has a block-diagonal structure. More restricted definitions of properness can also be given, whereby one or more pseudo-covariances are nonzero ($\mathbb{C}$-proper) [17]. This can be intuitively understood as rotation invariance along one or more of the quaternion axes, $\mathbb{Q}$-properness thus reflects rotation invariance along all the three imaginary axes.

Recall that the solution to the mean square error (MSE) estimator of a real-valued signal $y \in \mathbb{R}$ in terms of an observation x, that is

$$\hat{y} = E\{y|x\} \tag{7}$$

is given by $\hat{y} = \mathbf{h}^T \mathbf{x}$, where $\mathbf{h}$ is a coefficient vector and $\mathbf{x}$ the regressor vector. As a generalization, the MSE estimator for a quaternion-valued signal $y \in \mathbb{H}$ can be written in terms of the MSE estimators of its respective components, given by

$$\hat{y}_a = E\{y_a|q_a, q_b, q_c, q_d\} \quad \hat{y}_b = E\{y_b|q_a, q_b, q_c, q_d\}$$
$$\hat{y}_c = E\{y_c|q_a, q_b, q_c, q_d\} \quad \hat{y}_d = E\{y_d|q_a, q_b, q_c, q_d\} \tag{8}$$

such that

$$\begin{aligned}
\hat{y} &= \hat{y}_a + \iota\hat{y}_b + J\hat{y}_c + \kappa\hat{y}_d \\
&= E\{y_a|q_a, q_b, q_c, q_d\} + \iota E\{y_b|q_a, q_b, q_c, q_d\} \\
&\quad + J E\{y_c|q_a, q_b, q_c, q_d\} + \kappa E\{y_d|q_a, q_b, q_c, q_d\}.
\end{aligned} \tag{9}$$

Observe that by using the relations (5), the MSE estimator of y can be equivalently written as

$$\begin{aligned}
\hat{y} &= E\{y|q, q^\iota, q^J, q^\kappa\} + \iota E\{y^\iota|q, q^\iota, q^J, q^\kappa\} \\
&\quad + J E\{y^J|q, q^\iota, q^J, q^\kappa\} + \kappa E\{y^\kappa|q, q^\iota, q^J, q^\kappa\}
\end{aligned} \tag{10}$$

which for jointly normal independent components results in the widely linear estimator [18], [22]

$$\begin{aligned}
y &= \mathbf{h}^H \mathbf{q} + \mathbf{g}^H \mathbf{q}^\iota + \mathbf{u}^H \mathbf{q}^J + \mathbf{v}^H \mathbf{q}^\kappa \\
&= \boldsymbol{\omega}^{aH} \mathbf{q}^a
\end{aligned} \tag{11}$$

where the augmented weight vector $\boldsymbol{\omega}^a = [\mathbf{h}^T, \mathbf{g}^T, \mathbf{u}^T, \mathbf{v}^T]^T$. Thus, the linear estimator in (11) is the optimal estimator for the generality of quaternion-valued signals, both proper and improper.

C. Overview of the $\mathbb{HR}$ Calculus

A real-valued cost function, typically the error power, is a common performance criterion across statistical signal processing problems. However, the ways to calculate gradients of a real function of quaternion variables are only just being addressed. In a similar fashion to the complex $\mathbb{CR}$ calculus framework, where a real-valued function is defined based on the conjugate coordinates $\mathbf{z}$ and $\mathbf{z}^*$ [11], [13], [29], in the context of $\mathbb{HR}$ calculus [20] the mapping $f(\mathbf{q})$: $\mathbb{H}^N \mapsto \mathbb{R}$ can be considered as a function of the orthogonal quaternion basis vectors $\mathbf{q}, \mathbf{q}^\iota, \mathbf{q}^J$, and $\mathbf{q}^\kappa$, such that

$$f(\mathbf{q}, \mathbf{q}^\iota, \mathbf{q}^J, \mathbf{q}^\kappa): \mathbb{H}^N \times \mathbb{H}^N \times \mathbb{H}^N \times \mathbb{H}^N \mapsto \mathbb{R}. \tag{12}$$

Likewise, the duality between a quaternion function f and its real-valued equivalent g (vectors in $\mathbb{R}^4$) can be expressed as

$$\begin{aligned}
f(\mathbf{q}) &= f(\mathbf{q}, \mathbf{q}^\iota, \mathbf{q}^J, \mathbf{q}^\kappa) \\
&= f_a(\mathbf{q}_a, \mathbf{q}_b, \mathbf{q}_c, \mathbf{q}_d) + \iota f_b(\mathbf{q}_a, \mathbf{q}_b, \mathbf{q}_c, \mathbf{q}_d) \\
&\quad + J f_c(\mathbf{q}_a, \mathbf{q}_b, \mathbf{q}_c, \mathbf{q}_d) + \kappa f_d(\mathbf{q}_a, \mathbf{q}_b, \mathbf{q}_c, \mathbf{q}_d) \\
&= g(\mathbf{q}_a, \mathbf{q}_b, \mathbf{q}_c, \mathbf{q}_d).
\end{aligned} \tag{13}$$

Then, based on the isomorphism between $\mathbb{R}^4$ and $\mathbb{H}$ and by considering the components of the quaternion variable $\mathbf{q}$ in the orthogonal bases in (5), a relation can be established between the derivatives of $g \in \mathbb{R}^4$ and those taken directly with respect to the quaternion basis variables. Within the $\mathbb{HR}$ calculus, the $\mathbb{HR}$ derivatives are given by [20]

$$\begin{aligned}
\frac{\partial f}{\partial \mathbf{q}} &= \frac{1}{4}\left(\frac{\partial f}{\partial \mathbf{q}_a} - \iota\frac{\partial f}{\partial \mathbf{q}_b} - J\frac{\partial f}{\partial \mathbf{q}_c} - \kappa\frac{\partial f}{\partial \mathbf{q}_d}\right) \\
\frac{\partial f}{\partial \mathbf{q}^\iota} &= \frac{1}{4}\left(\frac{\partial f}{\partial \mathbf{q}_a} - \iota\frac{\partial f}{\partial \mathbf{q}_b} + J\frac{\partial f}{\partial \mathbf{q}_c} + \kappa\frac{\partial f}{\partial \mathbf{q}_d}\right) \\
\frac{\partial f}{\partial \mathbf{q}^J} &= \frac{1}{4}\left(\frac{\partial f}{\partial \mathbf{q}_a} + \iota\frac{\partial f}{\partial \mathbf{q}_b} - J\frac{\partial f}{\partial \mathbf{q}_c} + \kappa\frac{\partial f}{\partial \mathbf{q}_d}\right) \\
\frac{\partial f}{\partial \mathbf{q}^\kappa} &= \frac{1}{4}\left(\frac{\partial f}{\partial \mathbf{q}_a} + \iota\frac{\partial f}{\partial \mathbf{q}_b} + J\frac{\partial f}{\partial \mathbf{q}_c} - \kappa\frac{\partial f}{\partial \mathbf{q}_d}\right)
\end{aligned} \tag{14}$$

and the $\mathbb{HR}^*$ derivatives can then readily be written from (14) by using the property that, for real functions of quaternion variables, $(\partial f/\partial \mathbf{q})^* = \partial f/\partial \mathbf{q}^*$, leading to

$$\begin{aligned}
\frac{\partial f}{\partial \mathbf{q}^*} &= \frac{1}{4}\left(\frac{\partial f}{\partial \mathbf{q}_a} + \iota\frac{\partial f}{\partial \mathbf{q}_b} + J\frac{\partial f}{\partial \mathbf{q}_c} + \kappa\frac{\partial f}{\partial \mathbf{q}_d}\right) \\
\frac{\partial f}{\partial \mathbf{q}^{\iota*}} &= \frac{1}{4}\left(\frac{\partial f}{\partial \mathbf{q}_a} + \iota\frac{\partial f}{\partial \mathbf{q}_b} - J\frac{\partial f}{\partial \mathbf{q}_c} - \kappa\frac{\partial f}{\partial \mathbf{q}_d}\right) \\
\frac{\partial f}{\partial \mathbf{q}^{J*}} &= \frac{1}{4}\left(\frac{\partial f}{\partial \mathbf{q}_a} - \iota\frac{\partial f}{\partial \mathbf{q}_b} + J\frac{\partial f}{\partial \mathbf{q}_c} - \kappa\frac{\partial f}{\partial \mathbf{q}_d}\right) \\
\frac{\partial f}{\partial \mathbf{q}^{\kappa*}} &= \frac{1}{4}\left(\frac{\partial f}{\partial \mathbf{q}_a} - \iota\frac{\partial f}{\partial \mathbf{q}_b} - J\frac{\partial f}{\partial \mathbf{q}_c} + \kappa\frac{\partial f}{\partial \mathbf{q}_d}\right).
\end{aligned} \tag{15}$$

Similar to the conjugate derivatives in (15), involution-wise derivatives are also applicable to real-valued functions of quaternion variables, and are given by

$$\left(\frac{\partial f}{\partial \mathbf{q}}\right)^\alpha = \frac{\partial f}{\partial \mathbf{q}^\alpha}, \qquad \alpha = \{\iota, J, \kappa\}. \tag{16}$$

Recently, in [20], it has been shown that in the quaternion domain, the direction of steepest descent (maximum rate of change of $f(\mathbf{q})$) is given by the derivative with respect to $\mathbf{q}^*$, i.e., $(\partial f/\partial \mathbf{q}^*)$. This can be seen as a natural generalization of Brandwood's result for functions of complex variables [30]. Finally, note that, while we have considered real-valued functions of quaternion variables in the above discussion, the $\mathbb{HR}$ calculus framework can be equally utilized for the analysis of general quaternion-valued functions.

III. QUATERNION FastICA (Q-FastICA) ALGORITHM

Consider the standard ICA model

$$\acute{x} = \mathbf{A}\mathbf{s} \tag{17}$$

where observed mixtures $\acute{x} \in \mathbb{H}^N$ are a weighted sum of N_s latent sources $\mathbf{s} \in \mathbb{H}^{N_s}$ in a noise-free environment, and the

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

4

IEEE TRANSACTIONS ON NEURAL NETWORKS

rows of $\mathbf{A} \in \mathbb{H}^{N \times N_s}$ form the respective mixing parameters. While no knowledge of the mixing process is available, the sources are assumed statistically independent, for convenience, they have zero mean and unit variance and for generality no assumption is made regarding the ι-, $\jmath$-, and κ-variances. The mixing matrix $\mathbf{A}$ is assumed square ($N = N_s$), well-conditioned, and invertible.

We shall now show that for a quaternion random vector $\mathbf{q} \in \mathbb{H}^N$, its whitening matrix $\mathbf{V}$ is given by

$$\mathbf{V} = \mathbf{\Lambda}^{-\frac{1}{2}}\mathbf{E}^H \tag{18}$$

where $\mathbf{\Lambda}$ is the diagonal matrix of right eigenvalues and $\mathbf{E}$ is the matrix of the corresponding eigenvectors of the covariance matrix of $\mathbf{q}$.

Indeed, by expressing the covariance matrix in terms of the quaternion right eigenvalue decomposition $C_{\mathbf{qq}} = E\{\mathbf{qq}^H\} = \mathbf{E}\mathbf{\Lambda}\mathbf{E}^H$ [19], [31], the covariance matrix of the whitened random vector $\mathbf{p} = \mathbf{Vq}$ then becomes

$$E\left\{\mathbf{pp}^H\right\} = \mathbf{V}E\left\{\mathbf{qq}^H\right\}\mathbf{V}^H$$
$$= \mathbf{\Lambda}^{-\frac{1}{2}}\mathbf{E}^H\left(\mathbf{E}\mathbf{\Lambda}\mathbf{E}^H\right)\mathbf{E}\mathbf{\Lambda}^{-\frac{1}{2}} = \mathbf{I} \tag{19}$$

where $\mathbf{I}$ is the identity matrix.

Using the above result as a preprocessing step in ICA algorithms, the quaternion mixture $\acute{x}$ is whitened such that[4]

$$E\left\{\mathbf{xx}^H\right\} = \mathbf{M}E\left\{\mathbf{ss}^H\right\}\mathbf{M}^H = \mathbf{I} \tag{20}$$

where $\mathbf{x} = \mathbf{V}\acute{x} = \mathbf{VAs}$ and $\mathbf{M} \triangleq \mathbf{VA}$ is a new unitary mixing matrix containing the whitening matrix $\mathbf{V}$, given in (18). We aim to obtain a demixing matrix $\mathbf{W}$ such that $\mathbf{W}^H\mathbf{x}$ is an estimate of the original sources, albeit subject to the scaling, phase, and permutation ambiguities. Then for the nth source estimate, we have

$$y_n = \mathbf{w}_n^H\mathbf{x} = \mathbf{w}_n^H\mathbf{Ms} = \mathbf{u}^H\mathbf{s} = e^{\boldsymbol{\xi}\varphi}s_m \tag{21}$$

where $\mathbf{w}_n$ is the nth column of the demixing matrix $\mathbf{W}$, $\mathbf{u}$ is a vector with a single nonzero value given by $e^{\boldsymbol{\xi}\varphi}$ at the nth entry signifying an arbitrary direction within $\mathbb{H}$, φ is an unknown and arbitrary angle, and $\boldsymbol{\xi} = \left(\iota q_b + \jmath q_c + \kappa q_d / \sqrt{q_b^2 + q_c^2 + q_d^2}\right)$ is the unit pure quaternion vector.[5] Finally, note that, by constraining the demixing vector $\mathbf{w}_n$ to unit norm, the estimated source y_n is of unit variance

$$E\left\{y_n y_n^*\right\} = \mathbf{w}_n^H E\left\{\mathbf{xx}^H\right\}\mathbf{w}_n = \mathbf{w}_n^H\mathbf{w}_n = 1 \tag{22}$$

while the matrix $\mathbf{W}$ becomes unitary.

A. Newton-Update-Based ICA Algorithm

The proposed q-FastICA algorithm is based on the maximization of negentropy of the separated sources, following on the previous implementations of the FastICA algorithm in the real and complex domains [1], [7], [10]. This is achieved

---

[4]The result in [19] can be seen as the Takagi factorization of quaternion matrices.

[5]A pure *imaginary* quaternion is referred to as the imaginary or vector part of a quaternion variable.

by choosing an appropriate nonlinear function $G(y)$, so as to make a suitable approximation of the negentropy function.

In [25], three such distinct quaternion nonlinearities were identified: a nonlinear function applied component-wise to y (split-quaternion function), applied to the components of the Cayley–Dickson form of y (split-complex function), or applied directly to y (full-quaternion function). It was also shown that the full-quaternion nonlinearity resulted in the best separation performance. The difficulty in applying a full quaternion nonlinearity in this paper stems from the fact that under the stringent analyticity conditions of the Cauchy–Riemann–Fueter [32] equations, the only analytic function in $\mathbb{H}$ is a constant. As an alternative, local analyticity conditions may be considered in the calculation of the derivatives [24], [33]. However, the assumptions therein may not be valid for general nonlinear functions. For continuity, following the existing complex FastICA (c-FastICA) algorithms which all use a real-valued nonlinear function, we shall also employ a real-valued smooth and even nonlinearity $G : \mathbb{R} \mapsto \mathbb{R}$, while using the $\mathbb{HR}$ calculus and widely linear modeling to implement an augmented Newton method so as to employ the full information available within general $\mathbb{Q}$-improper mixtures.

The real-valued q-FastICA cost function, which approximates negentropy [1], is then defined as

$$\mathcal{J}(\mathbf{w}, \mathbf{w}^\iota, \mathbf{w}^\jmath, \mathbf{w}^\kappa) = E\left\{G(|\mathbf{w}^H\mathbf{x}|^2)\right\} \tag{23}$$

where the cost function $\mathcal{J}$ is written in terms of the four basis vectors, to emphasize the widely linear model. Based on (11), the optimization problem based on (23) can be stated as

$$\mathbf{w}_{\text{opt}} = \arg\max_{\|\mathbf{w}\|_2^2 = 1} \mathcal{J}(\mathbf{w}, \mathbf{w}^\iota, \mathbf{w}^\jmath, \mathbf{w}^\kappa) \tag{24}$$

where the demixing vector is normalized to avoid very small values of $\mathbf{w}$, while keeping the unit variances of the extracted sources.

The solution of this constrained optimization problem is found through the method of Lagrange multipliers and by utilizing the Newton method to perform a fast iterative search for the optimal value $\mathbf{w}_{\text{opt}}$. In summary, the q-FastICA algorithm for the estimation of one possibly noncircular source is expressed in its augmented form as

$$\mathbf{w}^a(k+1) = \mathbf{w}^a(k) - (\mathbf{H}_{\mathbf{ww}}^a)^{-1}\nabla_{\mathbf{w}^{a*}}\mathcal{L}$$
$$\lambda(k+1) = \lambda(k) + \mu\nabla_{\mathbf{w}^{a*}}\mathcal{L}$$
$$\mathbf{w}(k+1) \leftarrow \frac{\mathbf{w}(k+1)}{\|\mathbf{w}(k+1)\|_2} \tag{25}$$

where the augmented demixing vector $\mathbf{w}^a = [\mathbf{w}, \mathbf{w}^\iota, \mathbf{w}^\jmath, \mathbf{w}^\kappa]^T$, $\mathcal{L}$ is the Lagrangian function and λ is the Lagrange parameter updated via a gradient ascent method with step size μ. The vector $\nabla_{\mathbf{w}^{a*}}\mathcal{L}$ and matrix $\mathbf{H}_{\mathbf{ww}}^a$ are respectively, the augmented gradient vector and Hessian matrix of the Lagrangian function. The algorithm is summarized in Algorithm 1 and the full derivation is provided in the Appendix.

The estimation of multiple sources can be performed one by one through a deflationary procedure, where the nth estimated source is recovered either sequentially, by the following Gram–Schmidt orthogonalization procedure or, simultaneously, via a

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

JAVIDI *et al.*: FastICA ALGORITHM FOR QUATERNION VALUED SIGNALS

5

Algorithm 1: The q-FastICA algorithm

1: Preprocessing: Decorrelate the input data **x**, which may be achieved using the whitening matrix **V** (18), and center the data by removing their mean;

2: Initialize the demixing vector $\mathbf{w}_1$ by assigning random values or the identity matrix;

3: Calculate the augmented gradient vector $\nabla_{\mathbf{w}^{a*}}$ and Hessian matrix $\mathbf{H}^a_{\mathbf{ww}}$ of the Lagrangian according to (41);

4: Update the augmented demixing vector as
$$\mathbf{w}^a \leftarrow \mathbf{w}^a - (\mathbf{H}^a_{\mathbf{ww}})^{-1} \nabla_{\mathbf{w}^{a*}} \mathcal{L}$$

5: Update the Lagrange parameter as
$$\lambda \leftarrow \lambda + \mu \nabla_{\mathbf{w}^{a*}} \mathcal{L}$$

6: Normalize the updated demixing vector
$$\mathbf{w} \leftarrow \mathbf{w}/\|\mathbf{w}\|_2$$

7: If the algorithm has not converged, go to step 3.

---

symmetric orthogonalization method. Orthogonalization procedures in the quaternion domain follow from the already established results [34].

Computational complexity of the proposed q-FastICA is approximately $\mathcal{O}(4160N^3 + 32N^2T)$, whereas that of c-FastICA in the long vector mode is approximately $\mathcal{O}(32N^3 + 48N^2T)$, symbols $N = N_s$ and T denote, respectively, the number of mixtures/sources and the number of samples for each source.

IV. SIMULATIONS AND DISCUSSION

A. Benchmark Simulations

The performance of the q-FastICA algorithm is first assessed through simulations using synthetic 4-D signal codes located on the edges of geometric polytopes [35] with varying degree of $\mathbb{Q}$-improperness. To assess the degree of $\mathbb{Q}$-improperness of the generated sources, we define a circularity measure based on the ratio of the complementary variances to the standard variance, expressed as [18]

$$r = \frac{\left|E\{qq^{\imath*}\}\right| + \left|E\{qq^{\jmath*}\}\right| + \left|E\{qq^{\kappa*}\}\right|}{3E\{qq^*\}}, \quad r \in [0, 1]. \quad (26)$$

This way, a measure of $r = 0$ indicates a $\mathbb{Q}$-proper source, while for a highly $\mathbb{Q}$-improper source $r = 1$.

For the q-FastICA algorithm with symmetric orthogonalization, the standard performance index (PI) measure was used, given by

$$PI = 10 \log_{10} \left(\frac{1}{N} \sum_{i=1}^{N} \left(\sum_{j=1}^{N} \frac{|b_{ij}|}{\max\{|b_{i1}|, \ldots, |b_{iN}|\}} - 1 \right) \right.$$
$$\left. + \frac{1}{N} \sum_{j=1}^{N} \left(\sum_{i=1}^{N} \frac{|b_{ij}|}{\max\{|b_{1j}|, \ldots, |b_{Nj}|\}} - 1 \right) \right) \quad (27)$$

where $\mathbf{B}^H = \mathbf{W}^H \mathbf{VA}$ and $b_{ij} = (\mathbf{B})_{ij}$ and a PI of less than -10 dB signifies good separation performance.

For simulations, 5000 samples of four polytope sources were mixed using a randomly generated quaternion-valued
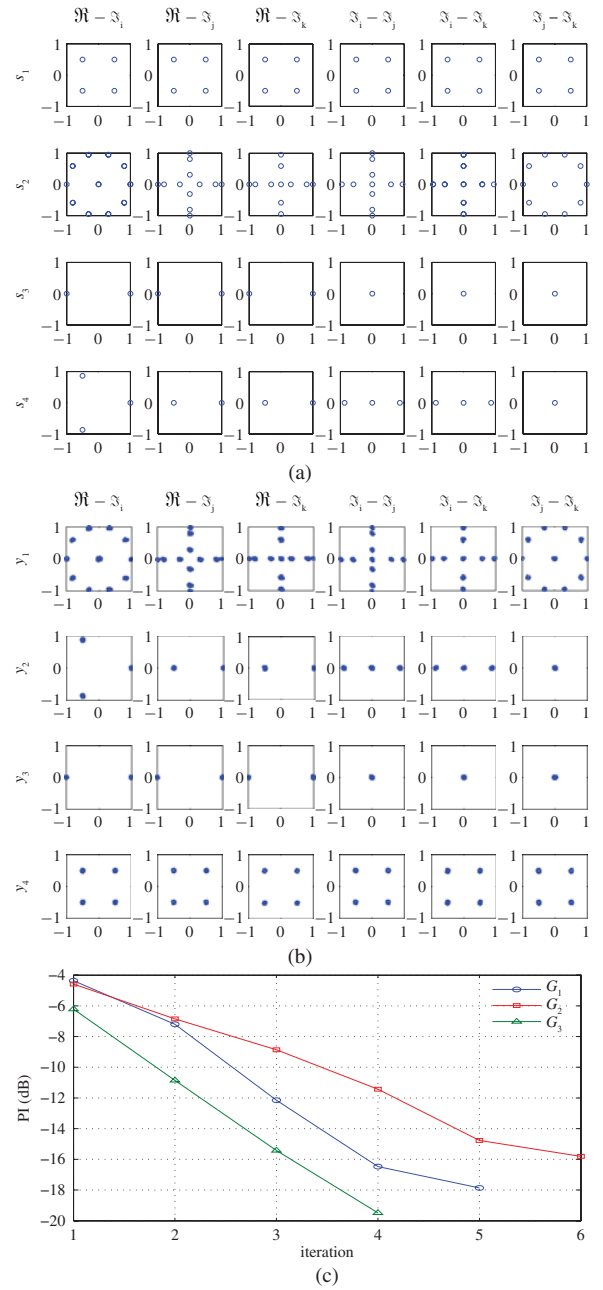

Fig. 1. Performance of the q-FastICA algorithm for the separation of four sources using a symmetric orthogonalization procedure. (a) Scatter plot of the quaternion sources, properties given in Table I. (b) Scatter plot of the estimated sources, with nonlinearity $G_1(y)$. (c) PI at each iteration of the ICA procedure for nonlinearities $G_1(y) = \log \cosh(y)$, $G_2(y) = \sqrt{0.1 + y}$ and $G_3(y) = \log(0.1 + y)$.

4×4 mixing matrix. The observed mixtures were then decorrelated and processed using the q-FastICA algorithm (25), with symmetric orthogonalization. To estimate the sources simultaneously, Table I describes the source properties, while scatter plot performance graphs are given in Fig. 1(a). Sources $s_1(k)$ to $s_4(k)$ were, respectively, generated from cubic, 5-point dicyclic, 2-point cyclic, and 3-point cyclic groups, source $s_3(k)$ had a high degree of $\mathbb{Q}$-improperness $r = 1$, the source $s_4(k)$ had the value of $r = 0.3351$, and the other two sources were $\mathbb{Q}$-proper. For performance comparison,

TABLE I

SOURCE PROPERTIES FOR SIMULATIONS ON BENCHMARK SIGNALS BASED ON SYMMETRIC ORTHOGONALIZATION

Source	Polytope	$\mathbb{Q}$-improperness measure (r)
$s_1(k)$	Cubic	0.0104
$s_2(k)$	Dicyclic (5 point)	0.0089
$s_3(k)$	Cyclic (2 point)	1.0000
$s_4(k)$	Cyclic (3 point)	0.3351

TABLE II

SEPARATION OF SOURCES WITH DIFFERENT MEASURES OF $\mathbb{Q}$-IMPROPERNESS. PI (27) IS GIVEN IN DECIBELS

			Quaternion infomax ICA		
	r_{s_1}	r_{s_2}	Full-$\mathbb{H}$	Split-$\mathbb{C}$	q-FastICA
'CASE 1'	0.01	0.01	-5.57	-3.11	-16.69
'CASE 2'	0.34	0.02	-4.57	-2.51	-18.16
'CASE 3'	0.35	1.00	-3.11	-2.21	-17.35

Fig. 2. Performance of the proposed q-FastICA algorithm (25) and the quaternion Infomax ICA in separating sources with varying degrees of $\mathbb{Q}$-improperness, whose properties are given in Table II.

the nonlinearity G was chosen as in [7], with $G_1(y) = \log \cosh(|y|^2)$, $G_2(y) = \sqrt{0.1 + |y|^2}$ and $G_3(y) = \log(0.1 + |y|^2)$. The demixing matrix $\mathbf{W}$ was initialized randomly, and the step sizes were set as $\mu_1 = 1$, $\mu_2 = 0.1$, $\mu_3 = 0.5$, and $\lambda = 5$ for the gradient ascent update algorithm. As shown in Fig. 1(c), the algorithm successfully separated all the four sources with the respective PI values of -17.87, -15.8086, and -19.4882 dB. Fig. 1(b) illustrates the scatter plots of the normalized estimated sources for the case with nonlinearity G_1, note that sources were estimated in a random order.

B. Effect of $\mathbb{Q}$-Improperness on Performance

In the next stage, the effect of the $\mathbb{Q}$-impropriety of sources on the performance of the algorithm when separating using a symmetric orthogonalizsation method was assessed over three cases for the mixture of 5000 samples of a $\mathbb{Q}$-proper and $\mathbb{Q}$-improper polytope signal, given in Table II and also shown in Fig. 2. For comparison, the quaternion Infomax ICA algorithm [25] was utilized with both full-quaternion and split-complex nonlinearity, where $G(y) = \log \cosh(y)$. Observe that the q-FastICA algorithm had a consistent performance in all three cases with an average PI of -17.40 dB, outperforming the quaternion Infomax ICA whose performance further deteriorated for $\mathbb{Q}$-improper sources, having an average PI of -4.42 and -2.61, respectively, for the full-quaternion and split-complex nonlinearity.

C. EEG Artifact Extraction

In an EEG recording session, each channel comprises a mixture of an EEG signal corresponding to the neural activity, and electrical activity pertaining to artifacts such as movement of the head, line noise, and eye blinks. In the modeling of the EEG signal, the artifacts, both external and biological, were considered statistically independent from pure EEG [36]–[38]; and the usefulness of the real-valued FastICA algorithm in the extraction of eyeblink artifacts was studied in [39].

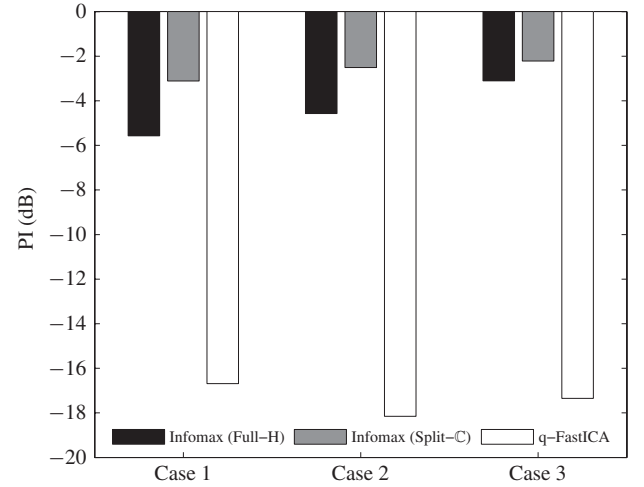In our experimental setup, data was sampled at 4.8 kHz for 30 s from 12 electrodes placed symmetrically on the scalp according to the 10–20 system, as shown in Fig. 3(f), with the reference and ground electrodes placed, respectively, on the right earlobe and forehead. The electrodes used were AF7, AF8, AF3, AF4, ML, MR, C3, C4, PO7, PO8, PO3, and PO4, where the ML and MR electrodes were placed, respectively, on the left and right mastoid. In addition, the voltage difference between the two pairs of electrodes placed above and to the side of the eye sockets measured the electrooculogram (EOG), i.e., the electrical activity due to eye blinks and eye movement. The left and right EOG was combined into a quadrivariate signal as a reference to assess the performance of the q-FastICA algorithm in artifact removal from EEG.

The three quaternion-valued EEG signals were formed from four symmetric electrodes from the frontal (AF7, AF8, AF3, AF4), central (ML, MR, C3, C4) and occipital (PO7, PO8, PO3, PO4) regions of the head. The so-constructed $\mathbb{Q}$-improper quaternion signals were

$$x_1(k) = \text{AF8}(k) + \iota \text{AF4}(k) + \jmath \text{AF7}(k) + \kappa \text{AF3}(k)$$
$$x_2(k) = \text{MR}(k) + \iota \text{C3}(k) + \jmath \text{ML}(k) + \kappa \text{C4}(k)$$
$$x_3(k) = \text{PO8}(k) + \iota \text{PO4}(k) + \jmath \text{PO3}(k) + \kappa \text{PO7}(k) \quad (28)$$

and the observed EEG mixtures were then represented as $\mathbf{x} = [x_1(k), x_2(k), x_3(k)]^T$. The corresponding degrees of $\mathbb{Q}$-impropriety were, respectively, 0.8902, 0.6824, and 0.8932, measured according to (26).

In this scheme, the q-FastICA algorithm (25) was first utilized to estimate the source signals, with the step size $\mu = 1$ and initial Lagrange parameter $\lambda = 5$ was chosen empirically, while following our earlier analysis, the nonlinearity was chosen to be $G(y) = \log \cosh(y)$. Next, the estimated source pertaining to the EOG artifact was selected through the examination of the kurtosis values of the components of the separated sources. Pure EEG signals typically have near-zero kurtosis values, while those belonging to EOG artifacts have super-Gaussian distributions and thus large kurtosis values [40], this being attributed the the sparse nature of eye blinks.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

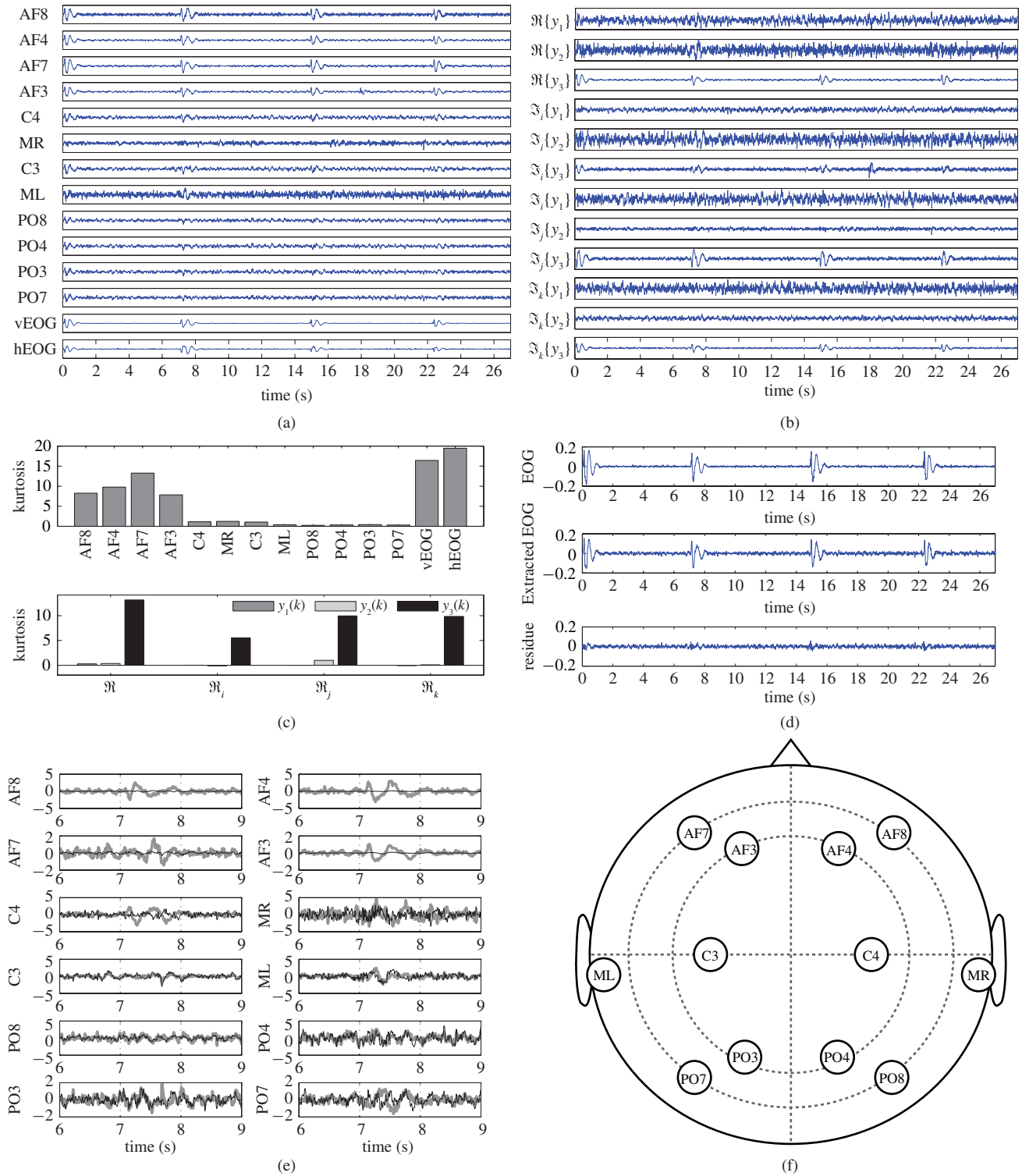JAVIDI *et al.*: FastICA ALGORITHM FOR QUATERNION VALUED SIGNALS

7

Fig. 3. Removal of EOG artifact from an EEG recording using the proposed q-FastICA algorithm. (a) Recorded EEG and EOG channels. (b) Components of the estimated sources. (c) *Top:* Kurtosis values of the recorded EEG channels, *bottom:* Kurtosis values of each component of the estimates. (d) Original and reconstructed EOG signals, along with the residual estimation error. (e) Original recorded EEG (thick gray line) and clean EEG mixture after artifact removal (thin black line), shown between 6s–9s. (f) Placement of the EEG recording electrodes.

A waveform of the original recorded channels and the components of the quaternion-valued separated sources are depicted, respectively, in Fig. 3(a) and (b). The occurrences of eye blinks can be seen at the beginning of the recording, then at around 7, 15, and 22 s, with the effect of the EOG artifact more prominent on the frontal lobe channel and less severe in the central and occipital channels. By inspection, the separated EOG artifact can be seen in the components of the third extracted source $y_3(k)$, that is $\Re\{y_3(k)\}, \Im_\iota\{y_3(k)\}, \Im_J\{y_3(k)\}, and \Im_\kappa\{y_3(k)\}$, and is verified

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

8   IEEE TRANSACTIONS ON NEURAL NETWORKS
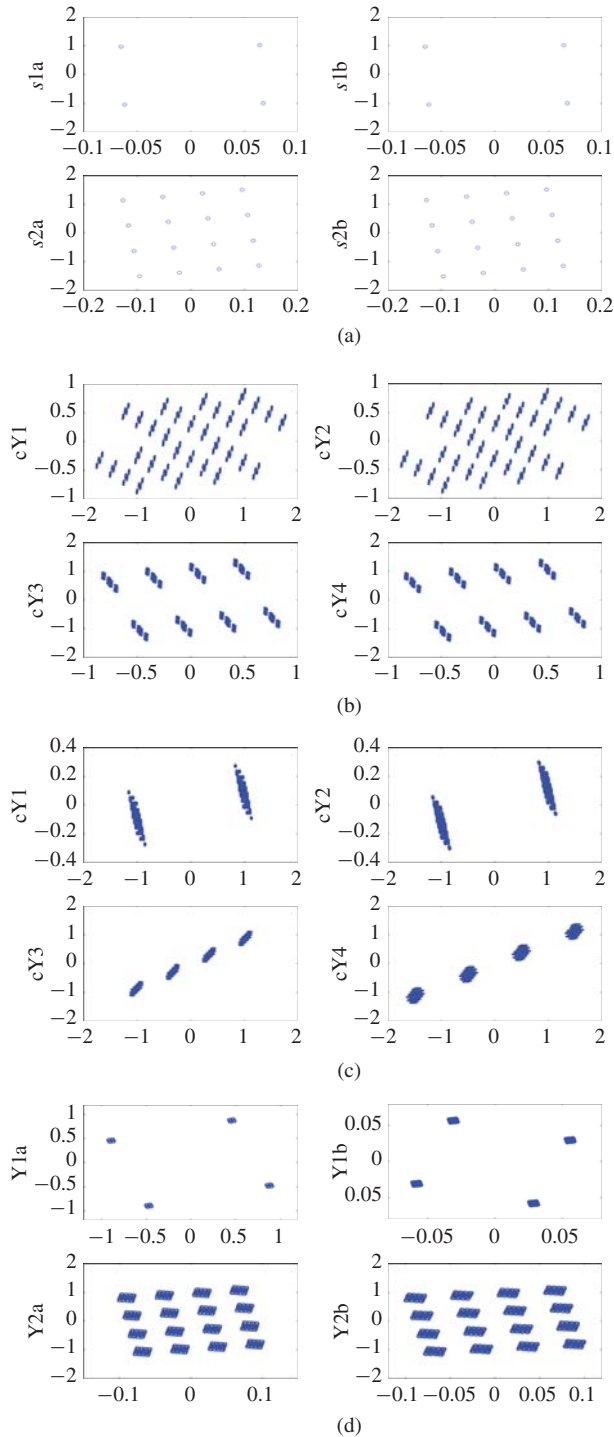
Fig. 4. Communication codes in quaternion mixing. (a) Communication sources. (b) c-FastICA estimates (long vector). (c) c-FastICA estimates. (d) q-FastICA estimates.
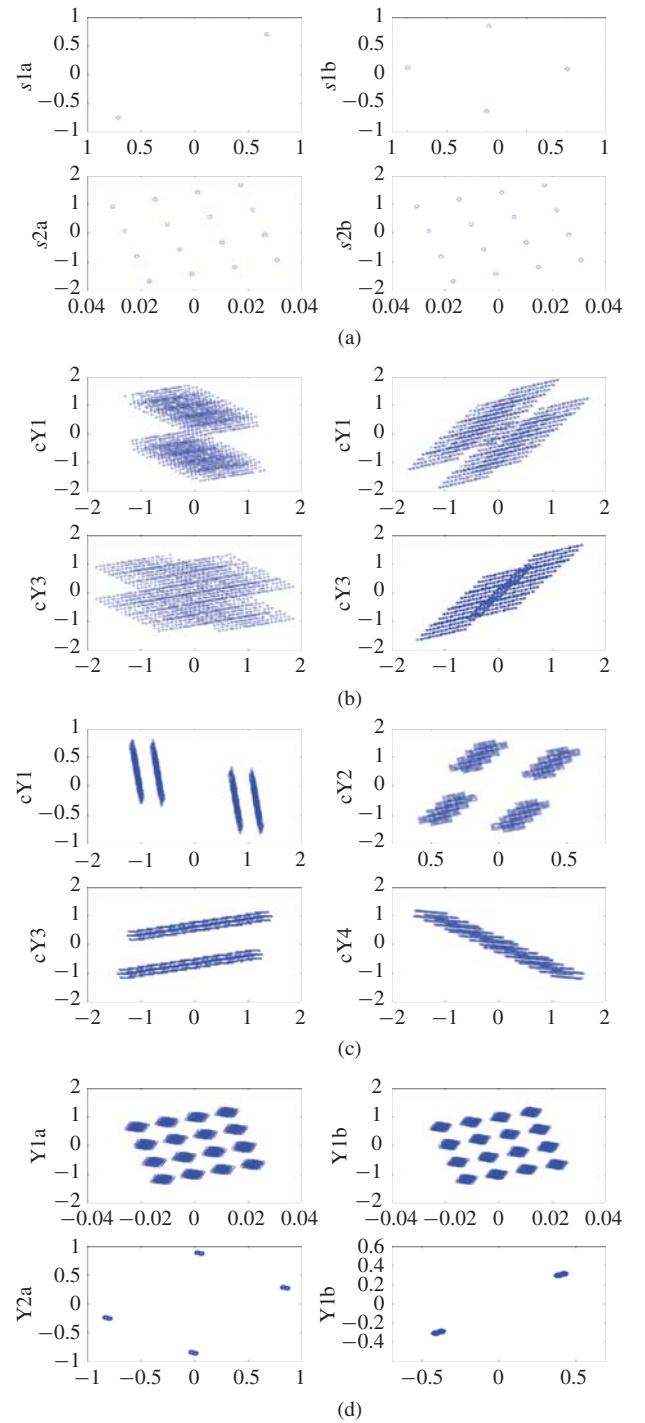
Fig. 5. Communication codes in quaternion mixing when the constellation signals $s1a \neq s1b$. (a) Communication sources. (b) c-FastICA estimates (long vector). (c) c-FastICA estimates. (d) q-FastICA estimates.

through comparison of the kurtosis values of each component Fig. 3(c). While most of the estimated sources had a near-zero measure of kurtosis, the real and imaginary components of $y_3(k)$ exhibited, in comparison, very large kurtosis values.

To study the effectiveness of the algorithm in removing the artifact, the components of $y_3(k)$ were reconstructed to form the EOG artifact by averaging the four channels of $y_3(k)$ and were then compared to the original quadrivariate

EOG recording (which is also the average of the two pairs of electrodes placed near the eye sockets). Fig. 3(d) depicts the EOG extraction performance, along with the residual error of the estimation process, having a small MSE of 1.21×10^{-4}. By excluding the artifact components contained in $y_3(k)$, the clean EEG mixture was reconstructed, a 3-s window between 6 and 9 s for each reconstructed channel is shown in Fig. 3(e), where the effect of the EOG present at 7 s was diminished in all the channels.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

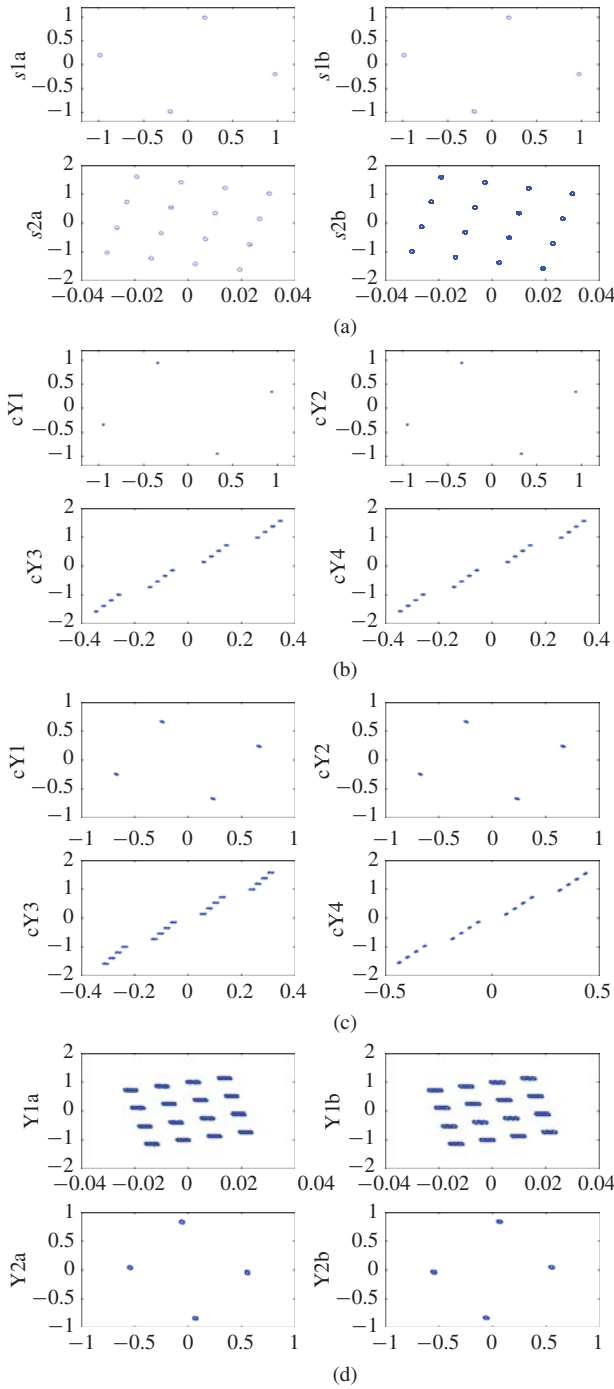JAVIDI *et al.*: FastICA ALGORITHM FOR QUATERNION VALUED SIGNALS

9

Fig. 6. Communication codes in complex mixing. (a) Communication sources. (b) c-FastICA estimates (long vector). (c) c-FastICA estimates. (d) q-FastICA estimates.

D. Alamouti-Based Communication Systems

We next present the application of q-FastICA in the context of a practical communication problem of Alamouti coding, which was shown to enhance the reliability of data communications. For a single user, the model of two transmit antennas and one antenna receiver can be expressed as [41]

$$\begin{bmatrix} \acute{x}_a \\ \acute{x}_b \end{bmatrix} = \begin{bmatrix} s_a & -s_b^* \\ s_b & s_a^* \end{bmatrix} \begin{bmatrix} a_\alpha \\ a_\beta \end{bmatrix} \in \mathbb{C} \qquad (29)$$

$$\acute{x} = as \in \mathbb{H} \qquad (30)$$

where $\acute{x}_a$ and $\acute{x}_b$ are two consecutive observed signals at the receiver, s_a and s_b denote the transmitted complex symbols to be recovered, and a_α and a_β denote the channel responses between the two transmit antennas and receiver. The complex mixing problem in (29) can be formulated as a quaternion source separation problem in (30), where $\acute{x} = \acute{x}_a + \acute{x}_b J$ and $s = s_a + s_b J$ [42]. The single-user quaternion model can be generalized for a multiuser communication system as in (17), where each mixing matrix entry $a_{ij\epsilon}$ represents the channel between the receiver of the ith user and the ϵth transmit antenna of the jth user, $\epsilon \in \{\alpha, \beta\}$.

For clarity, a 2×2 source separation in $\mathbb{H}$ (31) can be expressed in $\mathbb{C}$ (32) as

$$\begin{bmatrix} \acute{x}_1 \\ \acute{x}_2 \end{bmatrix} = \begin{bmatrix} a_{11} a_{12} \\ a_{21} a_{22} \end{bmatrix} \begin{bmatrix} s_1 \\ s_2 \end{bmatrix} \in \mathbb{H} \qquad (31)$$

corresponding to

$$\begin{bmatrix} \acute{x}_{1a} \\ \acute{x}_{1b} \end{bmatrix} = \begin{bmatrix} s_{1a} - s_{1b}^* \\ s_{1b} \;\; s_{1a}^* \end{bmatrix} \begin{bmatrix} a_{11\alpha} \\ a_{11\beta} \end{bmatrix} + \begin{bmatrix} s_{2a} - s_{2b}^* \\ s_{2b} \;\; s_{2a}^* \end{bmatrix} \begin{bmatrix} a_{12\alpha} \\ a_{12\beta} \end{bmatrix} \in \mathbb{C}$$

$$\begin{bmatrix} \acute{x}_{2a} \\ \acute{x}_{2b} \end{bmatrix} = \begin{bmatrix} s_{2a} - s_{2b}^* \\ s_{2b} \;\; s_{2a}^* \end{bmatrix} \begin{bmatrix} a_{22\alpha} \\ a_{22\beta} \end{bmatrix} + \begin{bmatrix} s_{1a} - s_{1b}^* \\ s_{1b} \;\; s_{1a}^* \end{bmatrix} \begin{bmatrix} a_{21\alpha} \\ a_{21\beta} \end{bmatrix} \in \mathbb{C}. \qquad (32)$$

We have considered three scenarios: 1) the pair of symbols $s_{ia} = s_{ib}$ corresponding to the ith user were equal; 2) the pair of symbols $s_{1a} \neq s_{1b}$ of the first user were not equal; and 3) the mixing matrix $\mathbf{A}$ in (31) was complex valued. The sources were communication constellations such as binary phase shift keying (BPSK), quadrature PSK (QPSK), and 16-quadrature amplitude modulation (QAM) for these three scenarios as shown in Figs. 4(a), 5(a), and 6(a). For all the scenarios, q-FastICA was compared with c-FastICA [10] to solve this special complex-valued BSS problem. Two methodologies were considered for using c-FastICA: 1) the long vector approach in which the ith quaternion observation $\acute{x}_i = \acute{x}_{ia} + \acute{x}_{ib}J$ were split into two complex observations $\acute{x}_{ia}$ and $\acute{x}_{ib}$, which were then concatenated into a "long" complex-valued vector, 2) the conventional approach whereby the two quaternion observations were split into four complex observations, i.e., $\{\acute{x}_1, \acute{x}_2\} \in \mathbb{H} \equiv \{\acute{x}_{1a}, \acute{x}_{1b}, \acute{x}_{2a}, \acute{x}_{2b}\} \in \mathbb{C}$. For the first scenario in Fig. 4, observe that the estimates of the long vector method are different from those by the conventional method. However, both estimates were far from being close to the original complex sources. On the other hand, the q-FastICA gave reasonable estimates of the sources. The nature of dependence between s_{ia} and s_{ib} explains why c-FastICA failed to recover the sources. Moreover, the model of Alamouti-based communication BSS problem in (29) is not best suited for the conventional ICA model. As for scenario two in Fig. 5, it was expected that the c-FastICA estimates would improve, as now we have three independent sources, instead of just two. Again, the long-vector c-FastICA technique did not estimate successfully any of the three sources, however, conventional c-FastICA recovered the QPSK source. In practice, it is highly unlikely that s_{ia} and s_{ib} are different, because that would mean the user is alternating between BPSK and QPSK for each consecutive symbol period. For the sake of comparison, we have also considered complex-valued mixing in the context of Alamouti-based communication systems in

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

10

IEEE TRANSACTIONS ON NEURAL NETWORKS

the third scenario in Fig. 6. This scenario is also unlikely due to the fact that the channel response a_{ib} of the second transmitter antenna for each ith user is zero. Both long vector and conventional c-FastICA recovered an improved version of the QPSK source (vis-a-vis their estimates of scenario 2). However, they did not succeed in reconstructing the 16-QAM source, in contrast to q-FastICA.

V. CONCLUSION

An ICA algorithm suitable for the blind separation of both $\mathbb{Q}$-proper and $\mathbb{Q}$-improper sources has been introduced. The well-known negentropy-based cost function has been utilized to estimate independent quaternion-valued sources, while an augmented Newton method implementation has allowed the extension of the FastICA methodology to the quaternion domain. The performance of the proposed widely q-FastICA algorithm in both deflationary and simultaneous separation, using benchmark quaternion polytope signals, has been illustrated. The algorithm has also been shown to be effective in the removal of ocular artifacts from the EEG and in Alamouti-based communication systems.

APPENDIX I
SOME BACKGROUND RESULTS FROM $\mathbb{HR}$ CALCULUS

Some results used in the derivation of the q-FastICA algorithm (25) are discussed below.

A. Chain Rule in the $\mathbb{HR}$ Calculus

For a composite quaternion function $F \circ G = F(G(q)) : \mathbb{H} \mapsto \mathbb{H}$, the chain rule is expressed as

$$\frac{\partial F}{\partial \xi} = \frac{\partial F}{\partial G}\frac{\partial G}{\partial \xi} + \frac{\partial F}{\partial G^\iota}\frac{\partial G^\iota}{\partial \xi} + \frac{\partial F}{\partial G^J}\frac{\partial G^J}{\partial \xi} + \frac{\partial F}{\partial G^\kappa}\frac{\partial G^\kappa}{\partial \xi} \quad (33)$$

and $\xi = \{q, q^\iota, q^J, q^\kappa\}$. To show this, the total differential of $F(\bar{q})$ can be written as [20]

$$dF = \frac{\partial F}{\partial \bar{q}}d\bar{q} + \frac{\partial F}{\partial \bar{q}^\iota}d\bar{q}^\iota + \frac{\partial F}{\partial \bar{q}^J}d\bar{q}^J + \frac{\partial F}{\partial \bar{q}^\kappa}d\bar{q}^\kappa \quad (34)$$

where the dummy variable $\bar{q} \triangleq G(q)$. Likewise, the total differential for $G(q)$ is given by

$$dG = \frac{\partial G}{\partial q}dq + \frac{\partial G}{\partial q^\iota}dq^\iota + \frac{\partial G}{\partial q^J}dq^J + \frac{\partial G}{\partial q^\kappa}dq^\kappa. \quad (35)$$

By substituting (35) into (34), and after rearranging the expressions, we obtain the total differential of F with respect to q as

$$dF = \left(\frac{\partial F}{\partial G}\frac{\partial G}{\partial q} + \frac{\partial F}{\partial G^\iota}\frac{\partial G^\iota}{\partial q} + \frac{\partial F}{\partial G^J}\frac{\partial G^J}{\partial q} + \frac{\partial F}{\partial G^\kappa}\frac{\partial G^\kappa}{\partial q}\right)dq$$

$$+ \left(\frac{\partial F}{\partial G}\frac{\partial G}{\partial q^\iota} + \frac{\partial F}{\partial G^\iota}\frac{\partial G^\iota}{\partial q^\iota} + \frac{\partial F}{\partial G^J}\frac{\partial G^J}{\partial q^\iota} + \frac{\partial F}{\partial G^\kappa}\frac{\partial G^\kappa}{\partial q^\iota}\right)dq^\iota$$

$$+ \left(\frac{\partial F}{\partial G}\frac{\partial G}{\partial q^J} + \frac{\partial F}{\partial G^\iota}\frac{\partial G^\iota}{\partial q^J} + \frac{\partial F}{\partial G^J}\frac{\partial G^J}{\partial q^J} + \frac{\partial F}{\partial G^\kappa}\frac{\partial G^\kappa}{\partial q^J}\right)dq^J$$

$$+ \left(\frac{\partial F}{\partial G}\frac{\partial G}{\partial q^\kappa} + \frac{\partial F}{\partial G^\iota}\frac{\partial G^\iota}{\partial q^\kappa} + \frac{\partial F}{\partial G^J}\frac{\partial G^J}{\partial q^\kappa} + \frac{\partial F}{\partial G^\kappa}\frac{\partial G^\kappa}{\partial q^\kappa}\right)dq^\kappa$$

where the derivatives $\partial F/\partial \xi$ are given by the terms within the brackets. The corresponding chain rule for the $\mathbb{HR}^*$ derivatives can be obtained similarly, and the result of (33) can be extended to vector-valued functions to form a generalized chain rule for the derivatives.

B. Augmented Quaternion Newton Method

The isomorphism between $\mathbb{R}^4$ and $\mathbb{H}$ allows for consideration of the duality between the derivatives in the two domains. This methodology was previously considered in [43] and resulted in the derivation of the augmented complex Newton method. The extension of this paper to the quaternion domain based on the involution bases was detailed in [20]. A short summary is presented below. For a function $f(\mathbf{q}) : \mathbb{H}^N \mapsto \mathbb{R}$, its augmented gradient $\nabla_{\mathbf{q}^{a*}} f = \partial f/\partial \mathbf{q}^{a*}$ and Hessian $\mathbf{H}_{\mathbf{qq}}^a = (\partial/\partial \mathbf{q}^{a*})(\partial f/\partial \mathbf{q}^{q*})^T$, where the augmented vector $\mathbf{q}^a = [\mathbf{q}^T, \mathbf{q}^{\iota T}, \mathbf{q}^{J T}, \mathbf{q}^{\kappa T}]^T$. The augmented Newton update can then be written as

$$\Delta \mathbf{q}^a = -(\mathbf{H}_{\mathbf{qq}}^a)^{-1} \cdot \nabla_{\mathbf{q}^{a*}} f \quad (36)$$

where $\Delta \mathbf{q}^a = \mathbf{q}^a(k+1) - \mathbf{q}^a(k)$ denotes the change in $\mathbf{q}^a$ in each consecutive update.

Finally, observe that the elements of the augmented Hessian matrix

$$\mathbf{H}_{\mathbf{qq}}^a = \begin{bmatrix} \mathbf{H}_{\mathbf{q}^*\mathbf{q}^*} & \mathbf{H}_{\mathbf{q}^{\iota*}\mathbf{q}^*} & \mathbf{H}_{\mathbf{q}^{J*}\mathbf{q}^*} & \mathbf{H}_{\mathbf{q}^{\kappa*}\mathbf{q}^*} \\ \mathbf{H}_{\mathbf{q}^*\mathbf{q}^{\iota*}} & \mathbf{H}_{\mathbf{q}^{\iota*}\mathbf{q}^{\iota*}} & \mathbf{H}_{\mathbf{q}^{J*}\mathbf{q}^{\iota*}} & \mathbf{H}_{\mathbf{q}^{\kappa*}\mathbf{q}^{\iota*}} \\ \mathbf{H}_{\mathbf{q}^*\mathbf{q}^{J*}} & \mathbf{H}_{\mathbf{q}^{\iota*}\mathbf{q}^{J*}} & \mathbf{H}_{\mathbf{q}^{J*}\mathbf{q}^{J*}} & \mathbf{H}_{\mathbf{q}^{\kappa*}\mathbf{q}^{J*}} \\ \mathbf{H}_{\mathbf{q}^*\mathbf{q}^{\kappa*}} & \mathbf{H}_{\mathbf{q}^{\iota*}\mathbf{q}^{\kappa*}} & \mathbf{H}_{\mathbf{q}^{J*}\mathbf{q}^{\kappa*}} & \mathbf{H}_{\mathbf{q}^{\kappa*}\mathbf{q}^{\kappa*}} \end{bmatrix} \quad (37)$$

can be written in terms of its first row by utilizing the involution property in (16) and noting that $((\cdot)^\alpha)^\beta = (\cdot)^\gamma$, $\alpha \neq \beta \neq \gamma = \{\iota, J, \kappa\}$.

APPENDIX II
DERIVATION OF THE AUGMENTED q-FASTICA UPDATE ALGORITHM

A. First and Second Derivatives of the Cost Function $\mathcal{J}(\mathbf{w})$

Firstly, by using the product rule, the derivatives of the involutions of $|y|^2 = yy^* = |\mathbf{w}^H \mathbf{x}|^2$ with respect to the conjugate demixing vector $\mathbf{w}^*$ are calculated as

$$\frac{\partial yy^*}{\partial \mathbf{w}^*} = \frac{\partial y}{\partial \mathbf{w}^*}y^* + y\frac{\partial y^*}{\partial \mathbf{w}^*} = \mathbf{x}y^* - \frac{1}{2}y\mathbf{x}^*$$

$$\frac{\partial (yy^*)^\iota}{\partial \mathbf{w}^*} = \frac{\partial y^\iota}{\partial \mathbf{w}^*}y^{\iota*} + y^\iota\frac{\partial y^{\iota*}}{\partial \mathbf{w}^*} = \frac{1}{2}y^\iota\mathbf{x}^{\iota*}$$

$$\frac{\partial (yy^*)^J}{\partial \mathbf{w}^*} = \frac{\partial y^J}{\partial \mathbf{w}^*}y^{J*} + y^J\frac{\partial y^{J*}}{\partial \mathbf{w}^*} = \frac{1}{2}y^J\mathbf{x}^{J*}$$

$$\frac{\partial (yy^*)^\kappa}{\partial \mathbf{w}^*} = \frac{\partial y^\kappa}{\partial \mathbf{w}^*}y^{\kappa*} + y^\kappa\frac{\partial y^{\kappa*}}{\partial \mathbf{w}^*} = \frac{1}{2}y^\kappa\mathbf{x}^{\kappa*}.$$

Then, by using the chain rule (33) and after simplification, we obtain the gradients of the cost function as

$$\nabla_{\mathbf{w}^*}\mathcal{J} = E\{2g(|y|^2)\mathbf{x}y^*\}$$
$$\nabla_{\mathbf{w}^{\iota*}}\mathcal{J} = E\{2g(|y|^2)\mathbf{x}y^{\iota*}\}$$
$$\nabla_{\mathbf{w}^{J*}}\mathcal{J} = E\{2g(|y|^2)\mathbf{x}y^{J*}\}$$
$$\nabla_{\mathbf{w}^{\kappa*}}\mathcal{J} = E\{2g(|y|^2)\mathbf{x}y^{\kappa*}\} \quad (38)$$

where g is the first derivative of G, this result can also be interpreted based on the involution property (16). After simplifications and considering the whiteness of $\mathbf{x}$, the second

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

JAVIDI *et al.*: FastICA ALGORITHM FOR QUATERNION VALUED SIGNALS
11

derivatives of the cost function $\mathcal{J}$ can then be calculated as

$$\frac{\partial}{\partial \mathbf{w}^*}\left(\frac{\partial \mathcal{J}}{\partial \mathbf{w}^*}\right)^T = E\{4g'(|y|^2)\mathbf{x}y^*\mathbf{x}^T y^* - g(|y|^2)\mathbf{I}\}$$

$$\frac{\partial}{\partial \mathbf{w}^{\iota*}}\left(\frac{\partial \mathcal{J}}{\partial \mathbf{w}^*}\right)^T = E\{2g'(|y|^2)(\mathbf{x}y^*)^\iota(\mathbf{x}^T y^*) + g(|y|^2)\mathbf{I}\}$$

$$\frac{\partial}{\partial \mathbf{w}^{J*}}\left(\frac{\partial \mathcal{J}}{\partial \mathbf{w}^*}\right)^T = E\{2g'(|y|^2)(\mathbf{x}y^*)^J(\mathbf{x}^T y^*) + g(|y|^2)\mathbf{I}\}$$

$$\frac{\partial}{\partial \mathbf{w}^{\kappa*}}\left(\frac{\partial \mathcal{J}}{\partial \mathbf{w}^*}\right)^T = E\{2g'(|y|^2)(\mathbf{x}y^*)^\kappa(\mathbf{x}^T y^*) + g(|y|^2)\mathbf{I}\}$$

$$(39)$$

where g' is the second derivative of G and the calculations of the remaining derivatives follow from property (16). Notice that the non-commutativity of the quaternion product prohibits further simplification of the derivatives in (39).

B. Augmented Newton Update

The Lagrangian function $\mathcal{L}$ for the optimization problem in (24) is given by

$$\mathcal{L}(\mathbf{w}, \lambda) = \mathcal{J}(\mathbf{w}) + \underbrace{\lambda(\mathbf{w}^H \mathbf{w} - 1)}_{\triangleq c} \qquad (40)$$

where $\lambda \in \mathbb{R}$ is the Lagrange parameter. We can use the Newton method (36) to find the extrema of (40), where

$$\frac{\partial \mathcal{L}}{\partial \mathbf{w}^{a*}} = \frac{\partial \mathcal{J}}{\partial \mathbf{w}^{a*}} + \frac{\partial c}{\partial \mathbf{w}^{a*}}$$

$$\frac{\partial}{\partial \mathbf{w}^{a*}}\left(\frac{\partial \mathcal{L}}{\partial \mathbf{w}^{a*}}\right)^T = \mathbf{H}^a_{\mathbf{ww}} + \frac{\partial}{\partial \mathbf{w}^{a*}}\left(\frac{\partial c}{\partial \mathbf{w}^{a*}}\right)^T. \qquad (41)$$

The augmented gradient and Hessian of $\mathcal{J}$ are then obtained using (38) and (39). The gradients of c are given by

$$\frac{\partial c}{\partial \mathbf{w}^*} = \lambda\left(\mathbf{w} - \frac{1}{2}\mathbf{w}^*\right) \qquad \frac{\partial c}{\partial \mathbf{w}^{\iota*}} = \frac{\lambda}{2}\mathbf{w}^*$$

$$\frac{\partial c}{\partial \mathbf{w}^{J*}} = \frac{\lambda}{2}\mathbf{w}^* \qquad \frac{\partial c}{\partial \mathbf{w}^{\kappa*}} = \frac{\lambda}{2}\mathbf{w}^*$$

and the Hessian can be calculated from

$$\frac{\partial}{\partial \mathbf{w}^*}\left(\frac{\partial c}{\partial \mathbf{w}^*}\right)^T = -\lambda\mathbf{I} \qquad \frac{\partial}{\partial \mathbf{w}^{\iota*}}\left(\frac{\partial c}{\partial \mathbf{w}^*}\right)^T = -\frac{\lambda}{2}\mathbf{I}$$

$$\frac{\partial}{\partial \mathbf{w}^{J*}}\left(\frac{\partial c}{\partial \mathbf{w}^*}\right)^T = -\frac{\lambda}{2}\mathbf{I} \qquad \frac{\partial}{\partial \mathbf{w}^{\kappa*}}\left(\frac{\partial c}{\partial \mathbf{w}^*}\right)^T = -\frac{\lambda}{2}\mathbf{I}.$$

The Newton update is obtained by substituting these results in (36). Finally, the Lagrange parameter λ is updated using a gradient ascent method, where at each iteration the demixing vector $\mathbf{w}$ is first updated via the augmented Newton method, followed by the update of λ using the current value of $\mathbf{w}$ and a normalization of the demixing vector [44], as in (25).

ACKNOWLEDGMENT

REFERENCES

[1] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*. New York: Wiley, 2001.

[2] A. Belouchrani, K. Abed-Meraim, J.-F. Cardoso, and E. Moulines, "A blind source separation technique using second-order statistics," *IEEE Trans. Signal Process.*, vol. 45, no. 2, pp. 434–444, Feb. 1997.

[3] J.-F. Cardoso and A. Souloumiac, "Blind beamforming for non-Gaussian signals," *IEE Proc. F Radar Signal Process.*, vol. 140, no. 6, pp. 362–370, Dec. 1993.

[4] A. J. Bell and T. J. Sejnowski, "An information-maximization approach to blind separation and blind deconvolution," *Neural Comput.*, vol. 7, no. 6, pp. 1129–1159, Nov. 1995.

[5] A. Hyvärinen, "Fast and robust fixed-point algorithms for independent component analysis," *IEEE Trans. Neural Netw.*, vol. 10, no. 3, pp. 626–634, May 1999.

[6] E. Oja and Z. Yuan, "The FastICA algorithm revisited: Convergence analysis," *IEEE Trans. Neural Netw.*, vol. 17, no. 6, pp. 1370–1381, Nov. 2006.

[7] E. Bingham and A. Hyvärinen, "A fast fixed-point algorithm for independent component analysis of complex valued signals," *J. Neural Syst.*, vol. 10, no. 1, pp. 1–8, Feb. 2000.

[8] S. C. Douglas, "Fixed-point algorithms for the blind separation of arbitrary complex-valued non-Gaussian signal mixtures," *EURASIP J. Appl. Signal Process.*, vol. 2007, no. 1, pp. 83–98, Jan. 2007.

[9] J. Eriksson and V. Koivunen, "Complex random vectors and ICA models: Identifiability, uniqueness, and separability," *IEEE Trans. Inf. Theory*, vol. 52, no. 3, pp. 1017–1029, Mar. 2006.

[10] M. Novey and T. Adali, "On extending the complex FastICA algorithm to noncircular sources," *IEEE Trans. Signal Process.*, vol. 56, no. 5, pp. 2148–2154, May 2008.

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

[12] P. J. Schreier and L. L. Scharf, *Statistical Signal Processing of Complex-Valued Data*. Cambridge, U.K.: Cambridge Univ. Press, 2010.

[13] K. Kreutz-Delgado, "The complex gradient operator and the $\mathbb{CR}$-calculus," Dept. Electr. Comput. Eng., Univ. California, San Diego, Tech. Rep. ECE275A, 2006, pp. 1–74.

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

[15] S. Sangwine and N. Le Bihan, "Quaternion polar representation with a complex modulus and complex argument inspired by the Cayley-Dickson form," *Adv. Appl. Clifford Algebras*, vol. 20, no. 1, pp. 111–120, Mar. 2010.

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

[17] N. Le Bihan and P. O. Amblard, "Detection and estimation of Gaussian proper quaternion valued random processes," in *Proc. 7th IMA Conf. Math. Signal Process.*, Cirencester, U.K., 2006, pp. 1–3.

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

[19] C. C. Took, D. Mandic, and F. Zhang, "On the unitary diagonalisation of a special class of quaternion matrices," *Appl. Math. Lett.*, vol. 24, no. 11, pp. 1806–1809, Nov. 2011.

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

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

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

[23] B. C. Ujang, C. C. Took, and D. P. Mandic, "Split quaternion nonlinear adaptive filtering," *Neural Netw.*, vol. 23, no. 3, pp. 426–434, Apr. 2010.

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

[25] N. Le Bihan and S. Buchholz, "Quaternionic independent component analysis using hypercomplex nonlinearities," in *Proc. 7th IMA Conf. Math. Signal Process.*, 2006, pp. 1–3.

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

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

[28] P. J. Schreier and L. L. Scharf, "Second-order analysis of improper complex random vectors and processes," *IEEE Trans. Signal Process.*, vol. 51, no. 3, pp. 714–725, Mar. 2003.

[29] W. Wirtinger, "Zur formalen theorie der funktionen von mehr komplexen veränderlichen," *Math. Annalen*, vol. 97, no. 1, pp. 357–375, Dec. 1927.

[30] D. H. Brandwood, "A complex gradient operator and its application in adaptive array theory," *IEE Proc. F: Commun., Radar Signal Process.*, vol. 130, no. 1, pp. 11–16, Feb. 1983.

[31] F. Zhang, "Quaternions and matrices of quaternions," *Linear Algebra Appl.*, vol. 251, pp. 21–57, Jan. 1997.

[32] A. Sudbery, "Quaternionic analysis," *Math. Proc. Cambridge Philos. Soc.*, vol. 85, no. 2, pp. 199–225, 1979.

[33] S. De Leo and P. P. Rotelli, "Quaternionic analyticity," *Appl. Math. Lett.*, vol. 16, no. 7, pp. 1077–1081, Oct. 2003.

[34] A. T. Erdogan, "On the convergence of ICA algorithms with symmetric orthogonalization," *IEEE Trans. Signal Process.*, vol. 57, no. 6, pp. 2209–2221, Jun. 2009.

[35] L. H. Zetterberg and H. Brändström, "Codes for combined phase and amplitude modulated signals in a 4-D space," *IEEE Trans. Commun.*, vol. 25, no. 9, pp. 943–950, Sep. 1977.

[36] M. K. I. Molla, T. Tanaka, T. M. Rutkowski, and A. Cichocki, "Separation of EOG artifacts from EEG signals using bivariate EMD," in *Proc. Int. Conf. Acoust. Speech Signal Process.*, Dallas, TX, Mar. 2010, pp. 562–565.

[37] A. Greco, N. Mammone, F. C. Morabito, and M. Versaci, "Semi-automatic artifact rejection procedure based on kurtosis, Renyi's entropy and independent component scalp maps," in *Proc. Int. Enform. Conf.*, 2005, pp. 22–26.

[38] P. S. Kumar, R. Arumuganathan, K. Sivakumar, and C. Vimal, "An adaptive method to remove ocular artifacts from EEG signals using wavelet transform," *J. Appl. Sci. Res.*, vol. 5, no. 7, pp. 711–745, 2009.

[39] R. N. Vigário, "Extraction of ocular artefacts from EEG using independent component analysis," *Electroencephal. Clinical Neurophysiol.*, vol. 103, no. 3, pp. 395–404, Sep. 1997.

[40] A. Delorme, S. Makeig, and T. Sejnowski, "Automatic artifact rejection for EEG data using high-order statistics and independent component analysis," in *Proc. Int. Workshop ICA*, 2001, pp. 457–462.

[41] S. M. Alamouti, "A simple transmit diversity technique for wireless communications," *IEEE J. Sel. Areas Commun.*, vol. 16, no. 8, pp. 1451–1458, Oct. 1998.

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

[43] A. Bos, "Complex gradient and Hessian," *IEE Proc. Vis., Image Signal Process.*, vol. 141, no. 6, pp. 380–383, Dec. 1994.

[44] S. Zhang and A. G. Constantinides, "Lagrange programming neural networks," *IEEE Trans. Circuits Syst. II: Analog Digital Signal Process.*, vol. 39, no. 7, pp. 441–452, Jul. 1992.

Soroush Javidi (S'09) received the M.Eng. and Ph.D. degrees in electrical and electronic engineering from Imperial College London, London, U.K., in 2006 and 2010, respectively.

His research has focused on the use of augmented complex statistics and widely linear models in practical applications, such as electroencephalograms and wind modeling. His current research interests include statistical signal processing and complex-valued blind source separation.

Clive Cheong Took (S'06–M'07) received the B.S. degree in telecommunication engineering from Kings College, London University, London, U.K., where he was the top departmental graduate in 2004, and the Ph.D. degree in blind signal processing from Cardiff University, Wales, U.K., in 2007.

He is currently with Imperial College London, London. His current research interests include adaptive, blind, and multidimensional signal processing with applications in biomedicine and renewable energy.

Danilo P. Mandic (M'99–SM'03) received the Ph.D. degree in electronic engineering from Imperial College London, London, U.K., in 1999.

He is a Professor of signal processing with Imperial College London and is working in the area of adaptive signal processing and multiscale modeling. He has been a Guest Professor with Katholieke Universiteit Leuven, Leuven, Belgium, and a Frontier Researcher at RIKEN, Tokyo, Japan. His publication record includes two research monographs entitled *Recurrent Neural Networks for Prediction* (1st ed., Aug. 2001) and *Complex Valued Nonlinear Adaptive Filters: Noncircularity, Widely Linear and Neural Models* (1st ed., Wiley, Apr. 2009), an edited book entitled *Signal Processing for Information Fusion* (Springer, 2008), and more than 200 publications on signal processing.

Prof. Mandic is a member of the IEEE Technical Committee on Theory and Methods for Signal Processing and an Associate Editor for the IEEE TRANSACTIONS ON NEURAL NETWORKS, and has been on the editorial boards of the IEEE TRANSACTIONS ON SIGNAL PROCESSING and the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS II. He has produced award winning papers and products resulting from his collaboration with industry.