

A Fast Independent Component Analysis Algorithm for Improper Quaternion Signals

Soroush Javidi and Danilo P. Mandic

Communications and Signal Processing Research Group,
Department of Electrical and Electronic Engineering,
Imperial College London, London, SW7 2AZ, U.K.
email: {soroush.javidi, d.mandic}@imperial.ac.uk
URL: <http://www.commsp.ee.ic.ac.uk/mandic>

Technical Report: TR-ICU-ICA-1070/06-10

Created: 08 June 2010, *Last Updated:* December 2010

Abstract

An extension of the FastICA algorithm is proposed for the blind separation of both \mathbb{Q} -proper and \mathbb{Q} -improper quaternion-valued signals. This is achieved by maximising a negentropy-based cost function, and is implemented using the Newton method in the augmented quaternion statistics framework. It is shown that the use of augmented statistics and the associated widely linear modeling provide theoretical and practical advantages over standard models. Simulations using both benchmark and real-world signals support the approach.

Index Terms

Independent Component Analysis (ICA), augmented quaternion statistics, quaternion widely linear modelling, quaternion noncircularity, quaternion blind source separation

I. INTRODUCTION

The independent component analysis (ICA) framework is a popular method for the separation of latent sources from an observed mixture, aided by the assumption of statistical independence of the sources [1]. In its standard form, the latent sources are assumed to be linearly mixed, and with additive noise present. For this scenario, algorithms proposed in the past two decades include those based on the utilisation of second- and higher-order statistics (SOBI and JADE algorithms [2], [3]), and those based on information theoretic criteria such as the maximisation of likelihood and minimisation of mutual information [4]. The FastICA algorithm [5], a fast converging algorithm based on the maximisation of non-Gaussianity and implemented using an approximative Newton optimisation method, has become a standard for the separation of both sub- and super-Gaussian sources. The algorithm was shown to exhibit cubic convergence for the deflationary separation process and local quadratic convergence for the symmetric orthogonalisation approach [6]. The FastICA algorithm was extended for the separation of complex circular sources in [7] and more recently it was generalised for the separation of noncircular complex sources [8]; this was achieved based on the fourth order moment and through utilisation of the strong uncorrelating transform [9]. A generalised FastICA algorithm for the separation of both circular and noncircular sources based on a negentropy-based FastICA algorithm was addressed in [10].

The recent progress in supervised and unsupervised adaptive signal processing algorithms in the complex domain [11], has been made possible owing to the advances in complex domain augmented statistics [12] and the analysis of non-analytic functions through the framework of $\mathbb{C}\mathbb{R}$ calculus (also known as Wirtinger calculus) [13]. In the same light, there has been recent interest in adaptive signal processing algorithms in the quaternion domain, a natural domain for the processing of three- and four-dimensional 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], and the so called $\mathbb{H}\mathbb{R}$ calculus, a unified framework for the analysis of non-analytic quaternion functions [19].

These advances have been exploited through widely linear modelling of quaternion signals, allowing us to incorporate the full second-order information and have led to the class of widely linear quaternion least mean square (WL-QLMS) algorithms [20]. In nonlinear signal models, both split- and fully-quaternionic nonlinear models have been successfully implemented [21]. In the study of unsupervised adaptive algorithms, a quaternion ICA algorithm based on likelihood maximisation and the concept of Infomax was proposed by Le Bihan and Buchholz in [22]. In their study, it was concluded that a fully-quaternion nonlinearity results in better separation performance.

In this paper, we propose a FastICA algorithm suitable for the separation of \mathbb{Q} -proper and \mathbb{Q} -improper quaternion-valued signals from an observed linear mixture. This is achieved by means of augmented quaternion statistics, widely linear modelling and $\mathbb{H}\mathbb{R}$ calculus, and based on the augmented Newton method, whereby at the cost of additional complexity we capture the complete statistical properties of the signals and ensure successful separation of latent sources. The performance of the

algorithm using synthetic \mathbb{Q} -proper and \mathbb{Q} -improper polytope signals in both deflationary and simultaneous separation scenarios is studied, and is followed by a real-world case study of electroencephalogram (EEG) artifact extraction.

This paper is organised as follows. In Section II the background on quaternion algebra, statistics and the \mathbb{H} calculus is provided. Section III introduces the ICA problem and proposes separation methodology for the generality of quaternionic signals. Simulations based on both synthetic sources and a real-world case study are given in Section IV. Section V concludes the paper.

II. PRELIMINARIES ON QUATERNION SIGNALS

A. Quaternion algebra

Consider the quaternion variable $\mathbf{q} = \mathbf{q}_a + \iota\mathbf{q}_b + j\mathbf{q}_c + \kappa\mathbf{q}_d \in \mathbb{H}$, where $\mathbf{q}_a, \mathbf{q}_b, \mathbf{q}_c$ and \mathbf{q}_d are real-valued scalars, and ι, j and κ are orthogonal unit vectors such that

$$\begin{aligned} \iota &= j = \kappa = \sqrt{-1} \\ \iota j &= \kappa \quad j\kappa = \iota \quad \kappa\iota = j \\ \iota j\kappa &= \iota^2 = j^2 = \kappa^2 = -1. \end{aligned} \quad (1)$$

These identities illustrate the non-commutative property of products in quaternion algebra, whereby $\mathbf{q}_1\mathbf{q}_2 \neq \mathbf{q}_2\mathbf{q}_1$. The number \mathbf{q} can also be written in terms of its real (scalar) part $\Re\{\mathbf{q}\} = \mathbf{q}_a$ and its vector part $\Im\{\mathbf{q}\} = \iota\Im_\iota\{\mathbf{q}\} + j\Im_j\{\mathbf{q}\} + \kappa\Im_\kappa\{\mathbf{q}\}$, such that $\mathbf{q} = \Re\{\mathbf{q}\} + \Im\{\mathbf{q}\}$. Alternatively, by adopting the Cayley-Dickson notation, \mathbf{q} can be constructed from a pair of complex quantities $\mathbf{z}_1 = \mathbf{q}_a + \iota\mathbf{q}_b$ and $\mathbf{z}_2 = \mathbf{q}_c + \iota\mathbf{q}_d$, such that $\mathbf{q} = \mathbf{z}_1 + \mathbf{z}_2j$, however in this article direct quaternionic notation will be used.

In the quaternion domain, we can consider three self-inverse mappings¹ or *involutions* [23] about the ι, j and κ axes,

$$\begin{aligned} \mathbf{q}^\iota &= -\iota\mathbf{q}\iota = \mathbf{q}_a + \iota\mathbf{q}_b - j\mathbf{q}_c - \kappa\mathbf{q}_d \\ \mathbf{q}^j &= -j\mathbf{q}j = \mathbf{q}_a - \iota\mathbf{q}_b + j\mathbf{q}_c - \kappa\mathbf{q}_d \\ \mathbf{q}^\kappa &= -\kappa\mathbf{q}\kappa = \mathbf{q}_a - \iota\mathbf{q}_b - j\mathbf{q}_c + \kappa\mathbf{q}_d \end{aligned} \quad (2)$$

which form the bases for augmented quaternion statistics [18]. Intuitively, an involution represents a rotation along each respective axis, while the conjugate operator $(\cdot)^*$ forms an involution along all three directions, where

$$\mathbf{q}^* = \mathbf{q}_a - \iota\mathbf{q}_b - j\mathbf{q}_c - \kappa\mathbf{q}_d. \quad (3)$$

The involutions have the property that $(\mathbf{q}_1\mathbf{q}_2)^\alpha = \mathbf{q}_1^\alpha\mathbf{q}_2^\alpha$, $\alpha = \{\iota, j, \kappa\}$, while $(\mathbf{q}_1\mathbf{q}_2)^* = \mathbf{q}_2^*\mathbf{q}_1^*$. Finally, the norm (modulus) of a quaternion variable \mathbf{q} is defined by

$$\|\mathbf{q}\|_2 = \sqrt{\mathbf{q}\mathbf{q}^*} = \sqrt{\mathbf{q}^*\mathbf{q}} = \sqrt{\mathbf{q}_a^2 + \mathbf{q}_b^2 + \mathbf{q}_c^2 + \mathbf{q}_d^2} \quad (4)$$

whereby 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 modelling

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 scalar and vector 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 $E\{\mathbf{q}\} = E\{\mathbf{q}_a\} + E\{\mathbf{q}_b\} + E\{\mathbf{q}_c\} + E\{\mathbf{q}_d\}$ and the quadrivariate covariance matrix of real-valued component vectors $\mathcal{C}_{\mathbf{q}\mathbf{q}}^r = E\{\mathbf{q}^r\mathbf{q}^{rT}\} \in \mathbb{R}^{4N \times 4N}$ describes the second-order relationship between the respective components of \mathbf{q} , where $\mathbf{q}^r = [\mathbf{q}_a, \mathbf{q}_b, \mathbf{q}_c, \mathbf{q}_d]^T$. Representing the components of $\mathcal{C}_{\mathbf{q}\mathbf{q}}^r$ by their equivalent quaternion counterparts allows for the complete second-order statistical information to be captured directly in \mathbb{H} [18]. This is achieved by considering the relation between the components of the quaternion variable \mathbf{q} and its involutions (2), given by

$$\begin{aligned} \mathbf{q}_a &= \frac{1}{4}(\mathbf{q} + \mathbf{q}^\iota + \mathbf{q}^j + \mathbf{q}^\kappa), & \mathbf{q}_b &= \frac{1}{4}(\mathbf{q} + \mathbf{q}^\iota - \mathbf{q}^j - \mathbf{q}^\kappa) \\ \mathbf{q}_c &= \frac{1}{4}(\mathbf{q} - \mathbf{q}^\iota + \mathbf{q}^j - \mathbf{q}^\kappa), & \mathbf{q}_d &= \frac{1}{4}(\mathbf{q} - \mathbf{q}^\iota - \mathbf{q}^j + \mathbf{q}^\kappa). \end{aligned} \quad (5)$$

In analogy to the complex domain² where both \mathbf{z} and \mathbf{z}^* are used to define the augmented statistics [24], [25], it can be shown that the bases $\mathbf{q}, \mathbf{q}^\iota, \mathbf{q}^j$ and \mathbf{q}^κ provide a suitable means to define the quaternion augmented statistics [18]. This way,

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

²In the complex domain, the real and imaginary components can be represented in terms of the conjugate coordinates \mathbf{z} and \mathbf{z}^* respectively as $\frac{1}{2}(\mathbf{z} + \mathbf{z}^*)$ and $\frac{1}{2j}(\mathbf{z} - \mathbf{z}^*)$.

the augmented random vector $\mathbf{q}^a = [\mathbf{q}^T, \mathbf{q}^{i^T}, \mathbf{q}^{j^T}, \mathbf{q}^{\kappa^T}]^T$ is used to define the augmented covariance matrix

$$\begin{aligned} \mathbf{C}_{\mathbf{q}}^a &= E\{\mathbf{q}^a \mathbf{q}^{aH}\} \\ &= \begin{bmatrix} \mathbf{C}_{\mathbf{q}\mathbf{q}} & \mathbf{C}_{\mathbf{q}^i} & \mathbf{C}_{\mathbf{q}^j} & \mathbf{C}_{\mathbf{q}^\kappa} \\ \mathbf{C}_{\mathbf{q}^i}^H & \mathbf{C}_{\mathbf{q}^i \mathbf{q}^i} & \mathbf{C}_{\mathbf{q}^i \mathbf{q}^j} & \mathbf{C}_{\mathbf{q}^i \mathbf{q}^\kappa} \\ \mathbf{C}_{\mathbf{q}^j}^H & \mathbf{C}_{\mathbf{q}^j \mathbf{q}^i} & \mathbf{C}_{\mathbf{q}^j \mathbf{q}^j} & \mathbf{C}_{\mathbf{q}^j \mathbf{q}^\kappa} \\ \mathbf{C}_{\mathbf{q}^\kappa}^H & \mathbf{C}_{\mathbf{q}^\kappa \mathbf{q}^i} & \mathbf{C}_{\mathbf{q}^\kappa \mathbf{q}^j} & \mathbf{C}_{\mathbf{q}^\kappa \mathbf{q}^\kappa} \end{bmatrix} \in \mathbb{H}^{4N \times 4N} \end{aligned} \quad (6)$$

which describes the complete second-order information available within a quaternion random vector. In (6), $\mathbf{C}_{\mathbf{q}^i}, \mathbf{C}_{\mathbf{q}^j}, \mathbf{C}_{\mathbf{q}^\kappa}$ are respectively termed the i -, j - and κ -covariance matrices $E\{\mathbf{q}\mathbf{q}^{\alpha H}\}$, $\alpha = \{i, j, \kappa\}$, while $\mathbf{C}_{\mathbf{q}\mathbf{q}} = E\{\mathbf{q}\mathbf{q}^H\}$ is the standard covariance matrix. The i -, j - and κ -covariance matrices are referred to as the complementary or pseudo-covariance matrices [25]. The concept of properness (rotation invariant pdf) can be extended from the complex to the quaternion domain and has been discussed in [16] and [17]. Following the involution-based augmented bases, a random vector is considered \mathbb{Q} -proper if it is not correlated with its involutions, or, $\mathbf{C}_{\mathbf{q}^i} = \mathbf{C}_{\mathbf{q}^j} = \mathbf{C}_{\mathbf{q}^\kappa} = \mathbf{0}$, and all cross-covariance matrices vanish, and is otherwise termed \mathbb{Q} -improper [18]. Therefore, for a \mathbb{Q} -proper random vector, the augmented covariance matrix (6) has a block-diagonal structure. More restricted definitions of properness can also be defined, whereby one or more pseudo-covariances are non-zero (\mathbb{C} -proper) [17], and can be intuitively understood as rotation invariance along one or more of the quaternion axes; \mathbb{Q} -properness thus reflects rotation invariance along all three 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 , expressed as $\hat{y} = E\{y|x\}$, is given by $\hat{y} = \mathbf{h}^T \mathbf{x}$, where \mathbf{h} is a coefficient vector and \mathbf{x} the regressor. As a generalisation, the MSE estimator for a quaternion-valued signal $y \in \mathbb{H}$ can then be written in terms of the MSE estimators of its respective components, given by

$$\begin{aligned} \hat{y}_a &= E\{y_a | q_a, q_b, q_c, q_d\} & \hat{y}_b &= E\{y_b | q_a, q_b, q_c, q_d\} \\ \hat{y}_c &= E\{y_c | q_a, q_b, q_c, q_d\} & \hat{y}_d &= E\{y_d | q_a, q_b, q_c, q_d\}, \end{aligned} \quad (7)$$

such that

$$\begin{aligned} \hat{y} &= \hat{y}_a + i\hat{y}_b + j\hat{y}_c + \kappa\hat{y}_d \\ &= E\{y_a | q_a, q_b, q_c, q_d\} + iE\{y_b | q_a, q_b, q_c, q_d\} \\ &\quad + jE\{y_c | q_a, q_b, q_c, q_d\} + \kappa E\{y_d | q_a, q_b, q_c, q_d\}. \end{aligned} \quad (8)$$

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

$$\begin{aligned} \hat{y} &= E\{y | q, q^i, q^j, q^\kappa\} + iE\{y | q, q^i, q^j, q^\kappa\} \\ &\quad + jE\{y | q, q^i, q^j, q^\kappa\} + \kappa E\{y | q, q^i, q^j, q^\kappa\}, \end{aligned} \quad (9)$$

and results in the widely linear estimator [18], [20]

$$\begin{aligned} y &= \mathbf{h}^H \mathbf{q} + \mathbf{g}^H \mathbf{q}^i \mathbf{u}^H + \mathbf{q}^j + \mathbf{v}^H \mathbf{q}^\kappa \\ &= \mathbf{w}^{aH} \mathbf{q}^a \end{aligned} \quad (10)$$

where the augmented weight vector $\mathbf{w}^a = [\mathbf{h}^T, \mathbf{g}^T, \mathbf{u}^T, \mathbf{v}^T]^T$. Thus (10) is the optimal estimator for the generality of quaternion-valued signals, both proper and improper.

C. An overview of $\mathbb{H}\mathbb{R}$ calculus

In signal processing problems, it is common to define a real-valued cost function, typically the error power. In a similar fashion to the $\mathbb{C}\mathbb{R}$ calculus framework where a function is defined based on the conjugate coordinates \mathbf{z} and \mathbf{z}^* [13], [26], in the context of $\mathbb{H}\mathbb{R}$ calculus [19], $f(\mathbf{q}) : \mathbb{H}^N \mapsto \mathbb{R}$ can be considered as a function of the orthogonal quaternion basis vectors $\mathbf{q}, \mathbf{q}^i, \mathbf{q}^j$ and \mathbf{q}^κ , such that

$$f(\mathbf{q}, \mathbf{q}^i, \mathbf{q}^j, \mathbf{q}^\kappa) : \mathbb{H}^N \times \mathbb{H}^N \times \mathbb{H}^N \times \mathbb{H}^N \mapsto \mathbb{R}. \quad (11)$$

Likewise, the duality between a quaternion function f and its real-valued equivalent g can be expressed as

$$\begin{aligned} f(\mathbf{q}) &= f(\mathbf{q}, \mathbf{q}^i, \mathbf{q}^j, \mathbf{q}^\kappa) \\ &= f_a(\mathbf{q}_a, \mathbf{q}_b, \mathbf{q}_c, \mathbf{q}_d) + i f_b(\mathbf{q}_a, \mathbf{q}_b, \mathbf{q}_c, \mathbf{q}_d) \\ &\quad + j f_c(\mathbf{q}_a, \mathbf{q}_b, \mathbf{q}_c, \mathbf{q}_d) + \kappa f_d(\mathbf{q}_a, \mathbf{q}_b, \mathbf{q}_c, \mathbf{q}_d) \\ &= g(\mathbf{q}_a, \mathbf{q}_b, \mathbf{q}_c, \mathbf{q}_d) \end{aligned} \quad (12)$$

Then, by considering the components of the quaternion variable \mathbf{q} and the orthogonal bases given in (5), a relation can be established between the derivatives taken with respect to the components of the quaternion variable and those taken directly

with respect to the quaternion basis variables, forming a fundamental result of $\mathbb{H}\mathbb{R}$ calculus. These relations, known as $\mathbb{H}\mathbb{R}$ derivatives, are given by [19], [27]

$$\begin{aligned}\frac{\partial f}{\partial \mathbf{q}} &= \frac{1}{4} \left(\frac{\partial f}{\partial \mathbf{q}_a} - \iota \frac{\partial f}{\partial \mathbf{q}_b} - j \frac{\partial f}{\partial \mathbf{q}_c} - \kappa \frac{\partial f}{\partial \mathbf{q}_d} \right) \\ \frac{\partial f}{\partial \mathbf{q}^\iota} &= \frac{1}{4} \left(\frac{\partial f}{\partial \mathbf{q}_a} - \iota \frac{\partial f}{\partial \mathbf{q}_b} + j \frac{\partial f}{\partial \mathbf{q}_c} + \kappa \frac{\partial f}{\partial \mathbf{q}_d} \right) \\ \frac{\partial f}{\partial \mathbf{q}^j} &= \frac{1}{4} \left(\frac{\partial f}{\partial \mathbf{q}_a} + \iota \frac{\partial f}{\partial \mathbf{q}_b} - j \frac{\partial f}{\partial \mathbf{q}_c} + \kappa \frac{\partial f}{\partial \mathbf{q}_d} \right) \\ \frac{\partial f}{\partial \mathbf{q}^\kappa} &= \frac{1}{4} \left(\frac{\partial f}{\partial \mathbf{q}_a} + \iota \frac{\partial f}{\partial \mathbf{q}_b} + j \frac{\partial f}{\partial \mathbf{q}_c} - \kappa \frac{\partial f}{\partial \mathbf{q}_d} \right).\end{aligned}\quad (13)$$

The so called $\mathbb{H}\mathbb{R}^*$ derivatives can then readily be written from (13) by using the property $(\frac{\partial f}{\partial \mathbf{q}})^* = \frac{\partial f}{\partial \mathbf{q}^*}$, where f is a real-valued function. Thus,

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

Similar to the conjugate derivatives property, an involution property is also applicable to real-valued functions, and is given by

$$\left(\frac{\partial f}{\partial \mathbf{q}} \right)^\alpha = \frac{\partial f}{\partial \mathbf{q}^\alpha}, \quad \alpha = \{\iota, j, \kappa\}.\quad (15)$$

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}^* , or $\frac{\partial f}{\partial \mathbf{q}^*}$. This can be seen as an extension of Brandwood's result for functions of complex variables [28], and it is thus natural to consider this gradient in the optimisation of cost functions. Finally, note that while we have considered real-valued functions in the above discussion, the $\mathbb{H}\mathbb{R}$ calculus framework can be equally utilised for the analysis of general quaternion-valued functions.

III. THE QUATERNION FASTICA ALGORITHM

Consider the standard ICA model

$$\hat{\mathbf{x}} = \mathbf{A}\mathbf{s}\quad (16)$$

whereby observed mixtures $\hat{\mathbf{x}} \in \mathbb{H}^N$ are a weighted sum of N_s latent sources $\mathbf{s} \in \mathbb{H}^{N_s}$ in a noise-free environment, and the 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 no assumption is made regarding the ι -, j - 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

$$\mathbf{V} = \mathbf{\Lambda}^{-1/2} \mathbf{E}^H,\quad (17)$$

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

To prove this, let us write the covariance matrix in terms of the quaternion right eigenvalue decomposition $\mathcal{C}_{\mathbf{q}\mathbf{q}} = E\{\mathbf{q}\mathbf{q}^H\} = \mathbf{E}\mathbf{\Lambda}\mathbf{E}^H$ [29]. The covariance matrix of the whitened random vector $\mathbf{p} = \mathbf{V}\mathbf{q}$ is then expressed as

$$\begin{aligned}E\{\mathbf{p}\mathbf{p}^H\} &= \mathbf{V}E\{\mathbf{q}\mathbf{q}^H\}\mathbf{V}^H \\ &= \mathbf{\Lambda}^{-1/2} \mathbf{E}^H (\mathbf{E}\mathbf{\Lambda}\mathbf{E}^H) \mathbf{E} \mathbf{\Lambda}^{-1/2} = \mathbf{I}\end{aligned}\quad (18)$$

where \mathbf{I} is the identity matrix. This result will be used for the whitening of the observed mixture $\hat{\mathbf{x}}$ in (16).

As a preprocessing step to aid the ICA algorithm, the quaternion mixture $\hat{\mathbf{x}}$ is whitened such that

$$E\{\mathbf{x}\mathbf{x}^H\} = \mathbf{M}E\{\mathbf{s}\mathbf{s}^H\}\mathbf{M}^H = \mathbf{I}\quad (19)$$

where $\mathbf{x} = \mathbf{V}\hat{\mathbf{x}} = \mathbf{V}\mathbf{A}\mathbf{s}$ and $\mathbf{M} \triangleq \mathbf{V}\mathbf{A}$ is the new unitary mixing matrix containing the whitening matrix \mathbf{V} , given in (17). We aim to obtain a demixing matrix \mathbf{W} such that $\mathbf{W}^H \mathbf{x}$ is an estimate of the original sources, albeit with a scaling, phase and permutation ambiguity. Then for the n th source estimate we have

$$y_n = \mathbf{w}_n^H \mathbf{x} = \mathbf{w}_n^H \mathbf{M}\mathbf{s} = \mathbf{u}^H \mathbf{s} = e^{\mu\varphi} s_m \quad (20)$$

where \mathbf{w}_n is the n th column of the demixing matrix \mathbf{W} , \mathbf{u} is a vector with a single non-zero value given by $e^{\mu\varphi}$ at the n th entry signifying an arbitrary direction within \mathbb{H} and $\boldsymbol{\mu} = \frac{(q_b + jq_c + \kappa q_d)}{\sqrt{q_b^2 + q_c^2 + q_d^2}}$ is the unit pure quaternion vector³. Finally, note that by constraining the demixing vector \mathbf{w}_n to unit norm, the estimated source y_n is of unit variance, that is

$$E\{y_n y_n^*\} = \mathbf{w}_n^H E\{\mathbf{x}\mathbf{x}^H\} \mathbf{w}_n = \mathbf{w}_n^H \mathbf{w}_n = 1 \quad (21)$$

while the matrix \mathbf{W} becomes unitary.

A. A Newton-update based ICA algorithm

The quaternion FastICA (q-FastICA) algorithm is based on the maximisation of the negentropy of the separated sources, following from previous implementations of the FastICA algorithm in the real and complex domains [1], [7], [10]. This is achieved by utilising an appropriate nonlinear function $G(y)$, so as to make a suitable approximation of the negentropy function. In [22], three distinct quaternion nonlinearities were identified whereby the nonlinear operation is split on each component of y (split-quaternion function), on the components of the Cayley-Dickson form of y (split-complex function), or applied directly on y (full-quaternion function). It was also shown that the full-quaternion nonlinearity resulted in the best separation performance. Under the stringent analyticity conditions of the Cauchy-Riemann-Feuter [30] 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 [31]. However, this depends on assumptions that may not be valid for general nonlinear functions. Thus, to avoid problems associated with the derivation of fully-quaternion nonlinearities, we shall utilise a real-valued smooth and even nonlinearity $G : \mathbb{R} \mapsto \mathbb{R}$, while implementing an augmented Newton method so as to employ the full information available within general \mathbb{Q} -improper mixtures.

The q-FastICA cost function is then defined as

$$\mathcal{J}(\mathbf{w}, \mathbf{w}^i, \mathbf{w}^j, \mathbf{w}^k) = E\{G(|\mathbf{w}^H \mathbf{x}|^2)\} \quad (22)$$

where the cost function \mathcal{J} is written in terms of the four basis vectors for emphasis on the equivalent notation. The optimisation problem based on (22) can then be stated as

$$\mathbf{w}_{opt} = \arg \max_{\|\mathbf{w}\|_2=1} \mathcal{J}(\mathbf{w}, \mathbf{w}^i, \mathbf{w}^j, \mathbf{w}^k) \quad (23)$$

where the demixing vector is normalised to avoid very small values of \mathbf{w} , while keeping the variance of the extracted sources equal to unity.

The solution of this constrained optimisation problem is found through the method of Lagrangian multipliers and by utilising the Newton method to perform a fast iterative search to the optimal value \mathbf{w}_{opt} . In summary, the quaternion FastICA algorithm for the estimation of one source is expressed in its augmented form as

$$\begin{aligned} \mathbf{w}^a(k+1) &= \mathbf{w}^a(k) - (\mathbf{H}_{\mathbf{w}\mathbf{w}}^a)^{-1} \nabla_{\mathbf{w}^a} \mathcal{L} \\ \lambda(k+1) &= \lambda(k) + \mu \nabla_{\lambda} \mathcal{L} \\ \mathbf{w}(k+1) &\leftarrow \frac{\mathbf{w}(k+1)}{\|\mathbf{w}(k+1)\|_2} \end{aligned} \quad (24)$$

where the augmented demixing vector $\mathbf{w}^a = [\mathbf{w}, \mathbf{w}^i, \mathbf{w}^j, \mathbf{w}^k]^T$, \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 $\mathbf{H}_{\mathbf{w}\mathbf{w}}^a$ are respectively the augmented gradient vector and Hessian matrix of the Lagrangian function. The full derivation is provided in the Appendix.

The estimation of multiple sources can be performed one by one through a deflationary procedure, where for the n th estimated source is given by the following Gram-Schmidt orthogonalisation procedure

$$\begin{aligned} \mathbf{w}_n(k+1) &\leftarrow \mathbf{w}_n(k+1) - \check{\mathbf{W}}\check{\mathbf{W}}^H \mathbf{w}_n(k+1) \\ \check{\mathbf{W}} &= [\mathbf{w}_1(k+1), \dots, \mathbf{w}_n(k+1)] \end{aligned} \quad (25)$$

or simultaneously via a symmetric orthogonalisation method

$$\mathbf{W}(k+1) \leftarrow (\mathbf{W}(k+1)\mathbf{W}^H(k+1))^{-1/2} \mathbf{W}(k+1), \quad (26)$$

where the orthogonalisation procedures in the quaternion domain follow from the already established results.

³A pure ‘imaginary’ quaternion is referred to as the imaginary or vector part of a quaternion variable.

TABLE I
SOURCE PROPERTIES FOR BENCHMARK SIMULATION WITH DEFLATIONARY APPROACH

Source	polytope	\mathbb{Q} -improperness measure (r)
$s_1(k)$	Cubic	0.0129
$s_2(k)$	Cyclic (2 point)	1.0000
$s_3(k)$	Cyclic (3 point)	0.3392
$s_4(k)$	5-Simplex	0.0071

IV. SIMULATIONS AND DISCUSSION

A. Benchmark simulations

The performance of the algorithm is first assessed through simulations using synthetic four dimensional signal codes located on the edges of geometric polytopes [32] with varying degree of \mathbb{Q} -improperness. To assess the degree of \mathbb{Q} -improperness of the generated sources, we define a measure based on the ratio of the complementary variances to the standard variance, expressed as

$$r = \frac{|E\{qq^{t*}\}| + |E\{qq^{j*}\}| + |E\{qq^{\kappa*}\}|}{3E\{qq^*\}}, \quad r \in [0, 1]. \quad (27)$$

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

The performance of the quaternion FastICA algorithm using the deflationary orthogonalisation was assessed using the Performance Index (PI) [33], which for $\mathbf{u}^H = \mathbf{w}^H \mathbf{V} \mathbf{A} = [u_1, \dots, u_N]^H$ is given as

$$PI = 10 \log_{10} \left(\frac{1}{N} \left(\sum_{i=1}^N \frac{|u_i|^2}{\max\{|u_1|^2, \dots, |u_N|^2\}} \right) \right) \quad (28)$$

and indicates the proximity of \mathbf{u} to a vector with a single non-zero element. For the deflationary approach, a PI of less than -20dB indicates good separation performance. For the q-FastICA algorithm with symmetric orthogonalisation, the full PI measure was used, given by

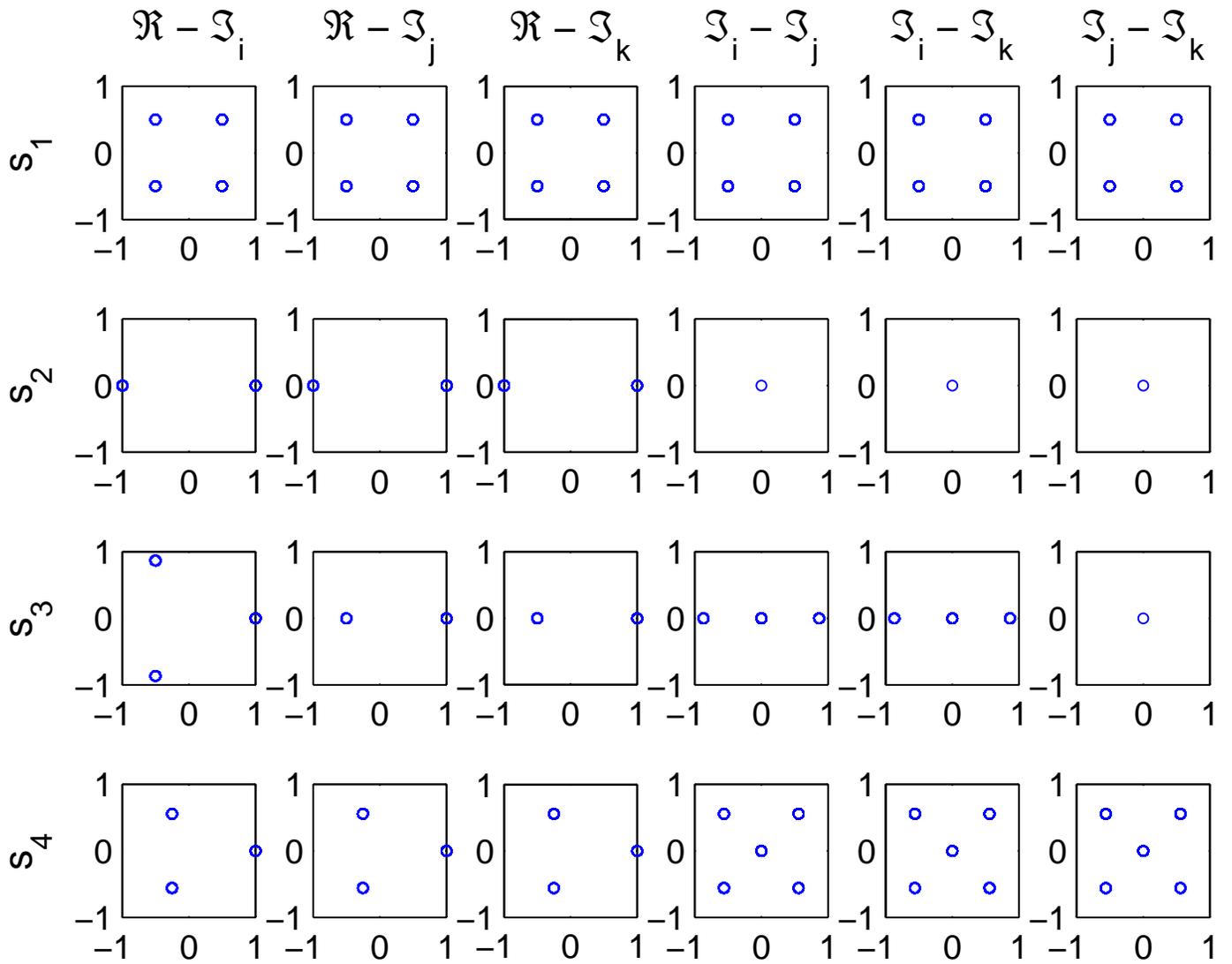
$$PI = 10 \log_{10} \left(\frac{1}{N} \sum_{i=1}^N \left(\sum_{j=1}^N \frac{|u_{ij}|}{\max\{|u_{i1}|, \dots, |u_{iN}|\}} - 1 \right) + \frac{1}{N} \sum_{j=1}^N \left(\sum_{i=1}^N \frac{|u_{ij}|}{\max\{|u_{1j}|, \dots, |u_{Nj}|\}} - 1 \right) \right). \quad (29)$$

where $\mathbf{U}^H = \mathbf{W}^H \mathbf{V} \mathbf{A}$ and $u_{ij} = (\mathbf{U})_{ij}$ and a PI less than -10dB signifies good separation performance.

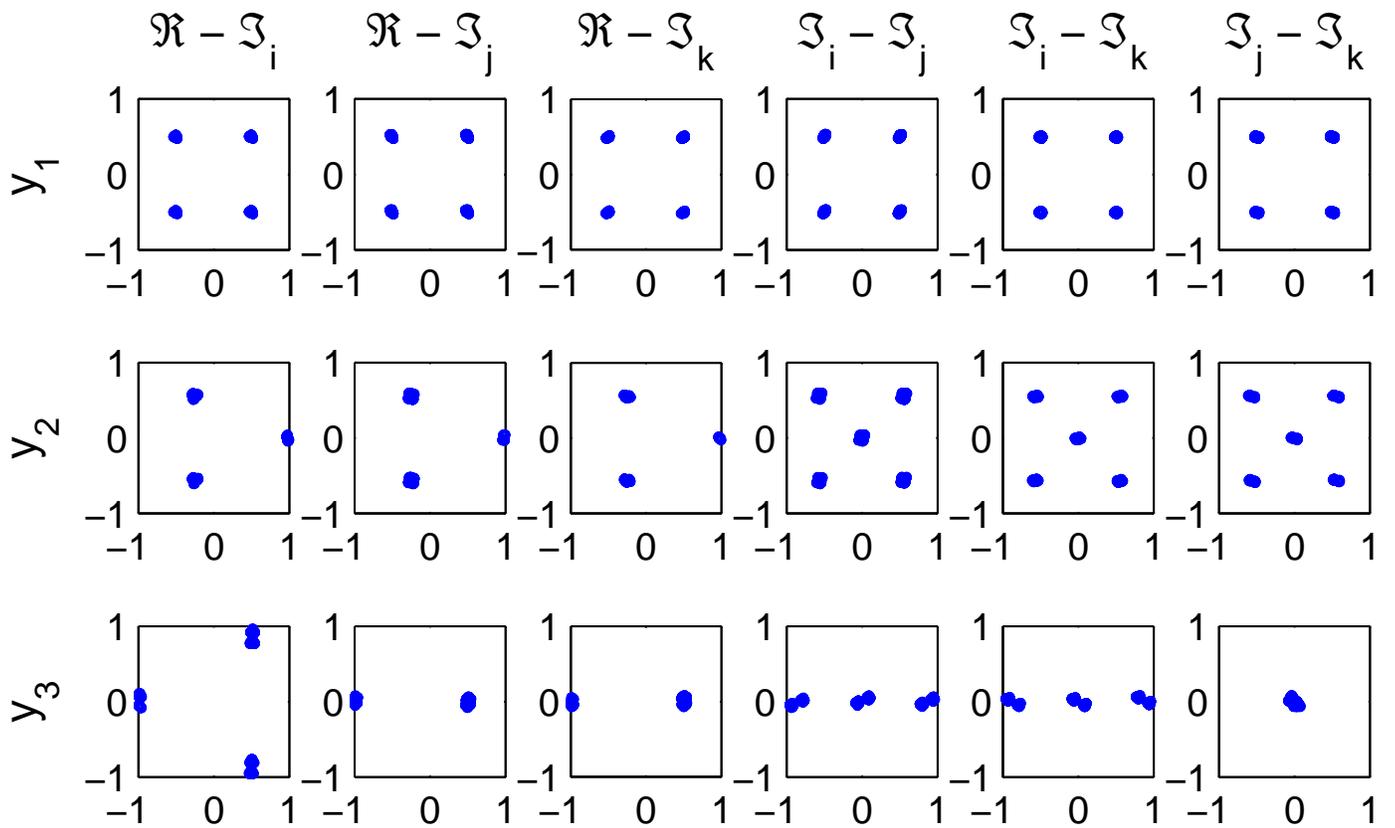
For simulations, 5000 samples of four polytope sources were mixed using a randomly generated quaternion-valued 4×4 mixing matrix. The observed mixtures were then whitened and processed using the q-FastICA algorithm (24), using the deflationary and symmetric orthogonalisation.

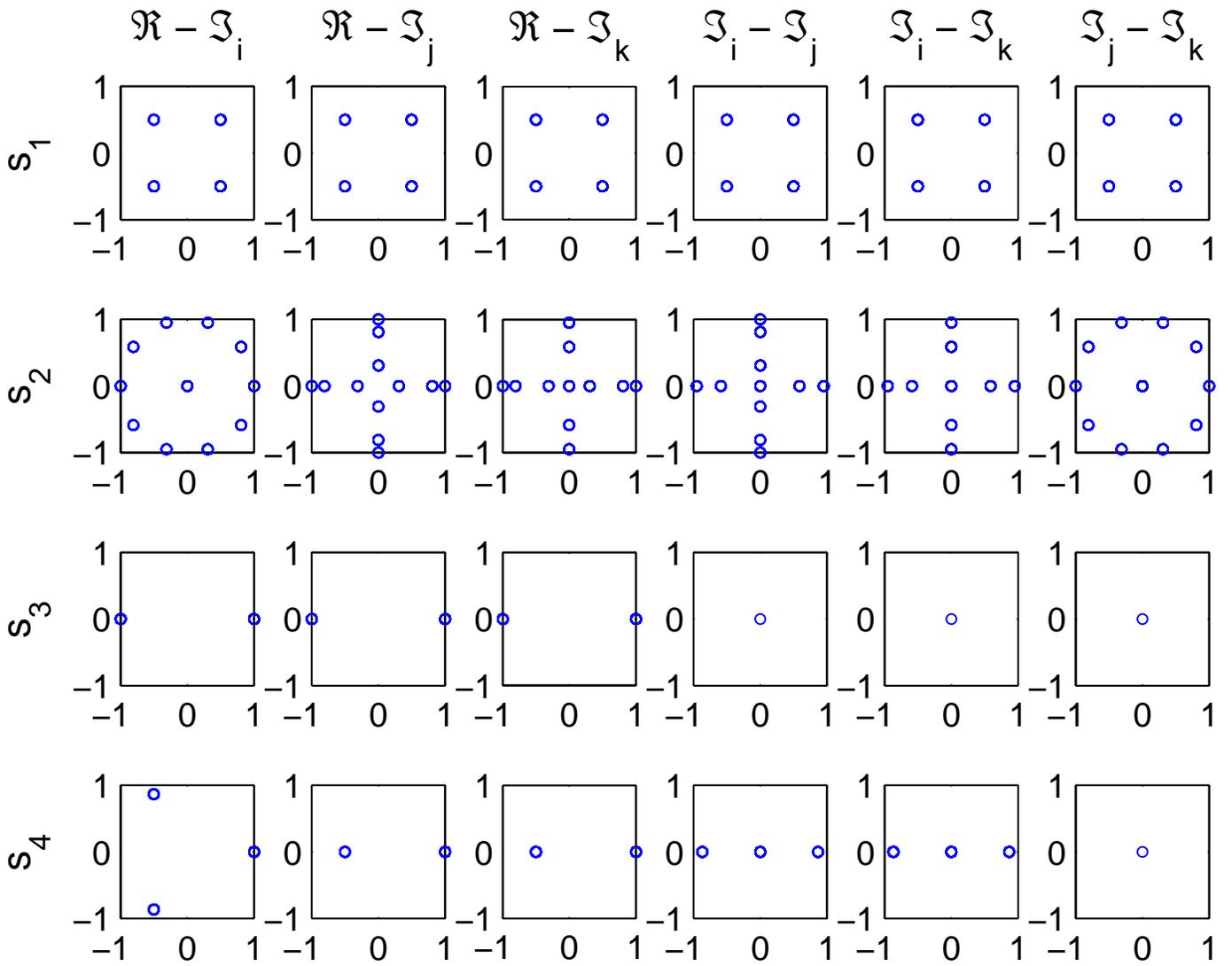
1) *Deflationary orthogonalisation:* The scatter plots of the four quaternion sources are shown in Fig. 1(a) and their properties are given in Table I. Source $s_1(k)$ was a cubic polytope, $s_2(k)$ and $s_3(k)$ were generated from cyclic groups with two and three points, and s_4 was a simplex with five vertices. The nonlinearity $G(y) = \log \cosh(y)$, the demixing vector \mathbf{w} was initialised randomly and the step-size of the gradient ascent update $\mu = 1$ and $\lambda = 5$. The scatter plot of the normalised estimated sources are given in Fig. 1(b) and the performance of the q-FastICA algorithm in the separation of each source and at each iteration stage is shown in Fig. 1(c). It can be seen that the algorithm was successful in estimating all the sources, converging to a solution in as few as four iterations. As expected from a deflationary orthogonalisation procedure, the performance of the algorithm degraded after each stage due to the accumulation of errors, with the final PI value for the first estimated source $y_1(k)$ of -39.93dB, while for $y_4(k)$ this value reduced to -26.28dB. Note that due to the symmetry of the signal codes, rotations of the extracted sources relative to the original source are not visible, and can only be observed in the scatter plot of $y_3(k)$.

2) *Symmetric orthogonalisation:* In this simulation, the sources were estimated simultaneously using the algorithm (24) and the orthogonalisation procedure (26). Table II describes the source properties; visual scatter plot representations are given in Fig. 2(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, the value of $r = 0.3351$ for $s_4(k)$, and the other two sources were \mathbb{Q} -proper. For performance comparison, the nonlinearity G was chosen as in [7], with $G_1(y) = \log \cosh(y)$, $G_2(y) = \sqrt{0.1 + y}$ and $G_3(y) = \log(0.1 + y)$. The demixing matrix \mathbf{W} was initialised randomly, and the step-sizes $\mu_1 = 1$, $\mu_2 = 0.1$, $\mu_3 = 0.5$ and $\lambda = 5$ for the gradient ascent update algorithm. As shown in Fig. 2(c), the algorithm successfully separated all the four sources with the respective PI values of -17.87dB, -15.8086 dB and -19.4882 dB. Fig. 2(b) depicts the scatter plots of the normalised estimated sources with nonlinearity G_1 , note that sources were estimated in a random order.



(a) The scatter plot of the quantum course properties given in Table I





(a) The scatter plot of the quaternion source properties given in Table I

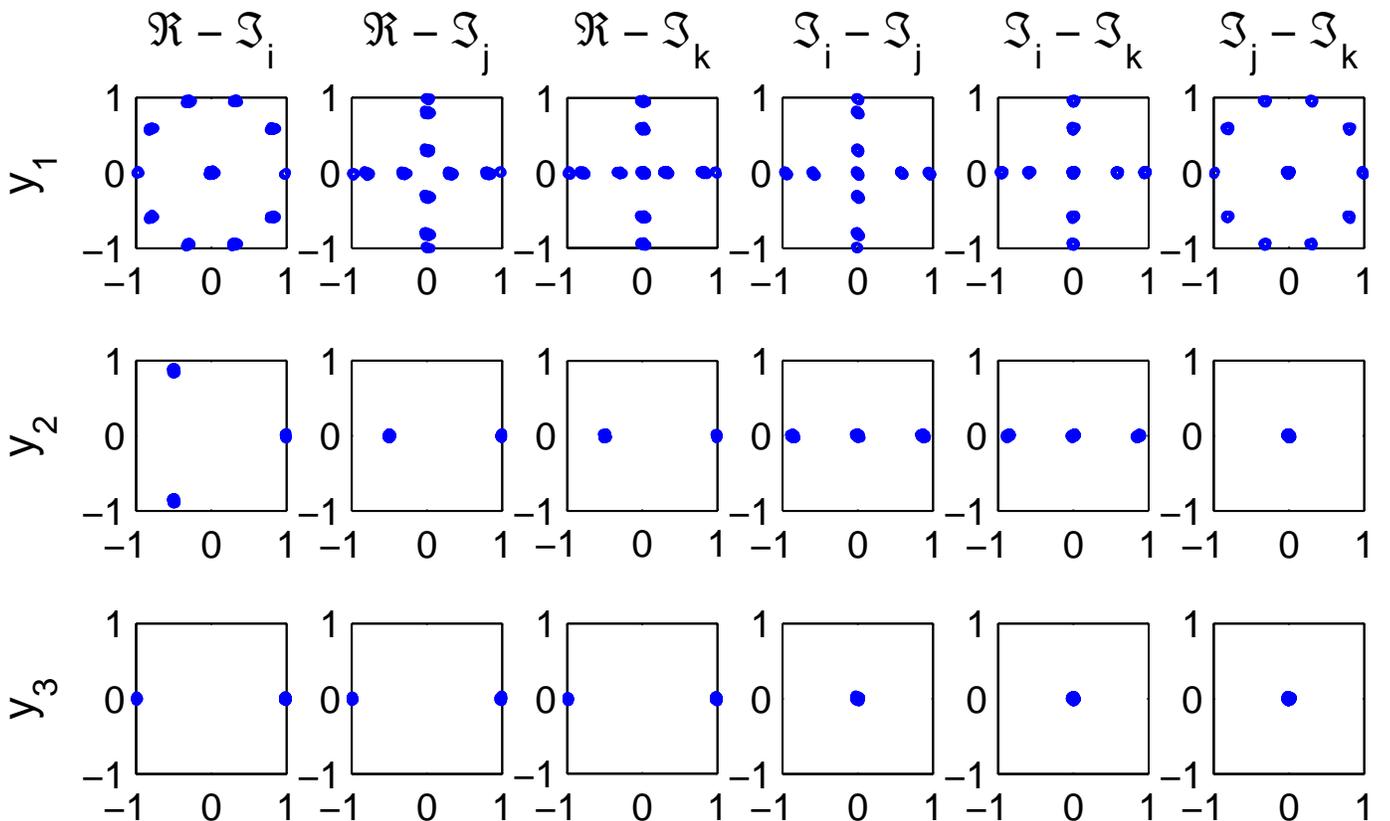


TABLE II
SOURCE PROPERTIES FOR BENCHMARK SIMULATION WITH SYMMETRIC ORTHOGONALISATION APPROACH

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

B. EEG artifact extraction

In a practical EEG recording session, each EEG recording channel consists of a superposition of a pure EEG signal corresponding to the collective neural activity within the brain, and electrical activity pertaining to distinctive artifacts such as movement of the head, line noise and eye blinks. In modelling the EEG signal, the artifacts, both external and biological, are considered statistically independent from the pure EEG recording [34]–[36]. The usefulness of the real-valued FastICA algorithm in the extraction of eyeblink artifacts was studied in [37].

In the experimental setup, data was sampled at 4.8kHz for 30s 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 the 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), that is, the electrical activity due to eye blinks and eye movement.

The 4-tuple 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 \mathbb{Q} -improper quaternion signals were constructed as

$$\begin{aligned}
 x_1(k) &= \text{AF8}(k) + i\text{AF4}(k) + j\text{AF7}(k) + \kappa\text{AF3}(k) \\
 x_2(k) &= \text{MR}(k) + i\text{C3}(k) + j\text{ML}(k) + \kappa\text{PO8}(k) \\
 x_3(k) &= \text{PO8}(k) + i\text{PO4}(k) + j\text{PO3}(k) + \kappa\text{PO7}(k)
 \end{aligned} \tag{30}$$

and the observed EEG mixture were then represented as $\mathbf{x} = [x_1(k), x_2(k), x_3(k)]^T$. The degree of \mathbb{Q} -impropriety of the signals were respectively 0.8902, 0.6824 and 0.8932, according to (27).

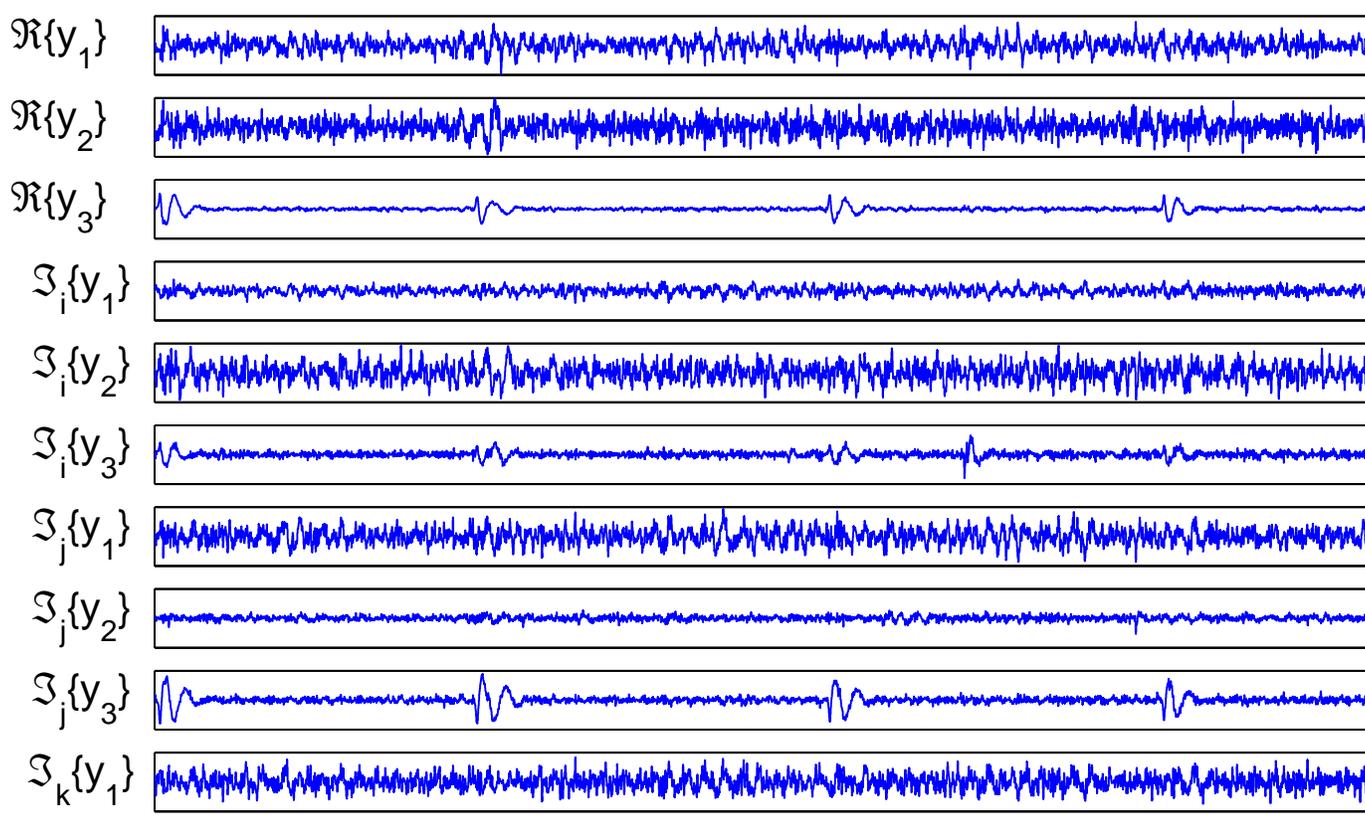
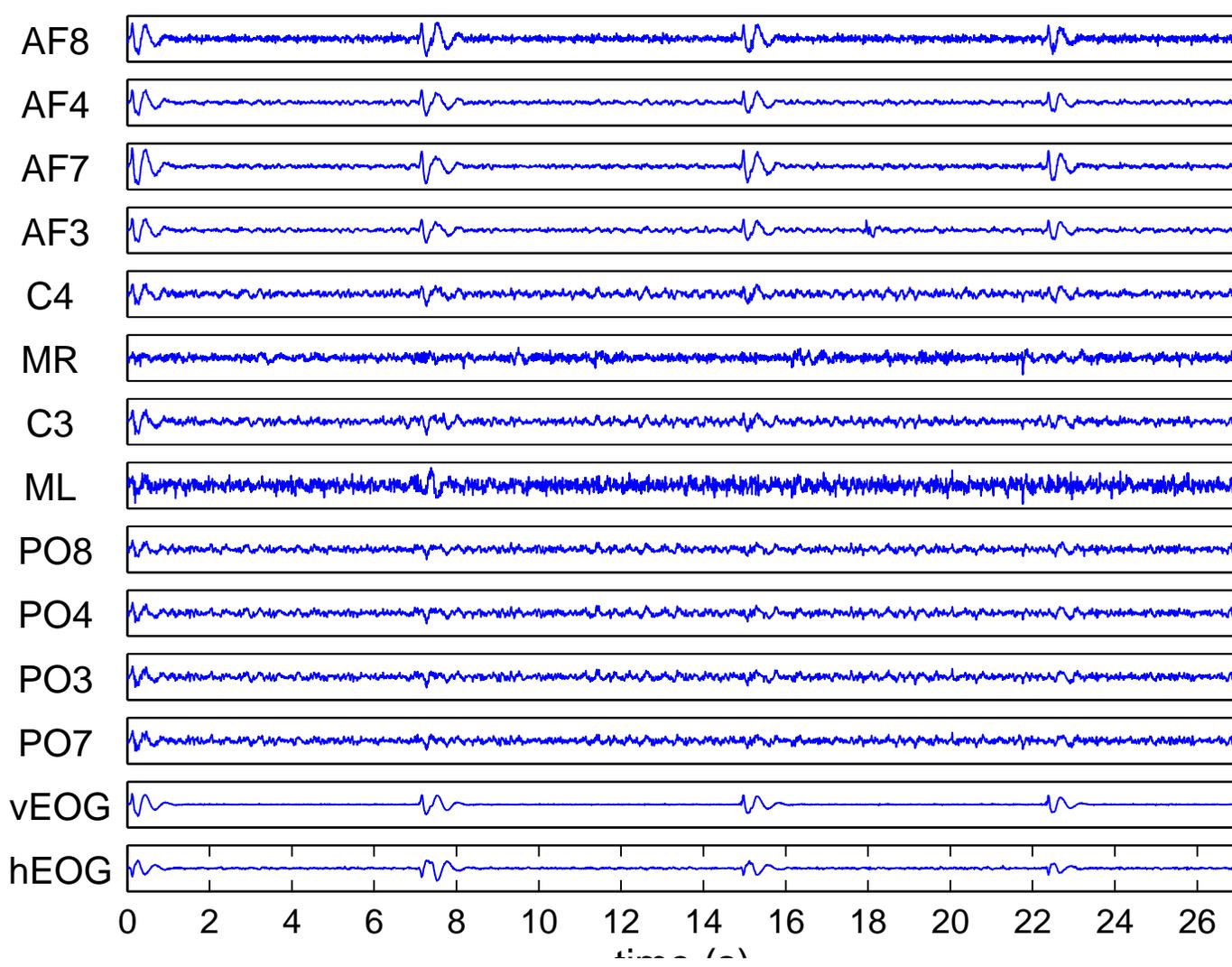
In this scheme, the quaternion FastICA algorithm (24) was first utilised to estimate the source signals, with the step-size $\mu = 1$ and initial Lagrange parameter $\lambda = 5$, while the nonlinearity was chosen as $G(y) = \log \cosh(y)$, to provide good overall performance. Next, the estimated source pertaining to the EOG artifact was selected through 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 [38], this being attributed the the sparse nature of eye blinks.

A time plot of the original recorded channels and the components of the quaternion-valued separated sources are depicted respectively in Fig. 3(a) and Fig. 3(b). The occurrence of the eye blinking can be seen at the beginning of the recording, then at around 7s, 15s and 22s, where the effect of the EOG artifact is more prominent on the frontal lobe channel, and less severe in the central and occipital channels. By visual inspection, the separated EOG artifact can be seen to span the components of the third extracted source $y_3(k)$, that is $\Re\{y_3(k)\}$, $\Im_i\{y_3(k)\}$, $\Im_j\{y_3(k)\}$, $\Im_\kappa\{y_3(k)\}$, and is confirmed through comparison of the kurtosis values of each components (Fig. 3(c)). While most estimated sources have a near-zero measure of kurtosis, the real and imaginary components of $y_3(k)$ have, 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 signal and then compared to the original combined EOG recording. Fig. 3(d) depicts both signals along with the residual error of the estimation process, having a mean square error of 1.21×10^{-4} . Also, by excluding the components of $y_3(k)$ the clean EEG mixture was reconstructed and a 3s window between 6s–9s for each reconstructed channel is shown in Fig. 3(e), where the effect of the EOG present at 7s was diminished in the channels.

V. CONCLUSIONS

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 utilised to estimate independent quaternion-valued sources, while an augmented Newton method implementation has allowed for the extension of the FastICA methodology to the quaternion domain. The performance of the quaternion FastICA (q-FastICA) algorithm in deflationary and simultaneous separation using benchmark quaternion polytope signals has been discussed, and the algorithm has been shown to be effective in the removal of ocular artifacts from EEG signals.



APPENDIX A
SOME RELEVANT RESULTS FROM $\mathbb{H}\mathbb{R}$ CALCULUS

Several results, used in the derivation of the q-FastICA algorithm (24) are discussed here.

A. Chain rule in $\mathbb{H}\mathbb{R}$ calculus

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

$$\frac{\partial F}{\partial \xi} = \frac{\partial F}{\partial G} \frac{\partial G}{\partial \xi} + \frac{\partial F}{\partial G^i} \frac{\partial G^i}{\partial \xi} + \frac{\partial F}{\partial G^j} \frac{\partial G^j}{\partial \xi} + \frac{\partial F}{\partial G^\kappa} \frac{\partial G^\kappa}{\partial \xi} \quad (31)$$

and $\xi = \{q, q^i, q^j, q^\kappa\}$. To show this, the total differential of $F(\bar{q})$ can be written as [19]

$$dF = \frac{\partial F}{\partial \bar{q}} d\bar{q} + \frac{\partial F}{\partial \bar{q}^i} d\bar{q}^i + \frac{\partial F}{\partial \bar{q}^j} d\bar{q}^j + \frac{\partial F}{\partial \bar{q}^\kappa} d\bar{q}^\kappa \quad (32)$$

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

$$dG = \frac{\partial G}{\partial q} dq + \frac{\partial G}{\partial q^i} dq^i + \frac{\partial G}{\partial q^j} dq^j + \frac{\partial G}{\partial q^\kappa} dq^\kappa \quad (33)$$

By substituting (33) into (32), and after rearranging the expressions, we obtain the total differential of F with respect to q as

$$\begin{aligned} dF &= \left(\frac{\partial F}{\partial G} \frac{\partial G}{\partial q} + \frac{\partial F}{\partial G^i} \frac{\partial G^i}{\partial q} + \frac{\partial F}{\partial G^j} \frac{\partial G^j}{\partial q} + \frac{\partial F}{\partial G^\kappa} \frac{\partial G^\kappa}{\partial q} \right) dq \\ &+ \left(\frac{\partial F}{\partial G} \frac{\partial G}{\partial q^i} + \frac{\partial F}{\partial G^i} \frac{\partial G^i}{\partial q^i} + \frac{\partial F}{\partial G^j} \frac{\partial G^j}{\partial q^i} + \frac{\partial F}{\partial G^\kappa} \frac{\partial G^\kappa}{\partial q^i} \right) dq^i \\ &+ \left(\frac{\partial F}{\partial G} \frac{\partial G}{\partial q^j} + \frac{\partial F}{\partial G^i} \frac{\partial G^i}{\partial q^j} + \frac{\partial F}{\partial G^j} \frac{\partial G^j}{\partial q^j} + \frac{\partial F}{\partial G^\kappa} \frac{\partial G^\kappa}{\partial q^j} \right) dq^j \\ &+ \left(\frac{\partial F}{\partial G} \frac{\partial G}{\partial q^\kappa} + \frac{\partial F}{\partial G^i} \frac{\partial G^i}{\partial q^\kappa} + \frac{\partial F}{\partial G^j} \frac{\partial G^j}{\partial q^\kappa} + \frac{\partial F}{\partial G^\kappa} \frac{\partial G^\kappa}{\partial q^\kappa} \right) dq^\kappa \end{aligned}$$

where the derivatives $\frac{\partial F}{\partial \xi}$ are given by the terms within the brackets, and form the chain rule. The chain rule for the $\mathbb{H}\mathbb{R}^*$ derivatives can be obtained similarly, and the result of (31) can be extended to vector-valued functions to form a generalised chain rule for the derivatives.

B. First and second derivatives of the cost function $\mathcal{J}(\mathbf{w})$

First, 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

$$\begin{aligned} \frac{\partial yy^*}{\partial \mathbf{w}^*} &= \frac{\partial y}{\partial \mathbf{w}^*} y^* + y \frac{\partial y^*}{\partial \mathbf{w}^*} = \mathbf{x}y^* - \frac{1}{2}y\mathbf{x}^* \\ \frac{\partial (yy^*)^i}{\partial \mathbf{w}^*} &= \frac{\partial y^i}{\partial \mathbf{w}^*} y^{i*} + y^i \frac{\partial y^{i*}}{\partial \mathbf{w}^*} = \frac{1}{2}y^i \mathbf{x}^{i*} \\ \frac{\partial (yy^*)^j}{\partial \mathbf{w}^*} &= \frac{\partial y^j}{\partial \mathbf{w}^*} y^{j*} + y^j \frac{\partial y^{j*}}{\partial \mathbf{w}^*} = \frac{1}{2}y^j \mathbf{x}^{j*} \\ \frac{\partial (yy^*)^\kappa}{\partial \mathbf{w}^*} &= \frac{\partial y^\kappa}{\partial \mathbf{w}^*} y^{\kappa*} + y^\kappa \frac{\partial y^{\kappa*}}{\partial \mathbf{w}^*} = \frac{1}{2}y^\kappa \mathbf{x}^{\kappa*}. \end{aligned}$$

Then by using the chain rule (31) and after simplification we obtain the gradients of the cost function as

$$\begin{aligned} \nabla_{\mathbf{w}^*} \mathcal{J} &= E\{2g(|y|^2)\mathbf{x}y^*\} \\ \nabla_{\mathbf{w}^{i*}} \mathcal{J} &= E\{2g(|y|^2)\mathbf{x}y^{i*}\} \\ \nabla_{\mathbf{w}^{j*}} \mathcal{J} &= E\{2g(|y|^2)\mathbf{x}y^{j*}\} \\ \nabla_{\mathbf{w}^{\kappa*}} \mathcal{J} &= E\{2g(|y|^2)\mathbf{x}y^{\kappa*}\} \end{aligned} \quad (34)$$

where g is the first derivative of G ; this result can also be interpreted based on the involution property (15). After simplifications and considering the whiteness of \mathbf{x} , the second derivatives of \mathcal{J} can then be calculated as

$$\begin{aligned}\frac{\partial}{\partial \mathbf{w}^*} \left(\frac{\partial \mathcal{J}}{\partial \mathbf{w}^*} \right)^T &= E\{4g'(|y|^2)\mathbf{x}y^*\mathbf{x}^T y^* - g(|y|^2)\mathbf{I}\} \\ \frac{\partial}{\partial \mathbf{w}^{i*}} \left(\frac{\partial \mathcal{J}}{\partial \mathbf{w}^*} \right)^T &= E\{2g'(|y|^2)(\mathbf{x}y^*)^i(\mathbf{x}^T y^*) + g(|y|^2)\mathbf{I}\} \\ \frac{\partial}{\partial \mathbf{w}^{j*}} \left(\frac{\partial \mathcal{J}}{\partial \mathbf{w}^*} \right)^T &= E\{2g'(|y|^2)(\mathbf{x}y^*)^j(\mathbf{x}^T y^*) + g(|y|^2)\mathbf{I}\} \\ \frac{\partial}{\partial \mathbf{w}^{\kappa*}} \left(\frac{\partial \mathcal{J}}{\partial \mathbf{w}^*} \right)^T &= E\{2g'(|y|^2)(\mathbf{x}y^*)^\kappa(\mathbf{x}^T y^*) + g(|y|^2)\mathbf{I}\},\end{aligned}\quad (35)$$

where g' is the second derivative of G and the calculations of the remaining derivatives follow from property (15). Finally, notice that the non-commutativity of the quaternion product prohibits further simplification of the derivatives in (35).

APPENDIX B THE AUGMENTED QUATERNION NEWTON METHOD

The duality between \mathbb{R}^4 and \mathbb{H} allows for the consideration of the relations between the derivatives in the two domains. This methodology was previously considered in [39] and resulted in the derivation of the augmented complex Newton method. The extension of this work to the quaternion domain based on the involution bases was detailed in [19], [27]. A short summary is presented below.

For a function $f(\mathbf{q}) : \mathbb{H}^N \mapsto \mathbb{R}$, its augmented gradient $\nabla_{\mathbf{q}^{a*}} f = \frac{\partial f}{\partial \mathbf{q}^{a*}}$ and Hessian $\mathbf{H}_{\mathbf{q}\mathbf{q}}^a = \frac{\partial}{\partial \mathbf{q}^{a*}} \left(\frac{\partial f}{\partial \mathbf{q}^{a*}} \right)^T$, where the augmented vector $\mathbf{q}^a = [\mathbf{q}^T, \mathbf{q}^{i*T}, \mathbf{q}^{j*T}, \mathbf{q}^{\kappa*T}]^T$. The augmented Newton update can then be written as

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

where $\Delta \mathbf{q}^a = \mathbf{q}^a(k+1) - \mathbf{q}^a(k)$ is the change in \mathbf{q}^a in each consecutive update. Finally, observe that the elements of the augmented Hessian matrix

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

can be written in terms of its first row by utilising the involution property (15) and noting that $((\cdot)^\alpha)^\beta = (\cdot)^\gamma, \alpha \neq \beta \neq \gamma = \{i, j, \kappa\}$.

APPENDIX C DERIVATION OF THE AUGMENTED Q-FASTICA UPDATE ALGORITHM

The Lagrangian function \mathcal{L} for the optimisation problem in (23) is given by

$$\mathcal{L}(\mathbf{w}, \lambda) = \mathcal{J}(\mathbf{w}) + \underbrace{\lambda(\mathbf{w}^H \mathbf{w} - 1)}_{\triangleq c} \quad (38)$$

where $\lambda \in \mathbb{R}$ is the Lagrange parameter. We utilise the Newton method (36) to find the extrema of (38), where

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial \mathbf{w}^{a*}} &= \frac{\partial \mathcal{J}}{\partial \mathbf{w}^{a*}} + \frac{\partial c}{\partial \mathbf{w}^{a*}} \\ \frac{\partial}{\partial \mathbf{w}^{a*}} \left(\frac{\partial \mathcal{L}}{\partial \mathbf{w}^{a*}} \right)^T &= \mathbf{H}_{\mathbf{w}\mathbf{w}}^a + \frac{\partial}{\partial \mathbf{w}^{a*}} \left(\frac{\partial c}{\partial \mathbf{w}^{a*}} \right)^T\end{aligned}\quad (39)$$

and the augmented gradient and Hessian of \mathcal{J} are obtained using (34) and (35). The gradients of c are then given by

$$\begin{aligned}\frac{\partial c}{\partial \mathbf{w}^*} &= \lambda(\mathbf{w} - \frac{1}{2}\mathbf{w}^*) & \frac{\partial c}{\partial \mathbf{w}^{i*}} &= \frac{\lambda}{2}\mathbf{w}^* \\ \frac{\partial c}{\partial \mathbf{w}^{j*}} &= \frac{\lambda}{2}\mathbf{w}^* & \frac{\partial c}{\partial \mathbf{w}^{\kappa*}} &= \frac{\lambda}{2}\mathbf{w}^*\end{aligned}$$

and the Hessian can be calculated from

$$\begin{aligned} \frac{\partial}{\partial \mathbf{w}^*} \left(\frac{\partial c}{\partial \mathbf{w}^*} \right)^T &= -\lambda \mathbf{I} & \frac{\partial}{\partial \mathbf{w}^{i*}} \left(\frac{\partial c}{\partial \mathbf{w}^*} \right)^T &= -\frac{\lambda}{2} \mathbf{I} \\ \frac{\partial}{\partial \mathbf{w}^{j*}} \left(\frac{\partial c}{\partial \mathbf{w}^*} \right)^T &= -\frac{\lambda}{2} \mathbf{I} & \frac{\partial}{\partial \mathbf{w}^{k*}} \left(\frac{\partial c}{\partial \mathbf{w}^*} \right)^T &= -\frac{\lambda}{2} \mathbf{I}. \end{aligned}$$

By substituting these results in (36), the Newton update for the Lagrangian is obtained. Finally, the Lagrange parameter λ is updated using a gradient ascent method, whereby 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 normalisation of the demixing vector [40], as in (24).

ACKNOWLEDGEMENT

We wish to thank Dr. Nicolas Le Bihan for providing the script to generate the class of polytope signal codes. We are also grateful to Dr. Clive Cheong-Took and Cyrus Jahanchahi for sharing insight and discussions on augmented quaternion statistics and $\mathbb{H}\mathbb{R}$ calculus.

REFERENCES

- [1] A. Hyvärinen, J. Karhunen, and E. Oja. *Independent Component Analysis*. Wiley, 2001.
- [2] A. Belouchrani, K. Abed-Meraim, J.-F. Cardoso, and E. Moulines. A blind source separation technique using second-order statistics. *IEEE Transactions on Signal Processing*, 45(2):434–444, 1997.
- [3] J. F. Cardoso and A. Souloumiac. Blind beamforming for non-Gaussian signals. *Radar and Signal Processing, IEE Proceedings F*, 140(6):362–370, 1993.
- [4] A. J. Bell and T. J. Sejnowski. An information-maximisation approach to blind separation and blind deconvolution. *Neural Computation*, 7:1129–1159, 1995.
- [5] A. Hyvärinen. Fast and robust fixed-point algorithms for independent component analysis. *IEEE Transactions on Neural Networks*, 10(3):626–634, May 1999.
- [6] E. Oja and Z. Yuan. The FastICA algorithm revisited: Convergence analysis. *IEEE Transactions on Neural Networks*, 17(6):1370–1381, November 2006.
- [7] E. Bingham and A. Hyvärinen. A fast fixed point algorithm for independent component analysis of complex valued signals. *Journal of Neural Systems*, 10:1–8, 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.*, 20072007(1):83–98, 2007.
- [9] J. Eriksson and V. Koivunen. Complex random vectors and ICA models: Identifiability, uniqueness, and separability. *IEEE Transactions on Information Theory*, 52(3):1017–1029, 2006.
- [10] M. Novey and T. Adali. On extending the complex FastICA algorithm to noncircular sources. *IEEE Transactions on Signal Processing*, 56(5):2148–2154, 2008.
- [11] D. P. Mandic and S. L. Goh. *Complex Valued Nonlinear Adaptive Filters: Noncircularity, Widely Linear and Neural Models*. Wiley, 2009.
- [12] P. J. Schreier and L. L. Scharf. *Statistical Signal Processing of Complex-Valued Data*. Cambridge University Press, 2010.
- [13] K. Kreutz-Delgado. The complex gradient operator and the $\mathbb{C}\mathbb{R}$ -calculus. *Dept. of Electrical and Computer Engineering, UC San Diego, Course Lecture Supplement No. ECE275A*, pages 1–74, 2006.
- [14] J. P. Ward. *Quaternions and Cayley numbers*. Kluwer Academic Publishers, 1997.
- [15] S. Sangwine and N. Le Bihan. Quaternion polar representation with a complex modulus and complex argument inspired by the Cayley-Dickson form. *Advances in Applied Clifford Algebras*, 20(1):111–120, March 2010.
- [16] N. N. Vakhania. Random vectors with values in quaternion Hilbert spaces. *Theory of Probability and its Applications*, 43(1):99–115, January 1999.
- [17] N. Le Bihan and P.O. Amblard. Detection and estimation of Gaussian proper quaternion valued random processes. In *7th IMA Conference on Mathematics in Signal Processing, Cirencester, UK*, 2006.
- [18] C. Cheong-Took and D. P. Mandic. Augmented second order statistics of quaternion random signals. *Signal Processing*, 2010 (accepted).
- [19] C. Jahanchahi, C. Cheong-Took, and D. P. Mandic. On $\mathbb{H}\mathbb{R}$ calculus, quaternion valued stochastic gradient, and adaptive three dimensional wind forecasting. In *International Joint Conference on Neural Networks*, 2010 (accepted).
- [20] C. Cheong-Took and D. P. Mandic. A quaternion widely linear adaptive filter. *IEEE Transactions on Signal Processing*, 2010 (accepted).
- [21] B. Che-Ujang, C. Cheong-Took, and D. P. Mandic. Split quaternion nonlinear adaptive filtering. *Neural Networks*, 23(3):426–434, April 2010.
- [22] N. Le Bihan and S. Buchholz. Quaternionic independent component analysis using hypercomplex nonlinearities. In *7th IMA Conference on Mathematics in Signal Processing*, 2006.
- [23] T. A. Ell and S. J. Sangwine. Quaternion involutions and anti-involutions. *Computers & Mathematics with Applications*, 53(1):137–143, January 2007.
- [24] B. Picinbono. Second-order complex random vectors and normal distributions. *IEEE Transactions on Signal Processing*, 44(10):2637–2640, 1996.
- [25] P.J. Schreier and L.L. Scharf. Second-order analysis of improper complex random vectors and processes. *IEEE Transactions on Signal Processing*, 51(3):714–725, 2003.
- [26] W. Wirtinger. Zur formalen theorie der funktionen von mehr komplexen veränderlichen. *Mathematische Annalen*, 97(1):357–375, December 1927.
- [27] D. P. Mandic. The $\mathbb{H}\mathbb{R}$ calculus and quaternion analyticity. Technical report, Imperial College, January 2010.
- [28] D. H. Brandwood. A complex gradient operator and its application in adaptive array theory. *IEE Proceedings F: Communications, Radar and Signal Processing*, 130(1):11–16, February 1983.
- [29] F. Zhang. Quaternions and matrices of quaternions. *Linear Algebra and its Applications*, 251:21–57, January 1997.
- [30] A. Sudbery. Quaternionic analysis. *Mathematical Proceedings of the Cambridge Philosophical Society*, 85(2):199–225, 1979.
- [31] S. De Leo and P. P. Rotelli. Quaternionic analyticity. *Applied Mathematics Letters*, 16(7):1077–1081, October 2003.
- [32] L. Zetterberg and H. Brandstrom. Codes for combined phase and amplitude modulated signals in a four-dimensional space. *IEEE Transactions on Communications*, 25(9):943–950, 1977.
- [33] A. Cichocki and S. Amari. *Adaptive Blind Signal and Image Processing, Learning Algorithms and Applications*. Wiley, 2002.
- [34] Md. K. I. Molla, T. Tanaka, T. M. Rutkowski, and A. Cichocki. Separation of EOG artifacts from EEG signals using bivariate EMD. In *ICASSP 2010*, pages 562–565, 2010.

- [35] 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 *International Enformatika Conference*, pages 22–26, 2005.
- [36] P. S. Kumar, R. Arumuganathan, K. Sivakumar, and C. Vimal. An adaptive method to remove ocular artifacts from EEG signals using wavelet transform. *Journal of Applied Sciences Research*, 5:711–745, 2009.
- [37] R. N. Vigário. Extraction of ocular artefacts from EEG using independent component analysis. *Electroencephalography and Clinical Neurophysiology*, 103(3):395–404, 1997.
- [38] A. Delorme, S. Makeig, and T. Sejnowski. Automatic artifact rejection for EEG data using high-order statistics and independent component analysis. In *International Workshop on ICA*, pages 457–462, 2001.
- [39] A. van den Bos. Complex gradient and Hessian. *IEE Proceedings of Vision, Image and Signal Processing*, 141(6):380–383, 1994.
- [40] S. Zhang and A. G. Constantinides. Lagrange programming neural networks. *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing*, 39(7):441–452, 1992.