Quaternion ICA From Second-Order Statistics

Javier Vía, *Member, IEEE*, Daniel P. Palomar, *Senior Member, IEEE*, Luis Vielva, and Ignacio Santamaría, *Senior Member, IEEE*

***Abstract*—This paper addresses the independent component analysis (ICA) of quaternion random vectors. In particular, we focus on the Gaussian case and therefore only consider the quaternion second-order statistics (SOS), which are given by the covariance matrix and three complementary covariance matrices. First, we derive the necessary and sufficient conditions for the identifiability of the quaternion ICA model, which are based on the definition of the *properness profile* of a quaternion random variable and more specifically on the concept of *rotationally equivalent properness profiles*. Second, we show that the maximum-likelihood (ML) approach to the quaternion ICA problem reduces to the approximated joint diagonalization (AJD) of the sample-mean estimates of the covariance and complementary covariance matrices. Unlike the complex case, these four matrices cannot be simultaneously diagonalized in general, and we have to resort to a particular AJD algorithm. The proposed technique, which can be seen as a quasi-Newton method, is based on the local approximation of the nonconvex ML-ICA cost function (a measure of the entropy loss due to the residual correlation among the estimated quaternion sources), and it provides a satisfactory solution of the quaternion ICA model. The performance of the proposed quaternion ML-ICA algorithm, as well as its relationship to the identifiability conditions, are illustrated by means of several numerical examples.**

***Index Terms*—Approximated joint diagonalization (AJD), blind source separation (BSS), independent component analysis (ICA), properness, properness-profile, propriety, quaternions, second-order circularity.**

I. Introduction

TRADITIONALLY, quaternion algebra [1], [2] has been extensively used in computer graphics [3] and aerospace applications [4], [5] due to its compact notation, moderate computational requirements and avoidance of singularities associated to 3×3 rotation matrices [3]. Moreover, the interest in quaternion signal processing has increased in the last years due to its applications in image processing [6]–[9], wind modeling [10], [11], and design (and processing) of space–time (and space–time-polarization [12]) block codes [13]–[17].

The increasing popularity of quaternion signal processing makes necessary the development of a statistical theory for quaternion random vectors, as well as the generalization of the classical multivariate statistical analysis techniques to the quaternion case. Thus, in [18] the authors have considered the quaternion extensions of principal component analysis (PCA), partial least squares (PLS), multiple linear regression (MLR), and canonical correlation analysis (CCA). However, the independent component analysis (ICA) [19] of quaternion random vectors has received limited attention [20], [21], even though it can be considered (together with PCA) as the most important multivariate statistical analysis technique.

This paper considers the ICA of quaternion random vectors. In particular, we focus on the case of Gaussian data and therefore only consider the second-order statistics (SOS) of the quaternion random vectors. The main goal of this paper consists in establishing the conditions for the identifiability of the ICA model, as well as in the derivation of a practical quaternion ICA algorithm, which not only will be a valuable tool for the statistical analysis of quaternion random vectors, but also will find direct application in some blind source separation (BSS) problems. As an example, in Section V, the proposed ICA algorithm is applied to the problem of blind decoding (or channel estimation) in multiuser systems based on the Alamouti code [22], [23].

A. Main Contributions of the Paper

After a brief review of quaternion algebra in Section II, the main contributions of this paper are presented in Sections III and IV. In particular, Section III introduces the necessary and sufficient conditions for the identifiability, from SOS, of the quaternion ICA model. The SOS of a quaternion random vector are given by the covariance matrix and three complementary covariance matrices (the cross-covariance between the quaternion vector and its involutions) [18], [24]. Therefore, the ICA problem amounts to finding the separation matrix diagonalizing these four matrices.

The analysis in Section III shows that the *properness profiles* play a key role in the identifiability analysis. The properness profile is a three-dimensional pure quaternion vector, which can be seen as the quaternion counterpart of the (scalar) circularity coefficients [25]–[27] of complex random vectors. Thus, we show that the quaternion ICA model is unambiguously identifiable up to the trivial ICA ambiguities (permutations and quaternion scale factors) and a set of arbitrary linear mixtures affecting those sources with *rotationally equivalent* properness profiles, i.e., properness profiles related by a three-dimensional rotation.

Manuscript received May 08, 2010; revised September 10, 2010 and November 29, 2010; accepted December 07, 2010. Date of publication December 20, 2010; date of current version March 09, 2011. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Visa Koivunen. This work was supported by the Spanish Government, Ministerio de Ciencia e Innovación (MICINN), under projects COMONSENS (CSD2008-00010, CONSOLIDER-INGENIO 2010 Program) and COSIMA (TEC2010-19545-C04-03) and by the Hong Kong RGC 618709 research grant. The material in this paper was presented in part at the 6th IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM), Israel, October 4–7, 2010.

J. Vía, L. Vielva, and I. Santamaría are with the Department of Communications Engineering, University of Cantabria, 39005 Santander, Spain (e-mail: jvia@gtas.dicom.unican.es; luis@gtas.dicom.unican.es; nacho@gtas.dicom.unican.es).

D. P. Palomar is with the Department of Electronic and Computer Engineering , Hong Kong University of Science and Technology, Kowloon, Hong Kong (e-mail: palomar@ust.hk).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSP.2010.2101065

In Section IV, we propose a practical quaternion ICA algorithm based on the maximum-likelihood (ML) approach for quaternion Gaussian data. Specifically, we show that the ML-ICA problem reduces to an approximated joint diagonalization (AJD) problem [28]–[34], whose cost function is a measure of the "mutual information" among several quaternion random variables. The proposed quaternion ML-ICA algorithm is based on the local quadratic approximation of this nonconvex cost function, and therefore it can be seen as a quasi-Newton method [35]. Interestingly, although the nonconvexity of the cost function could result in the existence of local minima, the proposed algorithm converges very fast to a satisfactory solution of the quaternion ICA problem, which is illustrated in Section V by means of several simulation examples.

B. Relationship With Previous Works

ICA has attracted a lot of attention in the last two decades [19]. However, the rigorous analysis of the complex ICA model is relatively recent [27], [36]–[38]. In particular, in [27] the authors proved that the complex ICA model can be unambiguously identified from the SOS, provided that the sources have different (scalar) improperness degrees, which are measured by the circularity coefficients (or canonical correlations [25], [26]). This paper can be seen as an extension of these results to the case of quaternion sources. However, we must note that the generalization is far from trivial. For instance, the identifiability conditions do not only consider the improperness degree, but also its profile, which is indicated by the properness profile. Moreover, unlike the complex case, there does not exist a *strong uncorrelating transform* [27] for quaternion vectors, and therefore the quaternion ICA problem cannot be solved in closed form.

To our best knowledge, the only previous works considering the separation of quaternionic sources are [20], [21], and [39]. In particular, in [20] the authors presented the quaternionic extension of the Infomax algorithm [40], whereas they proposed a supervised technique in [39] based on the application of a multidimensional multilayer perceptron. Finally, in [21], the authors consider a very particular convolutive mixture model based on biquaternions and para-unitary matrices, which in the case of instantaneous mixtures reduces to PCA.

Finally, AJD techniques have been used for ICA and BSS applications [28]–[34]. However, most AJD algorithms consider real data (although the complex extension is straightforward), and they are based on the minimization of the Frobenius norm of the off-diagonal matrix elements. The quaternion ML-ICA algorithm presented in this paper can be seen as an AJD method for quaternion matrices with the following particularities. First, it must jointly diagonalize (only) four quaternion matrices, three of which are generally not positive semidefinite. Second, the proposed cost function is a Kullback–Leibler (KL) divergence [41], which naturally appears in the ML formulation of the quaternion ICA problem and can be seen as a measure of the entropy loss due to the correlation among the estimated latent variables. This measure differs from the KL-based cost function considered in [28] and [29] for the case of nonstationary sources, where the authors minimize a sum of KL divergences involving positive semidefinite matrices. Thus, although related to previous AJD techniques, the proposed quaternion ML-ICA algorithm can be seen as a new AJD method particularly suited for the quaternion ICA case and not as a simple extension of previous AJD approaches to the case of quaternion matrices.

C. Notation

Throughout this paper, we will use boldfaced uppercase letters to denote matrices, boldfaced lowercase letters for column vectors, and lightfaced lowercase letters for scalar quantities. Superscripts $(\cdot)^*$, $(\cdot)^{\mathrm{T}}$, and $(\cdot)^{\mathrm{H}}$ denote quaternion (or complex) conjugate, transpose, and Hermitian (i.e., transpose and quaternion conjugate), respectively. The notation $\mathbf{A} \in \mathbb{R}^{n\times m}$ (respectively $\mathbf{A} \in \mathbb{C}^{n\times m}$ or $\mathbf{A} \in \mathbb{H}^{n\times m}$) means that $\mathbf{A}$ is a real (respectively complex or quaternion) $n \times m$ matrix. $\Re(\mathbf{A})$, $\mathrm{Tr}(\mathbf{A})$ and $|\mathbf{A}|$ denote the real part, trace, and determinant of matrix $\mathbf{A}$. The diagonal matrix with vector $\mathbf{a}$ along its diagonal is denoted as $\mathrm{diag}(\mathbf{a})$, and $\mathrm{vec}(\mathbf{A})$ is the columnwise vectorized version of matrix $\mathbf{A}$. The Kronecker product is denoted by $\otimes$, $\mathbf{I}_n$ is the identity matrix of dimension n, and $\mathbf{0}_{n\times m}$ is the $n \times m$ zero matrix. Finally, E is the expectation operator and, in general, $\mathbf{R}_{\mathbf{a},\mathbf{b}}$ is the cross-correlation matrix for vectors $\mathbf{a}$ and $\mathbf{b}$, i.e., $\mathbf{R}_{\mathbf{a},\mathbf{b}} = E\mathbf{a}\mathbf{b}^{\mathrm{H}}$.

II. Preliminaries

A. Quaternion Algebra

Here, the basic concepts on quaternion algebra are briefly reviewed. For an advanced reading on quaternions, we refer to [2], as well as to [42] and [43], for several important results on matrices of quaternions.

Quaternions are four-dimensional hypercomplex numbers invented by Hamilton [1]. A quaternion $x \in \mathbb{H}$ is defined as

$$x = r_1 + ir_i + jr_j + kr_k \tag{1}$$

where r_1, r_i, r_j, r_k are four real numbers and the imaginary units (i, j, k) satisfy

$$i^2 = j^2 = k^2 = ijk = -1 \tag{2}$$

which also implies

$$ij = k = -ji \tag{3}$$
$$jk = i = -kj \tag{4}$$
$$ki = j = -ik. \tag{5}$$

Quaternions form a skew field $\mathbb{H}$ [2], which means that they satisfy the axioms of a field except for the commutative law of the product, i.e., for $x, y \in \mathbb{H}$, $xy \neq yx$ in general. The conjugate of a quaternion x is $x^* = r_1 - ir_i - jr_j - kr_k$, and the conjugate of the product satisfies $(xy)^* = y^*x^*$. The inner product between two quaternions $x, y \in \mathbb{H}$ is defined[1] as xy^*, and two quaternions are orthogonal if and only if (iff) their scalar product (the real part of the inner product) is zero. The quaternion norm is defined as $|x| = \sqrt{xx^*} = \sqrt{r_1^2 + r_i^2 + r_j^2 + r_k^2}$, and it is easy to check that $|xy| = |x||y|$. The inverse of a quaternion $x \neq 0$ is $x^{-1} = x^*/|x|^2$, and we say that $\eta \in \mathbb{H}$ is a pure

[1]Other definitions of the quaternion inner product are possible; see for instance [2].

unit quaternion iff $\eta^2 = -1$ (i.e., iff $|\eta| = 1$ and its real part is zero). Quaternions also admit the Euler representation

$$x = |x|e^{\eta\theta} = |x|\left(\cos\theta + \eta\sin\theta\right) \tag{6}$$

where $\eta = (ir_i + jr_j + kr_k)/\sqrt{r_i^2 + r_j^2 + r_k^2}$ is a pure unit quaternion and $\theta = \arccos(r_1/|x|) \in \mathbb{R}$ is the angle (or argument) of the quaternion. Let us now introduce the rotation and involution operations.

Definition 1 (Quaternion Rotation [2]): Consider a nonzero quaternion $a = |a|e^{\eta\theta} = |a|\left(\cos\theta + \eta\sin\theta\right)$, then

$$x^{(a)} = axa^{-1} \tag{7}$$

represents a three-dimensional rotation of the imaginary part of x. In particular, the vector $[r_i, r_j, r_k]^{\mathrm{T}}$ is rotated an angle 2θ in the pure imaginary plane orthogonal to η.

Definition 2 (Quaternion Involution [2]): An involution is a quaternion rotation of angle π. Specifically, the involution of a quaternion x over a pure unit quaternion η is

$$x^{(\eta)} = \eta x\eta^{-1} = \eta x\eta^* = -\eta x\eta. \tag{8}$$

With the above definitions and given two quaternions $a, b \in \mathbb{H}$, it is easy to check the following properties:

$$x^{(a)^*} = (x^*)^{(a)} \qquad \forall x \in \mathbb{H} \tag{9}$$

$$\left(x^{(a)}\right)^{(b)} = x^{(ba)} \qquad \forall x \in \mathbb{H} \tag{10}$$

$$(xy)^{(a)} = x^{(a)}y^{(a)} \qquad \forall x, y \in \mathbb{H} \tag{11}$$

$$ab = b^{(a)}a = ba^{(b^*)}. \tag{12}$$

Here, we must point out that the real representation in (1) can be easily generalized to other orthogonal bases.[2] Specifically, we will consider an orthogonal system $\{1, \eta, \eta', \eta''\}$ given by

$$\begin{bmatrix} 1 \\ \eta \\ \eta' \\ \eta'' \end{bmatrix} = \begin{bmatrix} 1 & \mathbf{0}_{1\times 3} \\ \mathbf{0}_{3\times 1} & \mathbf{Q} \end{bmatrix} \begin{bmatrix} 1 \\ i \\ j \\ k \end{bmatrix} \tag{13}$$

where $\mathbf{Q} \in \mathbb{R}^{3\times 3}$ is a rotation matrix, (i.e., $\mathbf{Q}^{\mathrm{T}}\mathbf{Q} = \mathbf{I}_3$ and $|\mathbf{Q}| = 1$), which implies

$$\eta^2 = \eta'^2 = \eta''^2 = \eta\eta'\eta'' = -1. \tag{14}$$

Thus, any quaternion can be represented as

$$x = r_1 + \eta r_\eta + \eta' r_{\eta'} + \eta'' r_{\eta''} \tag{15}$$

where $[r_\eta, r_{\eta'}, r_{\eta''}] = [r_i, r_j, r_k]\,\mathbf{Q}^T$.

B. Augmented Covariance Matrix

Analogously to the case of complex vectors, the statistical analysis of a quaternion random vector $\mathbf{x} \in \mathbb{H}^{n\times 1}$ can be directly based on its real representation $\mathbf{r_x} = \left[\mathbf{r}_1^{\mathrm{T}}, \mathbf{r}_\eta^{\mathrm{T}}, \mathbf{r}_{\eta'}^{\mathrm{T}}, \mathbf{r}_{\eta''}^{\mathrm{T}}\right]^{\mathrm{T}}$. However, we can get more insight on the statistical properties by introducing the augmented quaternion vector[3] $\bar{\mathbf{x}} = \left[\mathbf{x}^{\mathrm{T}}, \mathbf{x}^{(\eta)\mathrm{T}}, \mathbf{x}^{(\eta')\mathrm{T}}, \mathbf{x}^{(\eta'')\mathrm{T}}\right]^{\mathrm{T}}$, which is related to the real representation as [18], [24]

$$\bar{\mathbf{x}} = 2\mathbf{T}_n\mathbf{r_x} \tag{16}$$

where

$$\mathbf{T}_n = \frac{1}{2}\begin{bmatrix} +1 & +\eta & +\eta' & +\eta'' \\ +1 & +\eta & -\eta' & -\eta'' \\ +1 & -\eta & +\eta' & -\eta'' \\ +1 & -\eta & -\eta' & +\eta'' \end{bmatrix} \otimes \mathbf{I}_n \tag{17}$$

is a unitary quaternion operator, i.e., $\mathbf{T}_n^{\mathrm{H}}\mathbf{T}_n = \mathbf{I}_{4n}$.

The second-order statistical information of the quaternion vector is given by the augmented covariance matrix

$$\mathbf{R}_{\bar{\mathbf{x}},\bar{\mathbf{x}}} = \begin{bmatrix} \mathbf{R}_{\mathbf{x},\mathbf{x}} & \mathbf{R}_{\mathbf{x},\mathbf{x}^{(\eta)}} & \mathbf{R}_{\mathbf{x},\mathbf{x}^{(\eta')}} & \mathbf{R}_{\mathbf{x},\mathbf{x}^{(\eta'')}} \\ \mathbf{R}^{(\eta)}_{\mathbf{x},\mathbf{x}^{(\eta)}} & \mathbf{R}^{(\eta)}_{\mathbf{x},\mathbf{x}} & \mathbf{R}^{(\eta)}_{\mathbf{x},\mathbf{x}^{(\eta'')}} & \mathbf{R}^{(\eta)}_{\mathbf{x},\mathbf{x}^{(\eta')}} \\ \mathbf{R}^{(\eta')}_{\mathbf{x},\mathbf{x}^{(\eta')}} & \mathbf{R}^{(\eta')}_{\mathbf{x},\mathbf{x}^{(\eta'')}} & \mathbf{R}^{(\eta')}_{\mathbf{x},\mathbf{x}} & \mathbf{R}^{(\eta')}_{\mathbf{x},\mathbf{x}^{(\eta)}} \\ \mathbf{R}^{(\eta'')}_{\mathbf{x},\mathbf{x}^{(\eta'')}} & \mathbf{R}^{(\eta'')}_{\mathbf{x},\mathbf{x}^{(\eta')}} & \mathbf{R}^{(\eta'')}_{\mathbf{x},\mathbf{x}^{(\eta)}} & \mathbf{R}^{(\eta'')}_{\mathbf{x},\mathbf{x}} \end{bmatrix} \tag{18}$$

where we can readily identify the covariance matrix $\mathbf{R}_{\mathbf{x},\mathbf{x}} = E\mathbf{x}\mathbf{x}^{\mathrm{H}}$ and three complementary covariance matrices $\mathbf{R}_{\mathbf{x},\mathbf{x}^{(\eta)}} = E\mathbf{x}\mathbf{x}^{(\eta)\mathrm{H}}$, $\mathbf{R}_{\mathbf{x},\mathbf{x}^{(\eta')}} = E\mathbf{x}\mathbf{x}^{(\eta')\mathrm{H}}$ and $\mathbf{R}_{\mathbf{x},\mathbf{x}^{(\eta'')}} = E\mathbf{x}\mathbf{x}^{(\eta'')\mathrm{H}}$.

Here, we must point out that the structure of $\mathbf{R}_{\bar{\mathbf{x}},\bar{\mathbf{x}}}$, i.e., the location of zero blocks (complementary covariance matrices), is invariant to linear transformations of the form $\mathbf{u} = \mathbf{F}_1^{\mathrm{H}}\mathbf{x}$ (with $\mathbf{F}_1$ a quaternion matrix) [18]. Analogously to the complex case, this ensures that the properness of a quaternion random vector will be preserved by quaternion linear transformations.

Finally, the following lemmas show that, given the complementary covariance matrices for an orthogonal basis $\{\eta, \eta', \eta''\}$, we can obtain the complementary covariance matrix $\mathbf{R}_{\mathbf{x},\mathbf{x}^{(\nu)}}$ (for all pure unit quaternions ν) as a quaternion linear combination of $\mathbf{R}_{\mathbf{x},\mathbf{x}^{(\eta)}}$, $\mathbf{R}_{\mathbf{x},\mathbf{x}^{(\eta')}}$, and $\mathbf{R}_{\mathbf{x},\mathbf{x}^{(\eta'')}}$.

Lemma 1: Given a quaternion random vector $\mathbf{x} \in \mathbb{H}^{n\times 1}$ and two different orthogonal bases $\{\eta, \eta', \eta''\}$ and $\{\nu, \nu', \nu''\}$, the corresponding augmented quaternion vectors are related as

$$\begin{bmatrix} \mathbf{x} \\ \mathbf{x}^{(\nu)} \\ \mathbf{x}^{(\nu')} \\ \mathbf{x}^{(\nu'')} \end{bmatrix} = \mathbf{\Gamma} \begin{bmatrix} \mathbf{x} \\ \mathbf{x}^{(\eta)} \\ \mathbf{x}^{(\eta')} \\ \mathbf{x}^{(\eta'')} \end{bmatrix} \tag{19}$$

where $\mathbf{\Gamma}$ is a unitary quaternion operator given by

$$\mathbf{\Gamma} = \begin{bmatrix} 1 & \mathbf{0}_{1\times 3} \\ \mathbf{0}_{3\times 1} & \mathbf{\Lambda}_\nu\mathbf{Q}\mathbf{\Lambda}_\eta^H \end{bmatrix} \otimes \mathbf{I}_n \tag{20}$$

$\mathbf{\Lambda}_\nu = \mathrm{diag}\left([\nu, \nu', \nu'']^{\mathrm{T}}\right)$ and $\mathbf{\Lambda}_\eta = \mathrm{diag}\left([\eta, \eta', \eta'']^{\mathrm{T}}\right)$.

Proof: Let us consider the pure unit quaternion $\nu = \nu_\eta\eta + \nu_{\eta'}\eta' + \nu_{\eta''}\eta''$, where $[\nu_\eta, \nu_{\eta'}, \nu_{\eta''}]$ is the first row of $\mathbf{Q}$. Thus, the involution of $\mathbf{x}$ over ν is

$$\begin{aligned} \mathbf{x}^{(\nu)} &= \nu\mathbf{x}\nu^* = \nu\mathbf{x}\left(\nu_\eta\eta^* + \nu_{\eta'}\eta'^* + \nu_{\eta''}\eta''^*\right) \\ &= \nu_\eta\nu\eta^*\mathbf{x}^{(\eta)} + \nu_{\eta'}\nu\eta'^*\mathbf{x}^{(\eta')} + \nu_{\eta''}\nu\eta''^*\mathbf{x}^{(\eta'')}. \end{aligned} \tag{21}$$

[2]The choice of a particular basis can be motivated by the specific application. For instance, in image-processing applications [44], it is frequent to align the "gray line" in RGB color space with the direction $\eta = (i + j + k)/\sqrt{3}$.

[3]From now on, we will use the notation $\mathbf{A}^{(a)}$ to denote the elementwise rotation (or involution for pure quaternion a) of matrix $\mathbf{A}$.

Repeating this procedure for ν' and ν'', we obtain the mapping between the augmented quaternion vectors in the two different bases. ■

Lemma 2: The augmented covariance matrices in two different orthogonal bases are related as

$$\mathbf{R}_{\bar{\mathbf{x}},\bar{\mathbf{x}}}(\{1,\nu,\nu',\nu''\}) = \mathbf{\Gamma}\mathbf{R}_{\bar{\mathbf{x}},\bar{\mathbf{x}}}(\{1,\eta,\eta',\eta''\})\mathbf{\Gamma}^{\mathrm{H}} \tag{22}$$

where the expressions in parentheses make the bases explicit.

Proof: This is a direct consequence of Lemma 1 and the definition of the augmented covariance matrix. ■

C. Properness of Quaternion Vectors

Unlike the complex case[4] [45]–[48], there exist different kinds of quaternion properness [18], [49], [50], which also have different implications on the optimal linear processing of a quaternion random vector [18]. In this paper, we will focus on the strongest kind on properness.

Definition 3 ($\mathbb{Q}$-Properness): A quaternion random vector $\mathbf{x}$ is $\mathbb{Q}$-proper iff the three complementary covariance matrices $\mathbf{R}_{\mathbf{x},\mathbf{x}^{(\eta)}}$, $\mathbf{R}_{\mathbf{x},\mathbf{x}^{(\eta')}}$, and $\mathbf{R}_{\mathbf{x},\mathbf{x}^{(\eta'')}}$ vanish.

Here, we must note that, as a direct consequence of Lemma 2, a quaternion random vector $\mathbf{x}$ is $\mathbb{Q}$-proper iff $\mathbf{R}_{\mathbf{x},\mathbf{x}^{(\nu)}} = \mathbf{0}_{n\times n}$ for all pure unit quaternions ν. Furthermore, we will say that $\mathbf{x}$ is $\mathbb{Q}$-improper iff it is not $\mathbb{Q}$-proper.

Analogously to the complex case, the optimal linear processing of a quaternion vector is in general *full-widely* linear [18], i.e., we must simultaneously operate on the quaternion vector and its three involutions. However, in the case of *jointly* $\mathbb{Q}$-proper vectors, the optimal linear processing takes the form of conventional linear processing $\mathbf{u} = \mathbf{F}_1^H\mathbf{x}$, i.e., we do not need to operate on the vector involutions [18]. In other words, we cannot expect any gain from the widely linear processing of $\mathbb{Q}$-proper vectors.

Finally, in [18] the authors have proposed several measures of the (scalar) improperness degree of a quaternion random vector (for the different kinds of properness). In particular, the $\mathbb{Q}$-improperness measure is given by

$$\mathcal{P}_{\mathbb{Q}} = -\frac{1}{2}\ln|\mathbf{\Phi}_{\mathbb{Q}}| \tag{23}$$

where[5] $\mathbf{\Phi}_{\mathbb{Q}} = \mathbf{D}_{\mathbb{Q}}^{-1/2}\mathbf{R}_{\bar{\mathbf{x}},\bar{\mathbf{x}}}\mathbf{D}_{\mathbb{Q}}^{-1/2}$ is defined as the $\mathbb{Q}$-coherence matrix

$$\begin{aligned}\mathbf{D}_{\mathbb{Q}} &= \mathrm{blkdiag}_n(\mathbf{R}_{\bar{\mathbf{x}},\bar{\mathbf{x}}})\\ &= \begin{bmatrix}\mathbf{R}_{\mathbf{x},\mathbf{x}} & \mathbf{0}_{n\times n} & \mathbf{0}_{n\times n} & \mathbf{0}_{n\times n}\\ \mathbf{0}_{n\times n} & \mathbf{R}_{\mathbf{x},\mathbf{x}}^{(\eta)} & \mathbf{0}_{n\times n} & \mathbf{0}_{n\times n}\\ \mathbf{0}_{n\times n} & \mathbf{0}_{n\times n} & \mathbf{R}_{\mathbf{x},\mathbf{x}}^{(\eta')} & \mathbf{0}_{n\times n}\\ \mathbf{0}_{n\times n} & \mathbf{0}_{n\times n} & \mathbf{0}_{n\times n} & \mathbf{R}_{\mathbf{x},\mathbf{x}}^{(\eta'')}\end{bmatrix}\end{aligned} \tag{24}$$

and $\mathrm{blkdiag}_n(\mathbf{A})$ denotes the block-diagonal matrix obtained from the $n\times n$ blocks in the diagonal of $\mathbf{A}$.

[4]A complex vector $\mathbf{x} \in \mathbb{C}^{n\times 1}$ is proper iff the complementary covariance matrix $\mathbf{R}_{\mathbf{x},\mathbf{x}^*} = E\mathbf{x}\mathbf{x}^T$ is zero.

[5]In this paper, $\mathbf{A}^{1/2}$ (respectively $\mathbf{A}^{-1/2}$) denotes the Hermitian square root of the Hermitian matrix $\mathbf{A}$ (respectively $\mathbf{A}^{-1}$).

III. ICA Problem and Identifiability Conditions

In this section, we present the general problem of ICA of quaternion vectors. More importantly, we derive the necessary and sufficient conditions for the blind identifiability of the ICA model in the case of jointly Gaussian sources or, equivalently, in the case of ICA methods exclusively based on SOS.

A. Problem Formulation

Consider a quaternion random vector $\mathbf{s} \in \mathbb{H}^{m\times 1}$ representing m source signals, which are mixed by a full-column rank mixing matrix $\mathbf{A} \in \mathbb{H}^{n\times m}$ $(n \geq m)$. Thus, we have the model

$$\mathbf{x} = \mathbf{A}\mathbf{s} \tag{25}$$

where $\mathbf{x} \in \mathbb{H}^{n\times 1}$ is a quaternion random vector representing the available observations. Here, we must point out that there exist other possible mixture models [20], such as

$$\mathbf{x} = \mathbf{A}\overleftarrow{\cdot}\mathbf{s} \tag{26}$$

where the notation $\overleftarrow{\cdot}$ means that we are using the right product. That is, the kth element of $\mathbf{x}$ is given by

$$x_k = \sum_l s_l a_{k,l} \tag{27}$$

where $a_{k,l}$ is the element in the kth row and lth column of $\mathbf{A}$ and s_l is the lth entry of $\mathbf{s}$. However, in this case we can make use of the property $(ab)^* = b^*a^*$ to write $\mathbf{x}^* = \mathbf{A}^*\mathbf{s}^*$. Thus, the mixture model based on right multiplications can be reformulated as (25) by simply taking the quaternion conjugate.

ICA is based on the crucial assumption that the elements of $\mathbf{s}$ are independent, which can be exploited to recover the (unknown) sources $\mathbf{s}$ and the (unknown) mixing matrix $\mathbf{A}$, from the observations $\mathbf{x}$. In general, ICA can make use of all the statistical information provided by the observations. However, in this paper we focus on methods solely based on SOS, which is justified by the fact that, unlike higher-order statistics, SOS can be accurately estimated from a moderate number of observations. Moreover, SOS provide all the statistical information in the fundamental case of Gaussian sources.

Before proceeding, we must take into account that, analogously to the well-known real and complex cases [19], [27], there exist two trivial ambiguities inherent to the quaternion ICA model. These ambiguities consist in a quaternion scale factor and a permutation of the sources and columns of the mixing matrix. Thus, as a direct consequence of the scale ambiguity, we can consider that the sources $\mathbf{s}$ (and the columns of $\mathbf{A}$) are normalized to satisfy $\mathbf{R}_{\mathbf{s},\mathbf{s}} = E\mathbf{s}\mathbf{s}^{\mathrm{H}} = \mathbf{I}_m$, and therefore all the second-order statistical information of the sources is given by the augmented covariance matrix

$$\mathbf{R}_{\bar{\mathbf{s}},\bar{\mathbf{s}}} = E\bar{\mathbf{s}}\bar{\mathbf{s}}^{\mathrm{H}} = \begin{bmatrix}\mathbf{I}_m & \mathbf{\Lambda}_\eta & \mathbf{\Lambda}_{\eta'} & \mathbf{\Lambda}_{\eta''}\\ \mathbf{\Lambda}_\eta^{(\eta)} & \mathbf{I}_m & \mathbf{\Lambda}_{\eta''}^{(\eta)} & \mathbf{\Lambda}_{\eta'}^{(\eta)}\\ \mathbf{\Lambda}_{\eta'}^{(\eta')} & \mathbf{\Lambda}_{\eta''}^{(\eta')} & \mathbf{I}_m & \mathbf{\Lambda}_{\eta}^{(\eta')}\\ \mathbf{\Lambda}_{\eta''}^{(\eta'')} & \mathbf{\Lambda}_{\eta'}^{(\eta'')} & \mathbf{\Lambda}_{\eta}^{(\eta'')} & \mathbf{I}_m\end{bmatrix} \tag{28}$$

where $\mathbf{\Lambda}_\eta$, $\mathbf{\Lambda}_{\eta'}$, $\mathbf{\Lambda}_{\eta''}$ are the diagonal (due to the independence of the elements in $\mathbf{s}$) complementary covariance matrices of the sources $\mathbf{s}$.

With this assumption and taking into account the property $(xy)^{(a)} = x^{(a)}y^{(a)}$, it is easy to see that the SOS of the observations are given by

$$\mathbf{R}_{\bar{\mathbf{x}},\bar{\mathbf{x}}} = E\bar{\mathbf{x}}\bar{\mathbf{x}}^{\mathrm{H}} = \bar{\mathbf{A}}\mathbf{R}_{\bar{\mathbf{s}},\bar{\mathbf{s}}}\bar{\mathbf{A}}^{\mathrm{H}} \tag{29}$$

where

$$\bar{\mathbf{A}} = \begin{bmatrix} \mathbf{A} & \mathbf{0}_{n\times m} & \mathbf{0}_{n\times m} & \mathbf{0}_{n\times m} \\ \mathbf{0}_{n\times m} & \mathbf{A}^{(\eta)} & \mathbf{0}_{n\times m} & \mathbf{0}_{n\times m} \\ \mathbf{0}_{n\times m} & \mathbf{0}_{n\times m} & \mathbf{A}^{(\eta')} & \mathbf{0}_{n\times m} \\ \mathbf{0}_{n\times m} & \mathbf{0}_{n\times m} & \mathbf{0}_{n\times m} & \mathbf{A}^{(\eta'')} \end{bmatrix}. \tag{30}$$

Thus, the ICA problem amounts to finding the mixing matrix $\mathbf{A}$ and the diagonal complementary covariance matrices of the sources ($\mathbf{\Lambda}_\eta$, $\mathbf{\Lambda}_{\eta'}$ and $\mathbf{\Lambda}_{\eta''}$) satisfying

$$\mathbf{A}\mathbf{A}^{\mathrm{H}} = \mathbf{R}_{\mathbf{x},\mathbf{x}} \tag{31}$$

$$\mathbf{A}\mathbf{\Lambda}_\eta\mathbf{A}^{(\eta)^{\mathrm{H}}} = \mathbf{R}_{\mathbf{x},\mathbf{x}^{(\eta)}} \tag{32}$$

$$\mathbf{A}\mathbf{\Lambda}_{\eta'}\mathbf{A}^{(\eta')^{\mathrm{H}}} = \mathbf{R}_{\mathbf{x},\mathbf{x}^{(\eta')}} \tag{33}$$

$$\mathbf{A}\mathbf{\Lambda}_{\eta''}\mathbf{A}^{(\eta'')^{\mathrm{H}}} = \mathbf{R}_{\mathbf{x},\mathbf{x}^{(\eta'')}}. \tag{34}$$

Obviously, once the mixing matrix $\mathbf{A}$ has been obtained, the independent sources can be recovered by means of its pseudoinverse. However, we must note that the solution of the ICA problem may not be unique, i.e., there could exist ambiguities (apart from the trivial ones) implying a residual mixture of the sources. In Section III-B, we establish the conditions for the uniqueness (up to quaternion scale factors and permutations) of the ICA solution.

B. Identifiability Conditions

Analogously to the case of complex vectors [27], the identifiability of the ICA model from SOS relies on the improperness of the sources. In order to analyze the identifiability of the quaternion ICA model, we start by introducing the following definitions.

Definition 4 (Properness Profile): The properness profile of a quaternion random variable s is defined as

$$\boldsymbol{\psi}_s = \begin{bmatrix} \psi_{s,\eta} \\ \psi_{s,\eta'} \\ \psi_{s,\eta''} \end{bmatrix} = \begin{bmatrix} \lambda_{s,\eta}\eta \\ \lambda_{s,\eta'}\eta' \\ \lambda_{s,\eta''}\eta'' \end{bmatrix} = \frac{1}{E|s|^2}\begin{bmatrix} Es\eta s^* \\ Es\eta' s^* \\ Es\eta'' s^* \end{bmatrix} \tag{35}$$

where, for a pure unit quaternion ν, $\lambda_{s,\nu} = Ess^{*(\nu)}/E|s|^2$ is the (normalized) complementary variance.

Definition 5 (Rotationally Equivalent Properness Profiles): The properness profiles of two quaternion random variables s_1, s_2 are rotationally equivalent iff they are related by a quaternion rotation, i.e., iff there exists a quaternion a such that

$$\boldsymbol{\psi}_{s_2} = \boldsymbol{\psi}_{s_1}^{(a)}. \tag{36}$$

The properness profile can be seen as the quaternion analog of the (scalar) circularity quotient [51], [52] of a complex random variable, whose polar representation provides the circularity coefficient (absolute value) [27] and circularity angle (argument). The following lemmas present three key properties of the properness profile.

Lemma 3: The properness profile is a pure quaternion vector, i.e., its real part is zero.

Proof: From the definition, it is easy to check that $\lambda_{s,\nu}$ satisfies $\lambda_{s,\nu}^* = \lambda_{s,\nu}^{(\nu)}$, which implies that it is orthogonal to ν. Thus, $\psi_{s,\nu} = \lambda_{s,\nu}\nu$ is a pure quaternion. ■

Lemma 4: The properness profile in a different orthogonal basis $\{\nu, \nu', \nu''\}$ can be obtained as

$$\boldsymbol{\psi}_s\left(\{\nu,\nu',\nu''\}\right) = \mathbf{Q}\boldsymbol{\psi}_s\left(\{\eta,\eta',\eta''\}\right) \tag{37}$$

where the expressions in parentheses make the bases explicit and $\mathbf{Q} \in \mathbb{R}^{3\times 3}$ is the real rotation matrix for the change of basis $[\nu,\nu',\nu''] = [\eta,\eta',\eta'']\mathbf{Q}^{\mathrm{T}}$.

Proof: Taking into account the relationship between the properness profile and the augmented covariance matrix, this can be seen as a direct consequence of Lemma 2. ■

Lemma 5: Given a quaternion random variable s_1 and a transformation $s_2 = as_1$, with $a \in \mathbb{H}$, we have

$$\boldsymbol{\psi}_{s_2} = \boldsymbol{\psi}_{s_1}^{(a)} \tag{38}$$

i.e., the properness profiles of s_1 and s_2 are rotationally equivalent.

Proof: For all pure unit quaternions ν, we have

$$\begin{aligned}\psi_{s_2,\nu} &= \frac{1}{E|s_2|^2}Es_2s_2^{*(\nu)}\nu \\ &= \frac{1}{|a|^2E|s_1|^2}Eas_1\nu s_1^*a^*\nu^*\nu \\ &= \frac{1}{|a|^2E|s_1|^2}aEs_1s_1^{*(\nu)}\nu a^* = \psi_{s_1,\nu}^{(a)}.\end{aligned} \tag{39}$$

■

Lemma 3 allows us to see the elements of the properness profile as points in a three-dimensional space. On the other hand, from Lemma 4 it is easy to prove that if two properness profiles are rotationally equivalent in some orthogonal basis, then they are rotationally equivalent in all the bases, i.e., the definition of rotationally equivalent properness profiles is independent of the orthogonal basis $\{\eta,\eta',\eta''\}$. Finally, Lemma 5 provides a clear example of a transformation (one of the trivial ambiguities in the ICA model) resulting in a rotation of the properness profile. Now, we can state the quaternion ICA identifiability conditions.

Theorem 1 (ICA Identifiability): Given the ICA model $\mathbf{x} = \mathbf{A}\mathbf{s}$, with independent entries in $\mathbf{s}$ and full-column rank $\mathbf{A}$, the sources $\mathbf{s}$ and the mixing matrix $\mathbf{A}$ can be recovered from the SOS of the observations up to the following ambiguities:

- a permutation and quaternion scale factor;
- a residual quaternion linear mixture affecting the sources with rotationally equivalent properness profiles.

Proof: See Appendix A. ■

Corollary 1: For complex sources in the plane $\{1,\eta\}$, Theorem 1 results in the conditions for the identifiability (from SOS) of the complex ICA model [27].

Proof: The properness profile of a complex source $s = r_1 + \eta r_\eta$ is given by

$$\boldsymbol{\psi}_s = [\eta, c_s\eta', c_s\eta'']^{\mathrm{T}} \tag{40}$$

where $c_s = Es^2/E|s|^2$ is the circularity quotient [51], [52] and $|c_s|$ is the circularity coefficient defined in [27]. Thus, it is easy to see that the properness profiles of two complex sources s_1, s_2 in the plane $\{1,\eta\}$ are rotationally equivalent iff they have the same circularity coefficient, i.e., $|c_{s_1}| = |c_{s_2}|$. In particular, for two complex sources s_1, s_2 with identical circularity coefficients and defining $a = (c_{s_2} c_{s_1}^*)^{1/2}$, we have $\boldsymbol{\psi}_{s_2} = \boldsymbol{\psi}_{s_1}^{(a)}$. ■

Theorem 1 shows that the properness profiles play a crucial role in the identifiability of the quaternion ICA model. As previously pointed out, the properness profile can be seen as the quaternion counterpart (including *phase information*) of the circularity coefficients [27] of complex random variables. However, we must remark three important differences with the complex case.

Remark 1: In the complex case, the ICA identifiability conditions can be reformulated in terms of the (scalar) improperness degree of the sources. That is, we can say that the complex ICA model is identifiable up to the trivial ambiguities and a complex linear mixture affecting those sources with identical improperness degrees, which are measured by the circularity coefficients (or canonical correlations) [25]–[27]. However, as a direct consequence of Theorem 1, two quaternion sources with the same (scalar) improperness degree [18] can be unambiguously recovered if their properness profiles are not rotationally equivalent. In other words, the quaternion ICA identifiability conditions do not only consider the (scalar) improperness degree of the sources, but also its profile, which is measured by the properness profile. As a matter of fact, it can be easily proven that a sufficient (but not necessary) condition for the identifiability of the quaternion ICA model consists in having quaternion sources with different (scalar) improperness degrees. As an example, consider two sources $s_1, s_2 \in \mathbb{H}$ with properness profiles $\boldsymbol{\Psi}_{s_1} = [0.5\eta, 0, 0]^{\mathrm{T}}$ and $\boldsymbol{\Psi}_{s_2} = [0, 0.5\eta', 0]^{\mathrm{T}}$, which are not rotationally equivalent. Then, it is clear that s_1 and s_2 can be unambiguously recovered, even though they have identical (scalar) improperness degrees.

Remark 2: As we will see in Section IV, in the general quaternion case, there does not exist a *strong uncorrelating transform* [27], i.e., given the covariance and complementary covariance matrices of a quaternion random vector $\mathbf{x}$, we cannot always find a solution $\{\mathbf{A}, \boldsymbol{\Lambda}_\eta, \boldsymbol{\Lambda}_{\eta'}, \boldsymbol{\Lambda}_{\eta''}\}$ of the system in (31)–(34). As an example, consider a quaternion random vector $\mathbf{x} \in \mathbb{H}^{2\times 1}$ with SOS

$$\begin{aligned}
\mathbf{R}_{\mathbf{x},\mathbf{x}} &= \mathbf{I}_2 \\
\mathbf{R}_{\mathbf{x},\mathbf{x}^{(\eta)}} &= \begin{bmatrix} 0.1 & 0 \\ 0 & 0.2 \end{bmatrix} \\
\mathbf{R}_{\mathbf{x},\mathbf{x}^{(\eta')}} &= \begin{bmatrix} 0.3 & -0.1 \\ -0.1 & 0.3 \end{bmatrix} \\
\mathbf{R}_{\mathbf{x},\mathbf{x}^{(\eta'')}} &= \begin{bmatrix} 0.3 & -0.1 \\ -0.1 & 0.3 \end{bmatrix}.
\end{aligned} \tag{41}$$

In this case, it is clear that we cannot simultaneously diagonalize the covariance and complementary covariance matrices. In other words, unlike the real and complex cases, not all the quaternion random vectors $\mathbf{x} \in \mathbb{H}^{n\times 1}$ can be represented as a quaternion linear combination of $m \leq n$ independent quaternion sources $\mathbf{s} \in \mathbb{H}^{m\times 1}$. However, in this paper we focus on quaternion random vectors satisfying the ICA model $\mathbf{x} = \mathbf{A}\mathbf{s}$.

Remark 3: In the complex case, the (real and scalar) circularity coefficients (which together form the circularity spectrum) are a maximal invariant of the complex random vector under the group of invertible complex-linear transformations [53]–[55]. Unfortunately, in the quaternion case, we do not have a similar result in terms of the properness profiles. If we focus on a quaternion random variable, it is easy to see that the properness profile is a maximal invariant under the group of non-null real scale factors. Furthermore, considering non-null quaternion scale factors and taking into account Lemma 5, we could introduce a rotation to obtain a maximal invariant given by a *rotated* properness profile with $\psi_{s,\eta}$ in the η-axis and $\psi_{s,\eta'}$ in the $\{\eta, \eta'\}$ plane. However, the problem is much more involved when we consider quaternion random vectors due to the fact that, in general, we cannot simultaneously diagonalize the sample-mean estimates of the covariance and complementary covariance matrices. Thus, only in the case of exactly jointly diagonalizable estimates of $\mathbf{R}_{\mathbf{x},\mathbf{x}}$, $\mathbf{R}_{\mathbf{x},\mathbf{x}^{(\eta)}}$, $\mathbf{R}_{\mathbf{x},\mathbf{x}^{(\eta')}}$, and $\mathbf{R}_{\mathbf{x},\mathbf{x}^{(\eta'')}}$, we could obtain a generalization of the circularity spectrum based on the *rotated* properness profiles of the *latent variables* $\mathbf{s}$.

IV. ML-ICA Algorithm for Quaternion Gaussian Vectors

In Section III, we have presented the necessary and sufficient conditions for the identifiability of the quaternion ICA model from SOS. Nevertheless, the derivation of practical algorithms from the proof of Theorem 1 is far from trivial. In this section, we propose an efficient quaternion ICA algorithm based on the ML approach for Gaussian vectors. The proposed technique consists in the AJD [28]–[34] of the covariance and complementary covariance matrices of the observations, which is achieved by means of a quasi-Newton method [35]. Thus, the proposed algorithm can be seen as the first AJD method for quaternion matrices.

A. Maximum-Likelihood Quaternion ICA

Let us consider T vector observations $\mathbf{x}[t] = \mathbf{A}\mathbf{s}[t]$ $(t = 0, \ldots, T-1)$ of the ICA model, where the $n \times m$ mixing matrix $\mathbf{A}$ is assumed to be full-column rank and the sources $\mathbf{s}[t]$ are i.i.d. zero-mean quaternion Gaussian vectors with independent elements. Let us also assume a nonsingular augmented covariance matrix $\mathbf{R}_{\bar{\mathbf{s}},\bar{\mathbf{s}}}$ and an augmented sample covariance matrix

$$\hat{\mathbf{R}}_{\bar{\mathbf{x}},\bar{\mathbf{x}}} = \frac{1}{T}\sum_{t=0}^{T-1} \bar{\mathbf{x}}[t]\bar{\mathbf{x}}^{\mathrm{H}}[t] \tag{42}$$

with rank $4m$, which obviously implies $T \geq 4m$. Under these assumptions and taking into account that we are considering a noiseless case, the signal subspace can be exactly recovered from the observations.[6] Therefore, the data can be projected onto the signal subspace, which allows us to reduce the ICA problem to the case of a square and nonsingular mixing matrix $\mathbf{A} \in \mathbb{H}^{m\times m}$. Thus, with a slight abuse of notation and assuming from now on that $\mathbf{A}$ is square, the probability density function (pdf) of the projected observations can be written as [18]

$$(\bar{\mathbf{x}}[0], \ldots, \bar{\mathbf{x}}[T-1]) = \prod_{t=0}^{T-1} \frac{\exp\left(-\frac{1}{2}\bar{\mathbf{x}}^{\mathrm{H}}[t]\mathbf{R}_{\bar{\mathbf{x}},\bar{\mathbf{x}}}^{-1}\bar{\mathbf{x}}[t]\right)}{\left(\frac{\pi}{2}\right)^{2m}|\mathbf{R}_{\bar{\mathbf{x}},\bar{\mathbf{x}}}|^{1/2}} \tag{43}$$

where $\mathbf{R}_{\bar{\mathbf{x}},\bar{\mathbf{x}}} = \bar{\mathbf{A}}\mathbf{R}_{\bar{\mathbf{s}},\bar{\mathbf{s}}}\bar{\mathbf{A}}^{\mathrm{H}}$ and $\bar{\mathbf{A}}$ has been defined in (30).

[6]In particular, the eigenvalue decomposition (EVD) [42], [43] of the estimated covariance matrix yields $\hat{\mathbf{R}}_{\mathbf{x},\mathbf{x}} = 1/T\sum_{t=0}^{T-1}\mathbf{x}[t]\mathbf{x}^{\mathrm{H}}[t] = \mathbf{U}\boldsymbol{\Sigma}\mathbf{U}^{\mathrm{H}}$, where $\mathbf{U} \in \mathbb{H}^{n\times m}$ is an orthogonal basis for the signal subspace.

Taking now the logarithm of the pdf and defining $\mathcal{R}$ and $\mathcal{D}$ as the sets of matrices with the structures in (28) and (30), the ML estimation problem is[7]

$$\underset{\substack{\bar{\mathbf{A}}\in\mathcal{D},\mathbf{R}_{\bar{\mathbf{s}},\bar{\mathbf{s}}}\in\mathcal{R}\\ \mathbf{R}_{\bar{\mathbf{x}},\bar{\mathbf{x}}}=\bar{\mathbf{A}}\mathbf{R}_{\bar{\mathbf{s}},\bar{\mathbf{s}}}\bar{\mathbf{A}}^{\mathrm{H}}}}{\text{maximize}} \; -\ln|\mathbf{R}_{\bar{\mathbf{x}},\bar{\mathbf{x}}}| - \Re\left[\mathrm{Tr}\left(\mathbf{R}_{\bar{\mathbf{x}},\bar{\mathbf{x}}}^{-1}\hat{\mathbf{R}}_{\bar{\mathbf{x}},\bar{\mathbf{x}}}\right)\right]. \tag{44}$$

Interestingly, the above problem can be seen as that of minimizing the KL divergence between two zero-mean quaternion Gaussian distributions with augmented covariance matrices $\hat{\mathbf{R}}_{\bar{\mathbf{x}},\bar{\mathbf{x}}}$ and $\mathbf{R}_{\bar{\mathbf{x}},\bar{\mathbf{x}}}$. Specifically, the ML-ICA problem can be rewritten as

$$\underset{\substack{\bar{\mathbf{A}}\in\mathcal{D},\mathbf{R}_{\bar{\mathbf{s}},\bar{\mathbf{s}}}\in\mathcal{R}\\ \mathbf{R}_{\bar{\mathbf{x}},\bar{\mathbf{x}}}=\bar{\mathbf{A}}\mathbf{R}_{\bar{\mathbf{s}},\bar{\mathbf{s}}}\bar{\mathbf{A}}^{\mathrm{H}}}}{\text{minimize}} \; D_{\mathrm{KL}}\left(\hat{\mathbf{R}}_{\bar{\mathbf{x}},\bar{\mathbf{x}}}\|\mathbf{R}_{\bar{\mathbf{x}},\bar{\mathbf{x}}}\right) \tag{45}$$

where the KL divergence is [18]

$$D_{\mathrm{KL}}\left(\hat{\mathbf{R}}_{\bar{\mathbf{x}},\bar{\mathbf{x}}}\|\mathbf{R}_{\bar{\mathbf{x}},\bar{\mathbf{x}}}\right) = \frac{1}{2}\ln\frac{|\mathbf{R}_{\bar{\mathbf{x}},\bar{\mathbf{x}}}|}{|\hat{\mathbf{R}}_{\bar{\mathbf{x}},\bar{\mathbf{x}}}|} + \frac{1}{2}\Re\left[\mathrm{Tr}\left(\mathbf{R}_{\bar{\mathbf{x}},\bar{\mathbf{x}}}^{-1}\hat{\mathbf{R}}_{\bar{\mathbf{x}},\bar{\mathbf{x}}}\right)\right] - 2m. \tag{46}$$

Here, we can exploit a well-known property of the KL divergence, which consists in its invariance under linear transformations [29], [41], [56]. In particular, for all invertible matrices $\bar{\mathbf{W}} \in \mathcal{D}$ we have

$$D_{\mathrm{KL}}\left(\bar{\mathbf{W}}\hat{\mathbf{R}}_{\bar{\mathbf{x}},\bar{\mathbf{x}}}\bar{\mathbf{W}}^{\mathrm{H}}\|\bar{\mathbf{W}}\mathbf{R}_{\bar{\mathbf{x}},\bar{\mathbf{x}}}\bar{\mathbf{W}}^{\mathrm{H}}\right) = D_{\mathrm{KL}}\left(\hat{\mathbf{R}}_{\bar{\mathbf{x}},\bar{\mathbf{x}}}\|\mathbf{R}_{\bar{\mathbf{x}},\bar{\mathbf{x}}}\right). \tag{47}$$

Thus, we can choose $\bar{\mathbf{W}}$ as the inverse of $\bar{\mathbf{A}}$, which reduces our optimization problem to

$$\underset{\substack{\bar{\mathbf{W}}\in\mathcal{D},\mathbf{R}_{\bar{\mathbf{s}},\bar{\mathbf{s}}}\in\mathcal{R}\\ \hat{\mathbf{R}}_{\bar{\mathbf{y}},\bar{\mathbf{y}}}=\bar{\mathbf{W}}\hat{\mathbf{R}}_{\bar{\mathbf{x}},\bar{\mathbf{x}}}\bar{\mathbf{W}}^{\mathrm{H}}}}{\text{minimize}} \; D_{\mathrm{KL}}\left(\hat{\mathbf{R}}_{\bar{\mathbf{y}},\bar{\mathbf{y}}}\|\mathbf{R}_{\bar{\mathbf{s}},\bar{\mathbf{s}}}\right) \tag{48}$$

where $\hat{\mathbf{R}}_{\bar{\mathbf{y}},\bar{\mathbf{y}}} = \bar{\mathbf{W}}\hat{\mathbf{R}}_{\bar{\mathbf{x}},\bar{\mathbf{x}}}\bar{\mathbf{W}}^{\mathrm{H}}$ can be seen as the sample-mean estimate of the augmented covariance matrix $\mathbf{R}_{\bar{\mathbf{y}},\bar{\mathbf{y}}} = E\bar{\mathbf{y}}\bar{\mathbf{y}}^{\mathrm{H}}$ and the augmented vector $\bar{\mathbf{y}} = \bar{\mathbf{W}}\bar{\mathbf{x}}$ contains the estimates of the original sources $\bar{\mathbf{s}}$. Summarizing, we have reformulated the ML-ICA problem as that of minimizing the KL divergence between an approximately *diagonalized* version of $\hat{\mathbf{R}}_{\bar{\mathbf{x}},\bar{\mathbf{x}}}$ and the *theoretical* augmented covariance matrix of the sources $\mathbf{R}_{\bar{\mathbf{s}},\bar{\mathbf{s}}}$.

B. Reformulation of the ML-ICA Problem

Unfortunately, although the sets $\mathcal{D}$ and $\mathcal{R}$ are convex, the cost function $D_{\mathrm{KL}}\left(\hat{\mathbf{R}}_{\bar{\mathbf{y}},\bar{\mathbf{y}}}\|\mathbf{R}_{\bar{\mathbf{s}},\bar{\mathbf{s}}}\right)$ of the ML-ICA problem presented in Section IV-A is not convex, which can be easily corroborated by considering the trivial ambiguities of the quaternion ICA model.[8]

In order to simplify the ML-ICA problem, here we will obtain the solution of (48) with respect to $\mathbf{R}_{\bar{\mathbf{s}},\bar{\mathbf{s}}}$, which can be easily done by reordering the rows and columns of the matrices $\mathbf{R}_{\bar{\mathbf{s}},\bar{\mathbf{s}}}$ and $\hat{\mathbf{R}}_{\bar{\mathbf{y}},\bar{\mathbf{y}}} = \bar{\mathbf{W}}\hat{\mathbf{R}}_{\bar{\mathbf{x}},\bar{\mathbf{x}}}\bar{\mathbf{W}}^{H}$. In particular, we define the vectors

$$\tilde{\mathbf{s}} = \begin{bmatrix}\bar{\mathbf{s}}_1\\ \vdots \\ \bar{\mathbf{s}}_m\end{bmatrix} = \mathbf{P}\bar{\mathbf{s}} \quad \tilde{\mathbf{y}} = \begin{bmatrix}\bar{\mathbf{y}}_1\\ \vdots \\ \bar{\mathbf{y}}_m\end{bmatrix} = \mathbf{P}\bar{\mathbf{y}} \tag{49}$$

where $\bar{\mathbf{s}}_k \in \mathbb{H}^{4\times 1}$ denotes the augmented vector for the kth source s_k (kth element of $\mathbf{s}$), $\bar{\mathbf{y}}_k \in \mathbb{H}^{4\times 1}$ is the augmented version of y_k (the kth element of $\mathbf{y} = \mathbf{W}\mathbf{x}$, with $\mathbf{W} = \mathbf{A}^{-1}$), and $\mathbf{P} \in \mathbb{R}^{4m\times 4m}$ is a permutation matrix. Thus, the reordered augmented covariance matrices are

$$\mathbf{R}_{\tilde{\mathbf{s}},\tilde{\mathbf{s}}} = \mathbf{P}\mathbf{R}_{\bar{\mathbf{s}},\bar{\mathbf{s}}}\mathbf{P}^{\mathrm{T}} = \begin{bmatrix} \mathbf{R}_{\bar{\mathbf{s}}_1,\bar{\mathbf{s}}_1} & \mathbf{0}_{4\times 4} & \cdots & \mathbf{0}_{4\times 4}\\ \mathbf{0}_{4\times 4} & \mathbf{R}_{\bar{\mathbf{s}}_2,\bar{\mathbf{s}}_2} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{0}_{4\times 4}\\ \mathbf{0}_{4\times 4} & \cdots & \mathbf{0}_{4\times 4} & \mathbf{R}_{\bar{\mathbf{s}}_m,\bar{\mathbf{s}}_m}\end{bmatrix} \tag{50}$$

$$\hat{\mathbf{R}}_{\tilde{\mathbf{y}},\tilde{\mathbf{y}}} = \mathbf{P}\hat{\mathbf{R}}_{\bar{\mathbf{y}},\bar{\mathbf{y}}}\mathbf{P}^{\mathrm{T}} = \begin{bmatrix} \hat{\mathbf{R}}_{\bar{\mathbf{y}}_1,\bar{\mathbf{y}}_1} & \hat{\mathbf{R}}_{\bar{\mathbf{y}}_1,\bar{\mathbf{y}}_2} & \cdots & \hat{\mathbf{R}}_{\bar{\mathbf{y}}_1,\bar{\mathbf{y}}_m}\\ \hat{\mathbf{R}}_{\bar{\mathbf{y}}_2,\bar{\mathbf{y}}_1} & \hat{\mathbf{R}}_{\bar{\mathbf{y}}_2,\bar{\mathbf{y}}_2} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots\\ \hat{\mathbf{R}}_{\bar{\mathbf{y}}_m,\bar{\mathbf{y}}_1} & \cdots & \cdots & \hat{\mathbf{R}}_{\bar{\mathbf{y}}_m,\bar{\mathbf{y}}_m}\end{bmatrix} \tag{51}$$

and defining the block-diagonal matrix

$$\hat{\mathbf{D}}_{\tilde{\mathbf{y}},\tilde{\mathbf{y}}} = \mathrm{blkdiag}_4(\hat{\mathbf{R}}_{\tilde{\mathbf{y}},\tilde{\mathbf{y}}}) = \begin{bmatrix} \hat{\mathbf{R}}_{\bar{\mathbf{y}}_1,\bar{\mathbf{y}}_1} & \mathbf{0}_{4\times 4} & \cdots & \mathbf{0}_{4\times 4}\\ \mathbf{0}_{4\times 4} & \hat{\mathbf{R}}_{\bar{\mathbf{y}}_2,\bar{\mathbf{y}}_2} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{0}_{4\times 4}\\ \mathbf{0}_{4\times 4} & \cdots & \mathbf{0}_{4\times 4} & \hat{\mathbf{R}}_{\bar{\mathbf{y}}_m,\bar{\mathbf{y}}_m}\end{bmatrix} \tag{52}$$

we can introduce the following decomposition

$$\hat{\mathbf{R}}_{\tilde{\mathbf{y}},\tilde{\mathbf{y}}} = \hat{\mathbf{D}}_{\tilde{\mathbf{y}},\tilde{\mathbf{y}}}^{1/2}\hat{\boldsymbol{\Phi}}_{\tilde{\mathbf{y}},\tilde{\mathbf{y}}}\hat{\mathbf{D}}_{\tilde{\mathbf{y}},\tilde{\mathbf{y}}}^{1/2} \tag{53}$$

where $\hat{\boldsymbol{\Phi}}_{\tilde{\mathbf{y}},\tilde{\mathbf{y}}} = \hat{\mathbf{D}}_{\tilde{\mathbf{y}},\tilde{\mathbf{y}}}^{-1/2}\hat{\mathbf{R}}_{\tilde{\mathbf{y}},\tilde{\mathbf{y}}}\hat{\mathbf{D}}_{\tilde{\mathbf{y}},\tilde{\mathbf{y}}}^{-1/2}$ is defined as the *coherence matrix*, which naturally appears in the multiset extension of quaternion CCA [18], [57], [58]. In particular, $\hat{\mathbf{R}}_{\tilde{\mathbf{y}},\tilde{\mathbf{y}}}$ contains all the information regarding the improperness and the correlation among the estimated sources (the elements in $\mathbf{y}$), whereas $\hat{\mathbf{D}}_{\tilde{\mathbf{y}},\tilde{\mathbf{y}}}$ only considers the individual improperness and variances. Thus, $\hat{\boldsymbol{\Phi}}_{\tilde{\mathbf{y}},\tilde{\mathbf{y}}}$ provides all the available information about the correlation among the estimated sources. Now, we can introduce the following theorem.

Theorem 2: Given an arbitrary augmented covariance matrix $\hat{\mathbf{R}}_{\bar{\mathbf{y}},\bar{\mathbf{y}}}$ and a matrix $\mathbf{R}_{\bar{\mathbf{s}},\bar{\mathbf{s}}} \in \mathcal{R}$, the Kullback–Leibler divergence $D_{\mathrm{KL}}(\hat{\mathbf{R}}_{\bar{\mathbf{y}},\bar{\mathbf{y}}}\|\mathbf{R}_{\bar{\mathbf{s}},\bar{\mathbf{s}}})$ can be decomposed as

$$\begin{aligned} D_{\mathrm{KL}}\left(\hat{\mathbf{R}}_{\bar{\mathbf{y}},\bar{\mathbf{y}}}\|\mathbf{R}_{\bar{\mathbf{s}},\bar{\mathbf{s}}}\right) &= D_{\mathrm{KL}}\left(\hat{\mathbf{R}}_{\tilde{\mathbf{y}},\tilde{\mathbf{y}}}\|\hat{\mathbf{D}}_{\tilde{\mathbf{y}},\tilde{\mathbf{y}}}\right) + D_{\mathrm{KL}}\left(\hat{\mathbf{D}}_{\tilde{\mathbf{y}},\tilde{\mathbf{y}}}\|\mathbf{R}_{\tilde{\mathbf{s}},\tilde{\mathbf{s}}}\right)\\ &= -\frac{1}{2}\ln|\hat{\boldsymbol{\Phi}}_{\tilde{\mathbf{y}},\tilde{\mathbf{y}}}| + D_{\mathrm{KL}}\left(\hat{\mathbf{D}}_{\tilde{\mathbf{y}},\tilde{\mathbf{y}}}\|\mathbf{R}_{\tilde{\mathbf{s}},\tilde{\mathbf{s}}}\right). \end{aligned} \tag{54}$$

[7]Note that, due to the noncommutativity of the quaternion product, we have to take the real part of $\mathrm{Tr}\left(\mathbf{R}_{\bar{\mathbf{x}},\bar{\mathbf{x}}}^{-1}\hat{\mathbf{R}}_{\bar{\mathbf{x}},\bar{\mathbf{x}}}\right)$. Alternatively, we could have written $\Re\left[\mathrm{Tr}\left(\mathbf{R}_{\bar{\mathbf{x}},\bar{\mathbf{x}}}^{-1}\hat{\mathbf{R}}_{\bar{\mathbf{x}},\bar{\mathbf{x}}}\right)\right] = \mathrm{Tr}\left(\mathbf{R}_{\bar{\mathbf{x}},\bar{\mathbf{x}}}^{-1/2}\hat{\mathbf{R}}_{\bar{\mathbf{x}},\bar{\mathbf{x}}}\mathbf{R}_{\bar{\mathbf{x}},\bar{\mathbf{x}}}^{-1/2}\right)$.

[8]Under perfect estimates $\hat{\mathbf{R}}_{\bar{\mathbf{x}},\bar{\mathbf{x}}} = \mathbf{R}_{\bar{\mathbf{x}},\bar{\mathbf{x}}}$, the solutions $(\bar{\mathbf{W}}, \mathbf{R}_{\bar{\mathbf{s}},\bar{\mathbf{s}}})$ associated to any permutation of the columns of the actual mixing matrix would result in zero cost. However, the linear combinations of the previous solutions will not satisfy the ICA model, and the associated cost (KL divergence) will be greater than zero.

Proof: First, we must note that the introduction of the permutation matrix does not change the KL divergence, i.e., $D_{\mathrm{KL}}\left(\hat{\mathbf{R}}_{\bar{\mathbf{y}},\bar{\mathbf{y}}}||\mathbf{R}_{\bar{\mathbf{s}},\bar{\mathbf{s}}}\right) = D_{\mathrm{KL}}\left(\hat{\mathbf{R}}_{\tilde{\mathbf{y}},\tilde{\mathbf{y}}}||\mathbf{R}_{\tilde{\mathbf{s}},\tilde{\mathbf{s}}}\right)$. Second, we can write

$$D_{\mathrm{KL}}\left(\hat{\mathbf{R}}_{\tilde{\mathbf{y}},\tilde{\mathbf{y}}}||\mathbf{R}_{\tilde{\mathbf{s}},\tilde{\mathbf{s}}}\right) = -\frac{1}{2}\ln\frac{|\hat{\mathbf{R}}_{\tilde{\mathbf{y}},\tilde{\mathbf{y}}}||\hat{\mathbf{D}}_{\tilde{\mathbf{y}},\tilde{\mathbf{y}}}|}{|\hat{\mathbf{D}}_{\tilde{\mathbf{y}},\tilde{\mathbf{y}}}||\mathbf{R}_{\tilde{\mathbf{s}},\tilde{\mathbf{s}}}|} + \frac{1}{2}\Re\left[\mathrm{Tr}\left(\mathbf{R}_{\tilde{\mathbf{s}},\tilde{\mathbf{s}}}^{-1}\hat{\mathbf{R}}_{\tilde{\mathbf{y}},\tilde{\mathbf{y}}}\right)\right] - 2m \quad (55)$$

and noting that $\mathrm{Tr}\left(\hat{\mathbf{D}}_{\tilde{\mathbf{y}},\tilde{\mathbf{y}}}^{-1}\hat{\mathbf{R}}_{\tilde{\mathbf{y}},\tilde{\mathbf{y}}}\right) = 4m$ and $\mathrm{Tr}\left(\mathbf{R}_{\tilde{\mathbf{s}},\tilde{\mathbf{s}}}^{-1}\hat{\mathbf{R}}_{\tilde{\mathbf{y}},\tilde{\mathbf{y}}}\right) = \mathrm{Tr}\left(\mathbf{R}_{\tilde{\mathbf{s}},\tilde{\mathbf{s}}}^{-1}\hat{\mathbf{D}}_{\tilde{\mathbf{y}},\tilde{\mathbf{y}}}\right)$, we have

$$\begin{aligned}
&D_{\mathrm{KL}}\left(\hat{\mathbf{R}}_{\tilde{\mathbf{y}},\tilde{\mathbf{y}}}||\mathbf{R}_{\tilde{\mathbf{s}},\tilde{\mathbf{s}}}\right)\\
&= -\frac{1}{2}\ln\frac{|\hat{\mathbf{R}}_{\tilde{\mathbf{y}},\tilde{\mathbf{y}}}|}{|\hat{\mathbf{D}}_{\tilde{\mathbf{y}},\tilde{\mathbf{y}}}|} + \frac{1}{2}\mathrm{Tr}\left(\hat{\mathbf{D}}_{\tilde{\mathbf{y}},\tilde{\mathbf{y}}}^{-1}\hat{\mathbf{R}}_{\tilde{\mathbf{y}},\tilde{\mathbf{y}}}\right) - 2m\\
&\quad - \frac{1}{2}\ln\frac{|\hat{\mathbf{D}}_{\tilde{\mathbf{y}},\tilde{\mathbf{y}}}|}{|\mathbf{R}_{\tilde{\mathbf{s}},\tilde{\mathbf{s}}}|} + \frac{1}{2}\Re\left[\mathrm{Tr}\left(\mathbf{R}_{\tilde{\mathbf{s}},\tilde{\mathbf{s}}}^{-1}\hat{\mathbf{D}}_{\tilde{\mathbf{y}},\tilde{\mathbf{y}}}\right)\right] - 2m\\
&= D_{\mathrm{KL}}\left(\hat{\mathbf{R}}_{\tilde{\mathbf{y}},\tilde{\mathbf{y}}}||\hat{\mathbf{D}}_{\tilde{\mathbf{y}},\tilde{\mathbf{y}}}\right) + D_{\mathrm{KL}}\left(\hat{\mathbf{D}}_{\tilde{\mathbf{y}},\tilde{\mathbf{y}}}||\mathbf{R}_{\tilde{\mathbf{s}},\tilde{\mathbf{s}}}\right)\\
&= -\frac{1}{2}\ln|\hat{\mathbf{\Phi}}_{\tilde{\mathbf{y}},\tilde{\mathbf{y}}}| + D_{\mathrm{KL}}\left(\hat{\mathbf{D}}_{\tilde{\mathbf{y}},\tilde{\mathbf{y}}}||\mathbf{R}_{\tilde{\mathbf{s}},\tilde{\mathbf{s}}}\right). \quad (56)
\end{aligned}$$

■

Theorem 2, which can be seen as a particular case of the Pythagorean theorem for exponential families of pdfs [56], [59], decomposes $D_{\mathrm{KL}}\left(\hat{\mathbf{R}}_{\bar{\mathbf{y}},\bar{\mathbf{y}}}||\mathbf{R}_{\bar{\mathbf{s}},\bar{\mathbf{s}}}\right)$ into the sum of two terms. The first term, which does not depend on the augmented covariance matrix $\mathbf{R}_{\bar{\mathbf{s}},\bar{\mathbf{s}}}$, can be seen as a measure of the residual correlation among the separated sources $\mathbf{y} = \mathbf{W}\mathbf{x}$. In particular, $-1/2\ln|\hat{\mathbf{\Phi}}_{\tilde{\mathbf{y}},\tilde{\mathbf{y}}}|$ represents the mutual information[9] among the quaternion random variables $y_1,\ldots,y_m$. The second term measures the divergence between the uncorrelated version $\left(\hat{\mathbf{D}}_{\bar{\mathbf{y}},\bar{\mathbf{y}}}\right)$ of $\hat{\mathbf{R}}_{\bar{\mathbf{y}},\bar{\mathbf{y}}}$ and the augmented covariance matrix of the sources $\mathbf{R}_{\bar{\mathbf{s}},\bar{\mathbf{s}}}$. Thanks to this decomposition, the ML-ICA problem in (48) can be rewritten as

$$\underset{\bar{\mathbf{W}}\in\mathcal{D}}{\text{minimize}} \quad \underset{\substack{\mathbf{R}_{\bar{\mathbf{s}},\bar{\mathbf{s}}}\in\mathcal{R}\\ \mathbf{R}_{\tilde{\mathbf{s}},\tilde{\mathbf{s}}}=\mathbf{P}\mathbf{R}_{\bar{\mathbf{s}},\bar{\mathbf{s}}}\mathbf{P}^{\mathrm{T}}}}{\text{infimum}} \quad -\frac{1}{2}\ln|\hat{\mathbf{\Phi}}_{\tilde{\mathbf{y}},\tilde{\mathbf{y}}}| + D_{\mathrm{KL}}\left(\hat{\mathbf{D}}_{\tilde{\mathbf{y}},\tilde{\mathbf{y}}}||\mathbf{R}_{\tilde{\mathbf{s}},\tilde{\mathbf{s}}}\right). \quad (57)$$

Furthermore, noting that $\hat{\mathbf{\Phi}}_{\tilde{\mathbf{y}},\tilde{\mathbf{y}}}$ is invariant to scaling factors on the rows of $\mathbf{W}$, we can assume without loss of generality that the diagonal elements of $\hat{\mathbf{D}}_{\tilde{\mathbf{y}},\tilde{\mathbf{y}}}$ are equal to 1. Thus, the solution with respect to $\mathbf{R}_{\tilde{\mathbf{s}},\tilde{\mathbf{s}}}$ is $\mathbf{R}_{\tilde{\mathbf{s}},\tilde{\mathbf{s}}} = \hat{\mathbf{D}}_{\tilde{\mathbf{y}},\tilde{\mathbf{y}}}$, and the ML-ICA problem reduces to

$$\underset{\bar{\mathbf{W}}\in\mathcal{D}}{\text{minimize}} \quad -\frac{1}{2}\ln\left|\hat{\mathbf{\Phi}}_{\tilde{\mathbf{y}},\tilde{\mathbf{y}}}\right|. \quad (58)$$

In other words, given a solution $(\bar{\mathbf{W}})$ of (58), we can always obtain a solution of (57) by scaling the rows of $\bar{\mathbf{W}}$ to satisfy $\mathrm{blkdiag}_1\left(\hat{\mathbf{D}}_{\tilde{\mathbf{y}},\tilde{\mathbf{y}}}\right) = \mathbf{I}_{4m}$ and taking $\mathbf{R}_{\tilde{\mathbf{s}},\tilde{\mathbf{s}}} = \hat{\mathbf{D}}_{\tilde{\mathbf{y}},\tilde{\mathbf{y}}}$. Finally, as one could expect, the optimization problem in (58) amounts to find the separation matrix $\mathbf{W}$ minimizing the correlation (dependence) among the estimated sources $\mathbf{y}$.

[9]Strictly speaking, the mutual information is only defined for two random variables. Here, mutual information refers to the entropy loss due to the dependence among the random variables [18], [19], [60].

C. Proposed Algorithm

So far, we have been able to find the optimal $\mathbf{R}_{\bar{\mathbf{s}},\bar{\mathbf{s}}}$ as a function of $\bar{\mathbf{W}}$, which reduces the ML-ICA problem to the minimization of the mutual information measure $-1/2\ln\left|\hat{\mathbf{\Phi}}_{\tilde{\mathbf{y}},\tilde{\mathbf{y}}}\right|$. However, the optimization problem in (58) is still nonconvex.[10] Here, we propose a practical ML-ICA algorithm based on the approximated joint diagonalization of the covariance and complementary covariance matrices of the estimated sources $\mathbf{y} = \mathbf{W}\mathbf{x}$. The main properties of the proposed technique are the following.

- As we have seen, the cost function $-1/2\ln\left|\hat{\mathbf{\Phi}}_{\tilde{\mathbf{y}},\tilde{\mathbf{y}}}\right|$ only depends on the data through the estimate $\hat{\mathbf{R}}_{\bar{\mathbf{x}},\bar{\mathbf{x}}}$. Therefore, although the proposed algorithm only provides the ML solution in the case of Gaussian data, its practical performance is determined by the accuracy of the estimate $\hat{\mathbf{R}}_{\bar{\mathbf{x}},\bar{\mathbf{x}}}$ and not by possible deviations from the Gaussianity assumption.
- Analogously to other AJD algorithms [30], the proposed method is based on a quadratic approximation of the cost function $-1/2\ln\left|\hat{\mathbf{\Phi}}_{\tilde{\mathbf{y}},\tilde{\mathbf{y}}}\right|$, and therefore it can be seen as a quasi-Newton technique [35] with computational cost of order $\mathcal{O}(m^2)$ per iteration. However, we must note that the application of a pure Newton method would require the inversion of a quaternion Hessian matrix of dimensions $[m(m+1)/2] \times [m(m+1)/2]$, which would result in a computational cost of order $\mathcal{O}(m^6)$ per iteration.
- At each iteration, the separation matrix is updated as $\mathbf{W} \leftarrow (\mathbf{I}_m + \mathbf{\Delta})\mathbf{W}$, where $\mathbf{\Delta} \in \mathbb{H}^{m\times m}$ is assumed to be a *small* quaternion matrix. In particular, this assumption will be exploited to simplify the approximated cost function and, in the final implementation of the algorithm, $\mathbf{\Delta}$ will be scaled (if necessary) to ensure the invertibility of $(\mathbf{I}_m + \mathbf{\Delta})$[30].
- Analogously to most AJD techniques [29]–[34], the proposed algorithm tries to solve a nonconvex optimization problem, and therefore it could suffer from local minima. However, we did not find this problem in our experiments, which makes us think that this case is highly unlikely. The rigorous analysis of convergence of the proposed algorithm, as well as that of most AJD methods, is an interesting topic for future research.
- Unlike most AJD algorithms [29]–[34], which are based on the minimization of the magnitudes of the off-diagonal terms, the proposed technique directly considers the minimization of the KL divergence given by $-1/2\ln\left|\hat{\mathbf{\Phi}}_{\tilde{\mathbf{y}},\tilde{\mathbf{y}}}\right|$. As we have seen, this cost function naturally appears in the quaternion ML-ICA problem and, to our best knowledge, it has been never used as an AJD criterion.[11]

In order to obtain the simplified cost function, we will start by introducing the matrix

$$\tilde{\mathbf{W}} = \mathbf{P}\bar{\mathbf{W}}\mathbf{P}^{\mathrm{T}} = \begin{bmatrix} \tilde{\mathbf{W}}_{1,1} & \cdots & \tilde{\mathbf{W}}_{1,m} \\ \vdots & \ddots & \vdots \\ \tilde{\mathbf{W}}_{m,1} & \cdots & \tilde{\mathbf{W}}_{m,m} \end{bmatrix} \quad (59)$$

[10]The nonconvexity of $-1/2\ln\left|\hat{\mathbf{\Phi}}_{\tilde{\mathbf{y}},\tilde{\mathbf{y}}}\right|$ can be again verified by considering the permutation ambiguities and their linear combinations.

[11]See [28] and [29] for an AJD method based on the minimization of a sum of KL divergences between positive semidefinite matrices and [33] for an approximation of KL divergences with weighted Frobenius norms.

where $\mathbf{P} \in \mathbb{R}^{4m\times 4m}$ is the permutation matrix defined in (49) (i.e., $\tilde{\mathbf{y}} = \mathbf{P}\bar{\mathbf{y}}$) and $\tilde{\mathbf{W}}_{k,l} \in \mathbb{H}^{4\times 4}$ is a diagonal matrix obtained from the element $w_{k,l}$ in the kth row and lth column of $\mathbf{W}$ as $\tilde{\mathbf{W}}_{k,l} = \mathrm{diag}\left(\left[w_{k,l}, w_{k,l}^{(\eta)}, w_{k,l}^{(\eta')}, w_{k,l}^{(\eta'')}\right]^{\mathrm{T}}\right)$.

Now, with similar definitions of $\delta_{k,l} \in \mathbb{H}$, $\tilde{\boldsymbol{\Delta}} \in \mathbb{H}^{4m\times 4m}$, and $\tilde{\boldsymbol{\Delta}}_{k,l} \in \mathbb{H}^{4\times 4}$ and introducing the operator $\mathrm{offdiag}_n(\mathbf{A}) = \mathbf{A} - \mathrm{blkdiag}_n(\mathbf{A})$ (which introduces zeros in the $n\times n$ diagonal blocks of $\mathbf{A}$), we are ready to introduce the quadratic approximation of $-1/2\ln\left|\hat{\boldsymbol{\Phi}}_{\tilde{\mathbf{y}},\tilde{\mathbf{y}}}\right|$ evaluated at $(\mathbf{I}_m+\boldsymbol{\Delta})\mathbf{W}$ for small $\boldsymbol{\Delta}$.

Lemma 6: Given a coherence matrix $\hat{\boldsymbol{\Phi}}_{\tilde{\mathbf{y}},\tilde{\mathbf{y}}}$ close to the identity, i.e., $\left\|\mathrm{offdiag}_4\left(\hat{\boldsymbol{\Phi}}_{\tilde{\mathbf{y}},\tilde{\mathbf{y}}}\right)\right\|^2 \ll 1$ and assuming $\|\boldsymbol{\Delta}\|^2 \ll 1$, the mutual information measure can be approximated by the following quadratic expression:

$$-\frac{1}{2}\ln\left|\hat{\boldsymbol{\Phi}}_{\tilde{\mathbf{y}},\tilde{\mathbf{y}}}\right| \simeq \frac{1}{4}\|\mathbf{J}(\boldsymbol{\Delta})\|^2 = \frac{1}{2}\sum_{k=1}^{m}\sum_{l=k+1}^{m}\|\mathbf{J}_{k,l}(\delta_{k,l},\delta_{l,k})\|^2 \tag{60}$$

where $\mathbf{J}(\boldsymbol{\Delta}) \in \mathbb{H}^{4m\times 4m}$ is a block matrix with 4×4 blocks, zero diagonal blocks, and off-diagonal blocks given by

$$\mathbf{J}_{k,l}(\delta_{k,l},\delta_{l,k}) = \hat{\mathbf{R}}_{\bar{\mathbf{y}}_k,\bar{\mathbf{y}}_k}^{-1/2}\hat{\mathbf{R}}_{\bar{\mathbf{y}}_k,\bar{\mathbf{y}}_l}\hat{\mathbf{R}}_{\bar{\mathbf{y}}_l,\bar{\mathbf{y}}_l}^{-1/2} + \hat{\mathbf{R}}_{\bar{\mathbf{y}}_k,\bar{\mathbf{y}}_k}^{-1/2}\tilde{\boldsymbol{\Delta}}_{k,l}\hat{\mathbf{R}}_{\bar{\mathbf{y}}_l,\bar{\mathbf{y}}_l}^{1/2} + \hat{\mathbf{R}}_{\bar{\mathbf{y}}_k,\bar{\mathbf{y}}_k}^{1/2}\tilde{\boldsymbol{\Delta}}_{l,k}^{\mathrm{H}}\hat{\mathbf{R}}_{\bar{\mathbf{y}}_l,\bar{\mathbf{y}}_l}^{-1/2}. \tag{61}$$

Proof: See Appendix B. ■

Thanks to the approximation in (60), the optimization problem to be solved in each iteration of the quasi-Newton AJD method is decoupled into $m(m-1)/2$ simpler problems. Specifically, the elements $\delta_{k,l}, \delta_{l,k}$ are obtained by solving the least squares (LS) problem

$$\underset{\delta_{k,l},\delta_{l,k}}{\mathrm{minimize}} \quad \|\mathbf{J}_{k,l}(\delta_{k,l},\delta_{l,k})\|^2 \tag{62}$$

whose solution is easily obtained by rewriting $\mathbf{J}_{k,l}(\delta_{k,l},\delta_{l,k})$ as a function of the eight real components of $\delta_{k,l}$ and $\delta_{l,k}$. In particular, using the real representations

$$\delta_{k,l} = [1,\eta,\eta',\eta'']\,\boldsymbol{\delta}_{k,l} \quad \delta_{l,k} = [1,\eta,\eta',\eta'']\,\boldsymbol{\delta}_{l,k} \tag{63}$$

with $\boldsymbol{\delta}_{k,l}, \boldsymbol{\delta}_{l,k} \in \mathbb{R}^{4\times 1}$, the above problem can be rewritten as

$$\underset{\boldsymbol{\delta}_{k,l},\boldsymbol{\delta}_{l,k}\in\mathbb{R}^{4\times 1}}{\mathrm{minimize}} \quad \left\|\boldsymbol{\phi}_{k,l} + \mathbf{C}_{k,l}\boldsymbol{\delta}_{k,l} + \mathbf{C}_{l,k}\boldsymbol{\delta}_{l,k}\right\|^2 \tag{64}$$

where $\boldsymbol{\phi}_{k,l} = \mathrm{vec}\left(\hat{\mathbf{R}}_{\bar{\mathbf{y}}_k,\bar{\mathbf{y}}_k}^{-1/2}\hat{\mathbf{R}}_{\bar{\mathbf{y}}_k,\bar{\mathbf{y}}_l}\hat{\mathbf{R}}_{\bar{\mathbf{y}}_l,\bar{\mathbf{y}}_l}^{-1/2}\right)$ and the pth column of $\mathbf{C}_{k,l} \in \mathbb{H}^{16\times 4}$ is, for $p = 1,\ldots,4$

$$\mathrm{vec}\left(\hat{\mathbf{R}}_{\bar{\mathbf{y}}_k,\bar{\mathbf{y}}_k}^{-1/2}\boldsymbol{\Omega}_p\hat{\mathbf{R}}_{\bar{\mathbf{y}}_l,\bar{\mathbf{y}}_l}^{1/2}\right), \qquad k < l \tag{65}$$

$$\mathrm{vec}\left(\hat{\mathbf{R}}_{\bar{\mathbf{y}}_k,\bar{\mathbf{y}}_k}^{1/2}\boldsymbol{\Omega}_p^{H}\hat{\mathbf{R}}_{\bar{\mathbf{y}}_l,\bar{\mathbf{y}}_l}^{-1/2}\right), \qquad k > l \tag{66}$$

with

$$\boldsymbol{\Omega}_1 = \mathbf{I}_4 \tag{67}$$

$$\boldsymbol{\Omega}_2 = \mathrm{diag}\left([\eta,\eta,-\eta,-\eta]\right) \tag{68}$$

$$\boldsymbol{\Omega}_3 = \mathrm{diag}\left([\eta',-\eta',\eta',-\eta']\right) \tag{69}$$

$$\boldsymbol{\Omega}_4 = \mathrm{diag}\left([\eta'',-\eta'',-\eta'',\eta'']\right). \tag{70}$$

Thus, defining the vector $\bar{\boldsymbol{\delta}}_{k,l} = \left[\boldsymbol{\delta}_{k,l}^{\mathrm{T}}, \boldsymbol{\delta}_{l,k}^{\mathrm{T}}\right]^{\mathrm{T}}$ and the matrix $\bar{\mathbf{C}}_{k,l} = [\,\mathbf{C}_{k,l} \quad \mathbf{C}_{l,k}\,]$, the solution of (62) can be obtained as

$$\bar{\boldsymbol{\delta}}_{k,l} = \mathbf{G}_{k,l}^{-1}\boldsymbol{\gamma}_{k,l} \tag{71}$$

where $\mathbf{G}_{k,l} = \Re\left(\bar{\mathbf{C}}_{k,l}^{\mathrm{H}}\bar{\mathbf{C}}_{k,l}\right)$ and $\boldsymbol{\gamma}_{k,l} = \Re\left(\bar{\mathbf{C}}_{k,l}^{\mathrm{H}}\boldsymbol{\phi}_{k,l}\right)$. Finally, we must note that the computational complexity of the proposed AJD method, which is summarized in Algorithm 1, is dominated by the inversion of the $m(m-1)/2$ matrices $\mathbf{G}_{k,l} \in \mathbb{R}^{8\times 8}$.

V. Simulation Results

In this section, the main results of the paper are illustrated by means of some simulation examples. In all the cases, the entries of the square mixing matrix $\mathbf{A} \in \mathbb{H}^{m\times m}$ have been generated as i.i.d. quaternion $\mathbb{Q}$-proper Gaussian random variables with zero mean and unit variance, and the sources are independent quaternion Gaussian random variables with zero mean, unit variance, and different properness profiles. The proposed quaternion ML-ICA algorithm has been limited to 50 iterations, and the threshold to ensure invertibility (see Algorithm 1) has been fixed to $\mu = 0.99$. The performance of the proposed algorithm is evaluated by means of the mutual information $J = -1/2\ln\left|\hat{\boldsymbol{\Phi}}_{\tilde{\mathbf{y}},\tilde{\mathbf{y}}}\right|$, the bit error rate (BER) of the communications system (see Section V-D), or by the residual mixture matrix

$$\mathbf{E} = \mathbf{W}\mathbf{A} \tag{72}$$

Algorithm 1 Quaternion ML-ICA

Input: Sample-mean estimate $\hat{\mathbf{R}}_{\tilde{\mathbf{x}},\tilde{\mathbf{x}}} \in \mathbb{H}^{4m\times 4m}$ and threshold parameter $\mu \in \mathbb{R}$.
Output: Separation matrix $\mathbf{W} \in \mathbb{H}^{m\times m}$ and augmented covariance of the estimated sources $\hat{\mathbf{R}}_{\tilde{\mathbf{y}},\tilde{\mathbf{y}}} \in \mathbb{H}^{4m\times 4m}$.
Initialize: $\mathbf{W} = \mathbf{I}_m$, $\hat{\mathbf{R}}_{\tilde{\mathbf{y}},\tilde{\mathbf{y}}} = \hat{\mathbf{R}}_{\tilde{\mathbf{x}},\tilde{\mathbf{x}}}$.
repeat
  for $k = 1,\ldots,m$ and $l = k+1,\ldots,m$ **do**
    Obtain $\bar{\boldsymbol{\delta}}_{k,l}$ from (71).
  end for
  if $\|\boldsymbol{\Delta}\| \geq \mu$ **then**
    $\boldsymbol{\Delta} \leftarrow \mu\boldsymbol{\Delta}/\|\boldsymbol{\Delta}\|$
  end if
  Update $\mathbf{W} \leftarrow (\mathbf{I}_m + \boldsymbol{\Delta})\mathbf{W}$.
  Update $\hat{\mathbf{R}}_{\tilde{\mathbf{y}},\tilde{\mathbf{y}}} \leftarrow \left(\mathbf{I}_{4m} + \tilde{\boldsymbol{\Delta}}\right)\hat{\mathbf{R}}_{\tilde{\mathbf{y}},\tilde{\mathbf{y}}}\left(\mathbf{I}_{4m} + \tilde{\boldsymbol{\Delta}}\right)^{\mathrm{H}}$.
until Convergence
Normalize (if wanted) the rows and columns of $\hat{\mathbf{R}}_{\tilde{\mathbf{y}},\tilde{\mathbf{y}}}$ (respectively the rows of $\mathbf{W}$) to obtain unit-variance sources.

where $\mathbf{W} \in \mathbb{H}^{m\times m}$ is the estimated separation matrix and $\mathbf{A}$ is the actual mixing matrix. In particular, after solving the possible permutation ambiguity, the residual mixture measure for the kth source is defined as

$$M_k = \frac{1}{|e_{k,k}|^2}\sum_{\substack{l=1\\ l\neq k}}^{m}|e_{k,l}|^2 \tag{73}$$

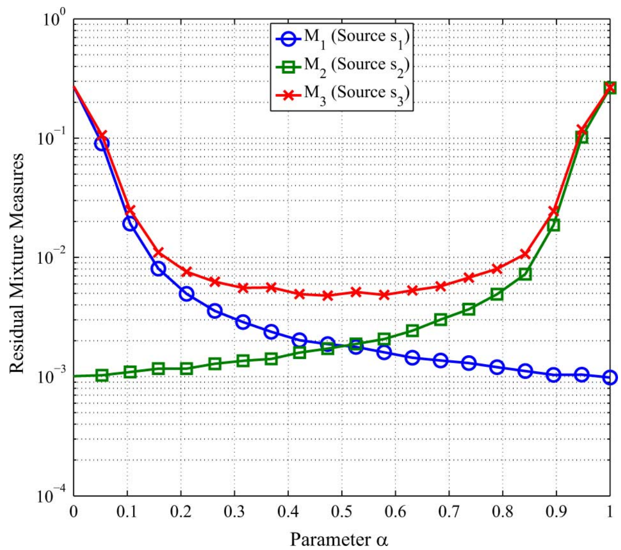

Fig. 1. Identifiability example. Three sources and $T = 100$ vector observations.

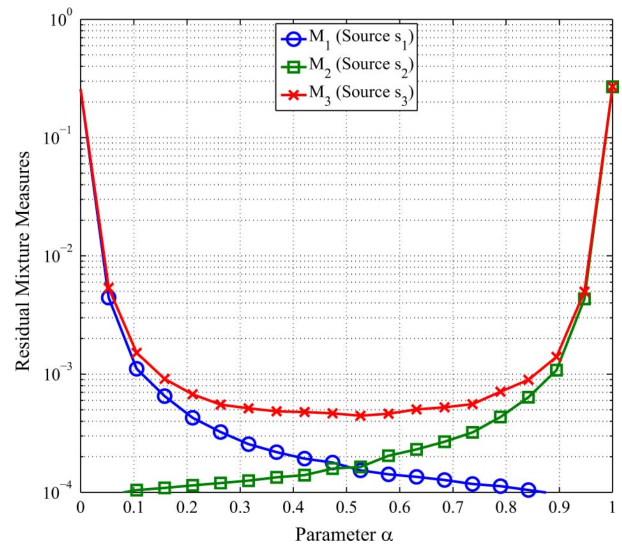

Fig. 2. Identifiability example. Three sources and $T = 1000$ vector observations.

where $e_{k,l}$ is the entry in the kth row and lth column of $\mathbf{E}$. Finally, the total residual mixture measure is

$$\bar{M} = \frac{1}{m}\sum_{k=1}^{m} M_k = \frac{1}{m}\left\|\mathrm{blkdiag}_1(\mathbf{E})^{-1}\mathbf{E}\right\|^2 - 1. \tag{74}$$

A. Identifiability Example

In the first example, we consider three independent sources with unit variance and properness profiles

$$\boldsymbol{\psi}_{s_1} = [0.1\eta'', 0.5\eta, 0.1\eta']^{\mathrm{T}} \tag{75}$$
$$\boldsymbol{\psi}_{s_2} = [0.1\eta'', 0.1\eta, 0.5\eta']^{\mathrm{T}} \tag{76}$$
$$\boldsymbol{\psi}_{s_3} = (1-\alpha)\boldsymbol{\psi}_{s_1} + \alpha\boldsymbol{\psi}_{s_2} \tag{77}$$

where $0 \leq \alpha \leq 1$ is a real parameter controlling the difference between $\boldsymbol{\psi}_{s_3}$ and the other properness profiles. In particular, for $\alpha = 0$, the properness profiles $\boldsymbol{\psi}_{s_1}$ and $\boldsymbol{\psi}_{s_3}$ are rotationally equivalent, whereas $\boldsymbol{\psi}_{s_2}$ and $\boldsymbol{\psi}_{s_3}$ are rotationally equivalent for $\alpha = 1$. Furthermore, although the two first sources have the same (scalar) improperness degree [18], their properness profiles are not rotationally equivalent, and therefore these sources can be unambiguously separated.

Figs. 1 and 2 show the residual mixture measure for the three sources as a function of the parameter α. The results have been obtained by averaging 1000 independent experiments for $T = 100$ and $T = 1000$ vector observations. As stated by Theorem 1, the only nontrivial ambiguities appear for the values of α resulting in rotationally equivalent properness profiles. Thus, there is a linear mixture of the sources s_1 and s_3 for $\alpha = 0$ and a mixture of s_2 and s_3 for $\alpha = 1$. Finally, it is also interesting to note that, from a practical point of view, the accuracy of the ICA method is controlled by a tradeoff between the number of observations T and the *distances* among the different properness profiles.

B. Convergence of the Quaternion ML-ICA Algorithm

In these examples, we consider $m = 10$ sources with randomly generated properness profiles.[12] In order to evaluate the possible convergence to local minima, the proposed algorithm has been initialized in 100 different points (equivalently, we have considered 100 different mixing matrices for the same sources). Additionally, we have also considered 100 independent experiments, with independently generated mixing matrices, properness profiles, and sources. The results for $T = 100$ and $T = 1000$ are shown in Figs. 3 and 4, where we can see that, despite the nonconvexity of the cost function $J = -1/2\ln\left|\hat{\boldsymbol{\Phi}}_{\tilde{\mathbf{y}},\tilde{\mathbf{y}}}\right|$, the proposed algorithm always converges to the same solution. As a heuristic explanation for this fact, we can think that, with high probability, the only minima of the cost function $-1/2\ln\left|\hat{\boldsymbol{\Phi}}_{\tilde{\mathbf{y}},\tilde{\mathbf{y}}}\right|$ are given by the *true* solution $\bar{\mathbf{W}}$ and the associated trivial ambiguities.

C. Overall Performance

In these examples, we evaluate the overall performance of the proposed algorithm for different numbers m of independent sources. Fig. 5 shows the results for the case of sources with random properness profiles, whereas Fig. 6 shows the results for the particular case of m sources with properness profiles

$$\boldsymbol{\psi}_{s_k} = \left[\frac{k-1}{m}\eta, 0, 0\right]^{\mathrm{T}}, \qquad k = 1, \ldots, m. \tag{78}$$

As can be seen, the proposed algorithm is able to recover the sources in both cases. However, the results are better in the case of randomly generated properness profiles. This is due to the fact that although the properness profiles in (78) are not rotationally equivalent, they are restricted to reside in the same *line*. Therefore, we can think that the properness profiles in (78) are *closer* to be rotationally equivalent than those generated at random.

[12]The four real components of each quaternion source follow a zero mean Gaussian distribution with covariance $\mathbf{B}\mathbf{B}^T$, where the entries of $\mathbf{B} \in \mathbb{R}^{4\times4}$ are i.i.d zero mean and unit variance random variables.

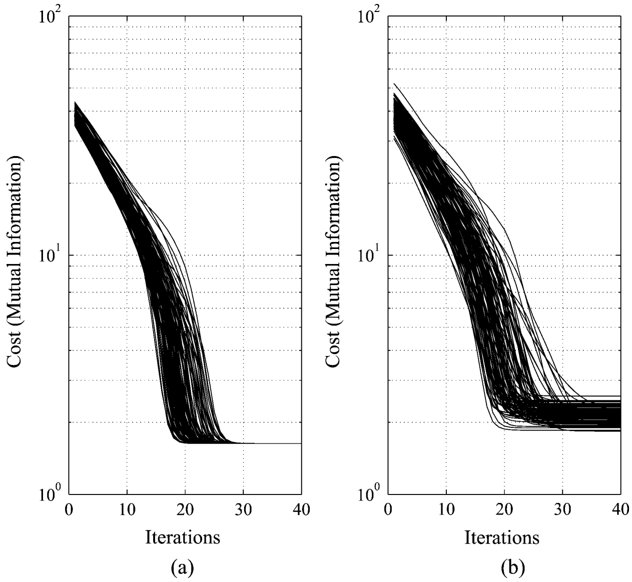

Fig. 3. Convergence of the ML-ICA algorithm. Ten sources with random properness profiles and $T = 100$ vector observations. (a) Fixed data and different initialization points. (b) Independent experiments.

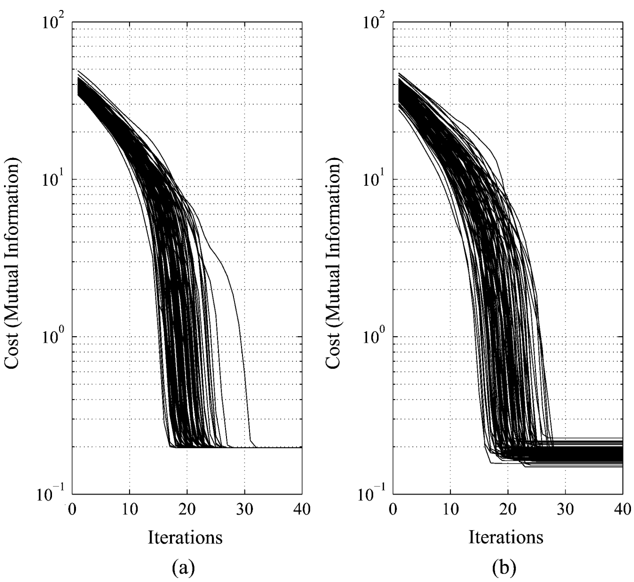

Fig. 4. Convergence of the ML-ICA algorithm. Ten sources with random properness profiles and $T = 1000$ vector observations. (a) Fixed data and different initialization points. (b) Independent experiments.

D. Blind Multiuser Decoding in Alamouti-Based Systems

In the final example, we illustrate the application of the proposed ML-ICA algorithm in a practical problem. In particular, we consider a multiuser wireless communications system based on Alamouti coding [22]. The basic signal model for a single user and a single-antenna receiver can be written as

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} s_1 & -s_2^* \\ s_2 & s_1^* \end{bmatrix} \begin{bmatrix} h_1 \\ h_2 \end{bmatrix} + \begin{bmatrix} n_1 \\ n_2 \end{bmatrix} \tag{79}$$

where the two rows represent two consecutive symbol periods, s_1 and s_2 are two complex information symbols (in the plane $\{1, \eta\}$), the columns of the 2×2 matrix are associated to one

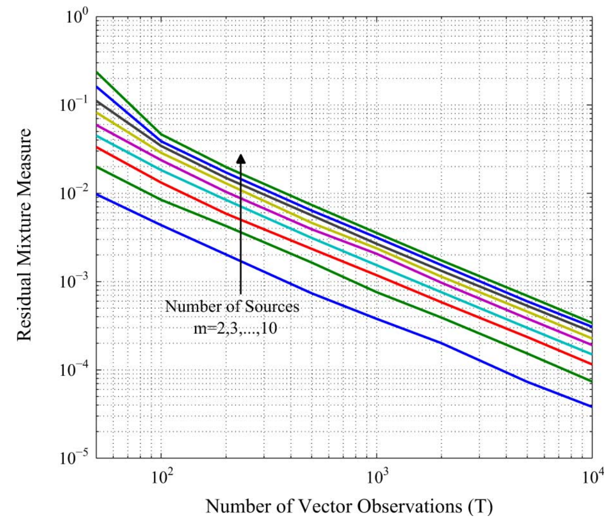

Fig. 5. Overall performance ($\bar{M}$) of the quaternion ML-ICA algorithm. Sources with random properness profiles.

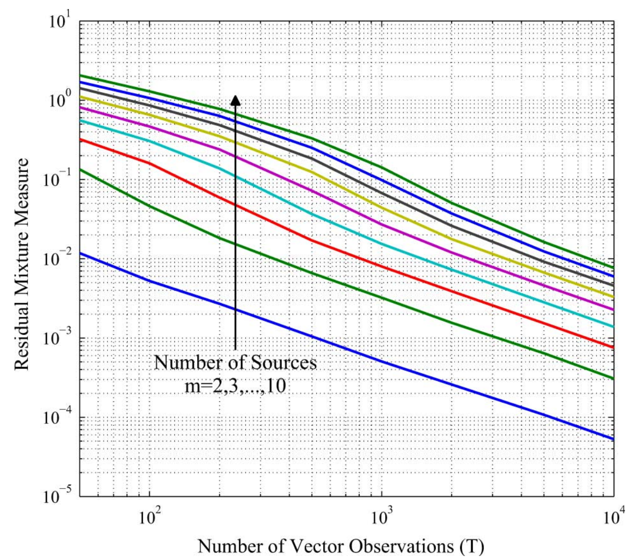

Fig. 6. Overall performance ($\bar{M}$) of the quaternion ML-ICA algorithm. Sources with properness profiles $\boldsymbol{\psi}_{s_k} = [\eta(k-1)/m, 0, 0]^T$, $k = 1, \ldots, m$.

transmit antenna, $h_k \in \mathbb{C}$ represents the channel response between the kth transmit antenna and the receive antenna, and n_1, n_2 are complex noise terms. Interestingly, the above equation can be compactly rewritten in terms of quaternions as

$$x = hs + n \tag{80}$$

where $x = x_1 + x_2\eta'$, $h = h_1 + h_2\eta'$, $s = s_1 + s_2\eta'$, and $n = n_1 + n_2\eta'$. Thus, if we consider a synchronous uplink channel with m users and a base station with m receive antennas, we obtain the model

$$\mathbf{x} = \mathbf{H}\mathbf{s} + \mathbf{n} \tag{81}$$

where $\mathbf{x} \in \mathbb{H}^{m \times 1}$ stacks the quaternion observations in each receive antenna, $\mathbf{H} \in \mathbb{H}^{m \times m}$ is the quaternion channel matrix, whose entry $h_{k,l}$ represents the quaternion channel between the

lth user and the kth receive antenna, $\mathbf{s} \in \mathbb{H}^{n\times 1}$ stacks the quaternion information symbols, and $\mathbf{n}$ is the quaternion noise.

In this experiment, we consider a multiuser system with $m = 2$ users transmitting with the same power. The channels are generated as i.i.d. zero-mean Rayleigh channels with unit variance, and the noise (n_1, n_2) is i.i.d. zero-mean jointly proper complex Gaussian noise. That is, the entries of $\mathbf{H}$ and $\mathbf{n}$ are independent zero-mean $\mathbb{Q}$-proper quaternion Gaussian random variables. The improperness of the quaternion sources in $\mathbf{s}$ is due to the complex improperness of the information symbols s_1, s_2, which in practice can be due to different reasons, such as the use of BPSK constellations, a power imbalance between the in-phase and quadrature branches of the antennas, or the introduction of correlations in order to solve the ambiguity problems associated to blind channel estimation methods for space–time block coded systems [61], [62]. In our case, the symbols of the first user are QPSK jointly complex proper, whereas the second user transmits QPSK symbols with a power imbalance between the in-phase and quadrature branches. Specifically, the power of the in-phase component is three times higher than that of the quadrature component. This results in a mixture of two quaternion sources, one of them $\mathbb{Q}$-proper and the other $\mathbb{Q}$-improper.

Our goal here consists in recovering, up to the trivial ambiguities, the sources $\mathbf{s}$ and the channel $\mathbf{H}$ from the observations $\mathbf{x}$. We must note that the model in (81) includes a noise term, which is not taken into account in our ICA model.[13] However, we will see that the direct application of the proposed ML-ICA algorithm provides satisfactory results.

For comparison purposes, we have evaluated the quaternion extension of the Infomax [40] algorithm proposed in [20], as well as the linear MMSE receiver with perfect channel knowledge. The obtained results for different numbers of vector observations are shown in Fig. 7, which illustrates the satisfactory performance of the ML-ICA method in the case of non-Gaussian data. Furthermore, we can see that the proposed technique outperforms the approach in [20], which is due to the fact that the ML-ICA algorithm is solely based on SOS, which can be accurately estimated from a limited number of vector observations. Thus, when the SOS provide sufficient information for solving the ICA problem, the proposed method outperforms the Infomax for small and moderate sample sizes. On the other hand, for non-Gaussian data and larger sample sizes, the Infomax should outperform any method exclusively based on SOS.

VI. Conclusion

This paper has addressed the independent component analysis (ICA) of quaternion random vectors. First, we have derived the necessary and sufficient conditions for the identifiability of the ICA model from the second-order statistics (SOS) of the observations, or equivalently, for the case of quaternion Gaussian data. In particular, we have introduced the concept of properness profile of a quaternion random variable, and we have shown that, excluding the trivial ICA ambiguities, the quaternion ICA model is unambiguously identifiable up to arbitrary linear mixtures affecting those sources with rotationally equivalent properness profiles, i.e., properness profiles related by a three-dimensional rotation. From a practical point of view, we have shown that the maximum-likelihood (ML) approach to the quaternion ICA problem reduces, in the Gaussian case, to the approximate joint diagonalization of the covariance matrix and three complementary covariance matrices of the observations. Thus, we have proposed a practical quaternion ML-ICA algorithm based on the local approximation of the cost function, which is a measure of the entropy loss due to the residual correlation among the separated sources. The proposed algorithm can be seen as a quasi-Newton method and, despite the nonconvexity of the ML-ICA cost function, it converges very fast to a solution of the quaternion ICA problem, which has been illustrated by means of several simulation examples.

[13]The extension of the considered ICA model to the case of noisy mixtures is an interesting topic for future research.

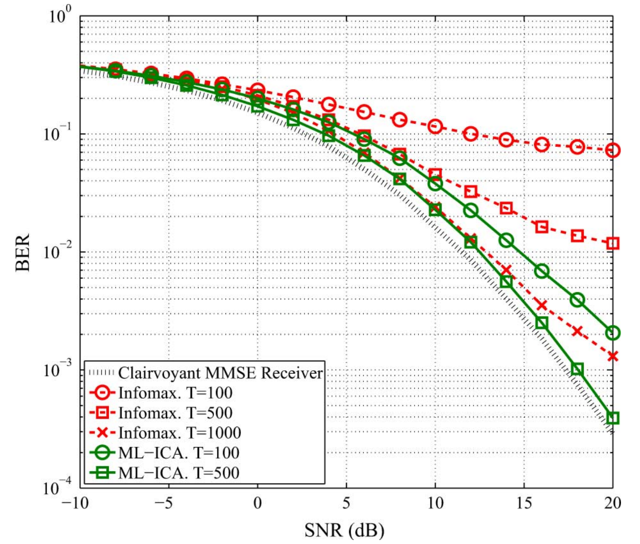

Fig. 7. Blind decoding in multiuser Alamouti systems based on quaternion ICA. Results for the MMSE receiver with perfect channel knowledge, the proposed ML-ICA algorithm, and the quaternion extension of the Infomax. Two-user system with QPSK constellations and different numbers T of received Alamouti blocks.

Appendix A
Proof of Theorem 1

Let us start by the trivial ambiguities. Given the ICA model $\mathbf{x} = \mathbf{A}\mathbf{s}$, it is easy to see that for all permutation $\mathbf{P} \in \mathbb{R}^{m\times m}$ and invertible diagonal matrices $\mathbf{\Phi} \in \mathbb{H}^{m\times m}$, we have

$$\mathbf{x} = \mathbf{A}\mathbf{s} = \mathbf{A}\mathbf{P}\mathbf{\Phi}\mathbf{\Phi}^{-1}\mathbf{P}^T\mathbf{s} = \tilde{\mathbf{A}}\tilde{\mathbf{s}} \tag{82}$$

where $\tilde{\mathbf{A}} = \mathbf{A}\mathbf{P}\mathbf{\Phi}$ and $\tilde{\mathbf{s}} = \mathbf{\Phi}^{-1}\mathbf{P}^{\mathrm{T}}\mathbf{s}$ is a quaternion vector with independent entries. Therefore, the permutation and scale factor ambiguities are unavoidable without exploiting some other properties of the sources or the mixing matrix.

As previously pointed out, the scale ambiguity allows us to focus on unit-variance sources. With this assumption and taking into account (31), it is easy to see that any solution $\hat{\mathbf{A}}$ of the ICA model is related to the true mixing matrix by

$$\hat{\mathbf{A}} = \mathbf{A}\mathbf{Q} \tag{83}$$

where $\mathbf{Q} \in \mathbb{H}^{m\times m}$ is a unitary matrix. Moreover, exploiting the trivial ambiguities, we can introduce a permutation and quaternion phase change (product by a unit quaternion) in the columns of $\hat{\mathbf{A}}$ to obtain an ambiguity matrix $\mathbf{Q}$ with real and positive diagonal elements.

Now, from (32)–(34) and Lemma 2, we can see that $\hat{\mathbf{A}} = \mathbf{AQ}$ is a solution of the ICA model iff

$$\mathbf{Q}^{\mathrm{H}}\boldsymbol{\Lambda}_\nu \mathbf{Q}^{(\nu)} = \hat{\boldsymbol{\Lambda}}_\nu \qquad \forall\nu \tag{84}$$

where $\boldsymbol{\Lambda}_\nu$ and $\hat{\boldsymbol{\Lambda}}_\nu$ are the diagonal complementary covariance matrices of the sources for the true $(\mathbf{A})$ and alternative $(\hat{\mathbf{A}})$ ICA solutions. Equivalently, the above equation can be rewritten as

$$\mathbf{Q}^{\mathrm{H}}\boldsymbol{\Psi}_\nu \mathbf{Q} = \hat{\boldsymbol{\Psi}}_\nu \qquad \forall\nu \tag{85}$$

where $\boldsymbol{\Psi}_\nu = \boldsymbol{\Lambda}_\nu \nu = \mathrm{diag}([\psi_{s_1,\nu}, \ldots, \psi_{s_m,\nu}])$ contains one of the elements of the properness profiles of all the sources and $\hat{\boldsymbol{\Psi}}_\nu$ is defined analogously.

Let us now focus on the first row and column of the ambiguity matrix $\mathbf{Q}$. In particular, we will write

$$\mathbf{Q} = \begin{bmatrix} q_1 & \mathbf{v}^{\mathrm{H}} \\ \mathbf{w} & \mathbf{Q}_{-1} \end{bmatrix} \tag{86}$$

where q_1 is a real and positive number, $\mathbf{v}, \mathbf{w} \in \mathbb{H}^{(m-1)\times 1}$, and $\mathbf{Q}_{-1} \in \mathbb{H}^{(m-1)\times(m-1)}$. Here, it is clear that the unitarity of $\mathbf{Q}$ implies

$$\mathbf{v}q_1 + \mathbf{Q}_{-1}^{\mathrm{H}}\mathbf{w} = \mathbf{0}_{(m-1)\times 1} \tag{87}$$
$$q_1^2 + \|\mathbf{w}\|^2 = 1 \tag{88}$$
$$\mathbf{w}\mathbf{w}^{\mathrm{H}} + \mathbf{Q}_{-1}\mathbf{Q}_{-1}^{H} = \mathbf{I}_{m-1}. \tag{89}$$

Analogously, the diagonal matrix $\boldsymbol{\Psi}_\nu$ can be written as $\boldsymbol{\Psi}_\nu = \mathrm{diag}([\psi_{s_1,\nu}, \boldsymbol{\psi}_{s_{-1},\nu}^{\mathrm{T}}]^{\mathrm{T}})$, where $\boldsymbol{\psi}_{s_{-1},\nu} = [\psi_{s_2,\nu}, \ldots, \psi_{s_m,\nu}]$. Thus, considering the first column of $\hat{\boldsymbol{\Psi}}_\nu$, we can see that (85) implies

$$\mathbf{v}\psi_{s_1,\nu}q_1 + \mathbf{Q}_{-1}^{\mathrm{H}}\mathrm{diag}(\boldsymbol{\psi}_{s_{-1},\nu})\mathbf{w} = \mathbf{0}_{(m-1)\times 1} \qquad \forall\nu. \tag{90}$$

Moreover, taking into account the property $ab = ba^{(b^*)}$ and noting that q_1 is a real scalar, the above equation can be rewritten as

$$\mathbf{v}q_1\psi_{s_1,\nu} + \mathbf{Q}_{-1}^{\mathrm{H}}\mathrm{diag}(\mathbf{w})\boldsymbol{\psi}_{s_{-1},\nu}^{(\mathbf{w}^*)} = \mathbf{0}_{(m-1)\times 1} \qquad \forall\nu \tag{91}$$

where, with a slight abuse of notation, $\boldsymbol{\psi}_{s_{-1},\nu}^{(\mathbf{w}^*)}$ denotes the elementwise rotation of the entries in $\boldsymbol{\psi}_{s_{-1},\nu}$. Now, defining $\mathbf{1} \in \mathbb{R}^{(m-1)\times 1}$ as the vector of ones and using (87), we have

$$\mathbf{Q}_{-1}^{\mathrm{H}}\mathrm{diag}(\mathbf{w})\left(\boldsymbol{\psi}_{s_{-1},\nu}^{(\mathbf{w}^*)} - \mathbf{1}\psi_{s_1,\nu}\right) = \mathbf{0}_{(m-1)\times 1} \qquad \forall\nu. \tag{92}$$

Additionally, noting that $q_1 > 0$, the combination of (88) and (89) ensures that $\mathbf{Q}_{-1}$ is invertible, which allows us to rewrite the above equation as

$$\mathrm{diag}(\mathbf{w})\left(\boldsymbol{\psi}_{s_{-1},\nu}^{(\mathbf{w}^*)} - \mathbf{1}\psi_{s_1,\nu}\right) = \mathbf{0}_{(m-1)\times 1} \qquad \forall\nu \tag{93}$$

or equivalently, for $k = 2, \ldots, m$

$$q_{k,1}(\psi_{s_k,\nu}^{(q_{k,1}^*)} - \psi_{s_1,\nu}) = 0 \qquad \forall\nu \tag{94}$$

where $q_{k,1}$ is the kth element in the first column of $\mathbf{Q}$. Therefore, since the above equation holds for all ν, we can conclude that if the properness profiles $\boldsymbol{\psi}_{s_1}$ and $\boldsymbol{\psi}_{s_k}$ are not rotationally equivalent, then $q_{k,1} = 0$.

Finally, following the same reasoning for the remaining rows and columns of $\mathbf{Q}$, we can see that, excluding the trivial ambiguities, the only possible indeterminacies are given by a unitary quaternion matrix affecting the sources with rotationally equivalent properness profiles. In fact, assuming a set of K sources $\mathbf{s} = [s_1, \ldots, s_K]^{\mathrm{T}}$ with rotationally equivalent properness profiles

$$\boldsymbol{\psi}_{s_1} = \boldsymbol{\psi}_{s_2}^{(a_2)} = \cdots = \boldsymbol{\psi}_{s_K}^{(a_K)} \tag{95}$$

we can easily see that the associated matrix $\boldsymbol{\Psi}_\nu = \mathrm{diag}([\psi_{s_1,\nu}, \ldots, \psi_{s_K,\nu}])$ can be written as

$$\boldsymbol{\Psi}_\nu = \mathrm{diag}(\mathbf{a})\psi_{s_1,\nu}\mathrm{diag}(\mathbf{a})^{\mathrm{H}} \qquad \forall\nu \tag{96}$$

where $\mathbf{a} = [1, a_2/|a_2|, \ldots, a_K/|a_K|]$. Therefore, any linear transformation of the form

$$\hat{\mathbf{s}} = \tilde{\mathbf{Q}}\mathrm{diag}(\mathbf{a})^{\mathrm{H}}\mathbf{s} \tag{97}$$

with $\tilde{\mathbf{Q}} \in \mathbb{R}^{K\times K}$ a real unitary matrix will satisfy the ambiguity condition in (85), i.e., the indeterminacy affecting the sources with rotationally equivalent properness profiles cannot be avoided without exploiting some additional property of the sources or the mixing matrix.

Appendix B
Proof of Lemma 6

Let us start by noting that the matrix update $\tilde{\mathbf{W}} \leftarrow \left(\mathbf{I}_{4m} + \tilde{\boldsymbol{\Delta}}\right)\tilde{\mathbf{W}}$ yields a cost function $J = -1/2\ln\left|\hat{\boldsymbol{\Phi}}_{\tilde{\mathbf{y}},\tilde{\mathbf{y}}}\right|$ for the tth iteration of the algorithm

$$J = -\frac{1}{2}\ln\left|\left(\mathbf{I}_{4m} + \tilde{\boldsymbol{\Delta}}\right)\hat{\mathbf{R}}_{\tilde{\mathbf{y}},\tilde{\mathbf{y}}}\left(\mathbf{I}_{4m} + \tilde{\boldsymbol{\Delta}}\right)^{\mathrm{H}}\right| + \frac{1}{2}\ln\left|\mathrm{blkdiag}_4\left[\left(\mathbf{I}_{4m} + \tilde{\boldsymbol{\Delta}}\right)\hat{\mathbf{R}}_{\tilde{\mathbf{y}},\tilde{\mathbf{y}}}\left(\mathbf{I}_{4m} + \tilde{\boldsymbol{\Delta}}\right)^{\mathrm{H}}\right]\right| \tag{98}$$

where $\hat{\mathbf{R}}_{\tilde{\mathbf{y}},\tilde{\mathbf{y}}}$ is the reordered augmented covariance matrix of the estimated sources after the $(t-1)$th iteration. Thus, the cost function can be written as

$$J = -\frac{1}{2}\ln\left|\hat{\mathbf{R}}_{\tilde{\mathbf{y}},\tilde{\mathbf{y}}} + \tilde{\boldsymbol{\Delta}}\hat{\mathbf{R}}_{\tilde{\mathbf{y}},\tilde{\mathbf{y}}} + \hat{\mathbf{R}}_{\tilde{\mathbf{y}},\tilde{\mathbf{y}}}\tilde{\boldsymbol{\Delta}}^{\mathrm{H}} + \tilde{\boldsymbol{\Delta}}\hat{\mathbf{R}}_{\tilde{\mathbf{y}},\tilde{\mathbf{y}}}\tilde{\boldsymbol{\Delta}}^{\mathrm{H}}\right| + \frac{1}{2}\ln\left|\mathrm{blkdiag}_4\left(\hat{\mathbf{R}}_{\tilde{\mathbf{y}},\tilde{\mathbf{y}}} + \tilde{\boldsymbol{\Delta}}\hat{\mathbf{R}}_{\tilde{\mathbf{y}},\tilde{\mathbf{y}}} + \hat{\mathbf{R}}_{\tilde{\mathbf{y}},\tilde{\mathbf{y}}}\tilde{\boldsymbol{\Delta}}^{\mathrm{H}} + \tilde{\boldsymbol{\Delta}}\hat{\mathbf{R}}_{\tilde{\mathbf{y}},\tilde{\mathbf{y}}}\tilde{\boldsymbol{\Delta}}^{\mathrm{H}}\right)\right| \tag{99}$$

and using the matrix decomposition $\hat{\mathbf{R}}_{\tilde{\mathbf{y}},\tilde{\mathbf{y}}} = \hat{\mathbf{D}}_{\tilde{\mathbf{y}},\tilde{\mathbf{y}}}^{1/2}\hat{\boldsymbol{\Phi}}_{\tilde{\mathbf{y}},\tilde{\mathbf{y}}}\hat{\mathbf{D}}_{\tilde{\mathbf{y}},\tilde{\mathbf{y}}}^{1/2}$, we can obtain

$$J = -\frac{1}{2}\ln\left|\hat{\boldsymbol{\Phi}}_{\tilde{\mathbf{y}},\tilde{\mathbf{y}}} + \mathbf{S}\right| + \frac{1}{2}\ln\left|\mathbf{I}_{4m} + \mathrm{blkdiag}_4(\mathbf{S})\right| \tag{100}$$

where $\mathbf{S} = \tilde{\boldsymbol{\Theta}}\boldsymbol{\Phi}_{\tilde{\mathbf{y}},\tilde{\mathbf{y}}} + \boldsymbol{\Phi}_{\tilde{\mathbf{y}},\tilde{\mathbf{y}}}\tilde{\boldsymbol{\Theta}}^{H} + \tilde{\boldsymbol{\Theta}}\boldsymbol{\Phi}_{\tilde{\mathbf{y}},\tilde{\mathbf{y}}}\tilde{\boldsymbol{\Theta}}^{\mathrm{H}}$ and $\tilde{\boldsymbol{\Theta}} = \hat{\mathbf{D}}_{\tilde{\mathbf{y}},\tilde{\mathbf{y}}}^{-1/2}\tilde{\boldsymbol{\Delta}}\hat{\mathbf{D}}_{\tilde{\mathbf{y}},\tilde{\mathbf{y}}}^{1/2}$.

Now, writing $\hat{\mathbf{\Phi}}_{\tilde{\mathbf{y}},\tilde{\mathbf{y}}} = \mathbf{I}_{4m} + \mathrm{offdiag}_4\left(\hat{\mathbf{\Phi}}_{\tilde{\mathbf{y}},\tilde{\mathbf{y}}}\right)$, assuming that $\|\mathrm{offdiag}_4(\hat{\mathbf{\Phi}}_{\tilde{\mathbf{y}},\tilde{\mathbf{y}}})\|^2 \ll 1$ and noting that $\|\tilde{\mathbf{\Delta}}\|^2 \ll 1$ implies $\|\tilde{\mathbf{\Theta}}\|^2 \ll 1$, we can use the approximation

$$\begin{aligned}\mathbf{S} =& \tilde{\mathbf{\Theta}} + \tilde{\mathbf{\Theta}}^{\mathrm{H}} + \tilde{\mathbf{\Theta}}\tilde{\mathbf{\Theta}}^{\mathrm{H}} + \tilde{\mathbf{\Theta}}\mathrm{offdiag}_4\left(\hat{\mathbf{\Phi}}_{\tilde{\mathbf{y}},\tilde{\mathbf{y}}}\right)\\ &+ \mathrm{offdiag}_4\left(\hat{\mathbf{\Phi}}_{\tilde{\mathbf{y}},\tilde{\mathbf{y}}}\right)\tilde{\mathbf{\Theta}}^{\mathrm{H}} + \tilde{\mathbf{\Theta}}\mathrm{offdiag}_4\left(\hat{\mathbf{\Phi}}_{\tilde{\mathbf{y}},\tilde{\mathbf{y}}}\right)\tilde{\mathbf{\Theta}}^{\mathrm{H}}\\ \simeq& \tilde{\mathbf{\Theta}} + \tilde{\mathbf{\Theta}}^{\mathrm{H}}. \end{aligned} \tag{101}$$

Furthermore, using the second-order Taylor's expansion of the determinant logarithm, we can write

$$\begin{aligned} J \simeq& -\frac{1}{2}\mathrm{Tr}\left(\hat{\mathbf{\Phi}}_{\tilde{\mathbf{y}},\tilde{\mathbf{y}}} + \mathbf{S} - \mathbf{I}_{4m}\right) + \frac{1}{4}\left\|\hat{\mathbf{\Phi}}_{\tilde{\mathbf{y}},\tilde{\mathbf{y}}} + \mathbf{S} - \mathbf{I}_{4m}\right\|^2\\ &+ \frac{1}{2}\mathrm{Tr}(\mathbf{S}) - \frac{1}{4}\|\mathrm{blkdiag}_4(\mathbf{S})\|^2\\ =& \frac{1}{4}\left\|\mathrm{offdiag}_4\left(\hat{\mathbf{\Phi}}_{\tilde{\mathbf{y}},\tilde{\mathbf{y}}} + \mathbf{S}\right)\right\|^2\\ \simeq& \frac{1}{4}\left\|\mathrm{offdiag}_4\left(\hat{\mathbf{\Phi}}_{\tilde{\mathbf{y}},\tilde{\mathbf{y}}} + \tilde{\mathbf{\Theta}} + \tilde{\mathbf{\Theta}}^{\mathrm{H}}\right)\right\|^2 \end{aligned} \tag{102}$$

and taking into account $\tilde{\mathbf{\Theta}} = \hat{\mathbf{D}}_{\tilde{\mathbf{y}},\tilde{\mathbf{y}}}^{-1/2}\tilde{\mathbf{\Delta}}\hat{\mathbf{D}}_{\tilde{\mathbf{y}},\tilde{\mathbf{y}}}^{1/2}$, we obtain

$$J \simeq \frac{1}{4}\|\mathbf{J}(\mathbf{\Delta})\|^2 \tag{103}$$

where $\mathbf{J}(\mathbf{\Delta}) \in \mathbb{H}^{4m\times 4m}$ is given by

$$\begin{aligned}\mathbf{J}(\mathbf{\Delta}) =& \mathrm{offdiag}_4\left(\hat{\mathbf{\Phi}}_{\tilde{\mathbf{y}},\tilde{\mathbf{y}}}\right) + \hat{\mathbf{D}}_{\tilde{\mathbf{y}},\tilde{\mathbf{y}}}^{-1/2}\mathrm{offdiag}_4\left(\tilde{\mathbf{\Delta}}\right)\hat{\mathbf{D}}_{\tilde{\mathbf{y}},\tilde{\mathbf{y}}}^{1/2}\\ &+ \hat{\mathbf{D}}_{\tilde{\mathbf{y}},\tilde{\mathbf{y}}}^{1/2}\mathrm{offdiag}_4\left(\tilde{\mathbf{\Delta}}^{\mathrm{H}}\right)\hat{\mathbf{D}}_{\tilde{\mathbf{y}},\tilde{\mathbf{y}}}^{-1/2}. \end{aligned} \tag{104}$$

REFERENCES

[1] W. R. Hamilton, "On quaternions," in *Proc. Roy. Irish Acad.*, Nov. 11, 1844, pp. 1–16.
[2] J. P. Ward, *Quaternions and Cayley Numbers: Algebra and Applications*. Dordrecht, The Netherlands: Kluwer, 1997.
[3] A. J. Hanson, *Visualizing Quaternions*. San Mateo, CA: Morgan Kaufmann, 2005.
[4] L. Fortuna, G. Muscato, and M. Xibilia, "A comparison between HMLP and HRBF for attitude control," *IEEE Trans. Neural Netw.*, vol. 12, no. 2, pp. 318–328, Mar. 2001.
[5] J. B. Kuipers, *Quaternions and Rotation Sequences: A Primer With Applications to Orbits, Aerospace and Virtual Reality*. Princeton, NJ: Princeton Univ. Press, 2002.
[6] S.-C. Pei and C.-M. Cheng, "Color image processing by using binary quaternion-moment-preserving thresholding technique," *IEEE Trans. Image Process.*, vol. 8, no. 5, pp. 614–628, May 1999.
[7] C. Moxey, S. Sangwine, and T. Ell, "Hypercomplex correlation techniques for vector images," *IEEE Trans. Signal Process.*, vol. 51, no. 7, pp. 1941–1953, Jul. 2003.
[8] T. Bulow and G. Sommer, "Hypercomplex signals-a novel extension of the analytic signal to the multidimensional case," *IEEE Trans. Signal Process.*, vol. 49, no. 11, pp. 2844–2852, Nov. 2001.
[9] M. Felsberg and G. Sommer, "The monogenic signal," *IEEE Trans. Signal Process.*, vol. 49, no. 12, pp. 3136–3144, Dec. 2001.
[10] C. C. Took and D. P. Mandic, "The quaternion LMS algorithm for adaptive filtering of hypercomplex processes," *IEEE Trans. Signal Process.*, vol. 57, no. 4, pp. 1316–1327, Apr. 2009.
[11] N. ur Rehman and D. P. Mandic, "Empirical mode decomposition for trivariate signals," *IEEE Trans. Signal Process.*, vol. 58, no. 3, pp. 1059–1068, Mar. 2010.
[12] J. Seberry, K. Finlayson, S. Adams, T. Wysocki, T. Xia, and B. Wysocki, "The theory of quaternion orthogonal designs," *IEEE Trans. Signal Process.*, vol. 56, no. 1, pp. 256–265, Jan. 2008.
[13] J.-C. Belfiore and G. Rekaya, "Quaternionic lattices for space-time coding," in *Proc. IEEE Inf. Theory Workshop*, Apr. 2003, pp. 267–270.
[14] J.-C. Belfiore, G. Rekaya, and E. Viterbo, "The golden code: A 2 × 2 full-rate space-tie code with non-vanishing determinant," *IEEE Trans. Inf. Theory*, vol. 51, no. 4, pp. 1432–1436, Apr. 2005.
[15] M.-Y. Chen, H.-C. Li, and S.-C. Pei, "Algebraic identification for optimal nonorthogonality 4 × 4 complex space-time block codes using tensor product on quaternions," *IEEE Trans. Inf. Theory*, vol. 51, no. 1, pp. 324–330, Jan. 2005.
[16] C. Hollanti, J. Lahtonen, K. Ranto, R. Vehkalahti, and E. Viterbo, "On the algebraic structure of the silver code: A 2 × 2 perfect space-time block code," in *Proc. IEEE ITW*, May 2008, pp. 91–94.
[17] S. Sirianunpiboon, A. Calderbank, and S. Howard, "Bayesian analysis of interference cancellation for Alamouti multiplexing," *IEEE Trans. Inf. Theory*, vol. 54, no. 10, pp. 4755–4761, Oct. 2008.
[18] 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.
[19] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*. New York: Wiley, 2001.
[20] N. Le Bihan and S. Buchholz, "Quaternionic idependent component analysis using hypercomplex nonlinearities," presented at the IMA Conf. Math. Signal Process., Cirenceister, U.K., 2006.
[21] G. M. Menanno and N. L. Bihan, "Quaternion polynomial matrix diagonalization for the separation of polarized convolutive mixture," *Signal Process.*, vol. 90, no. 7, pp. 2219–2231, Jul. 2010.
[22] S. Alamouti, "A simple transmit diversity technique for wireless communications," *IEEE J. Sel. Areas Commun.*, vol. 16, no. 8, pp. 1451–1458, 1998.
[23] S. Barbarossa, *Multiantenna Wireless Communication Systems*. Boston, MA: Artech House, 2005.
[24] 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.
[25] P. Schreier, L. Scharf, and C. Mullis, "Detection and estimation of improper complex random signals," *IEEE Trans. Inf. Theory*, vol. 51, no. 1, pp. 306–312, Jan. 2005.
[26] P. Schreier, L. Scharf, and A. Hanssen, "A generalized likelihood ratio test for impropriety of complex signals," *IEEE Signal Process. Lett.*, vol. 13, no. 7, pp. 433–436, Jul. 2006.
[27] 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.
[28] D.-T. Pham, "Joint approximate diagonalization of positive definite hermitian matrices," *SIAM J. Matrix Anal. Appl.*, vol. 22, no. 4, pp. 1136–1152, 2000.
[29] D.-T. Pham and J.-F. Cardoso, "Blind separation of instantaneous mixtures of nonstationary sources," *IEEE Trans. Signal Process.*, vol. 49, no. 9, pp. 1837–1848, Sep. 2001.
[30] A. Ziehe, P. Laskov, G. Nolte, and K. R. Müller, "A fast algorithm for joint diagonalization with non-orthogonal transformations and its application to blind source separation," *J. Mach. Learn. Res.*, vol. 5, pp. 777–800, 2004.
[31] R. Vollgraf and K. Obermayer, "Quadratic optimization for simultaneous matrix diagonalization," *IEEE Trans. Signal Process.*, vol. 54, no. 9, pp. 3270–3278, Sep. 2006.
[32] X.-L. Li and X.-D. Zhang, "Nonorthogonal joint diagonalization free of degenerate solution," *IEEE Trans. Signal Process.*, vol. 55, no. 5, pp. 1803–1814, May 2007.
[33] P. Tichavsky and A. Yeredor, "Fast approximate joint diagonalization incorporating weight matrices," *IEEE Trans. Signal Process.*, vol. 57, no. 3, pp. 878–891, Mar. 2009.
[34] A. Souloumiac, "Nonorthogonal joint diagonalization by combining givens and hyperbolic rotations," *IEEE Trans. Signal Process.*, vol. 57, no. 6, pp. 2222–2231, Jun. 2009.
[35] J. Nocedal and S. J. Wright, *Numerical Optimization*, ser. Operations Research. New York: Springer-Verlag, 1999.
[36] M. Novey and T. Adali, "Complex ICA by negentropy maximization," *IEEE Trans. Neural Netw.*, vol. 19, no. 4, pp. 596–609, Apr. 2008.
[37] T. Adali, H. Li, M. Novey, and J.-F. Cardoso, "Complex ICA using nonlinear functions," *IEEE Trans. Signal Process.*, vol. 56, no. 9, pp. 4536–4544, Sep. 2008.
[38] H. Li and T. Adali, "A class of complex ICA algorithms based on the kurtosis cost function," *IEEE Trans. Neural Netw.*, vol. 19, no. 3, pp. 408–420, Mar. 2008.
[39] S. Buchholz and N. L. Bihan, "Polarized signal classification by complex and quaternionic multi-layer perceptrons," *Int. J. Neural Syst.*, vol. 18, no. 2, pp. 75–85, 2008.

[40] 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, 1995.
[41] T. Cover and J. Thomas*, Elements of Information Theory*. New York: Wiley, 1991.
[42] F. Zhang, "Quaternions and matrices of quaternions," *Linear Algebra Appl.*, vol. 251, pp. 21–57, 1997.
[43] N. L. Bihan and J. Mars, "Singular value decomposition of quaternion matrices: A new tool for vector-sensor signal processing," *Signal Process.*, vol. 84, no. 7, pp. 1177–1199, 2004.
[44] T. Ell and S. Sangwine, "Hypercomplex Fourier transforms of color images," *IEEE Trans. Image Process.*, vol. 16, no. 1, pp. 22–35, Jan. 2007.
[45] F. Neeser and J. Massey, "Proper complex random processes with applications to information theory," *IEEE Trans. Inf. Theory*, vol. 39, no. 4, pp. 1293–1302, Jul. 1993.
[46] A. van den Bos, "The multivariate complex normal distribution-a generalization," *IEEE Trans. Inf. Theory*, vol. 41, no. 2, pp. 537–539, Mar. 1995.
[47] B. Picinbono, "Second-order complex random vectors and normal distributions," *IEEE Trans. Signal Process.*, vol. 44, no. 10, pp. 2637–2640, Oct. 1996.
[48] P. Schreier and 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.
[49] N. N. Vakhania, "Random vectors with values in quaternion Hilbert spaces," *Theory Probab. Appl.*, vol. 43, no. 1, pp. 99–115, 1999.
[50] P. Amblard and N. Le Bihan, "On properness of quaternion valued random variables," in *Proc. IMA Conf. Math. Signal Process.*, Cirencester, U.K., 2004, pp. 23–26.
[51] E. Ollila, "On the circularity of a complex random variable," *IEEE Signal Process. Lett.*, vol. 15, pp. 841–844, 2008.
[52] J. Eriksson, E. Ollila, and V. Koivunen, "Essential statistics and tools for complex random variables," *IEEE Trans. Signal Process.*, vol. 58, no. 10, pp. 5400–5408, Oct. 2010.
[53] E. L. Lehmann and J. P. Romano*, Testing Statistical Hypothesis*, 3rd ed. New York: Springer, 2005.
[54] A. Walden and P. Rubin-Delanchy, "On testing for impropriety of complex-valued Gaussian vectors," *IEEE Trans. Signal Process.*, vol. 57, no. 3, pp. 825–834, Mar. 2009.
[55] P. J. Schreier and L. L. Scharf*, Statistical Signal Processing of Complex-Valued Data: The Theory of Improper and Noncircular Signals*. Cambridge, U.K.: Cambridge Univ. Press, 2010.
[56] J. Cardoso, "Unsupervised adaptive filtering, volume 1, blind source separation," in *Ch. Entropic Contrasts for Source Separation: Geometry and Stability*. New York: Wiley, 2000, vol. 1, pp. 139–190.
[57] J. Vía, I. Santamaría, and J. Pérez, "A learning algorithm for adaptive canonical correlation analysis of several data sets," *Neural Netw.*, vol. 20, no. 1, pp. 139–152, Jan. 2007.
[58] J. R. Kettenring, "Canonical analysis of several sets of variables," *Biometrika*, vol. 58, no. 3, pp. 433–451, 1971.
[59] S.-I. Amari and H. Nagaoka*, Methods of Information Geometry*, A. M. Society, Ed. Oxford, U.K.: Oxford Univ. Press, 1993.
[60] F. R. Bach and M. I. Jordan, "Kernel independent component analysis," *J. Mach. Learn. Res.*, vol. 3, pp. 1–48, 2002.
[61] J. Vía and I. Santamaría, "On the blind identifiability of orthogonal space-time block codes from second order statistics," *IEEE Trans. Inf. Theory*, vol. 54, no. 2, pp. 709–722, Feb. 2008.
[62] J. Vía and I. Santamaría, "Correlation matching approaches for blind OSTBC channel estimation," *IEEE Trans. Signal Process.*, vol. 56, no. 12, pp. 5950–5961, Dec. 2008.

Javier Vía (S'04–M'08) received the Telecommunication Engineer Degree and Ph.D. degree in electrical engineering from the University of Cantabria, Santander, Spain, in 2002 and 2007, respectively.

In 2002, he joined the Department of Communications Engineering, University of Cantabria, where he is currently an Assistant Professor. He has spent visiting periods with the Smart Antennas Research Group, Stanford University, Stanford, CA, and with the Department of Electronics and Computer Engineering, Hong Kong University of Science and Technology, Hong Kong. He has actively participated in several European and Spanish research projects. His current research interests include blind channel estimation and equalization in wireless communication systems, multivariate statistical analysis, quaternion signal processing, and kernel methods.

Daniel P. Palomar (S'99–M'03–SM'08) received the Electrical Engineering and Ph.D. degrees (both with honors) from the Technical University of Catalonia (UPC), Barcelona, Spain, in 1998 and 2003, respectively.

He is an Associate Professor with the Department of Electronic and Computer Engineering, Hong Kong University of Science and Technology (HKUST), Hong Kong, which he joined in 2006. He had previously held several research appointments, namely at King's College London (KCL), London, U.K.; UPC; Stanford University, Stanford, CA; the Telecommunications Technological Center of Catalonia (CTTC), Barcelona, Spain; the Royal Institute of Technology (KTH), Stockholm, Sweden; the University of Rome "La Sapienza," Rome, Italy; and Princeton University, Princeton, NJ. His current research interests include applications of convex optimization theory, game theory, and variational inequality theory to signal processing and communications.

Dr. Palomar is an Associate Editor of the IEEE TRANSACTIONS ON INFORMATION THEORY and has been an Associate Editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING, a Guest Editor of the *IEEE Signal Processing Magazine* 2010 Special Issue on "Convex Optimization for Signal Processing," a Guest Editor of the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS 2008 Special Issue on "Game Theory in Communication Systems," and the Lead Guest Editor of the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS 2007 Special Issue on "Optimization of MIMO Transceivers for Realistic Communication Networks." He serves on the IEEE Signal Processing Society Technical Committee on Signal Processing for Communications (SPCOM). He was the General Co-Chair of the 2009 IEEE Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP). He is a recipient of a 2004–2006 Fulbright Research Fellowship; the 2004 Young Author Best Paper Award by the IEEE Signal Processing Society; the 2002–2003 best Ph.D. prize in Information Technologies and Communications by the UPC; the 2002–2003 Rosina Ribalta first prize for the Best Doctoral Thesis in Information Technologies and Communications by the Epson Foundation, and the 2004 prize for the Best Doctoral Thesis in Advanced Mobile Communications by the Vodafone Foundation and COIT.

Luis Vielva was born in Santander, Spain, in 1966. He received the Licenciado and Ph.D. degrees in physics from the University of Cantabria, Santander, Spain, in 1989 and 1997, respectively.

In 1989, he joined the Department of Communications Engineering, University of Cantabria, where he is currently an Associate Professor. His current research interests include systems biology, convex optimization, and geometric algebra.

Ignacio Santamaría (M'96–SM'05) received the Telecommunication Engineer Degree and Ph.D. degree in electrical engineering from the Universidad Politécnica de Madrid (UPM), Madrid, Spain, in 1991 and 1995, respectively.

In 1992, he joined the Department of Communications Engineering, University of Cantabria, Santander, Spain, where he is currently a Full Professor. He has been a Visiting Researcher with the Computational NeuroEngineering Laboratory, University of Florida, Gainesville, and with the Wireless Networking and Communications Group , University of Texas at Austin. He has more than 100 publications in refereed journals and international conference papers. He has been involved in several national and international research projects on his current research interests including signal processing algorithms for wireless communication systems, MIMO systems, multivariate statistical techniques, and machine learning theories.

Dr. Santamaría is currently serving as a member of the Machine Learning for Signal Processing Technical Committee of the IEEE Signal Processing Society.