MAXIMUM LIKELIHOOD ICA OF QUATERNION GAUSSIAN VECTORS

Javier Vía[⋆] *Daniel P. Palomar*[†] *Luis Vielva*[⋆] *Ignacio Santamaría*[⋆]

[⋆]Dept. of Communications Engineering. University of Cantabria. Spain
[†]Dept. of Electronic and Computer Engineering. Hong Kong University of Science and Technology

ABSTRACT

This work considers the independent component analysis (ICA) of quaternion random vectors. In particular, we focus on the Gaussian case, and therefore the ICA problem is solved by exclusively exploiting the second-order statistics (SOS) of the observations. In the quaternion case, the SOS of a random vector are given by the covariance matrix and three complementary covariance matrices. Thus, quaternion ICA amounts to jointly diagonalizing these four matrices. Following a maximum likelihood (ML) approach, we show that the ML-ICA problem reduces to the minimization of a cost function, which can be interpreted as a measure of the entropy loss due to the correlation among the estimated sources. In order to solve the non-convex ML-ICA problem, we propose a practical quasi-Newton algorithm based on quadratic local approximations of the cost function. Finally, the practical performance and potential application of the proposed technique is illustrated by means of numerical examples.

Index Terms— Independent component analysis, quaternion, properness, maximum likelihood, second-order statistics

1. INTRODUCTION

The interest in quaternion signal processing has increased in the last years due to its applications in image processing [1], wind modeling [2], and design (and processing) of space-time block codes [3]. This increasing popularity 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 [4] 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) [5] of quaternion random vectors has received limited attention [6, 7], even though it can be considered (together with PCA) as the most important multivariate statistical analysis technique.

In a previous work [7], we have shown that under mild assumptions on the properness of the quaternion random vector, the ICA problem can be solved from the second-order statistics (SOS) of the observations. In this work, we focus on the derivation of a practical quaternion ICA algorithm for Gaussian data, which requires the joint diagonalization of the quaternion covariance matrix and three complementary covariance matrices [4]. In particular, following a maximum likelihood (ML) approach, we show that the ML-ICA problem reduces to the minimization of a non-convex cost function, which can be interpreted as the entropy loss due to the correlations among the estimated sources. In order to solve the non-convex optimization problem, we propose a quasi-Newton algorithm based on local quadratic approximations of the cost function. Finally, some numerical results illustrate the performance and practical application of the proposed algorithm.

This work was supported by the Spanish Government, Ministerio de Ciencia e Innovación (MICINN), under projects MultiMIMO (TEC2007-68020-C04-02), COSIMA (TEC2010-19545-C04-03) and COMONSENS (CSD2008-00010, CONSOLIDER-INGENIO 2010 Program). Additionally, the work of the second author was supported by the Hong Kong RGC 618709 research grant.

2. PRELIMINARIES

Throughout this paper we will use bold-faced upper case letters to denote matrices, bold-faced lower case letters for column vectors, and light-faced lower case letters for scalar quantities. Superscripts $(\cdot)^T$ and $(\cdot)^H$ denote transpose and Hermitian (i.e., transpose and quaternion conjugate), respectively. The notation $\mathbf{A} \in \mathbb{R}^{m\times n}$ (respectively $\mathbf{A} \in \mathbb{H}^{m\times n}$) means that $\mathbf{A}$ is a real (respectively quaternion) $m \times n$ matrix. $\Re(\mathbf{A})$, Tr$(\mathbf{A})$, and $|\mathbf{A}|$ denote the real part, trace and determinant of matrix $\mathbf{A}$. $\mathbf{I}_m$ is the identity matrix of dimension m, $\mathbf{0}_{m\times n}$ is the $m \times n$ zero matrix, and diag$(\mathbf{a})$ denotes the diagonal matrix with vector $\mathbf{a}$ along its diagonal. 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., $\mathbf{R}_{\mathbf{a},\mathbf{b}} = E\mathbf{a}\mathbf{b}^H$.

2.1. Quaternion Algebra

Quaternions are hypercomplex numbers defined by

$$x = r_1 + \eta r_\eta + \eta' r_{\eta'} + \eta'' r_{\eta''},$$

where $r_1, r_\eta, r_{\eta'}, r_{\eta''} \in \mathbb{R}$ are four real numbers, and the three imaginary units (η, η', η'') satisfy

$$\eta^2 = \eta'^2 = \eta''^2 = \eta\eta'\eta'' = -1,$$

which also implies $\eta\eta' = \eta''$, $\eta'\eta'' = \eta$, and $\eta''\eta = \eta'$.

Quaternions form a skew field $\mathbb{H}$ [8], which means that they satisfy the axioms of a field except 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 $x^* = r_1 - \eta r_\eta - \eta' r_{\eta'} - \eta'' r_{\eta''}$, and the inner product of two quaternions x, y is defined as xy^*. Two quaternions are orthogonal if and only if (iff) their scalar product (the real part of the inner product) is zero, and the norm of a quaternion x is $|x| = \sqrt{xx^*} = \sqrt{r_1^2 + r_\eta^2 + r_{\eta'}^2 + r_{\eta''}^2}$. Furthermore, we say that ν is a pure unit quaternion iff $\nu^2 = -1$ (i.e., iff $|\nu| = 1$ and its real part is zero). Finally, the involution of a quaternion x over a pure unit quaternion ν is defined as

$$x^{(\nu)} = -\nu \mathbf{x} \nu,$$

and it represents a rotation of angle π in the imaginary plane orthogonal to ν [8].

2.2. Second-Order Statistics of Quaternion Random Vectors

Let us define[1] $\bar{\mathbf{x}} = \left[\mathbf{x}^T, \mathbf{x}^{(\eta)^T}, \mathbf{x}^{(\eta')^T}, \mathbf{x}^{(\eta'')^T}\right]^T$ as the augmented quaternion vector, which allows us to simplify the second-order statistical analysis of the quaternion vector $\mathbf{x}$ [4]. In particular, the SOS of $\mathbf{x}$ are given by the augmented covariance matrix

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

where we can readily identify the covariance matrix $\mathbf{R}_{\mathbf{x},\mathbf{x}} = E\mathbf{x}\mathbf{x}^H$ and three complementary covariance matrices $\mathbf{R}_{\mathbf{x},\mathbf{x}^{(\eta)}} = E\mathbf{x}\mathbf{x}^{(\eta)^H}$, $\mathbf{R}_{\mathbf{x},\mathbf{x}^{(\eta')}} = E\mathbf{x}\mathbf{x}^{(\eta')^H}$ and $\mathbf{R}_{\mathbf{x},\mathbf{x}^{(\eta'')}} = E\mathbf{x}\mathbf{x}^{(\eta'')^H}$. These matrices have been used in [4] to introduce three different definitions of quaternion properness. However, in this paper we will focus on the strongest kind of quaternion properness ($\mathbb{Q}$-properness), and we will say that a quaternion random vector is proper iff (all) their complementary covariance matrices vanish.

3. INDEPENDENT COMPONENT ANALYSIS OF QUATERNION GAUSSIAN VECTORS

3.1. ICA Model

Consider a quaternion random vector $\mathbf{s} \in \mathbb{H}^{m\times 1}$ representing m independent source signals, which are mixed by a non-singular mixing matrix $\mathbf{A} \in \mathbb{H}^{m\times m}$. That is, we have the model $\mathbf{x} = \mathbf{A}\mathbf{s}$, where $\mathbf{x} \in \mathbb{H}^{m\times 1}$ is a quaternion random vector representing the available observations. Due to the trivial ambiguities of the ICA model (permutations and scale factors) [5], we can assume without loss of generality that the sources are unit-variance quaternion random variables with diagonal complementary covariance matrices $\mathbf{\Lambda}_\eta = E\mathbf{s}\mathbf{s}^{(\eta)^H}$, $\mathbf{\Lambda}_{\eta'} = E\mathbf{s}\mathbf{s}^{(\eta')^H}$, $\mathbf{\Lambda}_{\eta''} = E\mathbf{s}\mathbf{s}^{(\eta'')^H}$, i.e., we have

$$\mathbf{R}_{\bar{\mathbf{s}},\bar{\mathbf{s}}} = E\bar{\mathbf{s}}\bar{\mathbf{s}}^H = \begin{bmatrix} \mathbf{I}_m & \mathbf{\Lambda}_\eta & \mathbf{\Lambda}_{\eta'} & \mathbf{\Lambda}_{\eta''} \\ \mathbf{\Lambda}^{(\eta)}_\eta & \mathbf{I}_m & \mathbf{\Lambda}^{(\eta)}_{\eta''} & \mathbf{\Lambda}^{(\eta)}_{\eta'} \\ \mathbf{\Lambda}^{(\eta')}_{\eta'} & \mathbf{\Lambda}^{(\eta')}_{\eta''} & \mathbf{I}_m & \mathbf{\Lambda}^{(\eta')}_{\eta} \\ \mathbf{\Lambda}^{(\eta'')}_{\eta''} & \mathbf{\Lambda}^{(\eta'')}_{\eta'} & \mathbf{\Lambda}^{(\eta'')}_{\eta} & \mathbf{I}_m \end{bmatrix}. \quad (1)$$

With the above assumptions, and limiting our analysis to SOS-based techniques, the ICA problem amounts to finding the mixing matrix $\mathbf{A}$ and the complementary covariance matrices $\mathbf{\Lambda}_\eta, \mathbf{\Lambda}_{\eta'}, \mathbf{\Lambda}_{\eta''} = E\mathbf{s}\mathbf{s}^{(\eta'')^H}$ satisfying

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

3.2. ML-ICA of Quaternion Gaussian Vectors

Let us consider T vector observations $\mathbf{x}[t] = \mathbf{A}\mathbf{s}[t]$ ($t = 0, \ldots, T-1$) of the ICA model. Thus, assuming that the sources $\mathbf{s}[t]$ are i.i.d. zero-mean quaternion Gaussian vectors with independent elements of unit variance, it can be proved [9] that the ML estimation problem can be written as

$$\underset{\substack{\overline{\mathbf{W}}\in\mathcal{D}, \mathbf{R}_{\bar{\mathbf{s}},\bar{\mathbf{s}}}\in\mathcal{R} \\ \hat{\mathbf{R}}_{\bar{\mathbf{y}},\bar{\mathbf{y}}} = \overline{\mathbf{W}}\hat{\mathbf{R}}_{\bar{\mathbf{x}},\bar{\mathbf{x}}}\overline{\mathbf{W}}^H}}{\text{minimize}} \quad D_{\text{KL}}(\hat{\mathbf{R}}_{\bar{\mathbf{y}},\bar{\mathbf{y}}} \| \mathbf{R}_{\bar{\mathbf{s}},\bar{\mathbf{s}}}), \quad (2)$$

where $D_{\text{KL}}(\hat{\mathbf{R}}_{\bar{\mathbf{y}},\bar{\mathbf{y}}} \| \mathbf{R}_{\bar{\mathbf{s}},\bar{\mathbf{s}}})$ is the Kullback-Leibler divergence between two zero-mean quaternion Gaussian distributions with covariance matrices $\hat{\mathbf{R}}_{\bar{\mathbf{y}},\bar{\mathbf{y}}}$ and $\mathbf{R}_{\bar{\mathbf{s}},\bar{\mathbf{s}}}$ [4]

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

$\hat{\mathbf{R}}_{\bar{\mathbf{x}},\bar{\mathbf{x}}} = \frac{1}{T}\sum_{t=0}^{T-1}\bar{\mathbf{x}}[t]\bar{\mathbf{x}}^H[t]$ is the sample covariance matrix estimator (which is assumed to be non-singular),

$$\overline{\mathbf{W}} = \begin{bmatrix} \mathbf{W} & \mathbf{0}_{m\times m} & \mathbf{0}_{m\times m} & \mathbf{0}_{m\times m} \\ \mathbf{0}_{m\times m} & \mathbf{W}^{(\eta)} & \mathbf{0}_{m\times m} & \mathbf{0}_{m\times m} \\ \mathbf{0}_{m\times m} & \mathbf{0}_{m\times m} & \mathbf{W}^{(\eta')} & \mathbf{0}_{m\times m} \\ \mathbf{0}_{m\times m} & \mathbf{0}_{m\times m} & \mathbf{0}_{m\times m} & \mathbf{W}^{(\eta'')} \end{bmatrix}, \quad (3)$$

$\mathbf{W} = \mathbf{A}^{-1}$ is the estimated separation matrix, and $\mathcal{R}$ and $\mathcal{D}$ are the the sets of quaternion matrices with the structures in (1) and (3).

Summarizing, the ML-ICA problem reduces to the minimization of 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 $\mathbf{R}_{\bar{\mathbf{s}},\bar{\mathbf{s}}}$.

3.3. Reformulation of the ML-ICA Problem

Unfortunately, the cost function in (2) is non-convex. However, taking into account the invariance of the KL divergence under linear transformations [10], we can write $D_{\text{KL}}(\hat{\mathbf{R}}_{\bar{\mathbf{y}},\bar{\mathbf{y}}} \| \mathbf{R}_{\bar{\mathbf{s}},\bar{\mathbf{s}}}) = D_{\text{KL}}(\hat{\mathbf{R}}_{\tilde{\mathbf{y}},\tilde{\mathbf{y}}} \| \mathbf{R}_{\tilde{\mathbf{s}},\tilde{\mathbf{s}}})$, where

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

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

$\mathbf{P}$ is a permutation matrix, $\bar{\mathbf{s}}_k \in \mathbb{H}^{4\times 1}$ denotes the augmented vector for the k-th source s_k (k-th element of $\mathbf{s}$), $\bar{\mathbf{y}}_k \in \mathbb{H}^{4\times 1}$ is defined in a similar way, and $\mathbf{R}_{\bar{\mathbf{s}}_k,\bar{\mathbf{s}}_k}, \hat{\mathbf{R}}_{\bar{\mathbf{y}}_k,\bar{\mathbf{y}}_k} \in \mathbb{H}^{4\times 4}$ are the corresponding augmented covariance matrices.

With the above definitions, it is easy to prove that the matrix $\mathbf{R}_{\tilde{\mathbf{s}},\tilde{\mathbf{s}}}$ minimizing $D_{\text{KL}}(\hat{\mathbf{R}}_{\tilde{\mathbf{y}},\tilde{\mathbf{y}}} \| \mathbf{R}_{\tilde{\mathbf{s}},\tilde{\mathbf{s}}})$ is given by the block-diagonal version of $\hat{\mathbf{R}}_{\tilde{\mathbf{y}},\tilde{\mathbf{y}}}$, i.e.,

$$\mathbf{R}_{\tilde{\mathbf{s}},\tilde{\mathbf{s}}} = \hat{\mathbf{D}}_{\tilde{\mathbf{y}},\tilde{\mathbf{y}}} = \begin{bmatrix} \hat{\mathbf{R}}_{\bar{\mathbf{y}}_1,\bar{\mathbf{y}}_1} & \mathbf{0}_{4\times 4} & \cdots & \mathbf{0}_{4\times 4} \\ \mathbf{0}_{4\times 4} & \hat{\mathbf{R}}_{\bar{\mathbf{y}}_2,\bar{\mathbf{y}}_2} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{0}_{4\times 4} \\ \mathbf{0}_{4\times 4} & \cdots & \mathbf{0}_{4\times 4} & \hat{\mathbf{R}}_{\bar{\mathbf{y}}_m,\bar{\mathbf{y}}_m} \end{bmatrix},$$

[1] From now on, we will use the notation $\mathbf{A}^{(\nu)}$ to denote the element-wise involution of matrix $\mathbf{A}$.

which reduces the ML estimation problem to

$$\underset{\overline{\mathbf{W}}\in\mathcal{D}}{\text{minimize}} \quad -\frac{1}{2}\ln|\hat{\mathbf{\Phi}}_{\bar{\mathbf{y}},\bar{\mathbf{y}}}|, \tag{4}$$

where $\hat{\mathbf{\Phi}}_{\bar{\mathbf{y}},\bar{\mathbf{y}}} = \hat{\mathbf{D}}_{\bar{\mathbf{y}},\bar{\mathbf{y}}}^{-1/2}\hat{\mathbf{R}}_{\bar{\mathbf{y}},\bar{\mathbf{y}}}\hat{\mathbf{D}}_{\bar{\mathbf{y}},\bar{\mathbf{y}}}^{-1/2}$ is defined as the *coherence matrix*. Interestingly, $-\frac{1}{2}\ln|\hat{\mathbf{\Phi}}_{\bar{\mathbf{y}},\bar{\mathbf{y}}}|$ can be seen as a measure of the entropy loss due to the correlation among the separated sources (entries of $\mathbf{y}$). Thus, as one could expect, the optimization problem in (4) amounts to finding the separation matrix $\mathbf{W}$ minimizing the correlation (dependence) among the estimated sources $\mathbf{y}$.

4. PROPOSED ML-ICA ALGORITHM

In this subsection, we propose a practical quasi-Newton ML-ICA algorithm based on local quadratic approximations of the non-convex cost function $-\frac{1}{2}\ln|\hat{\mathbf{\Phi}}_{\bar{\mathbf{y}},\bar{\mathbf{y}}}|$. Analogously to other general joint diagonalization algorithms [11], the separation matrix is updated at each iteration 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.[2]

Let us start by introducing the matrix

$$\widetilde{\mathbf{W}} = \mathbf{P}\overline{\mathbf{W}}\mathbf{P}^T = \begin{bmatrix} \widetilde{\mathbf{W}}_{1,1} & \cdots & \widetilde{\mathbf{W}}_{1,m} \\ \vdots & \ddots & \vdots \\ \widetilde{\mathbf{W}}_{m,1} & \cdots & \widetilde{\mathbf{W}}_{m,m} \end{bmatrix},$$

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

Now, with similar definitions of $\delta_{k,l} \in \mathbb{H}$, $\widetilde{\mathbf{\Delta}} \in \mathbb{H}^{4m\times 4m}$ and $\widetilde{\mathbf{\Delta}}_{k,l} \in \mathbb{H}^{4\times 4}$, we are ready to introduce the following lemma, whose proof is omitted here due to the lack of space, but can be found in the journal paper [9].

Lemma 1 *Given a coherence matrix $\hat{\mathbf{\Phi}}_{\tilde{\mathbf{y}},\tilde{\mathbf{y}}}$ close to the identity, and assuming $\|\mathbf{\Delta}\|^2 \ll 1$, the cost function $-\frac{1}{2}\ln|\hat{\mathbf{\Phi}}_{\tilde{\mathbf{y}},\tilde{\mathbf{y}}}|$ can be approximated by the following quadratic expression:*

$$-\frac{1}{2}\ln|\hat{\mathbf{\Phi}}_{\tilde{\mathbf{y}},\tilde{\mathbf{y}}}| \simeq \frac{1}{2}\sum_{k=1}^{m}\sum_{l=k+1}^{m}\|\mathbf{J}_{k,l}(\delta_{k,l},\delta_{l,k})\|^2, \tag{5}$$

where

$$\begin{aligned}\mathbf{J}_{k,l}(\delta_{k,l},\delta_{l,k}) &= \hat{\mathbf{R}}_{\tilde{\mathbf{y}}_k,\tilde{\mathbf{y}}_k}^{-1/2}\hat{\mathbf{R}}_{\tilde{\mathbf{y}}_k,\tilde{\mathbf{y}}_l}\hat{\mathbf{R}}_{\tilde{\mathbf{y}}_l,\tilde{\mathbf{y}}_l}^{-1/2} \\ &+ \hat{\mathbf{R}}_{\tilde{\mathbf{y}}_k,\tilde{\mathbf{y}}_k}^{-1/2}\widetilde{\mathbf{\Delta}}_{k,l}\hat{\mathbf{R}}_{\tilde{\mathbf{y}}_l,\tilde{\mathbf{y}}_l}^{1/2} + \hat{\mathbf{R}}_{\tilde{\mathbf{y}}_k,\tilde{\mathbf{y}}_k}^{1/2}\widetilde{\mathbf{\Delta}}_{l,k}^H\hat{\mathbf{R}}_{\tilde{\mathbf{y}}_l,\tilde{\mathbf{y}}_l}^{-1/2}.\end{aligned}$$

Thanks to the approximation in (5), the optimization problem to be solved in each iteration of the proposed method is decoupled into $m(m-1)/2$ simpler problems. In particular, the elements $\delta_{k,l}, \delta_{l,k}$ are obtained by solving the least squares (LS) problem

$$\underset{\delta_{k,l},\delta_{l,k}}{\text{minimize}} \quad \|\mathbf{J}_{k,l}(\delta_{k,l},\delta_{l,k})\|^2, \tag{6}$$

whose solution is easily obtained by rewriting $\mathbf{J}_{k,l}(\delta_{k,l},\delta_{l,k})$ as a (quadratic) function of the eight real components of $\delta_{k,l}$ and $\delta_{l,k}$. Finally, we must note that the computational complexity of the proposed method, which is summarized in Algorithm 1, is dominated by the solution of the $m(m-1)/2$ LS problems in (6).

[2]In the final implementation of the algorithm $\mathbf{\Delta}$ will be scaled (if necessary) to ensure the invertibility of $(\mathbf{I}_m + \mathbf{\Delta})$.

Algorithm 1 Quaternion ML-ICA

Input: $\hat{\mathbf{R}}_{\tilde{\mathbf{x}},\tilde{\mathbf{x}}} \in \mathbb{H}^{4m\times 4m}$ and threshold $0 < \mu < 1$.
Output: Separation matrix $\mathbf{W} \in \mathbb{H}^{m\times m}$.
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 $\delta_{k,l}, \delta_{l,k}$ solving the LS problem in (6).
  end for
  if $\|\mathbf{\Delta}\| \geq \mu$ **then**
    $\mathbf{\Delta} \leftarrow \mu\mathbf{\Delta}/\|\mathbf{\Delta}\|$
  end if
  Update $\mathbf{W} \leftarrow (\mathbf{I}_m + \mathbf{\Delta})\mathbf{W}$.
  Update $\hat{\mathbf{R}}_{\tilde{\mathbf{y}},\tilde{\mathbf{y}}} \leftarrow (\mathbf{I}_{4m} + \widetilde{\mathbf{\Delta}})\hat{\mathbf{R}}_{\tilde{\mathbf{y}},\tilde{\mathbf{y}}}(\mathbf{I}_{4m} + \widetilde{\mathbf{\Delta}})^H$.
until Convergence

5. SIMULATION RESULTS

In this section, the performance of the proposed algorithm is illustrated by means of some simulation examples. In all the cases, the entries of the mixing matrix have been generated as i.i.d. quaternion proper Gaussian random variables with zero mean and unit variance, and the sources are independent quaternion random variables with zero mean, unit variance, and different complementary variances. 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 $\mu = 0.99$.

In the first example we consider $T = 100$ observations of a (square) mixture of $m = 10$ Gaussian sources with randomly generated SOS.[3] In order to evaluate the possible convergence to local minima, the proposed algorithm has been initialized in 100 different points. Additionally, we have also considered 100 independent experiments, with independently generated SOS and mixing matrices. The results are shown in Fig. 1, where we can see that, despite the non-convexity of the cost function $J = -\frac{1}{2}\ln|\hat{\mathbf{\Phi}}_{\tilde{\mathbf{y}},\tilde{\mathbf{y}}}|$, the proposed algorithm always converges to the same solution.

In the second 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 [12]. It is well-known [3] that the Alamouti signal model can be compactly written in terms of quaternions as $x = hs + n$, where x, h, s and n are quaternion scalars representing the received symbols (in two consecutive channel uses), the 2×1 multiple-input single-output (MISO) channel, the information symbols (s is constructed from two complex symbols), and the noise. Thus, if we consider a synchronous uplink channel with m users and a base station with m receive antennas, we obtain the model

$$\mathbf{x} = \mathbf{H}\mathbf{s} + \mathbf{n}.$$

In this experiment we consider a multiuser system with $m = 2$ users transmitting with the same power. The entries of $\mathbf{H}$ and $\mathbf{n}$ are independent zero-mean proper quaternion Gaussian random variables. 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

[3]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 $\mathbf{B} \in \mathbb{R}^{4\times 4}$ are i.i.d zero mean and unit variance random variables.

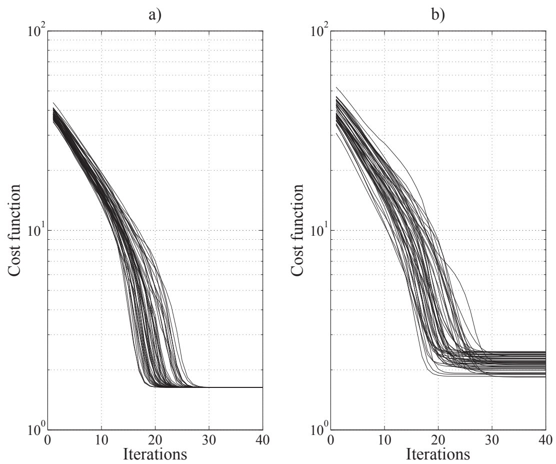

Fig. 1. Convergence of the ML-ICA algorithm. Ten sources and $T = 100$ vector observations. a) Fixed data and different initialization points. b) Independent experiments.

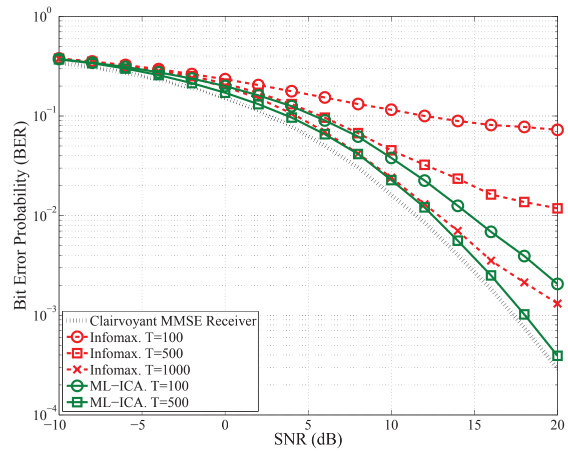

Fig. 2. Blind decoding in multiuser Alamouti systems based on quaternion ICA. Two-user system with QPSK constellations.

higher than that of the quadrature component. This results in a mixture of two independent quaternion sources, one of them proper and the other improper.

For comparison purposes, we have evaluated the quaternion extension of the Infomax algorithm [6,13], as well as the linear MMSE receiver with perfect channel knowledge. The obtained results for different numbers of vector observations are shown in Fig. 2, where we can see that the proposed method clearly outperforms the approach in [6]. This 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. Furthermore, we must note that this example illustrates the satisfactory performance of the proposed algorithm even in the case of non-Gaussian data.

6. CONCLUSIONS

In this paper we have presented an independent component analysis (ICA) algorithm for quaternion Gaussian vectors. In the quaternion case, the maximum-likelihood approach to the quaternion ICA problem reduces to the joint-diagonalization of the covariance and three complementary covariance matrices. In particular, the ML-ICA cost function can be seen as a measure of the entropy loss due to the correlation among the recovered sources and, although the ML-ICA problem is not convex, we have proposed a quasi-Newton algorithm which in practice provides very satisfactory results. Finally, some numerical examples have illustrated the performance and practical application of the proposed algorithm.

7. REFERENCES

[1] T. Bulow and G. Sommer, "Hypercomplex signals–a novel extension of the analytic signal to the multidimensional case," *IEEE Transactions on Signal Processing*, vol. 49, no. 11, pp. 2844–2852, Nov. 2001.

[2] C. C. Took and D. P. Mandic, "The quaternion LMS algorithm for adaptive filtering of hypercomplex processes," *IEEE Trans. on Sig. Proc.*, vol. 57, no. 4, pp. 1316–1327, Apr. 2009.

[3] S. Sirianunpiboon, A.R. Calderbank, and S.D. Howard, "Bayesian analysis of interference cancellation for Alamouti multiplexing," *IEEE Transactions on Information Theory*, vol. 54, no. 10, pp. 4755–4761, Oct. 2008.

[4] J. Vía, D. Ramírez, and I. Santamaría, "Properness and widely linear processing of quaternion random vectors," *IEEE Transactions on Information Theory*, vol. 56, no. 7, pp. 3502 –3515, July 2010.

[5] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*, Wiley Interscience, 2001.

[6] N. Le Bihan and S. Buchholz, "Quaternionic idependent component analysis using hypercomplex nonlinearities," in *IMA Conference on Mathemathics in Signal Processing*, Cirenceister, UK, 2006.

[7] J. Vía, L. Vielva, I. Santamaría, and D.P. Palomar, "Independent component analysis of quaternion gaussian vectors," in *6th IEEE Sensor Array and Multichannel Signal Processing Workshop*, Israel, Oct. 2010.

[8] J. P. Ward, *Quaternions and Cayley numbers: Algebra and applications*, Kluwer Academic, Dordrecht, Netherlands, 1997.

[9] J Vía, D. P. Palomar, L. Vielva, and I. Santamaría, "Quaternion ICA from second-order statistics," to appear in *IEEE Transactions on Signal Processing*, 2011.

[10] J. Cardoso, *Unsupervised Adaptive Filtering, Volume 1, Blind Source Separation*, vol. 1, chapter Entropic contrasts for source separation: geometry and stability, pp. 139–190, John Wiley & sons, 2000.

[11] 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.

[12] S.M. Alamouti, "A simple transmit diversity technique for wireless communications," *IEEE Journal on Selected Areas in Communications*, vol. 45, no. 9, pp. 1451–1458, 1998.

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