Quaternionic Independent Component Analysis using hypercomplex nonlinearities

By N. Le Bihan[†] and S. Buchholz[‡]
†: Laboratoire des Images et des Signaux, CNRS
ENSIEG, BP46, 38402 Saint-Martin d'Hères, FRANCE
‡: Cognitive Systems Group, Dept. of Computer Science,
CAU Kiel, 24098 Kiel, GERMANY
Nicolas.Le-Bihan@lis.inpg.fr, sbh@ks.informatik.uni-kiel.de

Abstract

We propose a quaternionic version of the Infomax algorithm to perform ICA on quaternion valued data. We introduce the three possible types of nonlinearities that can be used as activation functions and derive their differentiability properties. It is shown that only hypercomplex (fully quaternionic) nonlinearity can lead to the estimation of all possible classes of proper quaternion random variable. This fact is illlustrated on simulation data. It is shown that *fully quaternionic* Infomax algorithm is the only one that can perform blind separation of polarized signal corrupted by $\mathbb{H}$-proper (non-polarized) noise.

1. Quaternion random variables

Quaternions, denoted by $\mathbb{H}$, are 4D hypercomplex numbers. They form a non-commutative division algebra. A quaternion $q \in \mathbb{H}$ is given by $q = q_0 + q_1\mathtt{i} + q_2\mathtt{j} + q_3\mathtt{k}$, where $\{\mathtt{i}, \mathtt{j}, \mathtt{k}\}$ are pure imaginary numbers with multiplication rules: $\mathtt{i}^2 = \mathtt{j}^2 = \mathtt{k}^2 = -1$ and $\mathtt{ijk} = -1$. A useful way to consider quaternions is to look at them as complex numbers with complexified coefficients. This is known as the Cayley-Dickson notation: $q = z_1 + z_2\mathtt{j}$, where $z_1, z_2 \in \mathbb{C}$. For a more complete introduction to quaternions see *e.g.* (Ward, J.P. (1997)). Quaternion valued random variables (r.v.) have been studied in (Vakhania, N.N. (1998)) and (Amblard, P.O. & Le Bihan N. (2004)). As shown in these references, quaternion random variables include two interesting subclasses, called *proper* quaternion random variables.

1.1. *Properness*

There exists two levels of properness:

- A quaternion r.v. q is called $\mathbb{C}$-proper if: $q \stackrel{d}{=} e^{\eta\theta}q, \quad \forall\theta$ for one and only one imaginary unit $\eta = \mathtt{i}, \mathtt{j}$ or $\mathtt{k}$.
- A quaternion r.v. q is called $\mathbb{H}$-proper if: $q \stackrel{d}{=} e^{\eta\theta}q, \quad \forall\theta$ for any pure unit quaternion η.

We now present the link between properness and statistical properties of polarized signals.

1.2. *Proper random variables and polarization*

As known in Physics, a polarized signal can be described using a Jones vector (isomorphic to a quaternion). Quaternion models for polarized signals have recently been used in Polarization Mode Dispersion (PMD) and Polarization Dispersion Losses (PDL) treatment (Karlsson, M. & Petersson M. (2004)). Thus, a polarized signal $s[m]$ with m samples can be written as $s[m] = s_1[m] + s_2[m]\mathtt{j}$ where s_1 and s_2 are complex valued. If the signal is *purely* polarized, then $s[m]$ is deterministic. But if the signal is only partialy polarized, then $s[m]$ is random (Brosseau, C. (1998)). In this case, statistical description and processing is needed.

As shown in (Buchholz, S. & Le Bihan N. (2006)), if the two components of the signal are linked by a complex number (which phase and amplitude represents the polarization ellipsis), then its associated quaternion signal is $\mathbb{C}$-proper. If the two components are decorrelated, then the signal is $\mathbb{H}$-proper. Properness allows to distinguish between polarized and unpolarized signals (noise). Using this distinction, we aim at proposing an algorithm that is able to perform blind separation between different polarized signals or between polarized signals and noise.

2. Quaternionic Infomax

The Infomax algorithm was first proposed in (Bell, A.J., & Sejnowski, T.J. (1995)). Complex versions were introduced in (Calhoun, V. & Adalı, T. (2002), Adalı, T. & Kim, T. & Calhoun, V. (2004)). Here we present a quaternionic version of this algorithm.

Consider an observation vector data $\mathbf{x}$ that is a linear mixture of some quaternion valued sources $\mathbf{s}$:

$$\mathbf{x}[m] = \mathbf{A}\mathbf{s}[m] \qquad m = 0, 1, \ldots, M-1 \text{ (time index)} \tag{2.1}$$

where $\mathbf{x} \in \mathbb{H}^N$, $\mathbf{s} \in \mathbb{H}^N$ and $\mathbf{A} \in \mathbb{H}^{N\times N}$. Mixing matrix $\mathbf{A}$ is supposed to be constant in time and sources (and consequently observations) are supposed to be stationary. Note that due to non-commutativity of the quaternion product, other mixture models could be considered, depending on the Physics of the underlying problem (for example $\mathbf{x} = \mathbf{A}\mathbf{s}\mathbf{A}^\dagger$ or $\mathbf{x} = \mathbf{s}\mathbf{A}$, with $\dagger$ being the quaternion conjugation-transpotion operator). The Infomax algorithm maximizes the entropy of the output $\mathbf{y}[m]$ of a single layer neural network: $\mathbf{y}[m] = g(\mathbf{u}[m])$, where $g(.)$ is a quaternionic nonlinear function (*i.e.* $g : \mathbb{H} \to \mathbb{H}$). The vector $\mathbf{u}[m]$ is a weighted version of the input data: $\mathbf{u} = \mathbf{W}\mathbf{x}$ where $\mathbf{W} \in \mathbb{H}^{N\times N}$ is the weight matrix.

The entropy of the output vector $\mathbf{y} \in \mathbb{H}^N$ is given as:

$$H(\mathbf{y}) = -\mathbb{E}[\ln p(\mathbf{y})] = -\int_{-\infty}^{\infty} p(\mathbf{y}) \ln p(\mathbf{y}) d\mathbf{y} \tag{2.2}$$

Expectation is taken in the classical sense (see Amblard, P.O. & Le Bihan N. (2004)). The pdf of a quaternion random vector is in fact the joint pdf of its four vector components. So, the pdf of a quaternion random vector is: $p(\mathbf{y}) = p(\mathbf{q}_0, \mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3)$. The entropy of a quaternion valued random vector is the joint entropy of its components: $H(\mathbf{q}) \triangleq H(\mathbf{q}_0, \mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3)$. The calculations for the weight update are similar to the complex case:

$$\Delta\mathbf{W} = \frac{\partial H(\mathbf{y})}{\partial \mathbf{W}} \mathbf{W}^\dagger \mathbf{W} = \mu \left[\mathbf{I} + \varphi(\mathbf{u})\mathbf{u}^\dagger\right] \mathbf{W} \tag{2.3}$$

This means formal extension of the Infomax algorithm is straightforward from $\mathbb{C}$ to $\mathbb{H}$. The only point of matter is in the nonlinearity g. In the following we will only consider the $\tanh(\mathbf{u})$ nonlinearity, with different quaternionic versions. That is $\varphi = -2\tanh(\mathbf{u})$.

3. Quaternionic nonlinearities

In the work of Calhoun *et al.*, it is emphasized that it is possible to use two different definitions for the nonlinearity: the full complex and split. In the tanh case, the two possible definitions are namely the *split*: $g_s(u) = g_s(u_{Re} + \mathtt{i}u_{Im}) = \tanh(u_{Re}) + \mathtt{i}\tanh(u_{Im})$, and the *full*: $g_f(u) = g_f(u_{Re} + \mathtt{i}u_{Im}) = \tanh(u_{Re} + \mathtt{i}u_{Im})$. We now propose the extension of this work to the quaternion case.

3.1. *Nonlinearities definition*

In the quaternionic case, there are three possible levels of nonlinearity:

- $\mathbb{R}$-split: $g(u = u_0 + u_1\mathtt{i} + u_2\mathtt{j} + u_3\mathtt{k}) \triangleq \tanh(u_0) + \tanh(u_1)\mathtt{i} + \tanh(u_2)\mathtt{j} + \tanh(u_3)\mathtt{k}$
- $\mathbb{C}$-split: $g(u_0 + u_1\mathtt{i} + u_2\mathtt{j} + u_3\mathtt{k}) \triangleq \tanh(u_0 + u_1\mathtt{i}) + \mathtt{j}\tanh(u_2 + u_3\mathtt{i})$
- $\mathbb{H}$-full: $g(u_0 + u_1\mathtt{i} + u_2\mathtt{j} + u_3\mathtt{k}) \triangleq \tanh(u_0 + u_1\mathtt{i} + u_2\mathtt{j} + u_3\mathtt{k})$

Clearly, the $\mathbb{R}$-split nonlinearity processes each component separately, while the $\mathbb{C}$-split one process the two complex components independently. Those two types of nonlinearities are not able to take into account possible correlations between all the components. Only the $\mathbb{H}$-full nonlinearity is able do this, as shown in Section 4.

3.2. *Differentiability*

It is possible to define three types of differentiablity for quaternion valued functions. Consider such a function $g : \mathbb{H} \to \mathbb{H}$ for which $u \to g(u)$ when $u = u_0 + u_1\mathtt{i} + u_2\mathtt{j} + u_3\mathtt{k}$, and where $g = g_0 + g_1\mathtt{i} + g_2\mathtt{j} + g_3\mathtt{k} = g^{(1)} + g^{(2)}\mathtt{j}$. Then, it is possible to define:

- $\mathbb{R}$-differentiability: $[g]'_{\mathbb{R}} = \dfrac{\partial g}{\partial u} \triangleq \dfrac{\partial g}{\partial u_0} + \dfrac{\partial g}{\partial u_1}\mathtt{i} + \dfrac{\partial g}{\partial u_2}\mathtt{j} + \dfrac{\partial g}{\partial u_3}\mathtt{k}$
- $\mathbb{C}$-differentiability: $[g]'_{\mathbb{C}} = \dfrac{\partial g^{(1)}}{\partial v} + \mathtt{j}\dfrac{\partial g^{(2)}}{\partial w}$
- $\mathbb{H}$-differentiability: $[g]'_{\mathbb{H}} = \dfrac{dg}{du}$

In the Infomax context, where it is necessary to compute, for the nonlinear activation function g, the ratio $\frac{g''}{g'}$ (Bell, A.J., & Sejnowski, T.J. (1995)), using $\mathbb{R}$-differentiation is equivalent to process separately the four components. Choosing an approach using $\mathbb{C}$- or $\mathbb{H}$-differentiation allows to take into account the possible links between the components. Note that in the $\mathbb{C}$-differentiability, both complex components satisfy Cauchy-Riemann equations. In the $\mathbb{H}$-differentiability case, as explained in (Sudbery, A. (1979)), the condition imposed on g lead to the so-called Cauchy-Riemann-Fueter equation: $\frac{dg}{du} = \frac{\partial g}{\partial u_0} = -\mathtt{i}\frac{\partial g}{\partial u_1} = -\mathtt{j}\frac{\partial g}{\partial u_2} = -\mathtt{k}\frac{\partial g}{\partial u_3}$. All $\mathbb{H}$-differentiable functions are therefore of linear type: $f(q) = aq + b$ where $a, b \in \mathbb{H}$. However, just like in the complex case (Calhoun, V. & Adalı, T. (2002)), it is possible to overcome this problem while using full hypercomplex nonlinearities by assuming that singularities of a hypercomplex function (which is not $\mathbb{H}$-differentiable) have measure zero, which ensure that the update will not end up at such point.

4. Approximation and estimation properties

The purpose here is to present how well split- and fully- quaternionic approaches can "match" a certain distribution. We consider the case where there is only one weight and only a random variable (random vector with dimension 1). With these hypothese one has $y = g(wx)$. Then, remembering that it is possible to use vector representations for quaternion random variables, we choose the real representation from here, as a quaternion random variable q is completely described by the joint distribution of its four components, *i.e.* by the distribution of the vector: $\hat{\mathbf{q}} = [q_0\ q_1\ q_2\ q_3]^T$. So, from now we consider the following: $\hat{\mathbf{y}} = g(w\hat{\mathbf{x}})$, where $\hat{}$ denotes the real representation vector. The distribution of the output y can be linked to the distribution of the input x in the following way:

$$p(\hat{\mathbf{y}}) = \left.\frac{p(\hat{\mathbf{x}})}{|\mathbf{J}|}\right|_{\hat{\mathbf{x}}=w^{-1}g^{-1}(\hat{\mathbf{y}})} \tag{4.1}$$

where the terms of the Jacobian matrix $\mathbf{J}$ are given by $\mathbf{J}|_{ij} = \partial g_i/\partial x_j$ for $i, j = 0, 1, 2, 3$. We give the Jacobian expression for the three types of nonlinearities, namely the $\mathbb{R}$-split ($\mathbf{J}_{Rs}$), $\mathbb{C}$-split ($\mathbf{J}_{Cs}$) and $\mathbb{H}$-full ($\mathbf{J}_{Hf}$):

$$\mathbf{J}_{Rs} = \begin{pmatrix} \partial_0 g_0 & 0 & 0 & 0 \\ 0 & \partial_1 g_1 & 0 & 0 \\ 0 & 0 & \partial_2 g_2 & 0 \\ 0 & 0 & 0 & \partial_3 g_3 \end{pmatrix}; \mathbf{J}_{Cs} = \begin{pmatrix} \partial_0 g_0 & \partial_1 g_0 & 0 & 0 \\ \partial_0 g_1 & \partial_1 g_1 & 0 & 0 \\ 0 & 0 & \partial_2 g_2 & \partial_3 g_2 \\ 0 & 0 & \partial_2 g_3 & \partial_3 g_3 \end{pmatrix}; \mathbf{J}_{Hf} = \begin{pmatrix} \partial_0 g_0 & \partial_1 g_0 & \partial_2 g_0 & \partial_3 g_0 \\ \partial_0 g_1 & \partial_1 g_1 & \partial_2 g_1 & \partial_3 g_1 \\ \partial_0 g_2 & \partial_1 g_2 & \partial_2 g_2 & \partial_3 g_2 \\ \partial_0 g_3 & \partial_1 g_3 & \partial_2 g_3 & \partial_3 g_3 \end{pmatrix} \tag{4.2}$$

where ∂_α stands for $\partial/\partial x_\alpha$. The pattern of the Jacobian matrices allows to conclude that only the $\mathbb{H}$-full approach can take into account the largest set of possible relationships between the four components of a quaternion random variable. More specifically, in the $\mathbb{C}$-proper case, the first component is correlated with the third and fourth, which induces off-diagonal terms in the covariance matrix (see Amblard, P.O. & Le Bihan N. (2004) for details). As a consequence, only the $\mathbb{H}$-full nonlinearty in quaternionic Infomax can perform a thourought recovering of a $\mathbb{C}$-proper random variable. We illustrate this now on simulated signals.

5. Simulation results

We consider a simple example where two vector-sensors record a linear and instantaneous mixture of two random signals. The model for the recorded mixture is thus: $\mathbf{x}[m] = \mathbf{A}\mathbf{s}[m]$, where $\mathbf{x}[m] \in \mathbb{H}^2$, $\mathbf{s}[m] \in \mathbb{H}^2$ and $\mathbf{A} \in \mathbb{H}^{2\times 2}$. Source $s_1[m]$ is a $\mathbb{H}$-proper (non polarized, assumed as noise), Gaussian and i.i.d. signal. Source $s_2[m]$ is a random signal taking values on the edges of a 4D polytope (see Zetterberg, L. H. & Brändström (1977) for details). The Quaternionic Infomax results are presented in Figure 1. One can see that the $\mathbb{C}$-split approach is not able to separate the signal from noise while the $\mathbb{H}$-full approach lead to a better estimation of the signal and futhermore to the recovery of the $\mathbb{H}$-properness property for the noise.

6. Conclusion

We have proposed an extension of the Infomax algorithm to the quaternionic case. The choice of the nonlinearity has been demonstrated to be determinant in the separation result. Despite the lake of $\mathbb{H}$-differentiability, $\mathbb{H}$-full nonlinearities are the best choice to achieve the separation between polarized signals and noise. The proposed

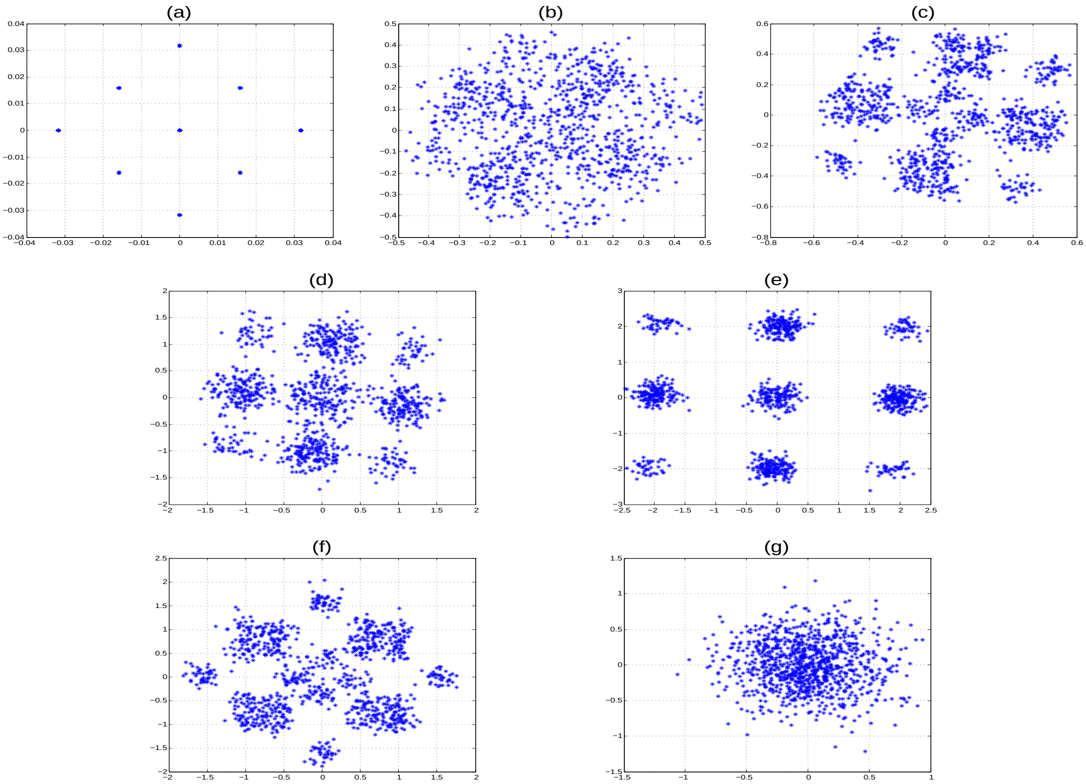

FIGURE 1. Original signal (a), mixture recorded on sensor 1 (b) and 2 (c), estimated signal with $\mathbb{C}$-split tanh (d) and $\mathbb{H}$-full (e) nonlinearities, estimated $\mathbb{H}$-proper noise with with $\mathbb{C}$-split tanh (f) and $\mathbb{H}$-full (g) nonlinearities

algorithm could be of interest in applications such as Optics, Electromagnetism or Seismic where polarized signals corrupted by noise are encountered and where ICA can help to recover the wavefield sources.

REFERENCES

ADALI, T. & KIM, T. & CALHOUN, V. 2004 Independent component analysis by complex nonlinearities, *IEEE International Conference on Acoustis, Speech and Signal Processing (ICASSP)* Montreal, Quebec, Canada.

AMBLARD, P.O. & LE BIHAN N. 2004 On properness of quaternion valued random variables, *IMA Conference on Mathematics in Signal Processing* Cirencester, UK.

BELL, A.J., & SEJNOWSKI, T.J. 1995 An information maximisation approach to blind separation and blind deconvolution, *Neural Computation* Vol. 7, No. 6, pp. 1129–1159

BROSSEAU, C. 1998 Fundamentals of Polarized Light: A Statistical Approach, *John Wiley and Sons*.

BUCHHOLZ, S. & LE BIHAN N. 2006 Optimal Separation of Polarized Signals by Quaternionic Neural Networks, *European Signal and Image Processing Conference (EUSIPCO)* Florence, Italy.

CALHOUN, V. & ADALI, T. 2002 Complex Infomax: convergence and approximation with complex nonlinearities, *IEEE Workshop on Neural Networks for Signal Processing (NNSP)* Martigny, Switzerland.

KARLSSON, M. & PETERSSON M. 2004 Quaternion approach to PMD and PDL phenomena in optical fiber systems, *IEEE Journal of Lightwave technology* Vol. 22, No. 4, pp. 1137–1146.

SUDBERY, A. 1979 Quaternionic analysis, *Math. Proc. Camb. Phil. Soc.* Vol. 85, pp. 199–225.

VAKHANIA, N.N. 1998 Random vectors with values in quaternion Hilbert spaces, *Th. Probab. Appl.* Vol. 43, No. 1, pp. 99–115.

WARD, J.P. 1997 Quaternions and Cayley numbers, *Kluwer Academic Publishers*.

ZETTERBERG, L. H. & BRÄNDSTRÖM, H. 1977 Codes for combined phase and amplitude modulated signals in a four-dimensional space *IEEE Trans. on Communications* Vol. 25, No. 9, pp. 943–950.