14th European Signal Processing Conference (EUSIPCO 2006), Florence, Italy, September 4-8, 2006, copyright by EURASIP

OPTIMAL SEPARATION OF POLARIZED SIGNALS BY QUATERNIONIC NEURAL NETWORKS

Sven Buchholz and Nicolas Le Bihan

Cognitive Systems Group
Department of Computer Science, CAU Kiel
24098 Kiel, Germany
phone: +49 431 880 7545; fax: +49 431 880 7550
email: sbh@ks.informatik.uni-kiel.de

Laboratoire des Images et des Signaux
CNRS UMR 5083
Grenoble, France
phone: +33 476 826 486; fax: +33 476 826 384
email: Nicolas.Le-Bihan@lis.inpg.fr

ABSTRACT

Statistical description of polarized signals is proposed in terms of proper quaternionic random processes. Within this framework, the intrinsic nature of such signals is captured well. Simulation results show the ability of quaternionic approach (statistical model and processing) to perform better separation of polarized signals than real–valued neural networks can do.

1. INTRODUCTION

In many applications such as seismology, electromagnetics, optics, communications, etc., the recorded signals are polarized. This property is due to the nature of the waves carrying the signals (elastic or electromagnetics waves). As a consequence, polarized signals describe the evolution of a vector with time, pointing into the direction of vibration of the medium pertubated by the wave. This vector is confined into a plane (the so-called *polarization plane*) and thus it is always possible (even for 3D signals) to describe a simply polarized signal in terms of two signals. Polarized signals are recorded with vector-sensors and so they are vector-valued signals.

A big challenge in polarized signal processing is to take advantage of the additional information provided by polarization without loosing this information in the processing. In order to do so, we adopt here a quaternionic model for polarized signals. This model was already used in [1] and here we add a statistical description to it. Using recent work on quaternion random variables [2], we express polarized signals as *proper* quaternion random processes. That way the intrinsic nature of polarized signal is captured well. Hence our model can be used advantageously for signal processing tasks. This is demonstrated for signal separation. For separation inside the quaternionic framework a quaternionic neural network is used and its superior performance over a standard real-valued network is demonstrated. Additionally, representation issues for optimal separation are discussed for the first time.

2. QUATERNIONS

Quaternions, denoted $\mathbb{H}$, are a 4D hypercomplex number system and form a noncommutative division algebra. The Cartesian notation of a quaternion q is given as:

$$q = q_0 + q_1\mathtt{i} + q_2\mathtt{j} + q_3\mathtt{k} \tag{1}$$

where $q_{0,1,2,3} \in \mathbb{R}$. Multiplication rules for imaginary units are:

$$\begin{array}{c} \mathtt{i}^2 = \mathtt{j}^2 = \mathtt{k}^2 = \mathtt{ijk} = -1 \\ \mathtt{ij} = -\mathtt{ji} = \mathtt{k} \end{array} \tag{2}$$

The real part of a quaternion $\Re(q) = q_0$ is a scalar, while its imaginary part $\Im(q) = q_1\mathtt{i} + q_2\mathtt{j} + q_3\mathtt{k}$ is a vector. A quaternion with a null real part is called *pure*. The conjugate of q is $\overline{q} = q_0 - q_1\mathtt{i} - q_2\mathtt{j} - q_3\mathtt{k}$, with the property: $\overline{pq} = \overline{q}\,\overline{p}$. The norm of q is $|q| = \sqrt{q_0^2 + q_1^2 + q_2^2 + q_3^2}$ and if $q \neq 0$, then its inverse is $q^{-1} = \overline{q}/|q|^2$. A quaternion is called *unit* if its norm equals 1. It is possible to write quaternions as complex numbers with complexified coefficients:

$$q = q^{(1)} + q^{(2)}\mathtt{j} \tag{3}$$

where $q^{(1)} = q_0 + q_1\mathtt{i}$ and $q^{(2)} = q_2 + q_3\mathtt{i}$. This way of writing quaternions is known as the *Cayley-Dickson* notation. The famous Euler formula generalizes to quaternions so that any quaternion $q \in \mathbb{H}$ can be written:

$$q = |q|(\cos\varphi + \mu\sin\varphi) = |q|e^{\mu\varphi} \tag{4}$$

where $\mu = (q_1\mathtt{i} + q_2\mathtt{j} + q_3\mathtt{k})/\sqrt{q_1^2 + q_2^2 + q_3^2}$ is a pure unit quaternion called the *axis* of q and φ is called the *angle* (or *argument*) of q. There are three canonical involutions defined on $\mathbb{H}$:

$$q_{\mathtt{i}} = -\mathtt{i}q\mathtt{i};\; q_{\mathtt{j}} = -\mathtt{j}q\mathtt{j};\; q_{\mathtt{k}} = -\mathtt{k}q\mathtt{k} \tag{5}$$

with the following properties ($\eta \in \{\mathtt{i},\mathtt{j},\mathtt{k}\}$):

$$\left\{ \begin{array}{rcr} \overline{q_\eta} & = & \overline{q}_\eta, \\ (q_\eta)_\eta & = & q, \\ (pq)_\eta & = & p_\eta q_\eta \end{array} \right. \tag{6}$$

Note that a more general definition for involutions on $\mathbb{H}$ can be stated as follows: $q_\eta = -\eta q \eta$, where $\eta^2 = -1$ and $\Re(\eta) = 0$. Here, we will make only use of the canonical involutions given in (5).

3. PROPER QUATERNION VALUED RANDOM VECTORS

The study of quaternion valued random vectors $\mathbf{q}$ consists in the study of the joint probability density function (pdf) of the four components. In the Gaussian case, the first and second order statistics (mean and variance) tell all the story. Here, we introduce basic definitions and properties of quaternion valued random vectors. As only Gaussian cases will be considered, we only present a second order study for quaternion valued random vectors. Available material on quaternionic random variables and vectors can be found in [2, 3]. Of main interest for us here is the concept of *proper* random variables and vectors.

3.1 Definitions and representations

One useful way to consider a quaternion random vector is to see it as a real valued random vector of size four times bigger. However, it is also possible to see it as a complex or quaternion valued vector of higher dimension. Thus, for a quaternion valued random vector $\mathbf{q} \in \mathbb{H}^N$ there exists three possible representations. Namely the real $\tilde{q} \in \mathbb{R}^{4N}$, complex $\hat{q} \in \mathbb{C}^{4N}$ and quaternion $\breve{q} \in \mathbb{H}^{4N}$ representations. Their expressions are:

$$\left\{ \begin{array}{l} \tilde{\mathbf{q}} = \left[\mathbf{q}_0^T \; \mathbf{q}_1^T \; \mathbf{q}_2^T \; \mathbf{q}_3^T\right]^T \\ \hat{\mathbf{q}} = \left[\mathbf{q}^{(1)T} \; \mathbf{q}^{(1)\dagger} \; \mathbf{q}^{(2)T} \; \mathbf{q}^{(2)\dagger}\right]^T \\ \breve{\mathbf{q}} = \left[\mathbf{q}^T \; \mathbf{q}_{\mathtt{i}}^T \; \mathbf{q}_{\mathtt{j}}^T \; \mathbf{q}_{\mathtt{k}}^T\right]^T \end{array} \right. \tag{7}$$

where † stands for conjugation-transposition. These representations allow to study the statistical relationships between the components of a quaternion random vector. Transition matrices to switch between them can be found in [2].

3.2 First and second order statistics

Expectation for $\mathbb{H}$-valued random vectors is naturally defined as:

$$\mathbb{E}[\mathbf{q}] = \mathbb{E}[\mathbf{q}_0] + \mathbb{E}[\mathbf{q}_1]\mathtt{i} + \mathbb{E}[\mathbf{q}_2]\mathtt{j} + \mathbb{E}[\mathbf{q}_3]\mathtt{k} \tag{8}$$

This is the definition of the *mean* of $\mathbf{q} \in \mathbb{H}^N$. Then, considering a centered random vector $\mathbf{q}$, its covariance matrix is given by:

$$\boldsymbol{\Lambda}_{\mathbf{q}} = \mathbb{E}[\mathbf{q}\mathbf{q}^\dagger] = \mathbb{E}[\mathbf{q}\overline{\mathbf{q}}^T] \tag{9}$$

Using the vector representations introduced before, the covariance matrix of $\mathbf{q}$ has also three representations:

$$\left\{ \begin{array}{l} \boldsymbol{\Lambda}_{\tilde{\mathbf{q}}} = \mathbb{E}[\tilde{\mathbf{q}}\tilde{\mathbf{q}}^T] \\ \boldsymbol{\Lambda}_{\hat{\mathbf{q}}} = \mathbb{E}[\hat{\mathbf{q}}\hat{\mathbf{q}}^\dagger] \\ \boldsymbol{\Lambda}_{\breve{\mathbf{q}}} = \mathbb{E}[\breve{\mathbf{q}}\breve{\mathbf{q}}^\dagger] \end{array} \right. \tag{10}$$

In the sequel, we will only make use of $\hat{\mathbf{q}}$. Thus we give here the explicit expression of its covariance matrix:

$$\boldsymbol{\Lambda}_{\hat{\mathbf{q}}} = \begin{bmatrix} \mathbf{K}_{\mathbf{q}^{(1)}} & \mathbf{C}_{\mathbf{q}^{(1)}} & \mathbf{K}_{\mathbf{q}^{(1)}q^{(2)}} & \mathbf{C}_{\mathbf{q}^{(1)}\mathbf{q}^{(2)}} \\ \mathbf{C}^*_{\mathbf{q}^{(1)}} & \mathbf{K}^*_{\mathbf{q}^{(1)}} & \mathbf{C}^*_{\mathbf{q}^{(1)}\mathbf{q}^{(2)}} & \mathbf{K}^*_{\mathbf{q}^{(1)}\mathbf{q}^{(2)}} \\ \mathbf{K}^\dagger_{\mathbf{q}^{(1)}\mathbf{q}^{(2)}} & \mathbf{C}^T_{\mathbf{q}^{(1)}\mathbf{q}^{(2)}} & \mathbf{K}_{\mathbf{q}^{(2)}} & \mathbf{C}_{\mathbf{q}^{(2)}} \\ \mathbf{C}^\dagger_{\mathbf{q}^{(1)}\mathbf{q}^{(2)}} & \mathbf{K}^T_{\mathbf{q}^{(1)}\mathbf{q}^{(2)}} & \mathbf{C}^*_{\mathbf{q}^{(2)}} & \mathbf{K}^*_{\mathbf{q}^{(2)}} \end{bmatrix} \tag{11}$$

where $\mathbf{K}_{\mathbf{vw}} = \mathbb{E}[\mathbf{v}\mathbf{w}^\dagger]$, $\mathbf{K}_{\mathbf{v}} = \mathbf{K}_{\mathbf{vv}}$, $\mathbf{C}_{\mathbf{vw}} = \mathbb{E}[\mathbf{v}\mathbf{w}^T]$ and $\mathbf{C}_{\mathbf{v}} = \mathbf{C}_{\mathbf{vv}}$, and with $\mathbf{v}, \mathbf{w} \in \mathbb{C}^N$. In the literature [4], $\mathbf{K}_{\mathbf{v}}$ is known as the covariance of $\mathbf{v}$ while $\mathbf{C}_{\mathbf{v}}$ is its *pseudo-covariance*. A complex random vector with vanishing *pseudo-covariance* is called *proper* (or also *circular*) [4, 6]. We now introduce the generalization of this property to the case of $\mathbb{H}$-valued random vectors.

3.3 Proper random vectors

As firstly described in [3] and generalized in [2] there exists two levels of properness.

3.3.1 $\mathbb{C}^\eta$-*properness*

A quaternion random vector $\mathbf{q} \in \mathbb{H}^N$ is called proper iff:

$$\mathbf{q} \overset{d}{=} e^{\eta\theta}\mathbf{q}, \; \forall\theta \tag{12}$$

for one and only one pure unit quaternion η, and where $\overset{d}{=}$ stands for "equality in distribution". A special case of interest for us in the sequel will be $\mathbb{C}^{\mathtt{i}}$-proper random vectors for which $\mathbf{q} \overset{d}{=} e^{\mathtt{i}\theta}\mathbf{q}$. As a consequence, the covariance matrix of a $\mathbb{C}^{\mathtt{i}}$-proper random vectors commutes with $\mathtt{i}$: $\boldsymbol{\Lambda}_{\mathbf{q}}\mathtt{i} = \mathtt{i}\boldsymbol{\Lambda}_{\mathbf{q}}$. This commutation induces a special structure in the covariance matrix (see [2, 3] for details), which in its complex representation reads

$$\boldsymbol{\Lambda}_{\hat{\mathbf{q}}} = \begin{bmatrix} \mathbf{K}_{\mathbf{q}(1)} & \mathbf{0} & \mathbf{0} & \mathbf{C}_{\mathbf{q}^{(1)}\mathbf{q}^{(2)}} \\ \mathbf{0} & \mathbf{K}^*_{\mathbf{q}^{(1)}} & \mathbf{C}^*_{\mathbf{q}^{(1)}\mathbf{q}^{(2)}} & \mathbf{0} \\ \mathbf{0} & \mathbf{C}^T_{\mathbf{q}^{(1)}\mathbf{q}^{(2)}} & \mathbf{K}_{\mathbf{q}^{(2)}} & \mathbf{0} \\ \mathbf{C}^\dagger_{\mathbf{q}^{(1)}\mathbf{q}^{(2)}} & \mathbf{0} & \mathbf{0} & \mathbf{K}^*_{\mathbf{q}^{(2)}} \end{bmatrix} \tag{13}$$

This suggests an equivalent definition for $\mathbb{C}^{\mathtt{i}}$-properness. Thus, a quaternion random vector is $\mathbb{C}^{\mathtt{i}}$-proper iff:

$$\left\{ \begin{array}{l} \mathbf{C}_{\mathbf{q}^{(1)}} = \mathbf{0} \\ \mathbf{C}_{\mathbf{q}^{(2)}} = \mathbf{0} \\ \mathbf{K}_{\mathbf{q}^{(1)}\mathbf{q}^{(2)}} = \mathbf{0} \end{array} \right. \tag{14}$$

One can see that $\mathbb{C}^{\mathtt{i}}$-properness for a quaternion random vector means that this vector can be seen as a pair of proper (in the complex sense) complex random vectors $\mathbf{q}^{(1)}$ and $\mathbf{q}^{(2)}$ with a non-null cross-*pseudo-covariance* but a vanishing cross-covariance. Note that the other possible cases (*i.e.* $\mathbb{C}^{\mathtt{j}}$ and $\mathbb{C}^{\mathtt{k}}$) lead to a different structure of the covariance matrix (different positions for the zeros).

3.3.2 $\mathbb{H}$-*properness*

A quaternion random vector $\mathbf{q} \in \mathbb{H}^N$ is called $\mathbb{H}$-proper iff:

$$\mathbf{q} \overset{d}{=} e^{\eta\theta}\mathbf{q}, \; \forall\theta \tag{15}$$

for any pure unit quaternion η. Once again, considering the classical basis for quaternions (*i.e.* $\{1, \mathtt{i}, \mathtt{j}, \mathtt{k}\}$), $\mathbb{H}$-properness is equivalent to having the following equalities verified:

$$\left\{ \begin{array}{l} \boldsymbol{\Lambda}_{\mathbf{q}}\mathtt{i} = \mathtt{i}\boldsymbol{\Lambda}_{\mathbf{q}} \\ \boldsymbol{\Lambda}_{\mathbf{q}}\mathtt{j} = \mathtt{j}\boldsymbol{\Lambda}_{\mathbf{q}} \end{array} \right. \tag{16}$$

These commutation rules induces a very special structure in the covariance matrix $\boldsymbol{\Lambda}_{\hat{\mathbf{q}}} = diag(\mathbf{K}_{\mathbf{q}})$, where $\mathbf{K}_{\mathbf{q}} = \mathbf{K}_{\mathbf{q}^{(1)}} = \mathbf{K}_{\mathbf{q}^{(2)}}$. This also suggests another equivalent definition for $\mathbb{H}$-properness:

$$\left\{ \begin{array}{l} \mathbf{C}_{\mathbf{q}^{(1)}} = \mathbf{0} \\ \mathbf{C}_{\mathbf{q}^{(2)}} = \mathbf{0} \\ \mathbf{K}_{\mathbf{q}^{(1)}\mathbf{q}^{(2)}} = \mathbf{0} \\ \mathbf{C}_{\mathbf{q}^{(1)}\mathbf{q}^{(2)}} = \mathbf{0} \end{array} \right. \tag{17}$$

Therefore a $\mathbb{H}$-proper random vector can be seen as a pair of proper (in the complex sense) complex random vectors $\mathbf{q}^{(1)}$ and $\mathbf{q}^{(2)}$ which are *jointly proper* (in the complex sense), *i.e.* their cross-covariance and cross-pseudo-covariance both vanish. In the $\mathbb{H}$-proper case, the two complex components $\mathbf{q}^{(1)}$ and $\mathbf{q}^{(2)}$ are thus *uncorrelated* [4].

3.4 Proper random variables and polarized signals

We now consider polarized random signals with possibly known polarization parameters ρ and ϕ. Consider the output of a two-component vector-sensor [5]. Such a sensor outputs two discrete signals $s_1(n)$ and $s_2(n)$ (with $n = 1,\dots,N$) that originate from vibrations in two orthogonal directions of 3D space. For such signal, we propose to use the model described in [1]. The output can be arranged in vectors $\mathbf{s}_{1,2} = [s_{1,2}(1)\ s_{1,2}(2)\ \dots\ s_{1,2}(N)]^T$. Thus the whole 2D vibration recorded on the sensor can be written in a vector $\mathbf{q}$ such as:

$$\mathbf{q} = \mathbf{z}_1 + \mathtt{j}\mathbf{z}_2 \qquad (18)$$

where $\mathbf{z}_1$ and $\mathbf{z}_2$ are the analytic signals of $\mathbf{s}_1$ and $\mathbf{s}_2$, respectively. Here, $\mathbf{s}_1$ and $\mathbf{s}_2$ are considered *i.i.d.* and Gaussian random processes. The real and imaginary parts of $\mathbf{z}_1$ and $\mathbf{z}_2$ have zero mean and same variance σ^2. It is well-known that the analytic signal is proper (in the complex sense) [6], which involves that $\mathbf{C}_{\mathbf{z}_1} = \mathbf{C}_{\mathbf{z}_2} = \mathbf{0}$. As a consequence, the variance of $\mathbf{z}_1$ and $\mathbf{z}_2$ equals $2\sigma^2$ (real and imaginary parts having the same variance and being decorrelated).

Now, assuming that the recorded signal is *polarized*, then there exists a phase shift and an amplitude ratio (both supposed constant along the time index here):

$$\mathbf{z}_1 = \mathscr{P}\mathbf{z}_2 = \rho e^{\mathtt{i}\phi}\mathbf{z}_2 \qquad (19)$$

where ρ and ϕ are the polarization parameters [1]. With the above assumptions, it is easy to verify that the covariance matrix of the random vector $\mathbf{q}$ made of the samples of a polarized signals has the following structure (complex notation):

$$\Lambda_{\hat{q}} = 2\sigma^2 \begin{bmatrix} \mathbf{I} & \mathbf{0} & \mathbf{0} & \rho e^{-\mathtt{i}\phi}\mathbf{I} \\ \mathbf{0} & \mathbf{I} & \rho e^{\mathtt{i}\phi}\mathbf{I} & \mathbf{0} \\ \mathbf{0} & \rho e^{-\mathtt{i}\phi}\mathbf{I} & \rho^2\mathbf{I} & 0 \\ \rho e^{\mathtt{i}\phi}\mathbf{I} & \mathbf{0} & 0 & \rho^2\mathbf{I} \end{bmatrix} \qquad (20)$$

Thus, the proposed modeling allows to consider a polarized random signal (with deterministic polarization parameter ρ and ϕ) as a $\mathbb{C}^{\mathtt{i}}$-proper quaternion valued random vector.

Now, consider the case where the recorded signal is not polarized. Then $\mathbf{z}_1$ and $\mathbf{z}_2$ are uncorrelated, so the signal can be seen as a $\mathbb{H}$-proper quaternion random vector. Its covariance will then be diagonal: $\Lambda_{\hat{q}} = 2\sigma^2\mathbf{I}_{4N\times 4N}$.

4. NEURAL ARCHITECTURE

Neural networks have found many applications in signal processing (see *e.g.*[7, 8] and references therein). The previous section established a theoretical link between polarized signals and proper $\mathbb{H}$–valued random variables. A quaternion valued neural network remains naturally inside our framework. Advantages of our proposed model should therefore result in an outperformance of such networks over real-valued networks, for example on a signal separation task. First we review technical details of a quaternion valued Multi-layer Perceptron [9]. Then the novel and relevant notion of isomorphic class labels is discussed.

4.1 Multi-layer Perceptrons (MLPs)

The atoms of neural networks are simple computational units that compute from input x an output according to $y = g(f(w,x))$. Thereby f is the so-called propagation function and g the so-called activation function. The weights w are free parameters that are adjusted trough learning. For nonlinear g, grouping neurons together in layers $\sum_i g(w_{ij}x_j + \theta_j)$ in a feed–forward and fully–connected manner yields the well-known Multi-layer Perceptron (MLP). MLPs are trained by supervised learning, *i.e.* on examples with target outputs. The most popular choice for g is the logistic function $\sigma(x) = (1+\exp(-x))^{-1}$ with output in $[0,1]$. Adapting weights is then done by minimizing an error function, say SSE, by gradient descent (backpropagation [10]). A quaternion valued MLP ($\mathbb{H}$–MLP)is obtained one by just using quaternionic entities instead of real ones. As activation function $\sigma(q)_{\mathbb{H}} = \sigma(q_0) + \sigma(q_1)\mathtt{i} + \sigma(q_2)\mathtt{j} + \sigma(q_3)\mathtt{k}$ is used [9]. Separation of two polarized signals is a classification task. A $\mathbb{H}$–MLP with one hidden layer of two neurons is therefore the smallest meaningful quaternion valued architecture in this case. We shall see that this minimal quaternionic network is also already sufficient. This is a direct consequence of the invariance of distributions under (12) and (15), respectively.

4.2 Isomorphic Class Labels

One issue of the $\mathbb{H}$-MLP remains for discussion. That is how to label the classes. This is one other consequence of the fundamental difference between real valued and quaternion valued neural networks. For a standard MLP neither the order of components of a input vector nor the order of components of a output vector has a semantical meaning. Permutations do not have any effect. Contrary, quaternions are tuples. Also, optimal separation of polarized signals by a $\mathbb{H}$-MLP relies on the preserving of structural information, which needs tuples of data beforehand. In a formal notion and in a wider context this is discussed for the broader class of networks with values in Clifford algebras in [11]. Here, there are four possibilities for labeling the "1" class:

$$\begin{aligned} \text{CL1} &: 1+0\mathtt{i}+0\mathtt{j}+0\mathtt{k} && (21)\\ \text{CL2} &: 0+1\mathtt{i}+0\mathtt{j}+0\mathtt{k} && (22)\\ \text{CL3} &: 0+0\mathtt{i}+1\mathtt{j}+0\mathtt{k} && (23)\\ \text{CL4} &: 0+0\mathtt{i}+0\mathtt{j}+1\mathtt{k} && (24) \end{aligned}$$

The latter three, refered to as imaginary labels, are isomorphic. Let $x = x_0 + x_1\mathtt{i} + x_2\mathtt{j} + x_3\mathtt{k}$ be fixed. Furthermore, let $r_0 + r_1\mathtt{i} + r_2\mathtt{j} + r_3\mathtt{k} = (a + b\mathtt{i} + c\mathtt{j} + d\mathtt{k})\,x$. Then yields

$$\begin{aligned} (-b+a\mathtt{i}-d\mathtt{j}+c\mathtt{k})\,x &= -r_1+r_0\mathtt{i}-r_3\mathtt{j}+r_2\mathtt{k} && (25)\\ (-c+d\mathtt{i}+a\mathtt{j}-b\mathtt{k})\,x &= -r_2+r_3\mathtt{i}+r_0\mathtt{j}-r_2\mathtt{k} && (26)\\ (-d-c\mathtt{i}+b\mathtt{j}+a\mathtt{k})\,x &= -r_3-r_2\mathtt{i}+r_1\mathtt{j}+r_0\mathtt{k}. && (27) \end{aligned}$$

Hence applying the appropriate isomorphism to every network parameter allows to construct an equivalent network having another imaginary class label than the original given one.

5. SIMULATIONS

This section reports results for separation of $\mathbb{C}^{\eta}$-proper vs $\mathbb{H}$-proper signals comparing the previously introduced two types of networks. Note that this task is more complicated than separation of $\mathbb{C}^{\eta}$-proper vs $\mathbb{C}^{\mu}$-proper signals (section 3.3). Simulations have been performed for synthetic data generated from the three four–dimensional distributions listed in table 1. From each of them 1000 points have been sampled. In each case the first 200 sample points have been used for training and the remaining 800 points for testing. Note that $\mathscr{D}_3$ is $\mathbb{H}$-proper and both $\mathscr{D}_1$ and $\mathscr{D}_2$ are $\mathbb{C}^{\mathtt{i}}$-proper, respectively (w.r.t. the canonical identification $\{\mathrm{a} \Leftrightarrow \mathtt{1}, \mathrm{b} \Leftrightarrow \mathtt{i}, \mathrm{c} \Leftrightarrow \mathtt{j}, \mathrm{d} \Leftrightarrow \mathtt{k}\}$. $\mathbb{C}^{\eta}$-proper data for $\eta \in \{\mathtt{j}, \mathtt{k}\}$ has been obtained from $\mathscr{D}_1, \mathscr{D}_2$ by simple permutation of components. For example, $\{\mathrm{a} \Leftrightarrow \mathtt{1}, \mathrm{b} \Leftrightarrow \mathtt{j}, \mathrm{c} \Leftrightarrow \mathtt{i}, \mathrm{d} \Leftrightarrow \mathtt{k}\}$ gives $\mathbb{C}^{\mathtt{j}}$-proper data. For the MLP all that way derived data is the same (section 4.2). Hence the number of setups reduces to two in this case.

	a	b	c	d
$\mathscr{D}_1$	$\mathscr{N}(0,1)$	$\mathscr{N}(0,1)$	$2a+4b$	$4a-2b$
$\mathscr{D}_2$	$\mathscr{N}(0,1)$	$\mathscr{N}(0,1)$	$1a+2b$	$2a-1b$
$\mathscr{D}_3$	$\mathscr{N}(0,1)$	$\mathscr{N}(0,1)$	$\mathscr{N}(0,1)$	$\mathscr{N}(0,1)$

Table 1: Distributions from which data for simulations has been generated. $\mathscr{N}(0,1)$ denotes one–dimensional normal distribution with mean 0 and variance 1.

5.1 Results for the MLP

Separation of $\mathscr{D}_1$ vs $\mathscr{D}_3$ turned out to be much more difficult than $\mathscr{D}_2$ vs $\mathscr{D}_3$ separation ($\mathscr{D}_3$ class was labeled as 0). Training in the latter case was always successful with 6 hidden nodes (3000 iterations, learning rate 0.1) yielding zero ore only misclassified pattern. Test performance varied between 97%-98% classification rate for different runs. Results for the other separation task are worse w.r.t. both efficiency and accuracy. With 20 hidden nodes (optimal learning parameters) a training error of 6-8 misclassified patterns was achieved. Here test performance varied between 93%-94% classification rate. Using more hidden nodes caused overfitting (worse test performance, see *e.g.*[7]) without reducing training error significantly. The lack of the MLP to actually identify structure (levels of properness) caused rather low performance in this particular task.

5.2 Results for the $\mathbb{H}$-MLP

$\mathscr{D}_3$ class was labeled as $0+0\mathtt{i}+0\mathtt{j}+0\mathtt{k}$ and for $\mathbb{C}^{\eta}$-proper data all four possible labels (21)-(24) have been tested yielding a total of 24 different setups. In 23 of these setups the $\mathbb{H}$-MLP, having only 2 hidden nodes, was successful with 0-5 misclassified training patterns and 93.5%-99.5% classification rate for test data. One particular solution for $\mathbb{C}^{\mathtt{i}}$-proper vs $\mathbb{H}$-proper separation using $\mathscr{D}_1$ and CL4 (24) for the $\mathbb{C}^{\mathtt{i}}$-proper class is listed in table 2. This solution corresponds to zero training error and 99.5% successful classification of test data. Note that the hidden weights w_{21}, w_{22} are basically just used for scaling. This means that separation has already been done in the hidden layer. Also note the symmetry of the $\mathtt{1}, \mathtt{j}$-components of the hidden layer weights. Same accuracy was obtained using the other class labels, which is always guaranteed theoretically for CL2, CL3 by the argumentation provided in section 4.2.

Weight	1	i	j	k
w_{11}	-9.350	-0.265	+9.374	+18.688
w_{12}	+8.925	+0.430	-8.955	-17.823
θ_{11}	+2.539	-2.369	-4.933	+1.275
θ_{12}	+5.350	-8.112	-6.532	+1.929
w_{21}	+3.965	+0.109	-0.115	-0.142
w_{22}	+3.878	+0.103	-0.022	-0.119
θ_{21}	-33.694	-20.531	-18.675	-23.595

Table 2: $\mathbb{C}^{\mathtt{i}}$-proper vs $\mathbb{H}$-proper separation using $\mathscr{D}_1$ and CL4 (24) for the $\mathbb{C}^{\mathtt{i}}$-proper class. Solution found by the quaternionic network yielding perfect separation (zero error) for the training sets.

Only for $\mathbb{C}^{\mathtt{j}}$-proper vs $\mathbb{H}$-proper signal separation different class labels caused non–equal performance. Output of the first hidden neuron and the second hidden neuron for a typical run using CL4 are shown in figure 1 and figure 2, respectively. Again, successful separation was already performed in the hidden layer. The $\mathbb{C}^{\mathtt{j}}$-proper class was solely coded in the $\mathtt{k}$-channel of the hidden neurons. Therfore output weights have been used again only for scaling. With CL3 used for coding the "same" solution is obtained. However, the $\mathbb{C}^{\mathtt{j}}$-proper class is then solely coded in the $\mathtt{j}$-channel. In both cases test classification rate was always in between 96%-97.5% for all runs. With CL2 coding this rate dropped down significantly to 93%-94.5%. This means the network was not able to reproduce the CL3, CL4 solutions, which can be done theoretically as outlined before. Actually, no other solution than the ones from CL3, CL4 could be observed when analyzing several runs. Switching class data in the output layer to the $\mathtt{i}$-channel from the $\mathtt{j}$-channel or $\mathtt{k}$-channel, respectively, needs a more complicated operation than scaling. Although theoretically possible for the $\mathbb{H}$-MLP, optimization becomes more complicated and the only locally convergent training algorithm may get trapped in local minima more likely. In the CL1 case, where there is also no other solution than the ones mentioned before, such operation does not exist at all. This therfore accounts for the only non–successful setting. Contrary, a CL1 solution well exists when casted as $\mathbb{C}^{\mathtt{i}}$-proper data, which can be seen from figure 3 (plot for the second hidden neuron is omitted again due to space limitations).

This plot shows a total different type of solution using two channels. Hence the difference of covariance matrices derived in section 3.3 also causes different solutions for separation. Summarizing all the simulations, either CL3 or CL4 should be used as class labels. Then optimal separation of $\mathbb{C}^{\eta}$-proper vs $\mathbb{H}$-proper signals is always possible by a $\mathbb{H}$-MLP having only two hidden neurons corresponding to a total of only 28 real parameters. Note that the gap to perfect 100% is both due to the gap between sample statistics and distribution statistics and also due to the local nature of training algorithm. The $\mathbb{H}$-MLP will usually outperform a standard MLP in terms of accuracy. The latter will always be less efficient in terms of complexity. In one of the simulations the standard MLP was not able to reach performance of the $\mathbb{H}$-MLP although having five times more parameters. Also, note that overfitting does not occur for the $\mathbb{H}$-MLP in any case.

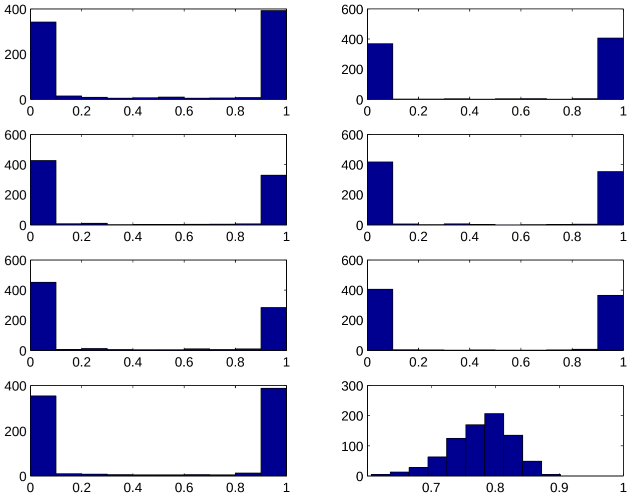

Figure 1: $\mathbb{C}^{\mathtt{j}}$-proper vs $\mathbb{H}$-proper separation using $\mathscr{D}_2$ and CL4 (24) for the $\mathbb{C}^{\mathtt{j}}$-proper class. Histograms of the output of the first hidden neuron. Left column shows response for $\mathbb{H}$-proper data. Right column shows response for $\mathbb{C}^{\mathtt{j}}$-proper data. Each row shows one component of the output.

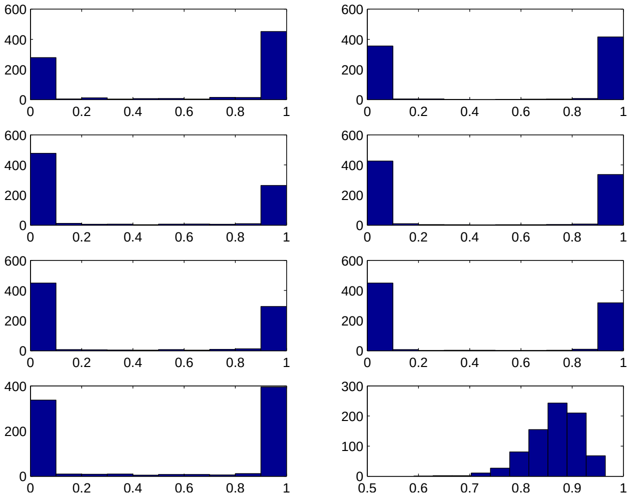

Figure 2: $\mathbb{C}^{\mathtt{j}}$-proper vs $\mathbb{H}$-proper separation using $\mathscr{D}_2$ and CL4 (24) for the $\mathbb{C}^{\mathtt{j}}$-proper class. Histograms of the output of the second hidden neuron.

6. CONCLUSION

We have presented a quaternionic approach for modeling polarized signals and their statistical properties. Staying inside the quaternionic framework by using a $\mathbb{H}$-MLP gave better results than using a real-valued MLP on signal separation tasks. Moreover, successful separation was achieved by a minimal quaternion architecture. Hence the presented results illustrate well the advantages of the quaternionic framework for processing of polarized signals.

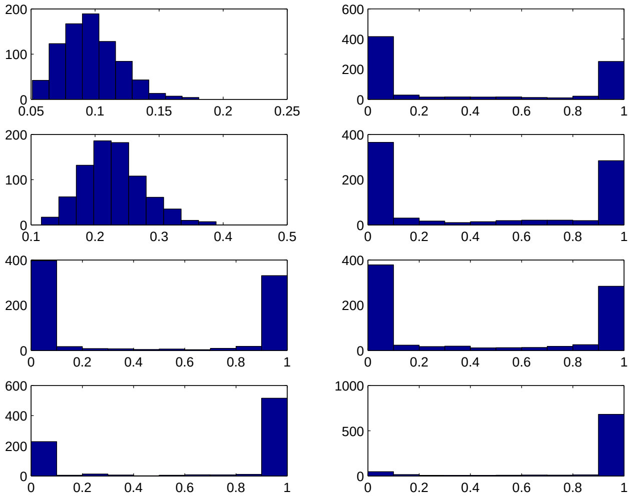

Figure 3: $\mathbb{C}^{\mathtt{i}}$-proper vs $\mathbb{H}$-proper separation using $\mathscr{D}_2$ and CL1 (24) for the $\mathbb{C}^{\mathtt{i}}$-proper class. Histograms of the output of the first hidden neuron.

REFERENCES

[1] S. Miron, N. Le Bihan and J. Mars, *High Resolution vector-sensor array processing using quaternions*, IEEE Workshop on Satistical Signal Processing, Bordeaux, 2005.

[2] P.-O. Amblard, and N. Le Bihan, *On properness of quaternion valued random variables*, 6$^{\text{th}}$ IMA Conference on Mathemathics in Signal Processing, Cirenceister, UK, 2004.

[3] N.N. Vakhania, *Random vectors with values in quaternion* Hilbert spaces, Theory of Probability and its Applications, Vol. 43, No. 1, pp. 99-115, 1999.

[4] F.D. Neeser, and J.L. Massey, *Proper Complex Random Process with applications to information theory*, IEEE Trans. on Information Theory, Vol. 39, No. 4, pp. 1293-1302, 1993.

[5] A. Nehorai, and E. Paldi, *Vector-sensor array processing for electromagnetic source localization*, IEEE Trans. on Signal Processing, Vol. 42, No. 2, pp. 376-398, 1994.

[6] B. Picinbono, *On circularity*, IEEE Trans. on Signal Processing, Vol. 42, No. 12, pp. 3473-3482, 1994.

[7] C.M. Bishop, *Neural Networks for Pattern Recognition*, Oxford University Press, Oxford, UK, 1995.

[8] T. Masters, *Signal and Image Processing with Neural Networks*, John Willey and Sons, 1994.

[9] P. Arena and L. Fortuna and G. Muscato and M.G. Xibilia, *Neural Networks in Multidimensional Domains*, LNCIS, Vol. 234, Springer–Verlag, 1998.

[10] D.E. Rumelhart and J.L. McClelland, *Parallel Distributed Processing*, MIT Press, Cambridge MA, 1986.

[11] S. Buchholz, *A Theory of Neural Computation with Clifford Algebras*, Technical Report Number 0504, CAU Kiel, Institut für Informatik, 2005.