# A QUATERNION KERNEL MINIMUM ERROR ENTROPY ADAPTIVE FILTER

Tokunbo Ogunfunmi, Carlo Safarian

Department of Electrical Engineering Santa Clara University Email :togunfunmi@scu.edu, csafarian@scu.edu

# ABSTRACT

In this paper, we develop a kernel adaptive filter for quaternion data based on minimum error entropy cost function. We apply generalized Hamilton-real (GHR) calculus that is applicable to Hilbert space for evaluating the cost function gradient to develop the quaternion kernel minimum error entropy (MEE) algorithm. The MEE algorithm minimizes Renyis quadratic entropy of the error between the filter output and desired response or indirectly maximizing the error information potential. Here, the approach is applied to quaternions for improving performance for biased or non-Gaussian signals compared with the minimum mean square error criterion of the kernel least mean square algorithm. Simulation results are used to verify the performance of the algorithm. Convergence is very fast and is shown to out-perform existing algorithms.

*Index Terms*— Adaptive filters, entropy, kernel least mean square (LMS) algorithm, quaternions

# 1. INTRODUCTION

Quaternion domain allows us to represent three or four dimensional signals in convenient way. The quaternion algebra reduces the number of parameters and computational complexity. Areas such as computer graphics, pattern recognition in images and motion tracking are significantly simplified using quaternions[1]-[5]. Some recent research results include the development of linear and widely linear filters based on quaternion data [1],[2] and the optimization in quaternion systems such as gradient learning algorithm [7]. The generalized Hamilton-real calculus (GHR) for the quaternion data is proposed in [6]. The GHR calculus simplified product and chain rules and allows us to calculate the quaternion based gradient and Hessian of cost function efficiently and use them for the learning algorithms. The quaternion reproducing kernel Hilbert spaces and its uniqueness is established in [8]. These provide a mathematical foundation to develop the quaternion value kernel learning algorithms. The reproducing property of the feature space replace the inner product of feature samples with kernel evaluation. The existence and uniqueness of quaternion reproducing kernel Hilbert space (QRKHS)

provide a theoretical basis for kernel algorithms operating in quaternion feature spaces. The quaternion kernel estimation is an emerging field and some algorithms are developed in quaternion domain [9]-[11]. The adaptive filtering cost function based on minimum square error (MSE) use only second order statistics and does not capture the probability of error distribution in the system. Another algorithm; Maximum Correntropy is based on local criterion and only cares about the local part of the error PDF falling within the kernel bandwidth. When the error modes are far from the origin, they fall outside the kernel bandwidth [16]. An information theoretic alternative is using Minimum Error Entropy as cost function and is expected to perform better with biased or non-Gaussian signals compared to MSE criteria adaptive filters. The Minimum Error Entropy has higher complexity than Maximum Correntropy but achieves robustness and efficiency by self adjusting the localness of the weighting function based on the error distribution [16]. The adaptive filter based on minimum error entropy (MEE) criteria is studied in [12]-[14].

In this paper, we describe a quaternion kernel adaptive filter based on minimum error entropy cost function that is referred to as the quaternion KMEE (QKMEE) algorithm. Section 2 covers the background material, Section 3 contains the algorithm derivation, Section 4 is simulation results and section 5 concludes the paper.

### 2. BACKGROUND

### 2.1. Quaternions and Properties

Quaternions are a 4-D associative, noncommutative, normed division algebra over the real numbers. The details about quaternions and the GHR calculus can be seen in [6][7] and [17].

### 2.2. Renyi Entropy and Parzen Window

Renyi's entropy definition such as the order- $\alpha$  Renyi's entropy is defined as [12]

$$H_{\alpha}(e) = \frac{1}{1-\alpha} \log \int_{-\infty}^{\infty} p_e^{\alpha}(e) de$$
(1)

where  $\alpha \in \mathbb{R}^+ \setminus \{1\}$  and  $p_e$  is probability distribution function of random variable e.

We can define order- $\alpha$  information potential  $V_{\alpha}$  as

$$V_{\alpha}(e) = \int_{-\infty}^{\infty} p_e^{\alpha}(e) de = \|p_e\|_{\alpha}^{\alpha}$$
(2)

where  $\|.\|^{\alpha}$  is standard norm- $\alpha$  in  $L_{\alpha}$ .

In practice the entropy function is not accessible since it is a function of the pdf of relative random variable e. With  $\alpha = 2$  the entropy can be estimated by using some specific method such as the *Parzen window* which is a good estimation of the *order–2 Renyi's entropy* function. For a set of N statistically independent random samples  $\{e_i\}_{i=1}^N$  of random variable e, the *Parzen window* computes the estimate of the probability distribution function  $p_e$  as

$$\hat{p_e}(e) = \frac{1}{N\sigma} \sum_{l=1}^{N} K(\frac{e-e_l}{\sigma}) = \frac{1}{N} \sum_{l=1}^{N} G_{\sqrt{2}\sigma}(e-e_l) \quad (3)$$

where K is the real value Gaussian Kernel and  $\sigma$  is the size of kernel and  $G_{\sqrt{2}\sigma}$  is defined as the following function

$$G_{\sqrt{2}\sigma}(e - e_l) = \frac{1}{\sqrt{2\pi}\sigma} \exp\{-\frac{(e - e_l)^2}{2\sigma^2}\}$$
(4)

The estimation of information potential  $\hat{V}(e)$  is given by

$$\hat{V}(e) = \frac{1}{N^2} \sum_{l_1=1}^{N} \sum_{l_2=1}^{N} G_{\sqrt{2}\sigma}(e_{l_1} - e_{l_2})$$
(5)

The global solution of maximization of the V(e) is the same as global solution of  $\hat{V}(e)$  with the Parzen window estimation and the global solution is achieved when all related errors are constant and the maximum value of V(e) is shown by V(0)or equally  $\hat{V}(0) = \frac{1}{\sqrt{2\sigma}}$ .

Gaussian-based kernel for quaternion data may be expressed as

$$\kappa_{\sigma}(X - Y) = \frac{4}{\sqrt{2\pi}\sigma} \exp\{\frac{-1}{2\sigma^2}(X_r - Y_r)^2 + (X_i - Y_i)^2 + (X_j - Y_j)^2 + (X_k - Y_k)^2\} \\ = \frac{4}{\sqrt{2\pi}\sigma} \exp\{\frac{-1}{2\sigma^2}|X - Y|^2\}$$

where X and Y are quaternion numbers  $\in \mathbb{H}$  in form of  $X = X_r + iX_i - jX_j - kX_k$  and  $Y = Y_r + iY_i - jY_j - kY_k$ . More details of the quaternion kernel is provided in [15].

Minimizing the error entropy can be done by maximizing the error information potential cost function J(n) in quaternion domain  $\mathbb{H}$  which can be defined as

$$J(n) = \frac{1}{N^2} \sum_{i,j=1}^{N} G_{q,\sqrt{2}\sigma}(e(n-i) - e(n-j))$$
(6)

where

$$G_{q,\sqrt{2}\sigma}(e_i - e_j) = \frac{4}{\sqrt{2\pi\sigma}} \exp\{-\frac{|e_i - e_j|^2}{2\sigma^2}\}$$
(7)

and  $e_i$  and  $e_i \in \mathbb{H}$ . Based on the error information potential based cost function in quaternion domain we develop *Quaternion Minimum Error Entropy algorithm* in next section.

### 3. QUATERNION MINIMUM ERROR ENTROPY ALGORITHM DERIVATION

For the quaternion kernel adaptive filter based on minimum entropy (QKMEE) with quaternion data, the goal is to maximize the information potential cost function J(n) (6). The filter can be expressed as  $y_n = \langle \Phi(\mathbf{u}_n), \mathbf{w}_n \rangle$ , which also can be written as

$$y_n = \mathbf{w}_n^H \varphi_n \tag{8}$$

where  $\varphi_n = \Phi(\mathbf{u}_n)$  and  $\Phi(.)$  is the kernel map to a quaternion RKHS [15]. Maximizing the information potential cost function J(n) (6) can be done with unconstrained optimization algorithm such as gradient ascent algorithm.

$$\mathbf{w}_{n+1} = \mathbf{w}_n + \eta \nabla_{\mathbf{w}_n^*} J(n) = \mathbf{w}_n + \eta \left(\frac{\partial J(n)}{\partial \mathbf{w}_n}\right)^H$$
$$= \mathbf{w}_n + \mu \left(\frac{\partial \left[\sum_{l,t=1}^N \exp\{-\frac{|e(n-l)-e(n-t)|^2}{2\sigma^2}\}\right]}{\partial \mathbf{w}_n}\right)^H$$
(9)

where  $\eta$  is adaptation step size and  $\mu = \eta \frac{1}{N^2} \frac{4}{\sqrt{2\pi\sigma}}$ , and  $e(n-l) = d(n-l) - \mathbf{w}_n^H \varphi_{n-l}$  as *posteriori errors* for each  $l: 1 \leq l \leq N$ .

To derive the gradient of cost function we define functions  $f: \mathbb{H} \to \mathbb{H}$  and  $g_{l,t}: \mathbb{H} \to \mathbb{R}$  as

$$f(x) = \exp(x) \tag{10}$$

$$g_{l,t}(\mathbf{w}_n) = -\frac{|e(n-l) - e(n-t)|^2}{2\sigma^2}$$
 (11)

To simplify the notation for function  $g_{l,t}$  in our derivative for a given l and  $t, 1 \le l \le N$  and  $1 \le t \le N$  we define  $g(\mathbf{w}_n) = g_{l,t}(\mathbf{w}_n) = -\frac{|e(n-l)-e(n-t)|^2}{2\sigma^2}$ .

with the above notation the equation (9) can be written as

$$\mathbf{w}_{n+1} = \mathbf{w}_n + \mu \left(\sum_{l=1}^N \sum_{t=1}^N \frac{\partial \left[f(g_{l,t}(\mathbf{w}_n))\right]}{\partial \mathbf{w}_n}\right)^H$$
(12)

For a given l and t, the partial derivative can be calculated with GHR chain rule as

$$\frac{\partial \left[ f(g_{l,t}(\mathbf{w}_n)) \right]}{\partial \mathbf{w}_n} = \frac{\partial \left[ f(g(\mathbf{w}_n)) \right]}{\partial \mathbf{w}_n} = \sum_{v \in \{1,i,j,k\}} \frac{\partial f}{\partial g^v} \frac{\partial g^v}{\partial \mathbf{w}_n}$$
(13)

Using HR derivative property and quaternion rotation, for  $\forall v \in \{i, j, k\}$  we can show that  $\frac{\partial f}{\partial g^v} = 0$ . Suppose v = i then

$$\frac{\partial f}{\partial g^{i}} = \frac{1}{4} \left( \frac{\partial f}{\partial g_{r}} - i \frac{\partial f}{\partial g_{i}} + j \frac{\partial f}{\partial g_{j}} + k \frac{\partial f}{\partial g_{k}} \right)$$

$$= \frac{1}{4} \left( \frac{\partial \exp(g)}{\partial g_{r}} - i \frac{\partial \exp(g)}{\partial g_{i}} + j \frac{\partial \exp(g)}{\partial g_{j}} + k \frac{\partial \exp(g)}{\partial g_{k}} \right)$$

$$= \frac{1}{4} \left( \exp(g) - i i \exp(g) + j j \exp(g) + k k \exp(g) \right)$$

$$= \frac{1}{4} \left( \exp(g) + \exp(g) - \exp(g) - \exp(g) \right) = 0$$
(14)

And if v = 1 then

$$\frac{\partial f}{\partial g} = \frac{1}{4} \left( \frac{\partial f}{\partial g_r} - i \frac{\partial f}{\partial g_i} - j \frac{\partial f}{\partial g_j} - k \frac{\partial f}{\partial g_k} \right)$$

$$= \exp(g)$$
(15)

By substituting (14) and (15) in (13) we can simplify (13) as follow

$$\sum_{v \in \{1,i,j,k\}} \frac{\partial f}{\partial g^v} \frac{\partial g^v}{\partial \mathbf{w}_n} = \exp(g) \frac{\partial g}{\partial \mathbf{w}_n}$$
(16)

$$g(\mathbf{w}_{n}) = -\frac{|e(n-l) - e(n-t)|^{2}}{2\sigma^{2}}$$

$$= \frac{-1}{2\sigma^{2}}|e(n-l) - e(n-t)|^{2}$$

$$= \frac{-1}{2\sigma^{2}}[(e(n-l) - e(n-t))(e(n-l) - e(n-t))^{*}]$$

$$= \frac{-1}{2\sigma^{2}}[(e(n-l) - e(n-t))(e^{*}(n-l) - e^{*}(n-t))]$$

$$= \frac{-1}{2\sigma^{2}}[|e(n-l)|^{2} + |e(n-t)|^{2} - e(n-l)e^{*}(n-t)]$$

$$- e(n-t)e^{*}(n-l)]$$
(17)

therefore by substituting (17) in (16) we can find partial derivative of g using GHR calculus as below

$$\frac{\partial g}{\partial \mathbf{w}_n} = \left(\frac{-1}{2\sigma^2}\right) \left[\frac{\partial |e(n-l)|^2}{\partial \mathbf{w}_n} + \frac{\partial |e(n-t)|^2}{\partial \mathbf{w}_n} - \frac{\partial e(n-l)e^*(n-l)}{\partial \mathbf{w}_n} - \frac{\partial e(n-t)e^*(n-l)}{\partial \mathbf{w}_n}\right]$$
(18)

by substituting  $e(n-l) = d(n-l) - \mathbf{w}_n^H \varphi_{n-l}$  in (18) and using GHR calculus, we can compute each partial derivative of (18) as

$$\frac{\partial |e(n-l)|^2}{\partial \mathbf{w}_n} = \frac{\partial e(n-l)e^*(n-l)}{\partial \mathbf{w}_n}$$

$$= e(n-l)\frac{\partial e^*(n-l)}{\partial \mathbf{w}_n} + \frac{\partial e(n-l)}{\partial \mathbf{w}_n^{e^*(n-l)}}e^*(n-l)$$
(19)

where

$$\frac{\partial e^*(n-l)}{\partial \mathbf{w}_n} = -\varphi_{n-l}^H \frac{\partial \mathbf{w}_n}{\partial \mathbf{w}_n} = -\varphi_{n-l}^H \tag{20}$$

and the second term of (19) can be calculated as

$$\frac{\partial e(n-l)}{\partial \mathbf{w}_{n}^{e^{*}(n-l)}}e^{*}(n-l) = -\frac{\partial \mathbf{w}_{n}^{H}\varphi_{n-l}}{\partial \mathbf{w}_{n}^{e^{*}(n-l)}}e^{*}(n-l)$$

$$= -\mathbf{w}_{n}^{H}\frac{\partial \varphi_{n-l}}{\partial \mathbf{w}_{n}^{e^{*}(n-l)}}e^{*}(n-l) - \frac{\partial \mathbf{w}_{n}^{H}}{\partial \mathbf{w}_{n}^{\varphi_{n-l}e^{*}(n-l)}}\varphi_{n-l}e^{*}(n-l)$$

$$= \frac{1}{2}(\varphi_{n-l}e^{*}(n-l))^{H}$$

$$= \frac{1}{2}e(n-l)\varphi_{n-l}^{H}$$
(21)

by substituting (20) and (21) in (19) we can obtain

$$\frac{\partial |e(n-l)|^2}{\partial \mathbf{w}_n} = -\frac{1}{2}e(n-l)\varphi_{n-l}^H \tag{22}$$

Using the same method, the other terms of (18) can be calculated. By substituting all partial derivatives , we can simplify (18) as below

$$\frac{\partial g}{\partial \mathbf{w}_{n}} = \left(\frac{-1}{2\sigma^{2}}\right) \left(-\frac{1}{2}e(n-l)\varphi_{n-l}^{H} - \frac{1}{2}e(n-t)\varphi_{n-t}^{H} + \frac{1}{2}e(n-l)\varphi_{n-t}^{H} + \frac{1}{2}e(n-t)\varphi_{n-l}^{H}\right) \\
= \left(\frac{1}{4\sigma^{2}}\right) \left[e(n-l) - e(n-t)\right] \left[\varphi_{n-l}^{H} - \varphi_{n-t}^{H}\right] \tag{23}$$

Therefore by substituting (23) in (16) we can obtain

$$\frac{\partial J(n)}{\partial \mathbf{w}_{n}} = \sum_{l=1}^{N} \sum_{t=1}^{N} \frac{\partial \left[f(g_{l,t}(\mathbf{w}_{n}))\right]}{\partial \mathbf{w}_{n}} \\
= \sum_{l=1}^{N} \sum_{t=1}^{N} \sum_{v \in \{1,i,j,k\}}^{N} \frac{\partial f}{\partial g^{v}} \frac{\partial g^{v}}{\partial \mathbf{w}_{n}} \\
= \left(\frac{1}{4\sigma^{2}}\right) \times \sum_{l=1}^{N} \sum_{t=1}^{N} \exp(g_{l,t}(\mathbf{w}_{n})) \\
\times \left[e(n-l) - e(n-t)\right] \left[\varphi_{n-l}^{H} - \varphi_{n-t}^{H}\right] \right)$$
(24)

Therefore the gradient of the cost function J(n) can be calculated based on the following equation

$$\nabla_{\mathbf{w}_{n}^{*}} J(n) = \left(\frac{\partial J(n)}{\partial \mathbf{w}_{n}}\right)^{H}$$
$$= \left(\frac{1}{4\sigma^{2}}\right) \times \left[\sum_{l=1}^{N} \sum_{t=1}^{N} \exp(g_{l,t}(\mathbf{w}_{n})) \right]$$
$$\times \left[e(n-l) - e(n-t)\right] \left[\varphi_{n-l}^{H} - \varphi_{n-t}^{H}\right]^{H}$$
(25)

by setting  $\mathbf{w}_0 = 0$  and replacing  $\exp(g)$  with its kernel equivalent  $\kappa_\sigma$  we can obtain filter output weight as

$$\mathbf{w}_{n} = \zeta \sum_{p=0}^{n-1} \sum_{l=1}^{N} \sum_{t=1}^{N} \left( \left[ \kappa_{\sigma} \left( e(p-l) - e(p-t) \right) \right] \times \left[ e(p-l) - e(p-t) \right] \left[ \varphi_{p-l}^{H} - \varphi_{p-t}^{H} \right] \right)^{H}$$
(26)

where  $\zeta = \mu \sqrt{2\pi}/16\sigma = \eta \frac{1}{4N^2\sigma^2}$ . By substituting the weight update in the  $y_n = \mathbf{w}_n^H \varphi_n$  and using properties of Quaternion Reproducing Kernel Hilbert Space (QRKHS) and the 'kernel trick' to replace the inner product of two vectors with quaternion kernel  $\bar{\kappa}_{\bar{\sigma}}$ , we can simplify the equation in kernel form as

$$y_n = \zeta \sum_{p=0}^{n-1} \sum_{l=1}^{N} \sum_{t=1}^{N} \left[ \kappa_{\sigma}(e(p-l) - e(p-t)) \right] \\ \times \left[ e(p-l) - e(p-t) \right] \left[ \bar{\kappa}_{\bar{\sigma}}(\mathbf{u_{p-l}}, \mathbf{u_n}) - \bar{\kappa}_{\bar{\sigma}}(\mathbf{u_{p-t}}, \mathbf{u_n}) \right]$$
(27)

#### 4. SIMULATION RESULTS

The Quat-KMEE (QKMEE) algorithm was simulated for a nonlinear channel with non-Gaussian noise versus Quat-KLMS[15].The channel consisted of the quaternion filter, i.e.,  $z(n) = g_1^*u(n) + g_2^*u^i(n) + g_3^*u^j(n) + g_4^*u^k(n)$  $+h_1^*u(n-1) + h_2^*u^i(n-1) + h_3^*u^j(n-1) + h_4^*u^k(n-1)$ and nonlineraity, i.e.,  $y(n) = z(n) + az^2(n) + bz^3(n) + v(n)$ 

where v(n) is added non-Gaussian noise described later. Coefficients g1, ..., g4, h1, ..., h4, a, b, and noise v(n) are all quaternion valued. The coefficients used were a = 0.075 + i0.35 + j0.1 - k0.05, b = -0.025 - i0.25 - j0.05 + k0.03, g1 = -0.40 + i0.30 + j0.15 - k0.45, h1 = 0.175 - i0.025 + j0.1 + k0.15, g2 = -0.35 - i0.15 - j0.05 + k0.20, h2 = 0.15 - i0.225 + j0.125 - k0.075, g3 = -0.10 - i0.40 + j0.20 - k0.05, h3 = +0.025 + i0.075 - j0.05 - k0.05, g4 = +0.35 + i0.10 - j0.10 - k0.15, h4 = -0.05 - i0.075 - j0.075 + k0.175.

For the tests, both input u(n) and noise v(n) were formed using impulsive Gaussian mixture models to form non-Gaussian signals [8]. The probability distributions used were  $p_u(i) = (0.85N(1.0, 0.01) + 0.15N(3.0, 0.01))$ +i(0.40N(0.5, 0.01)+0.60N(2.5, 0.01))+j(0.65N(3.5, 0.01)+0.35N(1.5, 0.01)) + k(0.25N(2.0, 0.01) + 0.75N(5.5, 0.01)) $p_{v}(i) = (0.90N(0.0, 0.01) + 0.10N(1.0, 0.01))$ +i(0.70N(3.0, 0.01)+0.30N(0.5, 0.01))+i(0.45N(1.0, 0.01)+0.55N(4.5, 0.01)) + k(0.80N(0.5, 0.01) + 0.20N(1.5, 0.01))where  $N(m_N, \sigma_N)$  denotes the normal (Gaussian) pdf with mean  $m_N$  and variance  $\sigma_N$ . The Quat-KMEE and Quat-KLMS simulation results for the nonlinear channel described are shown in Fig. 1. The convergence analysis results will be in future work and guided the choice of the parameters. The parameters for the Quat-KMEE were  $\eta = 0.35$ ,  $\bar{\sigma} = 2.24$ , and  $\sigma = 0.736$ , and for the Quat-KLMS  $\eta = 0.35$ ,  $\bar{\sigma} = 2.24$ were used. The results show improvement of Quat-KMEE for modeling nonlinear channel when the input noise is non-Gaussian compared with Quat-KLMS.

### Quat-KMEE vs Quat-KLMS (Non-Gaussian input, noise)



Fig. 1. Quat-KLMS and Quat-KMEE for non-Gaussian signal

### 5. CONCLUSION

We have shown the derivation and demonstration of convergence of a quaternion kernel adaptive algorithm based on minimum error entropy. The algorithm is based on information theoretic learning (ITL) cost function. The resulting algorithm is the Quat-KMEE algorithm using GHR calculus. A gradient is derived based on quaternion RKHS. Simulation results show the convergence curve of the mean square error of the new algorithm (QKMEE) versus the existing algorithm (QKLMS). The algorithm's convergence is very fast and outperforms the existing one QKLMS but has higher complexity. The QKMEE algorithm performed better with non-Gaussian signals compared to QKLMS which is based on the MSE criteria adaptive filters.

### 6. REFERENCES

- C. C. Took and D. P. Mandic, "A quaternion widely linear adaptive filter," in *Trans. Signal Process.* IEEE,2010, vol. 58, no. 8, pp. 44274431.
- [2] S. Javidi, C. C. Took, C. Jahanchahi, N. Le Bihan, and D. P. Mandic, "Blind extraction of improper quaternion sources," in *Proc. Inter. Conf. Acoust. Speech Signal Process.* IEEE,2011, pp. 30783711.
- [3] D. P. Mandic, C. Jahanchahi, and C. C. Took, "A quaternion gradient operator and its applications," in *Signal Process. Lett.* IEEE,2011, vol. 18, no. 1, pp. 4750.
- [4] A. Shilton and D. T. H. Lai, "Quaternionic and complexvalued support vector regression for equalization and function approximation," in *Proc.Inter.Joint Conf. Neural Netw.*, 2007, pp. 920925.
- [5] E. J. B. -Corrochano and N. A. -Daniel, "Clifford support vector machines for classification, regression, and recurrence," in *Trans. Neural Netw.* IEEE,2010, vol. 21, no. 11, pp. 17311746.
- [6] Dongpo Xu, Cyrus Jahanchahi, Clive C. Took and Danilo P. Mandic, "Quaternion Derivatives: The GHR Calculus," in Submitted exclusively to the London Mathematical Society doi:10.1112/0000/000000. 2014
- [7] Dongpo Xu, Yili Xia, Member, IEEE, and Danilo P. Mandic, "Optimization in Quaternion Dynamic Systems:Gradient, Hessian, and Learning Algorithms," in *Transactions on Neural Networks and Learning Systems*. IEEE,2016, VOL. 27, NO. 2, pp. 249 - 261.
- [8] Felipe A. Tobar and Danilo P. Mandic, "Quaternion Reproducing Kernel Hilbert Spaces:Existence and Uniqueness Conditions," in *Transactions on Information Theory* . IEEE, 2014, Volume: 60, Issue: 9, pp. 5736 - 5749.
- [9] Tokunbo Ogunfunmi and Thomas Paul, "The Quarternion Maximum Correntropy Algorithm," in *Transactions on Circuits and Systems II*. IEEE,2015, Volume: 62, Issue: 6, pp. 598 - 602.
- [10] F. A. Tobar and D. P. Mandic, "The quaternion kernel least squares," in *Proc. ICASSP*. IEEE,2013, pp. 61286132.
- [11] Q. Barthelemy, A. Larue, and J. I. Mars, "About QLMS derivations," in *Signal Process. Lett*. IEEE,2014,vol. 21, no. 2, pp. 240243.
- [12] Wallace A. Martins; Paulo S. R. Diniz; Yih-Fang Huang, "On the normalized minimum error-entropy adaptive algorithm: Cost function and update recursion," in 2010 First IEEE Latin American Symposium on Circuits and Systems (LASCAS). IEEE, 2010, pp. 1-4.

- [13] Deniz Erdogmus; Jose C. Principeg, "An error-entropy minimization algorithm for supervised training of nonlinear adaptive systems," in *Transactions on Signal Processing*. IEEE, 2002, Volume: 50, Issue: 7, pp. 1780 - 1786.
- [14] Pratik Prabhanjan Brahma, 'Prediction of Sunspot Number Using Minimum Error Entropy Cost Based Kernel Adaptive Filters," in 14th International Conference on Machine Learning and Applications (ICMLA). IEEE, 2015.
- [15] Thomas K. Paul and Tokunbo Ogunfunmi, "A Kernel Adaptive Algorithm for Quaternion-Valued Inputs," in *Transactions on Neural Networks and Learning Systems*. IEEE, 2015, Volume: 26, Issue: 10, pp. 2422 - 2439.
- [16] Weifeng Liu, P. P. Pokharel, J. C. Principe, "Error Entropy, Correntropy and M-Estimation," in 2006 16th IEEE Signal Processing Society Workshop on Machine Learning for Signal Processing. IEEE, 2006, pp. 179-184.
- [17] M. D. Jiang, Y. Li, and W. Liu, "Properties of a General Quaternion-Valued Gradient Operator and Its Application to Signal Processing," in *Frontiers of Information Technology and Electronic Engineering*. 2016, vol. 17, issue: 2, pp. 83-95.