THE QUATERNION KERNEL LEAST SQUARES

Felipe A. Tobar and Danilo P. Mandic, Fellow, IEEE

Electrical and Electronic Engineering Department
Imperial College London, UK

ABSTRACT

The quaternion kernel least squares algorithm (QKLS) is introduced as a generic kernel framework for the estimation of multivariate quaternion valued signals. This is achieved based on the concepts of quaternion inner product and quaternion positive definiteness, allowing us to define quaternion kernel regression. Next, the least squares solution is derived using the recently introduced $\mathbb{HR}$ calculus. We also show that QKLS is a generic extension of standard kernel least squares, and their equivalence is established for real valued kernels. The superiority of the quaternion-valued linear kernel with respect to its real-valued counterpart is illustrated for both synthetic and real-world prediction applications, in terms of accuracy and robustness to overfitting.

***Index Terms*—** Kernel least squares, quaternion estimation, quaternion kernels, body motion tracking.

1. INTRODUCTION

Kernel algorithms for estimation paradigms have attracted considerable attention over the last decade due to their enhanced function approximation ability [1–5]. The generic nature of kernel algorithms has also allowed for complex-valued extensions of real-valued kernels algorithms [6–8]; these inherit the advantages offered by the complex division algebra in general bivariate estimation applications. On the other hand, the recent progress in sensor technology and data acquisition systems has brought to light three- and four-dimensional sensor signals, in areas including wind prediction, econometric signal estimation and spatial motion tracking. These require appropriate learning strategies (in particular kernel algorithms) to match the multidimensional data natures. For three- and four-dimensional data, a natural extension of the existing real- and complex-valued kernel algorithms is a quaternion-valued kernel approach, in which the algorithm operates in a (left) vector space built upon the scalar input-output domain.

The quaternion set $\mathbb{H}$ is a four-dimensional vector space over the real field $\mathbb{R}$. A quaternion $q \in \mathbb{H}$ has one real and three imaginary parts, respectively denoted by $q_a, q_b, q_c, q_d \in \mathbb{R}$, and can therefore be expressed as $q = q_a + iq_b + jq_c + kq_d$, where i, j, k are the imaginary units. The properties of the quaternion normed division algebra provide physically meaningful representation and more accurate rotation and orientation modelling of 3D objects than real vectors. Thus, quaternion valued algorithms exhibit more degrees of freedom than their real-valued counterparts, and are particularly suited for rotation and orientation applications as well as unified 3D and 4D modelling [9, 10]. Our aim is to show that kernel regression also benefits from the enhanced dimensionality of quaternion valued kernels, and their inherent ability to both represent inter-dependence between signal components and to model highly coupled multidimensional data features.

Existing research on the development of kernels for quaternion signals only considers real-valued kernels [11, 12]. The design of quaternion-valued kernel will have to address: (i) the existence of quaternion valued reproducing kernel Hilbert spaces (QRKHS) and (ii) quaternion analyticity. We focus on the existence of QRKHS and its relationship with positive definite kernels based on extensions of both the Riesz representation theorem [13] and the Moore-Aronaszajn theorem [14] for quaternion left Hilbert spaces, as these (together with the Mercer's theorem for quaternion kernels [15]) guarantee theoretical consistency, physical meaning of parameters, and ease of implementation of quaternion kernel algorithms. The issue of quaternion analyticity arises in the usual formulation of quaternion estimation algorithms, which is based on a real-valued cost function of quaternion variable that is dependent of both the error and its Hermitian $(\cdot)^H$, $J = \mathbf{e}^H\mathbf{e}$. However, the derivative of a real function of quaternion variable is not defined in the standard Cauchy-Riemann-Fueter sense [16], and we need to resort to the $\mathbb{HR}$ calculus [17].

In this work, we first revisit the kernel least squares (KLS) [18–20] with a regularised cost function. We then generalise the real- and complex-valued RKHS framework by introducing the conditions for the existence and uniqueness of quaternion-valued RKHS and positive definite kernels. These represent the basis for the development of the optimal least squares solution for both real- and quaternion-valued kernels. Finally, we validate the proposed QKLS by comparing real- and quaternion-valued kernels, on the prediction of synthetic autoregressive processes and real-world 3D inertial body sensor signals.

2. THE KERNEL LEAST SQUARES ALGORITHM

Consider the sets $\mathbf{X} \subseteq \mathbb{R}^n$, $D \subseteq \mathbb{R}$ and the mapping:

$$f : \mathbf{x} \in \mathbf{X} \quad \longmapsto \quad y = f(\mathbf{x}) \in D. \tag{1}$$

Kernel regression algorithms [3–8,19–21] aim to approximate $y = f(\mathbf{x})$ by mapping the input $\mathbf{x}$ onto a reproducing kernel Hilbert space (RKHS) $\mathcal{H}$ according to $\mathbf{x} \longmapsto \phi(\mathbf{x})$, in order to yield the estimate of the function $f(\cdot)$ through a linear transformation in the feature space, that is

$$\hat{y} = \langle \phi(\mathbf{x}), \omega \rangle, \tag{2}$$

where $\omega \in \mathcal{H}$ is a weighting factor and the RKHS $\mathcal{H}$ is referred to as the feature space.

According to the signal subspace principle (SSP) [2], the weighting factor can be projected onto the sample feature space $\mathcal{H}_S = \text{span}\{\phi(\mathbf{x}_i)\}$, where $\{\mathbf{x}_i\}_{i=1,\dots,N}$ are referred to as *support vectors*. Therefore, if ϕ is chosen to be an expansion function of the reproducing kernel of $\mathcal{H}$, that is $K(\mathbf{x}_a, \mathbf{x}_b) = \phi^T(\mathbf{x}_a)\phi(\mathbf{x}_b)$, the kernel estimate can be expressed via the *kernel trick* [4] as

$$\hat{y} = \sum_{i=1}^{N} \alpha_i K(\mathbf{x}_i, \mathbf{x}). \tag{3}$$

Using the set of observations $\{y_j\}_{j=1,\dots,N}$ corresponding to the support vectors, the set of optimal least squares parameters $\mathbf{a} = [\alpha_1, ..., \alpha_N]^T$ can be found via the minimisation of the regularised cost function

$$J = \frac{1}{2} \sum_{j=1}^{N} \left(y_j - \sum_{i=1}^{N} \alpha_i K(\mathbf{x}_i, \mathbf{x}_j) \right)^2 + \frac{\rho}{2} \sum_{i=1}^{N} \alpha_i^2, \tag{4}$$

where ρ is the regularisation parameter. This concept is also known as kernel ridge regression [20].

Setting $\frac{\partial J}{\partial \mathbf{a}} = \mathbf{0}$ and solving for $\mathbf{a}$, we obtain the Wiener solution

$$\mathbf{a} = (\mathbf{K}\mathbf{K}^T + \rho \mathbf{I})^{-1} \mathbf{K} \mathbf{y}, \tag{5}$$

where $\mathbf{y} = [y_1, ..., y_N]^T$, $\mathbf{I}$ is the identity matrix, and $\mathbf{K}$ is the kernel matrix evaluated over the set of support vectors given by $\mathbf{K}_{ij} = K(\mathbf{x}_i, \mathbf{x}_j)$.

The estimate (3), where $\mathbf{a}$ is optimally obtained from (5), is the kernel extension of least squares regression and has been discussed in [18–21]. We will refer to this algorithm as the kernel least squares (KLS) algorithm.

3. QUATERNION KERNEL LEAST SQUARES

Recent results show clear advantage of kernel regression algorithms for complex-valued signals [6–8], in which the kernels themselves are also complex-valued. We next show that the quaternion-valued kernel will inherit this property by continuity, since both the real, complex and quaternion domain are normed division algebras. Our approach is introduced through rigorously addressing: 1) the existence of quaternion reproducing kernel Hilbert spaces, and 2) quaternion calculus, in which derivatives of conjugate quantities must be treated carefully. These issues are closely related to the non-commutative multiplication of the quaternion ring, the definition of left vector spaces and the properties of quaternion left Hilbert spaces.

Recall that for the real and complex valued kernels:

- The Riesz representation theorem [13] implies that for an evaluation functional $L_\mathbf{x}$, $\mathbf{x} \in \mathbf{X}$, defined over an RKHS $\mathcal{H}$, there exists a unique element $K_\mathbf{x}$ such that $L_\mathbf{x}(f) = \langle f, K_\mathbf{x} \rangle$, where $K(\mathbf{x}, \mathbf{y}) = K_\mathbf{x}(\mathbf{y})$ is referred to as the reproducing kernel of $\mathcal{H}$.
- Conversely, the Moore-Aronszajn theorem [14] states that for any given positive definite kernel K, there is a unique RKHS $\mathcal{H}$ having K as reproducing kernel.

These results provide the existence of a unique RKHS for any designed positive definite kernel, simplifying the requirements for a suitable RKHS into the design of a positive definite kernel.

3.1. Quaternion reproducing kernel Hilbert spaces

We now introduce a theoretical basis for some key concepts in quaternion kernel learning. As the quaternion division ring $\mathbb{H}$ is non-commutative, in this work we only consider the left vector spaces [22] to define both the quaternion Hilbert space and quaternion RKHS.

Definition 1 (Quaternion left Hilbert space). *A complete left vector space $\mathcal{H}$ is called a quaternion left Hilbert space if there exists a quaternion-valued function $\langle \cdot, \cdot \rangle : \mathcal{H} \times \mathcal{H} \longrightarrow \mathbb{H}$ with the following properties:*

1. $\langle \mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{y}, \mathbf{x} \rangle^*$
2. $\langle p\mathbf{x} + q\mathbf{y}, \mathbf{z} \rangle = p \langle \mathbf{x}, \mathbf{z} \rangle + q \langle \mathbf{y}, \mathbf{z} \rangle$
3. $\langle \mathbf{x}, p\mathbf{y} + q\mathbf{z} \rangle = \langle \mathbf{x}, \mathbf{y} \rangle p^* + \langle \mathbf{x}, \mathbf{z} \rangle q^*$
4. $\langle \mathbf{x}, \mathbf{x} \rangle \geq 0$ *and* $\langle \mathbf{x}, \mathbf{x} \rangle = 0 \iff \mathbf{x} = 0$.

We refer to the function $\langle \cdot, \cdot \rangle$ as the inner product, for which the induced norm is $\|\mathbf{x}\| = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle}$, while the conjugate operator is defined as $(\cdot)^*$.

Definition 2 (Quaternion reproducing kernel Hilbert space). *A quaternion left Hilbert space $\mathcal{H} = \{\psi : \mathbf{X} \rightarrow \mathbb{H}\}$ is a quaternion reproducing kernel Hilbert space (QRKHS) if the linear functional $L_\mathbf{x}(\psi) = \psi(\mathbf{x})$ is bounded $\forall \psi \in \mathcal{H}, \mathbf{x} \in \mathbf{X}$.*

We next define positive definiteness for quaternion valued kernels, a necessary requirement to establish the relationship between quaternion kernels and QRKHSs.

Definition 3 (Positive definite kernel). *A Hermitian kernel K, i.e. $K(\mathbf{x},\mathbf{y}) = K^*(\mathbf{y},\mathbf{x})$, is positive definite on $\mathbf{X}$ iff for any square-integrable function $\theta : \mathbf{X} \rightarrow \mathbb{H}$, $\theta \neq 0$, we have*

$$\int_{\mathbf{X}} \int_{\mathbf{X}} \theta^*(\mathbf{x}) K(\mathbf{x},\mathbf{y}) \theta(\mathbf{y}) d\mathbf{x} d\mathbf{y} > 0.$$

The QRKHS (Definition 2) based on a positive definite kernel inherits the inner product, completeness, and boundedness properties of real and complex valued QRKHSs.

Remark 1. *By continuity, the Riesz representation and Moore-Aronszajn theorems also hold for the above definitions of QRKHS and positive definiteness. Therefore, similarly to the real and complex cases, the theoretical requirement for the construction of a QRKHS can be simplified into that of the design of a positive definite quaternion-valued kernel.*

3.2. Quaternion least squares estimation

We propose a least squares estimator for the unknown scalar mapping $y = f(\mathbf{x})$ in (1) to be in the form $\hat{y} = \langle \phi(\mathbf{x}), \omega \rangle$. Projecting the weights ω onto the QRKHS spanned by the training samples $\{(\mathbf{x}_j, y_j)\}_{j=1,\dots,N}$ and a positive definite quaternion kernel K, we can use the quaternion inner product properties in Definition 1 to write

$$\hat{y} = \sum_{j=1}^{N} K(\mathbf{x}, \mathbf{x}_j) \alpha_j. \tag{6}$$

Denoting the training samples error vector by $\mathbf{e} = \mathbf{y} - \mathbf{K}\mathbf{a}$, the real valued cost function (4) can be expressed as

$$J = \mathbf{e}^H \mathbf{e} + \rho \mathbf{a}^H \mathbf{a}. \tag{7}$$

Furthermore, using the $\mathbb{HR}$ calculus in [17] to set the quaternion derivative $\frac{\partial J}{\partial \alpha_j} = \nabla_{\alpha_j} J = 0, \forall \alpha_j$, we obtain the following relationship for the quaternion least squares solution

$$\left(\nabla_{\alpha_j} \mathbf{e}^H\right) \mathbf{e} + \mathbf{e}^H \left(\nabla_{\alpha_j} \mathbf{e}\right) + \rho \left(\nabla_{\alpha_j} \mathbf{a}^H\right) \mathbf{a} + \rho \mathbf{a}^H \left(\nabla_{\alpha_j} \mathbf{a}\right) = 0.$$

The Wiener solution is obtained through evaluating these gradients according to [17]. In this way, $\nabla_{\alpha_j} \alpha_j = 1$, $\nabla_{\alpha_j} \alpha_j^* = -1/2$, and $\nabla_{\alpha_j} \alpha_i = \nabla_{\alpha_j} \alpha_i^* = 0, i \neq j$ to arrive at

$$\frac{1}{2} \mathbf{K}^H \mathbf{e} - (\mathbf{K}^H \mathbf{e})^* = \frac{1}{2} \rho \mathbf{a} - (\rho \mathbf{a})^*, \tag{8}$$

finally giving the coefficient vector for the Wiener solution in the form

$$\mathbf{a} = (\mathbf{K}^H \mathbf{K} + \rho \mathbf{I})^{-1} \mathbf{K}^H \mathbf{y}. \tag{9}$$

Remark 2. *The optimal least squares solution of the kernel regression problem in (9) has the same form as its real-valued counterpart in (5). These two forms are equivalent for real-valued kernels of quaternion variable.*

3.3. Kernel choice

The linear kernel is a standard in practical kernel estimation due to its robustness and ease of implementation. For quaternion valued signals, the quaternion linear kernel K_Q and its real-valued counterpart K_R are respectively given by

$$K_Q(\mathbf{x},\mathbf{y}) = 1 + \langle \mathbf{x},\mathbf{y} \rangle = 1 + \mathbf{x}^H \mathbf{y} \tag{10}$$

$$K_R(\mathbf{x},\mathbf{y}) = 1 + \langle \mathbf{x},\mathbf{y} \rangle_{\Re} = 1 + \Re\{\mathbf{x}^H \mathbf{y}\}, \tag{11}$$

where $\langle \mathbf{x},\mathbf{y} \rangle_{\Re}$ is the inner product of the real-valued isomorphisms of $\mathbf{x}$ and $\mathbf{y}$, and $\Re\{q\}$ denotes the real part of the quaternion q. To show that the quaternion linear kernel is positive semidefinite, combine (10) and Definition 3, and use Fubini's Theorem to give

$$\int_{\mathbf{X}^2} \theta^*(\mathbf{x})(1 + \mathbf{x}^H \mathbf{y}) \theta(\mathbf{y}) d\mathbf{x} d\mathbf{y} = \left\| \int_{\mathbf{X}} \theta(\mathbf{x}) d\mathbf{x} \right\|^2 + \left\| \int_{\mathbf{X}} \mathbf{x} \theta(\mathbf{x}) d\mathbf{x} \right\|^2.$$

Remark 3. *The quaternion linear kernel admits the modelling of statistical inter-dependence in its imaginary parts. Therefore, the proposed kernel in (10) has the ability to learn the relationship between the quadrivariate input variables, while preserving the mathematical simplicity of univariate kernel regression algorithms.*

4. SIMULATION RESULTS

The quaternion-valued linear kernel proposed in (10) was validated against its real-valued counterpart in (11) which is equivalent to the standard KLS (see Remark 2). The simulations were conducted in the least squares kernel regression setting for the prediction of autoregressive processes and multivariate real-world 3D inertial body sensor data.

We first considered the AR(1) process $x_{t+1} = Ax_t + Be_t$, where $x_t \in \mathbb{H}$, and e_t is a quaternion random variable whose components are uncorrelated and uniformly distributed in $[0, 1]$. Correlated and uncorrelated realisations of the process x_t were obtained by respectively setting $\Im\{A\} = \Im\{B\} = 0$ and by letting A, B to be full quaternions. Note from Fig. 1 that the difference in MSEs of kernel algorithms for the uncorrelated case remained fairly constant for different support vectors, whereas for the correlated case this difference increased with the number of support vectors, hence highlighting the ability of the QKLS to model coupled processes.

We next applied the proposed QKLS to multivariate prediction of inertial body sensors. Four accelerometers (placed at wrists and ankles) recorded the three Euler angles (Fig. 2), giving a total of 12 signals $\{\theta_s\}_{s=1,..,12}$ taking values in the range $[-\pi, \pi]$. For a more accurate representation, and to avoid discontinuities close to the limits of the signals range, each θ_s was mapped onto the pair $(\sin\theta_s, \cos\theta_s)$; this gave a

24-dimensional real signal, or equivalently, a *6-dimensional quaternion signal*. In this way, the input (two delayed signal samples) and output pairs were respectively elements of $\mathbb{H}^{12}$ and $\mathbb{H}^{6}$, while the training and validation sets were different Tai Chi sequences.

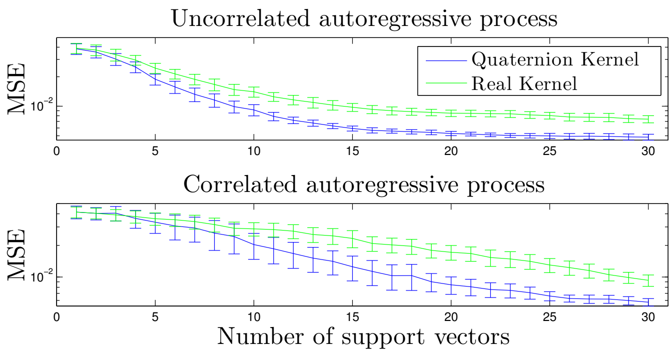

Fig. 1: MSE $\pm$ 0.5 standard deviations for kernel algorithms as a function of the number of support vectors in the estimation of both uncorrelated (top, $A = 0.6808$, $B = 0.1157$) and correlated (bottom, $A = 0.6808 + i0.07321 + j0.6222 - k0.2157$, $B = 0.1157 + i0.1208 + j0.8425 - k0.5121$) AR(1) processes.

Fig. 3 shows the averaged prediction MSE $\pm$ 0.5 standard deviations, over 10 independent realisations, as a function of the number of support vectors for different values of the regularisation parameter $\rho \in \{10^{-1}, 10^{-2}, 10^{-3}\}$. The support vectors and the validation set (50 samples) were randomly chosen, without repetition, for all realisations.

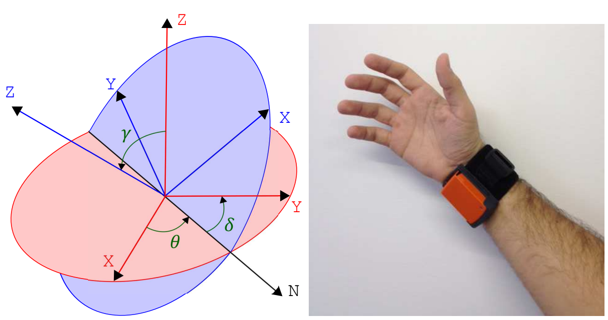

Fig. 2: Inertial body sensor setting. **[Left]** Fixed coordinate system (red), sensor coordinate system (blue) and Euler angles (green). **[Right]** A 3D inertial body sensor placed at the right wrist.

Observe that although the performances of both the quaternion and real kernel least squares were similar for training sets with fewer than 60 support vectors, the proposed QKLS with quaternion-valued linear kernel outperformed its real-valued counterpart as the number of support vectors (and therefore training samples) increased. Furthermore, although a lower regularization factor ρ provided improved estimation, the real-valued KLS algorithm suffered from overfitting and diverged due to the linear dependence of the feature samples. The quaternion kernel, on the contrary, proved more robust to the random choice of support vectors and exhibited virtually no overfitting and non-increasing MSE even for small regularisation factors. The better performance of QKLS for a larger number of support vectors can be explained by the inability of the real-valued kernel to model cross-coupling between data components (Remark 3). The inclusion of more support vectors allowed the quaternion kernel to span a space of higher dimensionality (compared to the real kernel) without leading to overfitting.

The superior performance of the quaternion kernel approach also suggests that the kernels for multivariate signal estimation should be of the same dimensionality as the data. This is in line with the background theory, as quaternions have advantage over real quadrivariate vectors in rotation and orientation modelling (needed to model body movement) and have thus become a standard in computer graphics.

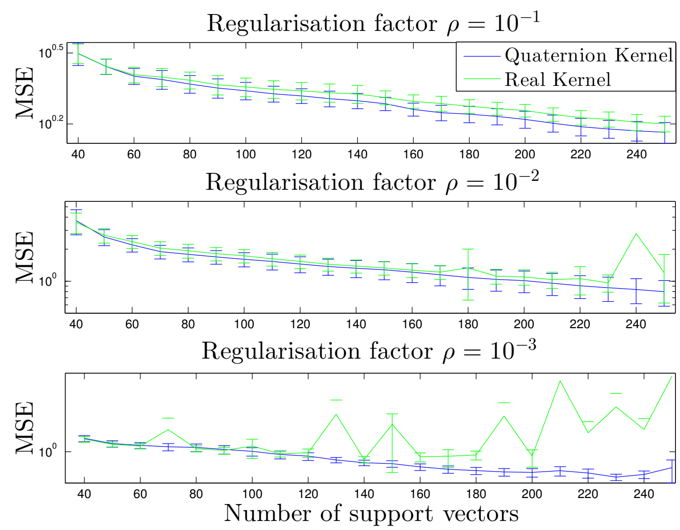

Fig. 3: MSE $\pm$ 0.5 standard deviations for kernel algorithms as a function of the number of support vectors for three different regularisation factors. Standard deviation bars which had negative values were omitted.

5. CONCLUSIONS

The quaternion kernel least squares (QKLS) algorithm has been introduced for the estimation of multidimensional real world processes. Necessary and sufficient conditions for the existence and uniqueness of quaternion RKHS, the backbone of quaternion kernel algorithms, have been illuminated by characterising a quaternion Hilbert space and its relationship with the standard real (and complex) RKHS theory. The least squares quaternion regression has been derived using the $\mathbb{HR}$ calculus, and it has been shown that, for real-valued kernel, its Wiener solution simplifies into the standard KLS solution. The enhanced performance of the QKLS algorithm, compared to the standard KLS, has been illustrated in terms of both the prediction MSE and robustness to overfitting, for prediction applications based on both synthetic autoregressive processes and real world multivariate body motion signals.

6. REFERENCES

[1] J. Kivinen, A. Smola, and R. Williamson, “Online learning with kernels,” *IEEE Trans. on Signal Processing*, vol. 52, no. 8, pp. 2165–2176, 2004.

[2] S.-Y. Kung, “Kernel approaches to unsupervised and supervised machine learning,” in *Advances in Multimedia Information Processing*. Springer Berlin / Heidelberg, 2009, vol. 5879, pp. 1–32.

[3] C. Richard, J. C. M. Bermudez, and P. Honeine, “Online prediction of time series data with kernels,” *IEEE Trans. on Signal Processing*, vol. 57, no. 3, pp. 1058–1067, 2009.

[4] W. Liu, J. C. Principe, and S. Haykin, *Kernel adaptive filtering: A comprehensive introduction*. Wiley, 2010.

[5] S. Van Vaerenbergh, M. Lazaro-Gredilla, and I. Santamaria, “Kernel recursive least-squares tracker for time-varying regression,” *IEEE Trans. on Neural Netwroks and Learning Systems*, vol. 23, no. 8, pp. 1313 –1326, 2012.

[6] A. Kuh and D. P. Mandic, “Applications of complex augmented kernels to wind profile prediction,” in *Proc. of IEEE ICASSP*, 2009, pp. 3581–3584.

[7] P. Bouboulis, S. Theodoridis, and M. Mavroforakis, “The augmented complex kernel LMS,” *IEEE Trans. on Signal Processing*, vol. 60, no. 9, pp. 4962–4967, 2012.

[8] F. Tobar, A. Kuh, and D. P. Mandic, “A novel augmented complex valued kernel LMS,” in *Proc. of the 7th IEEE Sensor Array and Multichannel Signal Processing Workshop*, 2012, pp. 481–484.

[9] B. Che Ujang, C. Took, and D. Mandic, “Quaternion-valued nonlinear adaptive filtering,” *IEEE Trans. on Neural Netwroks*, vol. 22, no. 8, pp. 1193 –1206, 2011.

[10] J. Via, D. Palomar, L. Vielva, and I. Santamaria, “Maximum likelihood ICA of quaternion Gaussian vectors,” in *Proc. of IEEE ICASSP*, 2011, pp. 4260 –4263.

[11] A. Shilton and D. Lai, “Quaternionic and complex-valued support vector regression for equalization and function approximation,” in *Proc. of IJCNN*, 2007, pp. 920–925.

[12] T. Ogunfunmi and T. Paul, “An alternative kernel adaptive filtering algorithm for quaternion-valued data,” in *Proc. of APSIPA*, 2012, pp. 1–5.

[13] A. Friedman, *Foundations of modern analysis*, 2nd ed. Dover, 1982.

[14] N. Aronszajn, “Theory of reproducing kernels,” *Trans. of the American Mathematical Society*, vol. 68, no. 3, pp. 337–404, 1950.

[15] A. Shilton, “Mercer’s theorem for quaternionic kernels,” The University of Melbournem Department of Electrical and Electronic Engineering, Tech. Rep., 2007.

[16] S. D. Leo and P. Rotelli, “Quaternion analyticity,” *Appl. Math. Lett*, vol. 16, no. 7, pp. 1077–1081, 2003.

[17] D. P. Mandic, C. Jahanchahi, and C. C. Took, “A quaternion gradient operator and its applications,” *IEEE Signal Processing Letters*, vol. 18, no. 1, pp. 47–50, 2011.

[18] V. Vapnik, *The nature of statistical learning theory*. Springer, 1995.

[19] H. Drucker, C. Burges, L. Kaufman, A. Smola, and V. Vapnik, “Support vector regression machines,” in *Proc. Advances in Neural Information Processing Systems 9*, 1996, pp. 155–161.

[20] C. Saunders, A. Gammerman, and V. Vovk, “Ridge regression learning algorithm in dual variables,” *Proc. of the 15th International Conference on Machine Learning*, vol. 75, no. 2, pp. 515–521, 1998.

[21] G. C. Cawley, N. L. C. Talbot, R. J. Foxall, S. R. Dorling, and D. P. Mandic, “Heteroscedastic kernel ridge regression,” *Neurocomputing*, vol. 57, pp. 105–124, 2004.

[22] N. Jacobson, *Basic algebra I*. Dover, 2009.