

Mercer's Theorem for Quaternionic Kernels

A. Shilton

Department of Electrical and Electronic Engineering
The University of Melbourne
mailto: apsh@ee.unimelb.edu.au

Abstract

An extension of Mercer's theorem to quaternionic valued kernel functions with applications in the field of machine learning is presented.

I. INTRODUCTION

In its best known form, Mercer's theorem may be stated as follows [1]:

Theorem 1: Let $K : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{C}$ be a continuous, non-negative definite, Hermitian kernel with eigenvalues $\{\lambda_\nu\}$ and eigenfunctions $\{\phi_\nu\}$. Then:

$$K(\mathbf{x}, \mathbf{y}) = \sum_{\nu=0}^{\infty} \frac{\phi_\nu(\mathbf{x}) \bar{\phi}_\nu(\mathbf{y})}{\lambda_\nu}$$

the series being uniformly and absolutely convergent in (\mathbf{x}, \mathbf{y}) .

A number of generalisations to Mercer's theorem may be found in the literature, in particular dealing with kernels $K : Y \times Y \rightarrow \mathbb{C}$ for various choices of Y . However there would appear to have been (to the best of the author's knowledge) no attempts made to extend Mercer's theorem to cover non-complex valued kernels. In the present paper we show how Mercer's theorem may be extended to cover one such family of kernels, namely the continuous quaternionic valued kernels $K : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{H}$. As quaternions provide a powerful tool for describing geometric problems [6] we anticipate that this extension will find applications in geometrical learning problems.

Throughout this paper the quaternionic division algebra [7], [3] will be denoted \mathbb{H} , the field of complex numbers \mathbb{C} and the completely ordered field of reals \mathbb{R} . The conjugate and norm of $x \in \mathbb{H}$ will be denoted \bar{x} and $|x|$ respectively, where $|x|^2 = \bar{x}x \in \mathbb{R} \setminus \mathbb{R}^-$. We define $\mathbb{R}^+ = (0, \infty)$ to be the set of positive reals, $\mathbb{R}^- = (-\infty, 0)$ the set of negative reals, \mathbb{N} the set of natural numbers including 0, $\mathbb{Z}_p = \{0, 1, \dots, p-1\}$ the set of integers modulo p (with the extensions $\mathbb{Z}_\infty = \mathbb{N}$, $\mathbb{Z}_0 = \emptyset$), $\mathbb{N}^\infty = \mathbb{N} \cup \{\infty\}$, and L^2 the set of square integrable quaternionic functions on \mathbb{R}^n :

$$L^2 = \left\{ \psi : \mathbb{R}^n \rightarrow \mathbb{H} \mid \int_{\mathbf{x} \in \mathbb{R}^n} |\psi(\mathbf{x})|^2 < \infty \right\}$$

II. MERCER'S THEOREM FOR QUATERNIONIC KERNELS

We begin by extending some standard notation to the quaternionic case. Given two functions $\phi, \psi : \mathbb{R}^n \rightarrow \mathbb{H}$ we define a binary operation $\langle \phi, \psi \rangle$:

$$\langle \phi, \psi \rangle = \int_{\mathbf{x} \in \mathbb{R}^n} \bar{\phi}(\mathbf{x}) \psi(\mathbf{x}) d\mathbf{x}$$

and by analogy with the usual definitions we say that a set of functions $\{\phi_\nu : \mathbb{R}^n \rightarrow \mathbb{H} \mid \nu \in \mathbb{Z}_m\}$ is *orthonormal* if $\langle \phi_\mu, \phi_\nu \rangle = \delta_{\mu\nu}$ for all $\mu, \nu \in \mathbb{Z}_m$, where $\delta_{\mu\nu}$ is the usual kronecker-delta symbol.

For all $\phi, \psi : \mathbb{R}^n \rightarrow \mathbb{H}$, $\lambda \in \mathbb{H}$:

$$\begin{aligned} \overline{\langle \phi, \psi \rangle} &= \langle \psi, \phi \rangle \\ \langle \psi, \phi \rangle \lambda &= \langle \psi, \phi \lambda \rangle \\ \bar{\lambda} \langle \psi, \phi \rangle &= \langle \psi \lambda, \phi \rangle \\ \langle \phi, \phi \rangle &\in \mathbb{R}^+ \cup \{0, \infty\} \end{aligned}$$

It should be noted that $\langle \phi, \psi \rangle$ this is not an inner product in the usual sense as \mathbb{H} is not a field. It is simply a notational convenience used for consistency with the standard form of Mercer's theorem.

We define a quaternionic kernel K to be any function $K : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{H}$. The *conjugate* K^\dagger of a kernel K is defined by $K^\dagger(\mathbf{x}, \mathbf{y}) = \bar{K}(\mathbf{y}, \mathbf{x})$. A quaternionic kernel may *operate* on a function $\phi : \mathbb{R}^n \rightarrow \mathbb{H}$ to obtain another function $\psi = K\phi : \mathbb{R}^n \rightarrow \mathbb{H}$ defined by:

$$\psi(\mathbf{x}) = \int_{\mathbf{y} \in \mathbb{R}^n} K(\mathbf{x}, \mathbf{y}) \phi(\mathbf{y}) d\mathbf{y}$$

where it may be seen that:

$$\langle K\psi, \phi \rangle = \langle \psi, K^\dagger \psi \rangle$$

Given two quaternionic kernels K and L we define the sum and product quaternionic kernels using:

$$\begin{aligned} (K + L)(\mathbf{x}, \mathbf{y}) &= K(\mathbf{x}, \mathbf{y}) + L(\mathbf{x}, \mathbf{y}) \\ (KL)(\mathbf{x}, \mathbf{y}) &= \int_{\mathbf{z} \in \mathbb{R}^n} K(\mathbf{x}, \mathbf{z}) L(\mathbf{z}, \mathbf{y}) d\mathbf{z} \end{aligned}$$

noting that $(K + L)\phi = K\phi + L\phi$ and $(KL)\phi = K(L\phi)$.

A *Hermitian* quaternionic kernel is a quaternionic kernel K for which $K(\mathbf{y}, \mathbf{x}) = \bar{K}(\mathbf{x}, \mathbf{y})$ for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, and hence $K^\dagger = K$. A *non-negative definite* quaternionic kernel is a quaternionic kernel K for which $\langle \phi, K\phi \rangle \in \mathbb{R} \setminus \mathbb{R}^-$ for all $\phi \in L^2$. Using the above results, it may be seen that any non-negative definite kernel will be Hermitian (but not vice-versa in general). For the remainder of the paper we will consider only Hermitian kernels.

Following the usual definitions, a homogeneous Fredholm integral equation of the second kind is an equation:

$$\phi = (K\phi)\lambda$$

wherein K is a known Hermitian quaternionic kernel function, and $\phi : \mathbb{R}^n \rightarrow \mathbb{H}$, $\lambda \in \mathbb{H}$ are unknown. For any solution (ϕ, λ) to the homogeneous Fredholm integral equation of the second kind we call ϕ an *eigenfunction* of K and λ an *eigenvalue* of K . A basis set of such solutions, with orthonormalised eigenfunctions, will be written $\{(\phi_\nu, \lambda_\nu) \mid \nu \in \mathbb{Z}_p, p \in \mathbb{N}^\infty\}$. Orthonormality may be achieved using the standard Gram-Schmidt procedure.

We now present some preliminary results:

Theorem 2: All eigenvalues λ_ν of a Hermitian quaternionic kernel K will be real, and furthermore all eigenvalues λ_ν of a non-negative definite Hermitian quaternionic kernel K will be positive.

Proof: To begin, suppose that K is Hermitian, but not necessarily non-negative definite. Let $\phi : \mathbb{R}^n \rightarrow \mathbb{H}$ be a normalised function (ie. $\langle \phi, \phi \rangle = 1$). Now, as K is Hermitian, $\langle \phi, K\phi \rangle = \langle K\phi, \phi \rangle = \overline{\langle \phi, K\phi \rangle}$ and hence $\langle \phi, K\phi \rangle \in \mathbb{R}$. Given this, $\langle \phi, K\phi \rangle \lambda = \langle \phi, (K\phi)\lambda \rangle = \langle \phi, \phi \rangle = 1$, and as $\langle \phi, K\phi \rangle \in \mathbb{R}$, $\langle \phi, K\phi \rangle \bar{\lambda} = \bar{\lambda} \langle K\phi, \phi \rangle = \langle (K\phi)\lambda, \phi \rangle = \langle \phi, \phi \rangle = 1$, from which we see that $\lambda = \bar{\lambda}$, and hence $\lambda \in \mathbb{R}$.

If K is non-negative definite and Hermitian, suppose that $\lambda_\mu < 0$ for some μ . Then, as the eigenfunctions are orthonormal, $\langle \phi_\mu, K\phi_\mu \rangle = \langle \phi_\mu, K\phi_\mu \rangle \lambda_\mu \lambda_\mu^{-1} = \langle \phi_\mu, (K\phi_\mu)\lambda_\mu \rangle \lambda_\mu^{-1} = \langle \phi_\mu, \phi_\mu \rangle \lambda_\mu^{-1} = \lambda_\mu^{-1} \in \mathbb{R}^-$. But this contradicts the assertion that K is non-negative definite, and so we may conclude that all eigenvalues must be positive reals in this case. ■

Theorem 3: Let $\{(\phi_\nu, \lambda_\nu) \mid \nu \in \mathbb{Z}_p, p \in \mathbb{N}^\infty\}$ be a basis set of eigenvalues and eigenfunctions associated with the non-negative definite Hermitian quaternionic kernel K , and let $m \in \mathbb{Z}_{p+1}$. Then:

$$K_m(\mathbf{x}, \mathbf{y}) = K(\mathbf{x}, \mathbf{y}) - \sum_{\nu \in \mathbb{Z}_m} \frac{\phi_\nu(\mathbf{x}) \bar{\phi}_\nu(\mathbf{y})}{\lambda_\nu}$$

is also non-negative definite.

Proof: Note that K_m is Hermitian. Now:

$$\psi = \sum_{\mu \in \mathbb{Z}_m} \phi_\mu \langle \phi_\mu, \psi \rangle + \psi_R$$

where as $\psi \in L^2$ it may be seen that $\langle \phi_\mu, \psi_R \rangle = 0$ for all $\mu \in \mathbb{Z}_m$. Hence:

$$\begin{aligned} \langle \psi, K\psi \rangle &= \left\langle \sum_{\mu \in \mathbb{Z}_m} \phi_\mu \langle \phi_\mu, \psi \rangle, K \sum_{\nu \in \mathbb{Z}_m} \phi_\nu \langle \phi_\nu, \psi \rangle \right\rangle + \langle \psi_R, K\psi_R \rangle \\ &+ \left\langle \sum_{\nu \in \mathbb{Z}_m} \phi_\nu \langle \phi_\nu, \psi \rangle, K\psi_R \right\rangle + \left\langle \psi_R, K \sum_{\nu \in \mathbb{Z}_m} \phi_\nu \langle \phi_\nu, \psi \rangle \right\rangle \\ &= \sum_{\mu, \nu \in \mathbb{Z}_m} \frac{\overline{\langle \phi_\mu, \psi \rangle} \langle \phi_\mu, \phi_\nu \rangle \langle \phi_\nu, \psi \rangle}{\lambda_\nu} + \langle \psi_R, K\psi_R \rangle \\ &+ 2\text{Re} \left(\left\langle \psi_R, K \sum_{\nu \in \mathbb{Z}_m} \phi_\nu \langle \phi_\nu, \psi \rangle \right\rangle \right) \\ &= \sum_{\nu \in \mathbb{Z}_m} \frac{|\langle \phi_\nu, \psi \rangle|^2}{\lambda_\nu} + \langle \psi_R, K\psi_R \rangle + 2 \sum_{\nu \in \mathbb{Z}_m} \text{Re} \left(\frac{\langle \psi_R, \phi_\nu \rangle \langle \phi_\nu, \psi \rangle}{\lambda_\nu} \right) \\ &= \sum_{\nu \in \mathbb{Z}_m} \frac{|\langle \phi_\nu, \psi \rangle|^2}{\lambda_\nu} + \langle \psi_R, K\psi_R \rangle \end{aligned}$$

Furthermore, defining $k(\mathbf{x}, \mathbf{y}) = \sum_{\nu \in \mathbb{Z}_m} \frac{\phi_\nu(\mathbf{x}) \bar{\phi}_\nu(\mathbf{y})}{\lambda_\nu}$:

$$\begin{aligned} \langle \psi, k\psi \rangle &= \sum_{\nu \in \mathbb{Z}_m} \left\langle \psi, \phi_\nu \frac{\langle \phi_\nu, \psi \rangle}{\lambda_\nu} \right\rangle \\ &= \sum_{\nu \in \mathbb{Z}_m} \frac{|\langle \phi_\nu, \psi \rangle|^2}{\lambda_\nu} \end{aligned}$$

and hence $\langle \psi, K_m \psi \rangle = \langle \psi_R, K \psi_R \rangle \geq 0$ (as K is non-negative definite), which proves the theorem. \blacksquare

Theorem 4: $K(\mathbf{x}, \mathbf{x}) \in \mathbb{R} \setminus \mathbb{R}^-$ for all $\mathbf{x} \in \mathbb{R}^n$ for all continuous, non-negative definite, Hermitian quaternionic kernels K .

Proof: For a continuous, non-negative definite Hermitian quaternionic kernel K we know that $\langle \phi, K\phi \rangle \geq 0$ for any $\phi \in L^2$. Consider the continuous positive real function:

$$\phi_\delta(\mathbf{x}) = \begin{cases} 0 & \text{if } |\mathbf{x} - \mathbf{x}_0| \geq \delta \\ \delta^2 - |\mathbf{x} - \mathbf{x}_0|^2 & \text{if } |\mathbf{x} - \mathbf{x}_0| < \delta \end{cases}$$

where $\delta \in \mathbb{R}^+$. It is clear that $\phi_\delta \in L^2$ and $\langle \phi_\delta, K\phi_\delta \rangle \geq 0$.

Suppose $K(\mathbf{x}_0, \mathbf{x}_0) \neq 0$ for some $\mathbf{x}_0 \in \mathbb{R}^n$. Then by continuity there exists $\delta \in \mathbb{R}^+$ such that $K(\mathbf{x}, \mathbf{y}) = q(\mathbf{x}, \mathbf{y}) K(\mathbf{x}_0, \mathbf{x}_0)$ where $q(\mathbf{x}, \mathbf{y}) > 0$ for all $\mathbf{x}, \mathbf{y} \in \{\mathbf{z} \in \mathbb{R}^n \mid |\mathbf{z} - \mathbf{x}_0| < \delta\}$, and hence $\text{sgn}(\langle \phi_\delta, K\phi_\delta \rangle) = \text{sgn}(K(\mathbf{x}_0, \mathbf{x}_0))$. It follows that for all continuous, non-negative definite Hermitian quaternionic kernels $K(\mathbf{x}, \mathbf{x}) \in \mathbb{R} \setminus \mathbb{R}^-$ for all $\mathbf{x} \in \mathbb{R}^n$. \blacksquare

This provides sufficient basis to begin considering the problem at hand, namely the extension of Mercer's theorem to the case of quaternionic kernels, viz.:

Theorem 5: Let $\{(\phi_\nu, \lambda_\nu) \mid \nu \in \mathbb{Z}_p, p \in \mathbb{N}^\infty\}$ be a basis set of eigenvalues and eigenfunctions of the continuous, non-negative definite, Hermitian quaternionic kernel K . Then:

$$K(\mathbf{x}, \mathbf{y}) = \sum_{\nu \in \mathbb{Z}_p} \frac{\phi_\nu(\mathbf{x}) \bar{\phi}_\nu(\mathbf{y})}{\lambda_\nu} \quad (1)$$

the series being uniformly and absolutely convergent in (\mathbf{x}, \mathbf{y}) .

Proof: Having established the precursor results and extended the notation appropriately, the proof of this theorem precisely follows that given in [1], page 197–199 (similar proofs may also be found in [4], [8], [2], [5]). It is summarised here for completeness.

Given that K is a continuous, non-negative definite, Hermitian quaternionic kernel, it follows that $K(\mathbf{x}, \mathbf{x}) \in \mathbb{R} \setminus \mathbb{R}^-$ for all $\mathbf{x} \in \mathbb{R}^n$. Theorem 3 implies that for all $m \in \mathbb{Z}_{p+1}$:

$$K_m(\mathbf{x}, \mathbf{y}) = K(\mathbf{x}, \mathbf{y}) - \sum_{\nu \in \mathbb{Z}_m} \frac{\phi_\nu(\mathbf{x}) \bar{\phi}_\nu(\mathbf{y})}{\lambda_\nu} \quad (2)$$

is also non-negative definite. Hence $K_m(\mathbf{x}, \mathbf{x}) \in \mathbb{R} \setminus \mathbb{R}^-$ for all $\mathbf{x} \in \mathbb{R}^n$, and:

$$\sum_{\nu \in \mathbb{Z}_p} \frac{|\phi_\nu(\mathbf{x})|^2}{\lambda_\nu} \leq K(\mathbf{x}, \mathbf{x}) \text{ for all } m \in \mathbb{Z}_{p+1}, \mathbf{x} \in \mathbb{R}^n$$

Taking the limit $m \rightarrow p$, we see that the series on the left-hand side of this expression must converge (absolutely) for all $\mathbf{x} \in \mathbb{R}^n$ and furthermore:

$$0 \leq \sum_{\nu \in \mathbb{Z}_p} \frac{|\phi_\nu(\mathbf{x})|^2}{\lambda_\nu} \leq K(\mathbf{x}, \mathbf{x}) \leq M = \max_{\mathbf{x} \in \mathbb{R}^n} K(\mathbf{x}, \mathbf{x}) \text{ for all } \mathbf{x} \in \mathbb{R}^n$$

Using this result and the Cauchy-Schwartz inequality we see that for all $i, j \in \mathbb{Z}_{p+1}$ where $i \leq j$:

$$\begin{aligned} \left(\sum_{\nu=i}^j \left| \frac{\phi_\nu(\mathbf{x}) \bar{\phi}_\nu(\mathbf{y})}{\lambda_\nu} \right| \right)^2 &\leq \left(\sum_{\nu=i}^j \frac{|\phi_\nu(\mathbf{x})|^2}{\lambda_\nu} \right) \left(\sum_{\nu=i}^j \frac{|\bar{\phi}_\nu(\mathbf{y})|^2}{\lambda_\nu} \right) \\ &\leq \left(\sum_{\nu=i}^j \frac{|\phi_\nu(\mathbf{x})|^2}{\lambda_\nu} \right) M \\ &\leq \epsilon M \text{ for all } i, j \geq N(\mathbf{x}, \epsilon) \end{aligned}$$

from which we see that the series (1) converges uniformly in \mathbf{y} for each \mathbf{x} , and by analogy converges uniformly in \mathbf{x} for each \mathbf{y} , and moreover the limit function in either case is the kernel K .¹ Hence the series of positive continuous functions in

¹Use is made here of a Lemma from [1], page 196, the proof of which trivially carries over to the present case.

(2) must have as a limit $K(\mathbf{x}, \mathbf{x})$, and therefore by Dini's theorem the convergence of the limit function must be uniform. It follows that the \mathbf{x} dependence of $N(\mathbf{x}, \epsilon)$ is extrinsic, and so the uniform and absolute convergence of the series (1) follows. ■

REFERENCES

- [1] James Alan Cochran, *The Analysis of Linear Integral Equations* (McGraw-Hill Book Company, 1972).
- [2] R. Courant and D. Hilbert, *Methods of Mathematical Physics* (John Wiley and sons, New York, 1937).
- [3] W. R. Hamilton, "On quaternions: or a new system of imaginaries in algebra", *Phil. Mag. ser. 3* **25** (1845) 210.
- [4] Harry Hochstadt, *Integral Equations* (Wiley-Interscience, 1973).
- [5] Konrad Jörgens, *Linear Integral Operators* (Pitman, London, 1970).
- [6] Jack B. Kuipers, *Quaternions and Rotation Sequences: A Primer with Applications to Orbits, Aerospace and Virtual Reality* (Princeton, 1999).
- [7] A. Sudbery, "Quaternionic analysis", *Cambridge Philosophical Society* **85** (1979) 199–225.
- [8] F. G. Tricomi, *Integral Equations* (Interscience, 1970).