Digital Signal Processing 21 (2011) 660–666

Contents lists available at ScienceDirect

Digital Signal Processing

www.elsevier.com/locate/dsp

On azimuthally symmetric 2-sphere convolution ☆

Rodney A. Kennedy [a,*], Tharaka A. Lamahewa [a,b], Liying Wei [a]

[a] Research School of Engineering, CECS, The Australian National University, Canberra, ACT, Australia
[b] Canberra Research Laboratories, National ICT Australia, Australia

ARTICLE INFO

Article history:
Available online 25 May 2011

Keywords:
Unit sphere
2-Sphere convolution
Left Convolution
2-Sphere filtering
SO(3) convolution

ABSTRACT

We consider the problem of azimuthally symmetric convolution of signals defined on the 2-Sphere. Applications of such convolution include but are not limited to: geodesy, astronomical data (such as the famous Wilkinson Microwave Anisotropy Probe data), and 3D beamforming/sensing. We review various definitions of convolution from the literature and show a nontrivial equivalence between different definitions. Some convolution formulations based on SO(3) are shown not to be well formed for applications and we demonstrate a simpler framework to understand, use and generalize azimuthally symmetric convolution.

1. Introduction

The processing of signals whose domain is the unit sphere, $\mathbb{S}^2 \triangleq \{\boldsymbol{x} \in \mathbb{R}^3: \|\boldsymbol{x}\| = 1\}$ is an increasingly active area of research with applications in geodesy, cosmology, and 3D beamforming/sensing [1–7]. This processing exhibits important differences from the processing of signals on euclidean domains such as: time based signals whose domain is the real line $\mathbb{R}$ or signals, and 2D or 3D images, whose domain is multidimensional but still euclidean. A most basic operation on signals is linear filtering, or convolution, and yet for signals on the unit sphere this is not consistently well defined with competing definitions and the prevalence of definitions that prove too restrictive.

The contributions of this paper are:

- We determine the relationship between various definitions of convolution for signals defined on $\mathbb{S}^2$ (corresponding to an azimuthally symmetric or isotropic convolution property). Different formulations in the literature are shown to be essentially identical.
- We develop a common framework to study these definitions. This clarifies that these definitions fall short of representing the most general form of convolution for signals defined on $\mathbb{S}^2$.

1.1. Notation and mathematical preliminaries

We consider signals in the complex Hilbert Space $L^2(\mathbb{S}^2)$, the space of square integrable or finite energy signals defined on the unit sphere $\mathbb{S}^2$ with inner product

$$\langle f, g\rangle \triangleq \int_{\mathbb{S}^2} f(\hat{\boldsymbol{x}})\overline{g(\hat{\boldsymbol{x}})}\, ds(\hat{\boldsymbol{x}}) = \int_0^{\pi}\int_0^{2\pi} f(\theta,\phi)\overline{g(\theta,\phi)}\sin\theta\, d\phi\, d\theta, \tag{1}$$

where, $f, g \in L^2(\mathbb{S}^2)$,

$$\hat{\boldsymbol{x}} \equiv \hat{\boldsymbol{x}}(\theta,\phi) \triangleq (\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta)' \in \mathbb{R}^3,$$

$$ds(\hat{\boldsymbol{x}}) \triangleq \sin\theta\, d\phi\, d\theta$$

with ϕ the longitude, θ the co-latitude, and $\overline{(\cdot)}$ is complex conjugation. The inner product induces a norm, $\|f\| = \langle f, f\rangle^{1/2}$. Here, the $\widehat{(\cdot)}$ denotes a unit vector, that is, $\|\hat{\boldsymbol{x}}\| = 1$. The north pole, denoted $\hat{\boldsymbol{\eta}}$, is at co-latitude $\theta = 0$ and satisfies $\|\hat{\boldsymbol{\eta}}\| = 1$.

The Hilbert Space $L^2(\mathbb{S}^2)$ is *separable* [8]. The archetype complete orthonormal sequence for this space is the set of spherical harmonics

$$Y_n^m(\theta,\phi) \triangleq \sqrt{\frac{2n+1}{4\pi}\frac{(n-|m|)!}{(n+|m|)!}}\, P_n^{|m|}(\cos\theta)e^{im\phi},$$

$$n = 0, 1, 2, \ldots, \quad m = -n, \ldots, n \tag{2}$$

where $P_n^m(\cos\theta)$ are the Associated Legendre Functions [9]. When $m = 0$, $P_n^m(\zeta) = P_n(\zeta)$ the regular Legendre polynomials in indeterminate $\zeta = \cos\theta$. The spherical harmonics have two indices,

☆ Work supported by the Australian Research Council Discovery Grant DP1094350.

* Corresponding author.
E-mail addresses: rodney.kennedy@anu.edu.au (R.A. Kennedy), tharaka.lamahewa@anu.edu.au (T.A. Lamahewa), liying.wei@cecs.anu.edu.au (L. Wei).

1051-2004/$ – see front matter
doi:10.1016/j.dsp.2011.05.002

m and n, but as a set can be represented using a single denumerable index and thus regarded as an orthonormal sequence [8].

1.1.1. Spherical harmonic representation

Any finite energy signal defined on the unit sphere, $f \in L^2(\mathbb{S}^2)$ being a function of longitude ϕ and co-latitude θ, can be represented

$$f(\hat{\boldsymbol{x}}) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} f_n^m Y_n^m(\hat{\boldsymbol{x}}) \tag{3}$$

where the Fourier coefficients are

$$\begin{aligned} f_n^m \triangleq \langle f, Y_n^m \rangle &= \int_{\mathbb{S}^2} f(\hat{\boldsymbol{x}}) \overline{Y_n^m(\hat{\boldsymbol{x}})}\, ds(\hat{\boldsymbol{x}}) \\ &= \int_0^{\pi} \int_0^{2\pi} f(\theta, \phi) \overline{Y_n^m(\theta, \phi)} \sin\theta \, d\phi \, d\theta. \end{aligned} \tag{4}$$

Note that the equality in (3) is understood in the sense of convergence in the mean, that is,

$$\lim_{N\to\infty} \left\| f(\hat{\boldsymbol{x}}) - \sum_{n=0}^{N} \sum_{m=-n}^{n} \langle f, Y_n^m \rangle Y_n^m(\hat{\boldsymbol{x}}) \right\|^2 = 0, \quad \forall f \in L^2\big(\mathbb{S}^2\big) \tag{5}$$

which expresses the completeness of the spherical harmonics [9, Theorem 2.7].

Finally, we note the Spherical Harmonic Addition Theorem [9] which we require later

$$\sum_{m=-n}^{n} Y_n^m(\hat{\boldsymbol{x}}) \overline{Y_n^m(\hat{\boldsymbol{y}})} = \frac{(2n+1)}{4\pi} P_n(\hat{\boldsymbol{x}} \cdot \hat{\boldsymbol{y}}) \tag{6}$$

where $\hat{\boldsymbol{x}} \cdot \hat{\boldsymbol{y}}$ is the dot/inner product between the unit vectors $\hat{\boldsymbol{x}}$ and $\hat{\boldsymbol{y}}$ in $\mathbb{R}^3$, and $P_n(\zeta)$ is the Legendre polynomial in indeterminate ζ of degree n.

1.1.2. Rotationally symmetric special case

As it is important in the sequel, we also define the linear subspace of $L^2(\mathbb{S}^2)$, denoted $\mathcal{H}^0(\mathbb{S}^2)$, retaining the inner product (1), by

$$\mathcal{H}^0\big(\mathbb{S}^2\big) \triangleq \big\{ f \in L^2\big(\mathbb{S}^2\big): f(\theta, \phi) = f(\theta) \big\} \tag{7}$$

which is the subspace of finite energy signals with azimuthal/rotational symmetry about the north pole axis $\hat{\boldsymbol{\eta}}$ (direction $\theta = 0$). The $m = 0$ spherical harmonics, $\{Y_n^0(\hat{\boldsymbol{x}})\}_{n=0}^{\infty}$, are complete in $\mathcal{H}^0(\mathbb{S}^2)$ (explained below). So any $g \in \mathcal{H}^0(\mathbb{S}^2)$ can be written

$$g(\hat{\boldsymbol{x}}) = \sum_{n=0}^{\infty} \langle g, Y_n^0 \rangle Y_n^0(\hat{\boldsymbol{x}}), \quad \forall g \in \mathcal{H}^0\big(\mathbb{S}^2\big) \tag{8}$$

where the Fourier coefficients are

$$\begin{aligned} \langle g, Y_n^0 \rangle &= \int_{\mathbb{S}^2} g(\hat{\boldsymbol{x}}) \overline{Y_n^0(\hat{\boldsymbol{x}})}\, ds(\hat{\boldsymbol{x}}) \\ &= \int_0^{\pi} \int_0^{2\pi} g(\theta) \overline{Y_n^0(\theta)} \sin\theta \, d\phi \, d\theta \\ &= \sqrt{\pi(2n+1)} \int_0^{\pi} g(\theta) P_n(\cos\theta) \sin\theta \, d\theta \end{aligned}$$

using (2). It is the completeness of the Legendre polynomials, $P_n(\zeta)$, on $|\zeta| \leqslant 1$, that implies completeness of the $m = 0$ spherical harmonics (also called the zonal harmonics).

Finally, a given function $f \in L^2(\mathbb{S}^2)$ has an orthogonal projection onto $\mathcal{H}^0(\mathbb{S}^2)$. This projection is represented by the (linear) projection operator $\mathcal{P}^0 f$ defined by

$$\big(\mathcal{P}^0 f\big)(\hat{\boldsymbol{x}}) \triangleq \sum_{n=0}^{\infty} \langle f, Y_n^0 \rangle Y_n^0(\hat{\boldsymbol{x}}), \quad \forall f \in L^2\big(\mathbb{S}^2\big). \tag{9}$$

The projection $\mathcal{P}^0 f$ can be interpreted as the average of f over rotations about the north pole axis.

2. Convolution and problem formulation

An approach to formulating convolution for signals on the sphere is to emulate the convolution operator familiar in euclidean spaces. Convolution on $\mathbb{R}^2$ is given by

$$(g \otimes f)(\boldsymbol{x}) \triangleq \int_{\mathbb{R}^2} g(\boldsymbol{x} - \boldsymbol{y}) f(\boldsymbol{y})\, d\boldsymbol{y}, \quad \boldsymbol{x} \in \mathbb{R}^2, \tag{10}$$

and can be used as a reference to develop different notions of convolution on the 2-sphere. We review two key definitions of 2-sphere convolution which have appeared in the literature.

2.1. SO(3) left convolution

Let $f, h \in L^2(\mathbb{S}^2)$. Using the binary operand symbol $\star$, "convolution on the 2-sphere" was defined as follows [10]

$$(f \star h)(\hat{\boldsymbol{x}}) \triangleq \frac{1}{2\pi} \int_{\boldsymbol{v} \in \mathrm{SO}(3)} h(\boldsymbol{v}\hat{\boldsymbol{\eta}}) f\big(\boldsymbol{v}^{-1}\hat{\boldsymbol{x}}\big)\, d\boldsymbol{v}, \quad \hat{\boldsymbol{x}} \in \mathbb{S}^2 \tag{11}$$

where $\hat{\boldsymbol{\eta}} \in \mathbb{S}^2$ is the north pole (co-latitude $\theta = 0$), and $\boldsymbol{v}$ is a 3D proper rotation element in SO(3) [11]. These rotations are conventionally parametrized by three Euler angles, α, β and γ, and therefore (11) is implicitly a triple integral. As it is important, in the later comparison, we elaborate on these notations.

Define

$$\boldsymbol{r}_\beta^{(y)} \triangleq \begin{pmatrix} \cos\beta & 0 & \sin\beta \\ 0 & 1 & 0 \\ -\sin\beta & 0 & \cos\beta \end{pmatrix}$$

and

$$\boldsymbol{r}_\gamma^{(z)} \triangleq \begin{pmatrix} \cos\gamma & -\sin\gamma & 0 \\ \sin\gamma & \cos\gamma & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{12}$$

Denote $\boldsymbol{v} = \boldsymbol{r}_\alpha^{(z)} \boldsymbol{r}_\beta^{(y)} \boldsymbol{r}_\gamma^{(z)}$ as the "zyz" rotation in SO(3), being the composition of rotations about the z-axis through $\gamma \in [0, 2\pi)$, y-axis through $\beta \in [0, \pi]$ and z-axis (again) though $\alpha \in [0, 2\pi)$, in that order.

To effect a general rotation of $f \in L^2(\mathbb{S}^2)$, through the Euler angles implicit in $\boldsymbol{v}$, is an operator which we denote $\mathcal{R}_{\boldsymbol{v}}$. The effect of this operator can be equivalently realized through an inverse rotation of the coordinate system, that is,

$$(\mathcal{R}_{\boldsymbol{v}} f)(\hat{\boldsymbol{x}}) = f\big(\boldsymbol{v}^{-1}\hat{\boldsymbol{x}}\big). \tag{13}$$

Some examples that use the left convolution (11) on 2-sphere are: filter bank designs [12], spherical image processing [13], rotation estimation for images defined on the 2-sphere [14] and 3D surface filtering/smoothing [11,15,16].

In (11) we have scaled the definition relative to the original given in [10]. In the spherical harmonic domain [10, Theorem 1] (up to a scaling), convolution is given by

$$(f \star h)_n^m = \sqrt{\frac{4\pi}{(2n+1)}} f_n^m h_n^0, \tag{14}$$

where

$$(f \star h)_n^m = \langle f \star h, Y_n^m \rangle, \quad f_n^m = \langle f, Y_n^m \rangle \quad \text{and} \quad h_n^0 = \langle h, Y_n^0 \rangle.$$

Eq. (14) is the "frequency domain" version of (11) and exhibits a multiplication property. Due to the $m = 0$ action on the second argument h in (14), this convolution definition is not commutative, $f \star h \neq h \star f$. Further, there is the scaling factor which depends on n. Later we show a geometric interpretation of Left Convolution using the projection (9) which explains these distinguishing features.

2.2. Isotropic convolution

Using another binary operand symbol $\circledast$, another definition of convolution comes from [17,18] and takes the form

$$(z \circledast f)(\hat{\boldsymbol{x}}) = \int_{\mathbb{S}^2} z(\hat{\boldsymbol{x}} \cdot \hat{\boldsymbol{y}}) f(\hat{\boldsymbol{y}})\, ds(\hat{\boldsymbol{y}}), \quad \hat{\boldsymbol{x}} \in \mathbb{S}^2, \tag{15}$$

which bears greater similarity to (10). Noting that $-1 \leqslant \hat{\boldsymbol{x}} \cdot \hat{\boldsymbol{y}} \leqslant 1$ then this definition can be viewed as a linear operator on $L^2(\mathbb{S}^2)$, mapping $f(\cdot) \in L^2(\mathbb{S}^2)$ to $(z \circledast f)(\cdot) \in L^2(\mathbb{S}^2)$ parametrized with the univariate function $z(\cdot) \in L^2([-1, +1])$. Here, $z(\cdot)$ is a kernel and can be regarded as defining a filter response. This filter response only depends on the angle (arc length) and not the relative orientation between the two spatial variables $\hat{\boldsymbol{x}}$ and $\hat{\boldsymbol{y}}$ (since $\hat{\boldsymbol{x}} \cdot \hat{\boldsymbol{y}} = \cos\xi$, where ξ is the angle between $\hat{\boldsymbol{x}}$ and $\hat{\boldsymbol{y}}$), and, therefore, is *isotropic*. For example, if we had to blur a spherical image with a rotationally symmetric filter then (15) can be used.

There are clear limitations in (15) when we wish to use non-rotationally symmetric filters and it is evident, despite the similarity with (10), that it cannot be the most general form of convolution on the sphere $\mathbb{S}^2$. However, being an azimuthally symmetric 2-Sphere convolution, it is a widely useful special case on which we focus in the paper.

2.3. Problem formulation

This paper poses the following questions:

(1) What is the relationship between various definitions of convolution for signals defined on $\mathbb{S}^2$?
(2) Can these be understood in a common framework? Do these notions represent the most general form of convolution for signals defined on $\mathbb{S}^2$?

3. Convolution on $\mathbb{S}^2$

3.1. Operator formulation of convolution

Our approach is to define convolution guided by the analog of frequency domain multiplication for convolution of time domain signals. To do so we develop the necessary theory based on linear operators. Consistent with other works we use the spherical harmonics to build the "frequency domain" description. Our formulation is developed from three equivalent pictures for convolution: an integral equation *operator* picture, a *binary operation on two signals* picture, and an *infinite matrix representation*.

3.2. Matrix representation of bounded operators

We expect convolution to be a special type of bounded linear operator on $L^2(\mathbb{S}^2)$. A linear operator $\mathcal{B}$ is *bounded* if there exists a constant B such that

$$\|\mathcal{B}f\| \leqslant B\|f\|, \quad \forall f \in L^2(\mathbb{S}^2). \tag{16}$$

As well as this *operator* picture for convolution we also develop other equivalent pictures. We can regard convolution as a *binary operation on two signals*, and we have a picture based on an *infinite matrix representation*. This last form provides the simplest framework to develop candidate definitions of convolution.

Any bounded linear operator, $\mathcal{B}$, on a separable[1] space admits an infinite matrix representation with respect to a given complete orthonormal sequence [8]. We previously noted that $L^2(\mathbb{S}^2)$ is separable and the spherical harmonics are a complete orthonormal sequence. Hence any bounded linear operator, $\mathcal{B}$ on $L^2(\mathbb{S}^2)$, admits an infinite matrix representation [8]

$$\mathbf{B}_{nq}^{mp} \triangleq \langle \mathcal{B}Y_q^p, Y_n^m \rangle \tag{17}$$

for all $n = 0, 1, 2, \ldots,\ m = -n, \ldots, n$ and $q = 0, 1, 2, \ldots,\ p = -q, \ldots, q$. This characterizes completely how the bounded linear operator maps input to outputs, with respect to the spherical harmonics. In (17) we should think of the pair p, q as a single input/column index, and pair m, n as a single output/row index. That is, the coefficient $\mathbf{B}_{nq}^{mp}$ represents how much of Y_q^p as an input gets projected along the Y_n^m direction of the output under $\mathcal{B}$.

So a bounded linear operator, $\mathcal{B}$, mapping $f \in L^2(\mathbb{S}^2)$ to $d \in L^2(\mathbb{S}^2)$, viz.,

$$d = \mathcal{B}f, \tag{18}$$

can be expressed in the spherical harmonic domain in terms of (17) as

$$d_n^m = \sum_{q=0}^{\infty} \sum_{p=-q}^{q} \mathbf{B}_{nq}^{mp} f_q^p, \tag{19}$$

where

$$d_n^m = \langle d, Y_n^m \rangle = \langle \mathcal{B}f, Y_n^m \rangle \quad \text{and} \quad f_q^p = \langle f, Y_q^p \rangle.$$

3.3. Harmonic multiplication on $\mathbb{S}^2$ – binary operation picture

The frequency domain for signals on $\mathbb{R}^2$ clearly reveals a simple characterization of convolution via multiplication. So by analogy we expect the spherical harmonic coefficient sequence, f_n^m defined in (4), to be pivotal in our definition of convolution on $\mathbb{S}^2$. This leads to the concept of harmonic multiplication and will be used later to study and interpret azimuthally symmetric convolution.

Definition 1 *(Harmonic multiplication on $\mathbb{S}^2$).* As a binary operation, using symbol $\odot$, the harmonic multiplication of two signals $f, g \in L^2(\mathbb{S}^2)$ is defined by

$$(g \odot f)(\hat{\boldsymbol{x}}) \triangleq \sum_{n=0}^{\infty} \sum_{m=-n}^{n} g_n^m f_n^m Y_n^m(\hat{\boldsymbol{x}}) \tag{20}$$

where $f_n^m \triangleq \langle f, Y_n^m \rangle$, and $g_n^m \triangleq \langle g, Y_n^m \rangle$.

In contrast to Left Convolution, (14), this definition, regarded as a binary operation, is commutative, that is,

$$g \odot f = f \odot g, \quad \forall f, g \in L^2(\mathbb{S}^2). \tag{21}$$

Harmonic multiplication should not be regarded as exclusively a form of convolution, but part of the framework to understand azimuthally symmetric convolution.

[1] By separable we mean the space contains a complete orthonormal sequence (of functions) which is dense.

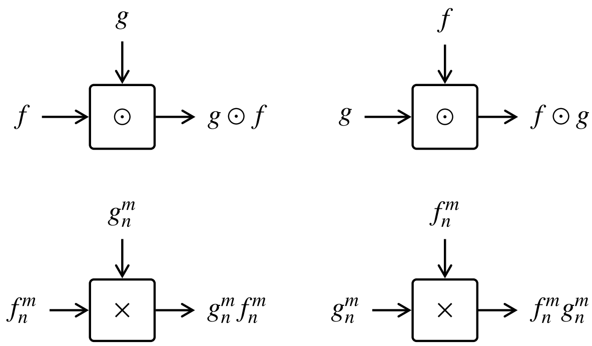

Fig. 1. Representation of Harmonic Multiplication: Top – convolution in spatial domain – f, g in $L^2(\mathbb{S}^2)$. Bottom – multiplication in discrete spherical harmonic "frequency" domain – f_n^m, g_n^m in ℓ^2.

3.4. *Harmonic multiplication on $\mathbb{S}^2$ – infinite matrix picture*

In the context of a bounded operator on $L^2(\mathbb{S}^2)$, we can regard the input as being $f \in L^2(\mathbb{S}^2)$, and regard $g \in L^2(\mathbb{S}^2)$ as parametrizing the operator, which leads to an alternative definition in terms of an infinite matrix (equivalent to Definition 1). If $\mathcal{B}$ were this operator parametrized by g, then we introduce the special notation $\mathcal{B} \equiv g\odot$, so $\mathcal{B}f = g \odot f$ which is consistent with the notation in Definition 1.

Definition 2 *(Harmonic multiplication on $\mathbb{S}^2$).* As an infinite matrix, harmonic multiplication on $L^2(\mathbb{S}^2)$ is a bounded linear operator, denoted $g\odot$, parametrized by $g \in L^2(\mathbb{S}^2)$, with infinite matrix representation defined by

$$\mathbf{G}_{nq}^{mp} \equiv \langle g \odot Y_q^p, Y_n^m \rangle \triangleq g_q^p \delta^{mp} \delta_{nq}, \tag{22}$$

where $g_n^m \triangleq \langle g, Y_n^m \rangle$, and $\delta^{mp}\delta_{nq}$ denotes the Kronecker delta.[2]

Using (22), with input $f \in L^2(\mathbb{S}^2)$ (having Fourier coefficients $f_n^m = \langle f, Y_n^m \rangle$), and output $b = g \odot f$ (having Fourier coefficients $b_n^m = \langle b, Y_n^m \rangle$), the spherical harmonic coefficients transform as

$$\begin{aligned} b_n^m &= \sum_{q=0}^{\infty} \sum_{p=-q}^{q} \mathbf{G}_{nq}^{mp} f_q^p \\ &= g_n^m f_n^m. \end{aligned} \tag{23}$$

The infinite matrix in (22) is diagonal. Fig. 1 summarizes the properties of harmonic multiplication.

3.5. *Harmonic multiplication on $\mathbb{S}^2$ – integral equation picture*

Definition 2 describes harmonic multiplication with an infinite matrix, which leads to the matrix equation in the spherical harmonic domain ("frequency domain") according to (23). This equation is similar to (14). The question is what is the corresponding spatial form of convolution analogous to (11) or (15)? This is resolved by the following theorem.

Theorem 1. *Harmonic multiplication on $\mathbb{S}^2$ is a Fredholm integral equation operator, $g\odot$, parametrized by $g \in L^2(\mathbb{S}^2)$, given by*

$$(g \odot f)(\hat{\boldsymbol{x}}) = \int_{\mathbb{S}^2} K_g(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}) f(\hat{\boldsymbol{y}})\, ds(\hat{\boldsymbol{y}}), \quad f \in L^2(\mathbb{S}^2) \tag{24}$$

and is compact with Hilbert–Schmidt kernel

$$K_g(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}) \triangleq \sum_{n=0}^{\infty} \sum_{m=-n}^{n} g_n^m Y_n^m(\hat{\boldsymbol{x}}) \overline{Y_n^m(\hat{\boldsymbol{y}})}, \tag{25}$$

where $g_n^m = \langle g, Y_n^m \rangle$. The eigenfunctions of (24) are $Y_n^m(\hat{\boldsymbol{x}})$ with corresponding eigenvalues g_n^m.

Proof. See Appendix A. □

4. Properties of convolution

4.1. *Isotropic convolution – complete characterization*

Eqs. (24) and (25) do not appear to bear a strong similarity to conventional convolution defined on $\mathbb{R}^2$, (10). However, consider the special case of the harmonic multiplication operator, written $\mathcal{Z}$, whose infinite matrix, (22), takes the form

$$\mathbf{Z}_{nq}^{mp} \triangleq \langle \mathcal{Z} Y_q^p, Y_n^m \rangle = z_n \delta^{mp} \delta_{nq} \tag{26}$$

which implies $g_n^m = z_n$, $\forall m, n$, or

$$g(\hat{\boldsymbol{x}}) = \sum_{n=0}^{\infty} z_n \sum_{m=-n}^{n} Y_n^m(\hat{\boldsymbol{x}}) \tag{27}$$

in the context of Theorem 1. In this special case the kernel (25) then simplifies to

$$\begin{aligned} K_z(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}) &\triangleq \sum_{n=0}^{\infty} \sum_{m=-n}^{n} z_n Y_n^m(\hat{\boldsymbol{x}}) \overline{Y_n^m(\hat{\boldsymbol{y}})} \\ &= \sum_{n=0}^{\infty} z_n \sum_{m=-n}^{n} Y_n^m(\hat{\boldsymbol{x}}) \overline{Y_n^m(\hat{\boldsymbol{y}})} \\ &= \frac{1}{4\pi} \sum_{n=0}^{\infty} z_n (2n+1) P_n(\hat{\boldsymbol{x}} \cdot \hat{\boldsymbol{y}}) \end{aligned} \tag{28}$$

where the last step follows from (6), the Spherical Harmonic Addition Theorem. Therefore, this special case is a form of isotropic convolution, (15), with the identification

$$z(\zeta) \triangleq \frac{1}{4\pi} \sum_{n=0}^{\infty} z_n (2n+1) P_n(\zeta), \quad -1 \leqslant \zeta \leqslant 1. \tag{29}$$

The kernel in (28) can be written $K_z(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}) = z(\hat{\boldsymbol{x}} \cdot \hat{\boldsymbol{y}})$.

Conversely, if we have the isotropic convolution in (15) we can show this is equivalent to determining the z_n in the matrix identity (26). That is, given the univariate kernel $z(\cdot)$ in (15), the completeness and orthogonality of the Legendre polynomials on the interval $[-1, +1]$,

$$\int_{-1}^{+1} P_n(\zeta) P_m(\zeta)\, d\zeta = \frac{2}{2n+1} \delta_{nm},$$

implies

$$\begin{aligned} z_n &= 2\pi \int_{-1}^{+1} z(\zeta) P_n(\zeta)\, d\zeta \\ &= 2\pi \int_{0}^{\pi} z(\cos\theta)\, P_n(\cos\theta) \sin\theta\, d\theta. \end{aligned} \tag{30}$$

[2] Equals one if and only if $m = p$ and $n = q$, and is zero otherwise.

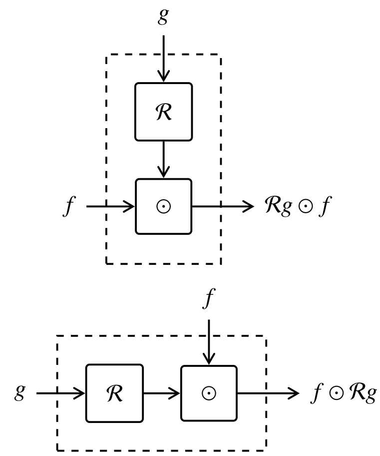

Fig. 2. Isotropic convolution synthesized from harmonic multiplication $\odot$ and operator $\mathcal{R}$ defined in (33). Note $(\mathcal{R}g)\odot f = f\odot(\mathcal{R}g)$.

Comments

(1) Convolution is thus isotropic, (15), *if and only if* the infinite matrix is diagonal and of the form (26). This is what we mean by complete characterization. Further, in the frequency domain isotropic convolution takes the form

$$\langle \mathcal{Z} f, Y_n^m\rangle = z_n f_n^m, \tag{31}$$

indicating it as a special case of harmonic multiplication (22).

(2) As an example, in [11],[3] a parametric form of smoothing (low pass filtering) based on spherical diffusion was introduced which acts on a given signal on $\mathbb{S}^2$ and corresponds to

$$z_n \triangleq e^{-n(n+1)kt} \tag{32}$$

in (26), where kt controls the spherical harmonic "bandwidth" of the filter. Of course, any well-formed sequence, z_n, that monotonically decreases to zero with n (sufficiently quickly), in the spirit of (32), will act as a low pass filter.

(3) Finally, this framework suggests that a more general form of Theorem 1 must be associated with convolution that has some degree of *anisotropy* which leads to considerations outside the scope of this paper. A broader class of integral equation kernels than (25) can be considered in the anisotropic case.

4.2. Interpretation of isotropic convolution

Isotropic Convolution, (15), convolves an input function $f \in L^2(\mathbb{S}^2)$ with univariate filter kernel $z \in L^2([-1,+1])$ to generate an output $f \in L^2(\mathbb{S}^2)$. This differs from Left Convolution, (11), which takes a "filter kernel" $h \in L^2(\mathbb{S}^2)$. To bring Isotropic Convolution and Left Convolution into a common framework, and show they are equivalent, we shall express Isotropic Convolution in an alternate way that uses harmonic multiplication.

Consider the two systems in Fig. 2 with the introduction of an operator, $\mathcal{R}$, defined through the infinite matrix

$$\mathbf{R}_{nq}^{mp} = \langle \mathcal{R} Y_q^p, Y_n^m\rangle \triangleq \delta^{0p}\delta_{nq} \tag{33}$$

which implies an input $Y_q^p(\cdot)$ maps to output $Y_n^m(\cdot)$ with gain one only when both $p=0$ and $q=n$, otherwise $Y_q^p(\cdot)$ maps to zero. The two systems are equivalent because

$$(\mathcal{R}g)\odot f = f\odot(\mathcal{R}g) \tag{34}$$

by the commutativity of the harmonic multiplication. In the upper system in Fig. 2, the kernel is parametrized by $\mathcal{R}g$ and the input is f. In the lower system in Fig. 2, g may be regarded as the input to the operator composition, $f\odot\mathcal{R}$. This lower system simplifies the analysis.

Using Definition 2, (22) with filter kernel f acting to parametrize the operator, and (33), the operator composition $f\odot\mathcal{R}$ has infinite matrix

$$\begin{aligned}\mathbf{J}_{ns}^{mr} &= \sum_{q=0}^{\infty}\sum_{p=-q}^{q}\mathbf{F}_{nq}^{mp}\mathbf{R}_{qs}^{pr}\\ &= \sum_{q=0}^{\infty}\sum_{p=-q}^{q} f_n^m\delta^{mp}\delta_{nq}\delta^{0r}\delta_{qs}\\ &= f_n^m\delta^{0r}\delta_{ns}.\end{aligned} \tag{35}$$

Then the Fourier coefficients of $f\odot(\mathcal{R}g)$ become

$$\begin{aligned}\sum_{s=0}^{\infty}\sum_{r=-s}^{s}\mathbf{J}_{ns}^{mr} g_s^r &= f_n^m\sum_{s=0}^{\infty}\sum_{r=-s}^{s}\delta^{0r}\delta_{ns} g_s^r\\ &= f_n^m g_n^0\end{aligned} \tag{36}$$

which, by (34), are the Fourier coefficients of $(\mathcal{R}g)\odot f$. This establishes the following theorem which summarizes these findings.

Theorem 2. *Consider two signals $f, g \in L^2(\mathbb{S}^2)$, and the bounded operator $\mathcal{R}$ with infinite matrix with respect to the spherical harmonics*

$$\mathbf{R}_{nq}^{mp} = \langle \mathcal{R} Y_q^p, Y_n^m\rangle \triangleq \delta^{0p}\delta_{nq}. \tag{37}$$

Then, with $\odot$ denoting the harmonic multiplication defined in (20),

$$\big((\mathcal{R}g)\odot f\big)(\hat{\boldsymbol{x}}) = \int_{\mathbb{S}^2} z(\hat{\boldsymbol{x}}\cdot\hat{\boldsymbol{y}}) f(\hat{\boldsymbol{y}})\, ds(\hat{\boldsymbol{y}}) \tag{38}$$

where the univariate isotropic convolution kernel is given by

$$z(\zeta) \triangleq \frac{1}{4\pi}\sum_{n=0}^{\infty} g_n^0 (2n+1) P_n(\zeta), \quad -1 \leqslant \zeta \leqslant +1 \tag{39}$$

with $g_n^0 = \langle g, Y_n^0\rangle$. Further, in the spherical harmonic coefficient domain

$$\langle (\mathcal{R}g)\odot f, Y_n^m\rangle = f_n^m g_n^0 \tag{40}$$

where $f_n^m = \langle f, Y_n^m\rangle$.

Comments

(1) The equivalent operator $(\mathcal{R}g)\odot$ is depicted in the dashed box in the upper system in Fig. 2. Here g parametrizes the operator, and $\mathcal{R}$ reduces it to $z(\cdot)$.

(2) The operator $\mathcal{R}$ has integral equation representation

$$(\mathcal{R}g)(\hat{\boldsymbol{x}}) = \int_{\mathbb{S}^2} R(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}) g(\hat{\boldsymbol{y}})\, ds(\hat{\boldsymbol{y}}) \tag{41}$$

with kernel

$$\begin{aligned}R(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}) &\triangleq \sum_{n=0}^{\infty}\sum_{m=-n}^{n} Y_n^m(\hat{\boldsymbol{x}})\sum_{q=0}^{\infty}\sum_{p=-q}^{q}\mathbf{R}_{nq}^{mp}\overline{Y_q^p(\hat{\boldsymbol{y}})}\\ &= \sum_{n=0}^{\infty}\overline{Y_n^0(\hat{\boldsymbol{y}})}\sum_{m=-n}^{n} Y_n^m(\hat{\boldsymbol{x}}).\end{aligned}$$

(3) The frequency domain description (40) is very similar to the one for Left Convolution, (14). In the next section we develop an analogous result for Left Convolution and give its geometric interpretation and some examples.

[3] Note that the definition of the spherical harmonics in [11] appears slightly different from our definition [9] but in fact is equivalent.

4.3. Interpretation of SO(3) left convolution

Redefine $\mathcal{R}$ in Fig. 2 to be

$$\mathbf{R}_{nq}^{mp} \equiv \langle \mathcal{R} Y_q^p, Y_n^m \rangle \triangleq \sqrt{\frac{4\pi}{(2n+1)}} \delta^{0p} \delta_{nq}. \tag{42}$$

In this case, by a simple variation on (29), we have

$$z(\zeta) \triangleq \sum_{n=0}^{\infty} g_n^0 \sqrt{\frac{(2n+1)}{4\pi}} P_n(\zeta), \quad -1 \leqslant \zeta \leqslant +1. \tag{43}$$

Letting $\zeta = \cos\theta$, and noting $g_n^0 = \langle g, Y_n^0 \rangle$

$$\begin{aligned} z(\cos\theta) &= \sum_{n=0}^{\infty} g_n^0 \sqrt{\frac{(2n+1)}{4\pi}} P_n(\cos\theta) \\ &= \sum_{n=0}^{\infty} \langle g, Y_n^0 \rangle Y_n^0(\theta, \phi) \\ &= (\mathcal{P}^0 g)(\cos\theta) \end{aligned} \tag{44}$$

by the definition of projection operator $\mathcal{P}^0$ in (9). Here note that the zonal spherical harmonics

$$Y_n^0(\theta, \phi) = Y_n^0(\theta) = \sqrt{\frac{(2n+1)}{4\pi}} P_n(\cos\theta) \tag{45}$$

are independent of ϕ and are naturally functions of $\cos\theta$.

We summarize this in the following theorem which is equivalent to [10, Theorem 1].

Theorem 3. *The Left Convolution of two signals $f, h \in L^2(\mathbb{S}^2)$, (11), is equivalent to isotropic convolution*

$$(f \star h)(\hat{\boldsymbol{x}}) = \int_{\mathbb{S}^2} z(\hat{\boldsymbol{x}} \cdot \hat{\boldsymbol{y}}) f(\hat{\boldsymbol{y}})\, ds(\hat{\boldsymbol{y}}) \tag{46}$$

where the isotropic convolution kernel is induced by $h \in L^2(\mathbb{S}^2)$ and given by

$$\begin{aligned} z(\cos\theta) &= (\mathcal{P}^0 h)(\cos\theta), \quad 0 \leqslant \theta \leqslant \pi \\ &= \sum_{n=0}^{\infty} \langle h, Y_n^0 \rangle Y_n^0(\theta), \end{aligned} \tag{47}$$

the co-latitude function of the orthogonal projection of h onto the subspace $\mathcal{H}^0(\mathbb{S}^2)$ of zonal spherical harmonics in $L^2(\mathbb{S}^2)$.

Further, in the spherical harmonic coefficient domain

$$\langle f \star h, Y_n^m \rangle = \sqrt{\frac{4\pi}{2n+1}} h_n^0 f_n^m \tag{48}$$

where $f_n^m = \langle f, Y_n^m \rangle$.

Comments

(1) This result shows the fundamental role the azimuthally symmetric function subspace plays in characterizing the action of Left Convolution. It shows that any component of the filter function h orthogonal to $\mathcal{H}^0(\mathbb{S}^2) \subset L^2(\mathbb{S}^2)$ plays no role in the final result.

(2) A standard example of such convolution, from [18], uses the von Mises distribution, centered at $\mu = 0$ and concentration parameter $\kappa \geqslant 0$

$$n(\theta) = \frac{1}{I_0(\kappa)} e^{\kappa \cos\theta} \tag{49}$$

for the co-latitude function $z(\cos\theta)$. The function on the sphere, h, in Left Convolution, would have projection $\mathcal{P}^0 h$, see (9), proportional to the well-known von Mises–Fisher distribution.

(3) Theorem 3 takes a given $h(\cdot)$ in $L^2(\mathbb{S}^2)$ and synthesizes a kernel $z(\cdot)$ on $L^2([-1,+1])$. Conversely, we can ask what $h(\cdot)$ in $L^2(\mathbb{S}^2)$ corresponds to a given $z(\cdot)$ on $L^2([-1,+1])$. The answer, by the orthogonality of the Legendre polynomials, is any $h \in L^2(\mathbb{S}^2)$ satisfying

$$h_n^0 = \langle h, Y_n^0 \rangle = \sqrt{\frac{2n+1}{4\pi}} \int_{-1}^{+1} z(\zeta) P_n(\zeta)\, d\zeta.$$

Because the convolution is isotropic, there is no need to start with a function on $L^2(\mathbb{S}^2)$ to define the convolution when the kernel $z(\cdot)$ on $L^2([-1,+1])$ is all that is needed.

In (46), the filtering kernel $z(\cdot)$ has an obvious geometric interpretation. It is of interest to know whether $h(\cdot)$ in $L^2(\mathbb{S}^2)$ in Theorem 3 has a geometric interpretation. It is clear that the Left Convolution depends only on the rotational symmetric part of h in $f \star h$. One might speculate that if h is purely rotationally symmetric itself then h might form directly the filtering kernel. Noting

$$Y_n^0(\theta, \phi) = \sqrt{\frac{(2n+1)}{4\pi}} P_n(\cos\theta) \tag{50}$$

is invariant to ϕ, then, from (47),

$$\begin{aligned} z(\cos\theta) &= \frac{1}{2\pi} \sum_{n=0}^{\infty} h_n^0 Y_n^0(\theta, \phi), \quad h_n^0 = \langle h, Y_n^0 \rangle \\ &= \frac{1}{2\pi} h(\theta), \quad \text{if} \quad h(\theta, \phi) = h(\theta). \end{aligned}$$

So indeed the filtering kernel $z(\cdot)$ is governed by the co-latitude profile of the symmetric part of h, apart from the extraneous factor of $1/2\pi$.

Comments

(1) The framework we developed in this paper indicates that SO(3) methods [10,11] are not a natural extension of conventional convolution defined on $\mathbb{R}^2$, given in (10). The simpler isotropic convolution formula, (15), has the equivalent action.

(2) It is clear from the earlier discussion that Left Convolution implicitly integrates over three Euler angles, in contrast to $\mathbb{R}^2$ convolution (10) which integrates over 2D translations in the plane. Integration over the third rotation is the source of the difficulties, essentially this is integration over $\gamma \in [0, 2\pi)$ through matrix $\boldsymbol{r}_\gamma^{(z)}$ in (12). This is precisely integrating over rotations about the north pole $\hat{\boldsymbol{\eta}} \in \mathbb{S}^2$ which zeroes the contribution from components with $m \neq 0$.

(3) There is an argument that rotations in SO(3) are the analog of translations in $\mathbb{R}^2$. This is strictly not correct. Under the actions of all elements of SO(3) a function in $L^2(\mathbb{S}^2)$ is to reposition to all possible locations on the 2-sphere in a proper sense (determinant 1). The analogy in the 2D would have the set of all 2D translations plus a rotation – a playing card flicked on the floor can also take an arbitrary orientation. However, 2D convolution is defined without the rotation and the orientation is fixed. Were rotation included in the $\mathbb{R}^2$ convolution (triple integral) then we would observe the same rotation averaging we see in Left Convolution.

5. Conclusions

We developed a general framework which introduced the notion of harmonic multiplication and projections on to the subspace of finite energy signals with azimuthal/rotational symmetry about the north pole axis. Using this we have shown a nontrivial equivalence between different competing notions of convolution on the sphere but all equivalent to convolution with a kernel which is azimuthally symmetric or isotropic. Our results indicate that a more general form of convolution that incorporates some degree of anisotropy requires more complete investigation and this is the focus of our ongoing work to develop a more extensive framework.

Appendix A. Proof of Theorem 1

Proof. Using (23)

$$b(\hat{\boldsymbol{x}}) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} b_n^m Y_n^m(\hat{\boldsymbol{x}}) \tag{A.1a}$$

$$= \sum_{n=0}^{\infty} \sum_{m=-n}^{n} g_n^m \langle f, Y_n^m \rangle Y_n^m(\hat{\boldsymbol{x}}) \tag{A.1b}$$

$$= \sum_{n=0}^{\infty} \sum_{m=-n}^{n} g_n^m \int_{\mathbb{S}^2} f(\hat{\boldsymbol{y}}) \overline{Y_n^m(\hat{\boldsymbol{y}})}\, ds(\hat{\boldsymbol{y}}) Y_n^m(\hat{\boldsymbol{x}}) \tag{A.1c}$$

$$= \int_{\mathbb{S}^2} K_g(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}) f(\hat{\boldsymbol{y}})\, ds(\hat{\boldsymbol{y}}) \tag{A.1d}$$

where $K_g(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}})$ is given by (25). So (24) is the spatial form of the convolution and (25) is the convolution kernel. Further, by orthonormality of the spherical harmonics and Parseval's Theorem

$$\int_{\mathbb{S}^2} \left| K_g(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}) \right|^2 ds(\hat{\boldsymbol{x}})\, ds(\hat{\boldsymbol{y}}) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} \left| g_n^m \right|^2$$

$$= \int_{\mathbb{S}^2} \left| g(\hat{\boldsymbol{x}}) \right|^2 ds(\hat{\boldsymbol{x}}). \tag{A.2}$$

Therefore, $g \in L^2(\mathbb{S}^2)$ implies $K_g(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}) \in L^2(\mathbb{S}^2 \times \mathbb{S}^2)$, that is, the kernel is Hilbert–Schmidt and the operator is compact [8].

Since $(\mathcal{G} Y_n^m)(\hat{\boldsymbol{x}}) = g_n^m Y_n^m(\hat{\boldsymbol{x}})$, the eigenfunctions of (24) are the spherical harmonics $Y_n^m(\hat{\boldsymbol{x}})$, with corresponding eigenvalues g_n^m. Hence the eigenfunctions of $\mathcal{G}$ form a complete orthonormal system in $L^2(\mathbb{S}^2)$ since the spherical harmonics have this property [9, p. 25]. This establishes the Spectral Theorem for this class of (non self-adjoint) compact operators.

Given convolutions on $\mathbb{S}^2$ have the diagonal infinite matrix form in (22) then applying multiple convolutions is equivalent to a single convolution (closure) and convolutions are associative. An inverse may not exist given one of the g_n^m might be zero. This is the *semi-group* property. Commutativity as a property does hold. □

References

[1] F.J. Simons, F.A. Dahlen, M.A. Wieczorek, Spatiospectral concentration on a sphere, SIAM Rev. 48 (3) (2006) 504–536.
[2] K.M. Górski, E. Hivon, A.J. Banday, B.D. Wandelt, F.K. Hansen, M. Reinecke, M. Bartelmann, Healpix: A framework for high-resolution discretization and fast analysis of data distributed on the sphere, Astrophys. J. 622 (2) (2005) 759–771.
[3] C. Armitage, B.D. Wandelt, Deconvolution map-making for cosmic microwave background observations, Phys. Rev. D 70 (12) (2004) 123007.
[4] K.-W. Ng, Full-sky correlation functions for CMB experiments with asymmetric window functions, Phys. Rev. D 71 (8) (2005) 083009.
[5] B.D. Wandelt, K.M. Górski, Fast convolution on the sphere, Phys. Rev. D 63 (12) (2001) 123002.
[6] B. Rafaely, Plane-wave decomposition of the sound field on a sphere by spherical convolution, J. Acoust. Soc. America 116 (4) (2004) 2149–2157.
[7] Y. Wiaux, L. Jacques, P. Vielva, P. Vandergheynst, Fast directional correlation on the sphere with steerable filters, Astrophys. J. 652 (1) (2006) 820–832.
[8] L. Debnath, P. Mikusiński, Introduction to Hilbert Spaces with Applications, Academic Press, San Diego, CA, 1999.
[9] D. Colton, R. Kress, Inverse Acoustic and Electromagnetic Scattering Theory, 2nd edition, Springer-Verlag, Berlin, Germany, 1998.
[10] J.R. Driscoll, D.M. Healy, Computing Fourier transforms and convolutions on the 2-sphere, Adv. in Appl. Math. 15 (1994) 202–250.
[11] T. Bulow, Spherical diffusion for 3D surface smoothing, IEEE Trans. Pattern Anal. Machine Intell. 26 (12) (2004) 1650–1654.
[12] B.T.T. Yeo, O. Wanmei, P. Golland, On the construction of invertible filter banks on the 2-sphere, IEEE Trans. Image Process. 17 (3) (2008) 283–300, doi:10.1109/TIP.2007.915550.
[13] T. Bülow, Multiscale image processing on the sphere, in: L. Van Gool (Ed.), Pattern Recognition, in: Lecture Notes in Comput. Sci., vol. 2449, Springer, Berlin/Heidelberg, 2002, pp. 609–617.
[14] A. Makadia, K. Daniilidis, Rotation recovery from spherical images without correspondences, IEEE Trans. Pattern Anal. Machine Intell. 28 (7) (2006) 1170–1175, http://doi.ieeecomputersociety.org/10.1109/TPAMI.2006.150.
[15] T. Bülow, Spherical diffusion for 3D surface smoothing, in: First International Symposium on 3D Data Processing and Visualization and Transmission, 2002, pp. 449–458.
[16] K. Zhou, J. Bao, J. Shi, 3D surface filtering using spherical harmonics, Science Direct: Comput.-Aided Design 36 (4) (2004) 363–375, doi:10.1016/S0010-4485(03)00098-8.
[17] M. Tegmark, D.H. Hartmann, M.S. Briggs, C.A. Meegan, The angular power spectrum of BATSE 3B gamma-ray burst, Astrophys. J. 468 (1996) 214, arXiv:astro-ph/9510129v1.
[18] K.-I. Seon, Smoothing of an all-sky survey map with a Fisher–von Mises function, J. Korean Phys. Soc. 48 (3) (2006) 331–334.

Rodney A. Kennedy received his BE from the University of New South Wales, Australia, ME from the University of Newcastle, and PhD from the Australian National University. He worked 3 years for the Commonwealth Scientific and Industrial Research Organization (CSIRO) on the Australia Telescope Project. Since 2000 he has been Professor and is currently in the Research School of Engineering at the Australian National University. His research interests are in the fields of digital signal processing, digital and wireless communications, and acoustical signal processing.

Tharaka A. Lamahewa (M'06) received the BE degree from the University of Adelaide, South Australia, in 2000, and the PhD degree from the Australian National University in 2007. He was with Motorola Electronics Pvt Ltd., Singapore, for two years as a software design engineer. From 2006 to 2007, he was an algorithm design engineer with Nanoradio AB, Melbourne, Australia. He is now with the Applied Signal Processing Group, School of Engineering, Australian National University, Canberra, Australia. His research interests include wireless communications, underwater communications and tracking, and digital signal processing.

Liying Wei received the BE degree from Xidian University in 2002, the ME degree from Beijing University of Posts and Telecommunications in 2006, and the PhD degree from The Australian National University in 2011. She worked as a Research Fellow in the Applied Signal Processing Group, Research School of Information Science and Engineering (RSISE), the Australian National University from 2010 to 2011. Her current research interests include signal processing with applications in wireless communication, acoustic, medical imaging, and sensor networks.