Mercer's Theorem, Feature Maps, and Smoothing

Ha Quang Minh[1], Partha Niyogi[1], and Yuan Yao[2]

[1] Department of Computer Science, University of Chicago
1100 East 58th St, Chicago, IL 60637, USA
[2] Department of Mathematics, University of California, Berkeley
970 Evans Hall, Berkeley, CA 94720, USA
minh,niyogi@cs.uchicago.edu, yao@math.berkeley.edu

Abstract. We study Mercer's theorem and feature maps for several positive definite kernels that are widely used in practice. The smoothing properties of these kernels will also be explored.

1 Introduction

Kernel-based methods have become increasingly popular and important in machine learning. The central idea behind the so-called "kernel trick" is that a closed form Mercer kernel allows one to efficiently solve a variety of non-linear optimization problems that arise in regression, classification, inverse problems, and the like. It is well known in the machine learning community that kernels are associated with "feature maps" and a kernel based procedure may be interpreted as mapping the data from the original input space into a potentially higher dimensional "feature space" where linear methods may then be used. One finds many accounts of this idea where the input space X is mapped by a feature map $\Phi : X \to \mathcal{H}$ (where $\mathcal{H}$ is a Hilbert space) so that for any two points $x, y \in X$, we have $K(x, y) = \langle \phi(x), \phi(y) \rangle_{\mathcal{H}}$.

Yet, while much has been written about kernels and many different kinds of kernels have been discussed in the literature, much less has been explicitly written about their associated feature maps. In general, we do not have a clear and concrete understanding of what exactly these feature maps are. Our goal in this paper is to take steps toward a better understanding of feature maps by explicitly computing them for a number of popular kernels for a variety of domains. By doing so, we hope to clarify the precise nature of feature maps in very concrete terms so that machine learning researchers may have a better feel for them.

Following are the main points and new results of our paper:

1. As we will illustrate, feature maps and feature spaces are not unique. For a given domain X and a fixed kernel K on $X \times X$, there exist in fact infinitely many feature maps associated with K. Although these maps are essentially equivalent, in a sense to be made precise in Section 4.3, there are subtleties

that we wish to emphasize. For a given kernel K, the feature maps of K induced by Mercer's theorem depend fundamentally on the domain X, as will be seen in the examples of Section 2. Moreover, feature maps do not necessarily arise from Mercer's theorem, examples of which will be given in Section 4.2. The importance of Mercer's theorem, however, goes far beyond the feature maps that it induces: the eigenvalues and eigenfunctions associated with K play a central role in obtaining error estimates in learning theory, see for example [8], [4]. For this reason, the determination of the spectrum of K, which is highly nontrivial in general, is crucially important in its own right. Theorems 2 and 3 in Section 2 give the complete spectrum of the polynomial and Gaussian kernels on S^{n-1}, including sharp rates of decay of their eigenvalues. Theorem 4 gives the eigenfunctions and a recursive formula for the computation of eigenvalues of the polynomial kernel on the hypercube $\{-1,1\}^n$.

2. One domain that we particularly focus on is the unit sphere S^{n-1} in $\mathbb{R}^n$, for several reasons. First, it is a special example of a compact Riemannian manifold and the problem of learning on manifolds has attracted attention recently, see for example [2], [3]. Second, its symmetric and homogeneous nature allows us to obtain complete and explicit results in many cases. We believe that S^{n-1} together with kernels defined on it is a fruitful source of examples and counterexamples for theoretical analysis of kernel-based learning. We will point out that intuitions based on low dimensions such as $n = 2$ in general do not carry over to higher dimensions - Theorem 5 in Section 3 gives an important example along this line. We will also consider the unit ball B^n, the hypercube $\{-1,1\}^n$, and $\mathbb{R}^n$ itself.
3. We will also try to understand the smoothness property of kernels on S^{n-1}. In particular, we will show that the polynomial and Gaussian kernels define Hilbert spaces of functions whose norms may be interpreted as smoothness functionals, similar to those of splines on S^{n-1}. We will obtain precise and sharp results on this question in the paper. This is the content of Section 5. The smoothness implications allow us to better understand the applicability of such kernels in solving smoothing problems.

Notation: For $X \subset \mathbb{R}^n$ and μ a Borel measure on X, $L^2_\mu(X) = \{f : X \to \mathbb{C} : \int_X |f(x)|^2 d\mu(x) < \infty\}$. We will also use $L^2(X)$ for $L^2_\mu(X)$ and dx for $d\mu(x)$ if μ is the Lebesgue measure on X. The surface area of the unit sphere S^{n-1} is denoted by $|S^{n-1}| = \frac{2\pi^{\frac{n}{2}}}{\Gamma(\frac{n}{2})}$.

2 Mercer's Theorem

One of the fundamental mathematical results underlying learning theory with kernels is Mercer's theorem. Let X be a closed subset of $\mathbb{R}^n$, $n \in \mathbb{N}$, μ a Borel measure on X, and $K : X \times X \to \mathbb{R}$ a symmetric function satisfying: for any finite set of points $\{x_i\}_{i=1}^N$ in X and real numbers $\{a_i\}_{i=1}^N$

$$\sum_{i,j=1}^N a_i a_j K(x_i, x_j) \geq 0 \tag{1}$$

K is said to be a positive definite kernel on X. Assume further that

$$\int_X \int_X K(x,t)^2 d\mu(x) d\mu(t) < \infty \tag{2}$$

Consider the induced integral operator $L_K : L^2_\mu(X) \to L^2_\mu(X)$ defined by

$$L_K f(x) = \int_X K(x,t) f(t) d\mu(t) \tag{3}$$

This is a self-adjoint, positive, compact operator with a countable system of nonnegative eigenvalues $\{\lambda_k\}_{k=1}^\infty$ satisfying $\sum_{k=1}^\infty \lambda_k^2 < \infty$. L_K is said to be Hilbert-Schmidt and the corresponding $L^2_\mu(X)$-normalized eigenfunctions $\{\phi_k\}_{k=1}^\infty$ form an orthonormal basis of $L^2_\mu(X)$. We recall that a Borel measure μ on X is said to be strictly positive if the measure of every nonempty open subset in X is positive, an example being the Lebesgue measure in $\mathbb{R}^n$.

Theorem 1 (Mercer). *Let $X \subset \mathbb{R}^n$ be closed, μ a strictly positive Borel measure on X, K a continuous function on $X \times X$ satisfying (1) and (2). Then*

$$K(x,t) = \sum_{k=1}^\infty \lambda_k \phi_k(x) \phi_k(t) \tag{4}$$

where the series converges absolutely for each pair $(x,t) \in X \times X$ and uniformly on each compact subset of X.

Mercer's theorem still holds if X is a finite set $\{x_i\}$, such as $X = \{-1,1\}^n$, K is pointwise-defined positive definite and $\mu(x_i) > 0$ for each i.

2.1 Examples on the Sphere S^{n-1}

We will give explicit examples of the eigenvalues and eigenfunctions in Mercer's theorem on the unit sphere S^{n-1} for the polynomial and Gaussian kernels. We need the concept of spherical harmonics, a modern and authoritative account of which is [6]. Some of the material below was first reported in the kernel learning literature in [9], where the eigenvalues for the polynomial kernels with $n = 3$, were computed. In this section, we will carry out computations for a general $n \in \mathbb{N}$, $n \geq 2$.

Definition 1 (Spherical Harmonics). *Let $\Delta_n = -\left[\frac{\partial^2}{\partial x_1^2} + \ldots + \frac{\partial^2}{\partial x_n^2}\right]$ denote the Laplacian operator on $\mathbb{R}^n$. A homogeneous polynomial of degree k in $\mathbb{R}^n$ whose Laplacian vanishes is called a homogeneous harmonic of order k. Let $\mathcal{Y}_k(n)$*

denote the subspace of all homogeneous harmonics of order k on the unit sphere S^{n-1} in $\mathbb{R}^n$. The functions in $\mathcal{Y}_k(n)$ are called spherical harmonics of order k. We will denote by $\{Y_{k,j}(n;x)\}_{j=1}^{N(n,k)}$ any fixed orthonormal basis for $\mathcal{Y}_k(n)$ where $N(n,k) = \dim \mathcal{Y}_k(n) = \frac{(2k+n-2)(k+n-3)!}{k!(n-2)!}$, $k \geq 0$.

Theorem 2. *Let $X = S^{n-1}$, $n \in \mathbb{N}$, $n \geq 2$. Let μ be the uniform probability distribution on S^{n-1}. For $K(x,t) = \exp(-\frac{||x-t||^2}{\sigma^2})$, $\sigma > 0$*

$$\lambda_k = e^{-2/\sigma^2}\sigma^{n-2} I_{k+n/2-1}(\frac{2}{\sigma^2})\Gamma(\frac{n}{2}) \tag{5}$$

for all $k \in \mathbb{N} \cup \{0\}$, where I denotes the modified Bessel function of the first kind, defined below. Each λ_k occurs with multiplicity $N(n,k)$ with the corresponding eigenfunctions being spherical harmonics of order k on S^{n-1}. The λ_k's are decreasing if $\sigma \geq (\frac{2}{n})^{1/2}$. Furthermore

$$(\frac{2e}{\sigma^2})^k \frac{A_1}{(2k+n-2)^{k+\frac{n-1}{2}}} < \lambda_k < (\frac{2e}{\sigma^2})^k \frac{A_2}{(2k+n-2)^{k+\frac{n-1}{2}}} \tag{6}$$

for A_1, A_2 depending on σ and n given below.

Remark 1. $A_1 = e^{-2/\sigma^2 - 1/12}\frac{1}{\sqrt{\pi}}(2e)^{\frac{n}{2}-1}\Gamma(\frac{n}{2})$, $A_2 = e^{-2/\sigma^2 + 1/\sigma^4}\frac{1}{\sqrt{\pi}}(2e)^{\frac{n}{2}-1}\Gamma(\frac{n}{2})$. For $\nu, z \in \mathbb{C}$, $I_\nu(z) = \sum_{j=0}^{\infty} \frac{1}{j!\Gamma(\nu+j+1)}\left(\frac{z}{2}\right)^{\nu+2j}$.

Theorem 3. *Let $X = S^{n-1}$, $n \in \mathbb{N}$, $n \geq 2$, and $d \in \mathbb{N}$. Let μ be the uniform probability distribution on S^{n-1}. For $K(x,t) = (1 + \langle x,t \rangle)^d$, the nonzero eigenvalues of $L_K : L^2_\mu(X) \to L^2_\mu(X)$ are*

$$\lambda_k = 2^{d+n-2}\frac{d!}{(d-k)!}\frac{\Gamma(d+\frac{n-1}{2})\Gamma(\frac{n}{2})}{\sqrt{\pi}\Gamma(d+k+n-1)} \tag{7}$$

for $0 \leq k \leq d$. Each λ_k occurs with multiplicity $N(n,k)$, with the corresponding eigenfunctions being spherical harmonics of order k on S^{n-1}. Furthermore, the λ_k's form a decreasing sequence and

$$\frac{B_1}{(k+d+n-2)^{2d+n-\frac{3}{2}}} < \lambda_k < \frac{B_2}{(k+d+n-2)^{d+n-\frac{3}{2}}} \tag{8}$$

where $0 \leq k \leq d$, for B_1, B_2 depending on d, n given below.

Remark 2. $B_1 = e^d(2e)^{d+n-2}d!\frac{\Gamma(d+\frac{n-1}{2})\Gamma(\frac{n}{2})}{2\pi\sqrt{\pi}e^{1/6}d^{d+\frac{1}{2}}}$, $B_2 = e^d(2e)^{d+n-2}d!\frac{\Gamma(d+\frac{n-1}{2})\Gamma(\frac{n}{2})}{\sqrt{2\pi}}$.

2.2 Example on the Hypercube $\{-1,1\}^n$

We will now give an example with the hypercube $\{-1,1\}^n$. Let $\mathcal{M}_k = \{\alpha = (\alpha_i)_{i=1}^n, \alpha_i \in \{0,1\}, |\alpha| = \alpha_1 + \cdots + \alpha_n = k\}$, then the set $\{x^\alpha\}_{\alpha \in \mathcal{M}_k, 0 \leq k \leq n}$, consists of multilinear mononomials $\{1, x_1, x_1x_2, \ldots, x_1 \ldots x_n\}$.

Theorem 4. *Let $X = \{-1,1\}^n$. Let $d \in \mathbb{N}$, $d \leq n$ be fixed. Let $K(x,t) = (1+\langle x,t\rangle)^d$ on $X \times X$. Let μ be the uniform distribution on X, then the nonzero eigenvalues λ_k^d's of $L_K : L^2_\mu(X) \to L^2_\mu(X)$ satisfy*

$$\lambda_k^{d+1} = k\lambda_{k-1}^d + \lambda_k^d + (n-k)\lambda_{k+1}^d \tag{9}$$

$$\lambda_0^d \geq \lambda_1^d \geq \ldots \geq \lambda_{d-1}^d = \lambda_d^d = d! \tag{10}$$

and $\lambda_k^d = 0$ for $k > d$. The corresponding $L^2_\mu(X)$-normalized eigenfunctions for each λ_k are $\{x^\alpha\}_{\alpha \in \mathcal{M}_k}$.

Example 1 ($d = 2$). The recurrence relation (9) is nonlinear in two indexes and hence a closed analytic expression for λ_k^d is hard to find for large d. It is straightforward, however, to write a computer program for computing λ_k^d. For $d = 2$

$$\lambda_0^2 = n+1 \;\; \lambda_1^2 = 2 \;\; \lambda_2^2 = 2$$

with corresponding eigenfunctions 1, $\{x_1, \ldots, x_n\}$, and $\{x_1x_2, x_1x_3, \ldots, x_{n-1}x_n\}$, respectively.

2.3 Example on the Unit Ball B^n

Except for the homogeneous polynomial kernel $K(x,t) = \langle x,t\rangle^d$, the computation of the spectrum of L_K on the unit ball B^n is much more difficult analytically than that on S^{n-1}. For $K(x,t) = (1+\langle x,t\rangle)^d$ and small values of d, it is still possible, however, to obtain explicit answers.

Example 2 ($X = B^n$, $K(x,t) = (1+\langle x,t\rangle)^2$). Let μ be the uniform measure on B^n. The eigenspace spanned by $\{x_1, \ldots, x_n\}$ corresponds to the eigenvalue $\lambda_1 = \frac{2}{(n+2)}$. The eigenspace spanned by $\{||x||^2 Y_{2,j}(n; \frac{x}{||x||})\}_{j=1}^{N(n,2)}$ corresponds to the eigenvalue $\lambda_2 = \frac{2}{(n+2)(n+4)}$. The eigenvalues that correspond to *span*$\{1, ||x||^2\}$ are

$$\lambda_{0,1} = \frac{(n+2)(n+5) + \sqrt{D}}{2(n+2)(n+4)} \quad \lambda_{0,2} = \frac{(n+2)(n+5) - \sqrt{D}}{2(n+2)(n+4)}$$

where $D = (n+2)^2(n+5)^2 - 16(n+4)$.

3 Unboundedness of Normalized Eigenfunctions

It is known that the L^2_μ-normalized eigenfunctions $\{\phi_k\}$ are generally unbounded, that is in general

$$\sup_{k\in\mathbb{N}} ||\phi_k||_\infty = \infty$$

This was first pointed out by Smale, with the first counterexample given in [14]. This phenomenon is very common, however, as the following result shows.

Theorem 5. *Let $X = S^{n-1}$, $n \geq 3$. Let μ be the Lebesgue measure on S^{n-1}. Let $f : [-1,1] \to \mathbb{R}$ be a continuous function, giving rise to a Mercer kernel $K(x,t) = f(\langle x,t\rangle)$ on $S^{n-1} \times S^{n-1}$. If infinitely many of the eigenvalues of $L_K : L^2_\mu(S^{n-1}) \to L^2_\mu(S^{n-1})$ are nonzero, then for the set of corresponding L^2_μ-normalized eigenfunctions $\{\phi_k\}_{k=1}^\infty$*

$$\sup_{k\in\mathbb{N}} ||\phi_k||_\infty = \infty \tag{11}$$

Remark 3. This is in sharp contrast with the case $n = 2$, where we will show that

$$\sup_k ||\phi_k||_\infty \leq \frac{1}{\sqrt{\pi}}$$

with the supremum attained on the functions $\{\frac{\cos k\theta}{\sqrt{\pi}}, \frac{\sin k\theta}{\sqrt{\pi}}\}_{k\in\mathbb{N}}$. Theorem 5 applies in particular to the Gaussian kernel $K(x,t) = \exp(-\frac{||x-t||^2}{\sigma^2})$. Hence care needs to be taken in applying analysis that requires $C_K = \sup_k ||\phi_k||_\infty < \infty$, for example [5].

4 Feature Maps

4.1 Examples of Feature Maps via Mercer's Theorem

A natural feature map that arises immediately from Mercer's theorem is

$$\Phi_\mu : X \to \ell^2 \quad \Phi_\mu(x) = (\sqrt{\lambda_k}\phi_k(x))_{k=1}^\infty \tag{12}$$

where if only $N < \infty$ of the eigenvalues are strictly positive, then $\Phi_\mu : X \to \mathbb{R}^N$. This is the map that is often covered in the machine learning literature.

Example 3 ($n = d = 2$,$X = S^{n-1}$). Theorem 3 gives the eigenvalues $(3\pi, 2\pi, \frac{\pi}{2})$, with eigenfunctions $(\frac{1}{\sqrt{2\pi}}, \frac{x_1}{\sqrt{\pi}}, \frac{x_2}{\sqrt{\pi}}, \frac{2x_1x_2}{\sqrt{\pi}}, \frac{x_1^2-x_2^2}{\sqrt{\pi}}) = (\frac{1}{\sqrt{2\pi}}, \frac{\cos\theta}{\sqrt{\pi}}, \frac{\sin\theta}{\sqrt{\pi}}, \frac{\sin 2\theta}{\sqrt{\pi}}, \frac{\cos 2\theta}{\sqrt{\pi}})$, where $x_1 = \cos\theta$, $x_2 = \sin\theta$, giving rise to the feature map

$$\Phi_\mu(x) = (\sqrt{\frac{3}{2}}, \sqrt{2}x_1, \sqrt{2}x_2, \sqrt{2}x_1x_2, \frac{x_1^2 - x_2^2}{\sqrt{2}})$$

Example 4 ($n = d = 2$, $X = \{-1,1\}^2$). Theorem 4 gives

$$\Phi_\mu(x) = (\sqrt{3}, \sqrt{2}x_1, \sqrt{2}x_2, \sqrt{2}x_1x_2)$$

Observation 1 *(i) As our notation suggests, Φ_μ depends on the particular measure μ that is in the definition of the operator L_K and thus is not unique. Each measure μ gives rise to a different system of eigenvalues and eigenfunctions $(\lambda_k, \phi_k)_{k=1}^\infty$ and therefore a different Φ_μ.*

(ii) In Theorem 3 and 2, the multiplicity of the λ_k's means that for each choice of orthonormal bases of the space $\mathcal{Y}_k(n)$ of spherical harmonics of order k, there is a different feature map. Thus are infinitely many feature maps arising from the uniform probability distribution on S^{n-1} alone.

4.2 Examples of Feature Maps not via Mercer's Theorem

Feature maps do not necessarily arise from Mercer's theorem. Consider any set X and any pointwise-defined, positive definite kernel K on $X \times X$. For each $x \in X$, let $K_x : X \to \mathbb{R}$ be defined by $K_x(t) = K(x,t)$ and

$$\mathcal{H}_K = \overline{\text{span}\{K_x : x \in X\}} \tag{13}$$

be the Reproducing Kernel Hilbert Space (RKHS) induced by K, with the inner product $\langle K_x, K_t \rangle_K = K(x,t)$, see [1]. The following feature map is then immediate:

$$\Phi_K : X \to \mathcal{H}_K \quad \Phi_K(x) = K_x \tag{14}$$

In this section we discuss, via examples, two other methods for obtaining feature maps. Let $X \subset \mathbb{R}^n$ be any subset. Consider the Gaussian kernel $K(x,t) = \exp(-\frac{||x-t||^2}{\sigma^2})$ on $X \times X$, which admits the following expansion

$$K(x,t) = \exp(-\frac{||x-t||^2}{\sigma^2}) = e^{-\frac{||x||^2}{\sigma^2}} e^{-\frac{||t||^2}{\sigma^2}} \sum_{k=0}^{\infty} \frac{(2/\sigma^2)^k}{k!} \sum_{|\alpha|=k} C_\alpha^k x^\alpha t^\alpha \tag{15}$$

where $C_\alpha^k = \frac{k!}{(\alpha_1)!\dots(\alpha_n)!}$, which implies the feature map: $\Phi_g : X \to \ell^2$ where

$$\Phi_g(x) = e^{-\frac{||x||^2}{\sigma^2}} (\sqrt{\frac{(2/\sigma^2)^k C_\alpha^k}{k!}} x^\alpha)_{|\alpha|=k,k=0}^{\infty}$$

Remark 4. The standard polynomial feature maps in machine learning, see for example ([7], page 28), are obtained exactly in the same way.

Consider next a special class of kernels that is widely used in practice, called **convolution kernels**. We recall that for a function $f \in L^1(\mathbb{R}^n)$, its Fourier transform is defined to be

$$\hat{f}(\xi) = \int_{\mathbb{R}^n} f(x) e^{-i\langle \xi, x \rangle} dx$$

By Fourier transform computation, it may be shown that if $\mu : \mathbb{R}^n \to \mathbb{R}$ is even, nonnegative, such that $\mu, \sqrt{\mu} \in L^1(\mathbb{R}^n)$, then the kernel $K : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ defined by

$$K(x,t) = \int_{\mathbb{R}^n} \mu(u) e^{-i\langle x-t, u \rangle} du \tag{16}$$

is continuous, symmetric, positive definite. Further more, for any $x, t \in \mathbb{R}^n$

$$K(x,t) = \frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} \widehat{\sqrt{\mu}}(x-u) \widehat{\sqrt{\mu}}(t-u) du \tag{17}$$

The following then is a feature map of K on $X \times X$

$$\Phi_{conv} : X \to L^2(\mathbb{R}^n) \tag{18}$$

$$\Phi_{conv}(x)(u) = \frac{1}{(2\pi)^{\frac{n}{2}}}\widehat{\sqrt{\mu}}(x-u)$$

For the Gaussian kernel $e^{-\frac{||x-t||^2}{\sigma^2}} = (\frac{\sigma}{2\sqrt{\pi}})^n \int_{\mathbb{R}^n} e^{-\frac{\sigma^2||u||^2}{4}} e^{-i\langle x-t,u\rangle} du$ and

$$(\Phi_{conv}(x))(u) = (\frac{2}{\sigma\sqrt{\pi}})^{\frac{n}{2}} e^{-\frac{2||x-u||^2}{\sigma^2}}$$

One can similarly obtain feature maps for the inverse multiquadric, exponential, or B-spline kernels. The identity $e^{-\frac{||x-t||^2}{\sigma^2}} = (\frac{4}{\pi\sigma^2})^{\frac{n}{2}} \int_{\mathbb{R}^n} e^{-\frac{2||x-u||^2}{\sigma^2}} e^{-\frac{2||t-u||^2}{\sigma^2}} du$ can also be verified directly, as done in [10], where implications of the Gaussian feature map $\Phi_{conv}(x)$ above are also discussed.

4.3 Equivalence of Feature Maps

It is known ([7], page 39) that, given a set X and a pointwise-defined, symmetric, positive definite kernel K on $X \times X$, all feature maps from X into Hilbert spaces are essentially equivalent. In this section, we will make this statement precise. Let $\mathcal{H}$ be a Hilbert space and $\Phi : X \to \mathcal{H}$ be such that $\langle \Phi_x, \Phi_t\rangle_{\mathcal{H}} = K(x,t)$ for all $x, t \in X$, where $\Phi_x = \Phi(x)$. The evaluation functional $L_x : \mathcal{H} \to \mathbb{R}$ given by $L_x v = \langle v, \Phi_x\rangle_{\mathcal{H}}$, where x varies over X, defines an inclusion map

$$L_\Phi : \mathcal{H} \to \mathbb{R}^X \quad (L_\Phi v)(x) = \langle v, \Phi_x\rangle_{\mathcal{H}}$$

where $\mathbb{R}^X$ denotes the vector space of pointwise-defined, real-valued functions on X. Observe that as a vector space of functions, $\mathcal{H}_K \subset \mathbb{R}^X$.

Proposition 1. *Let $\mathcal{H}_\Phi = \overline{\text{span}\{\Phi_x : x \in X\}}$, a subspace of $\mathcal{H}$. The restriction of L_Φ on $\mathcal{H}_\Phi$ is an isometric isomorphism between $\mathcal{H}_\Phi$ and $\mathcal{H}_K$.*

Proof. First, L_Φ is bijective from $\mathcal{H}_\Phi$ to the image $L_\Phi(\mathcal{H}_\Phi)$, since $\ker L_\Phi = \mathcal{H}_\Phi^\perp$. Under the map L_Φ, for each $x, t \in X$, $(L_\Phi \Phi_x)(t) = \langle \Phi_x, \Phi_t\rangle = K_x(t)$, thus $K_x \equiv L_\Phi \Phi_x$ as functions on X. This implies that $\text{span}\{K_x : x \in X\}$ is isomorphic to $\text{span}\{\Phi_x : x \in X\}$ as vector spaces. The isometric isomorphism of $\mathcal{H}_\Phi = \overline{\text{span}\{\Phi_x : x \in X\}}$ and $\mathcal{H}_K = \overline{\text{span}\{K_x : x \in X\}}$ then follows from $\langle \Phi_x, \Phi_t\rangle_{\mathcal{H}} = K(x,t) = \langle K_x, K_t\rangle_K$. This completes the proof.

Remark 5. Each choice of Φ is thus equivalent to a factorization of the map $\Phi_K : x \to K_x \in \mathcal{H}_K$, that is the following diagram is commutative

$$\begin{tikzcd} x \in X \arrow[rr, "\Phi_K"] \arrow[dr, "\Phi"'] & & K_x \in \mathcal{H}_K \\ & \Phi_x \in \mathcal{H}_\Phi \arrow[ur, "L_\Phi"'] & \end{tikzcd} \qquad (19)$$

We will call $\Phi_K : x \to K_x \in \mathcal{H}_K$ the *canonical feature map* associated with K.

5 Smoothing Properties of Kernels on the Sphere

Having discussed feature maps, we will in this section analyze the smoothing properties of the polynomial and Gaussian kernels and compare them with those of spline kernels on the sphere S^{n-1}. In the spline smoothing problem on S^1 as described in [11], one solves the minimization problem

$$\frac{1}{m}\sum_{i=1}^{m}(f(x_i)-y_i)^2+\lambda\int_0^{2\pi}(f^{(m)}(t))^2dt \tag{20}$$

for $x_i \in [0,2\pi]$ and $f \in W_m$, where $J_m(f)=\int_0^{2\pi}(f^{(m)}(t))^2dt$ is the square norm of the RKHS

$$W_m^0=\{f: f, f^{'}, \ldots, f^{(m-1)} \text{ absolutely continuous}, f^{(m)} \in L^2[0,2\pi], \\ f^{(j)}(0)=f^{(j)}(2\pi), j=0,1,\ldots,m-1\}$$

The space W_m is then the RKHS defined by

$$W_m=\{1\}\oplus W_m^0=\{f: ||f||_K^2=\tfrac{1}{4\pi^2}\left(\int_0^{2\pi}f(t)dt\right)^2+\int_0^{2\pi}(f^{(m)}(t))^2dt<\infty\}$$

induced by a kernel K. *One particular feature of spline smoothing, on S^1, S^2, or $\mathbb{R}^n$, is that in general the RKHS W_m does not have a closed form kernel K that is efficiently computable.* This is in contrast with the RKHS that are used in kernel machine learning, all of which correspond to closed-form kernels that can be evaluated efficiently. It is not clear, however, whether the norms in these RKHS correspond to smoothness functionals. In this section, we will show that for the polynomial and Gaussian kernels on S^{n-1}, they do.

5.1 The Iterated Laplacian and Splines on the Sphere S^{n-1}

Splines on S^{n-1} for $n=2$ and $n=3$, as treated by Wahba [11], [12], can be generalized to any $n \geq 2$, $n \in \mathbb{N}$, via the iterated Laplacian (also called the Laplace-Beltrami operator) on S^{n-1}. The RKHS corresponding to W_m in (20) is a subspace of $L^2(S^{n-1})$ described by

$$\mathcal{H}_K=\{f: ||f||_K^2=\frac{1}{|S^{n-1}|^2}\left(\int_{S^{n-1}}f(x)dx\right)^2+\int_{S^{n-1}}f(x)\Delta^m f(x)dx<\infty\}$$

The Laplacian Δ on S^{n-1} has eigenvalues $\lambda_k=k(k+n-2)$, $k \geq 0$, with corresponding eigenfunctions $\{Y_{k,j}(n;x)\}_{j=1}^{N(n,k)}$, which form an orthonormal basis in the space $\mathcal{Y}_k(n)$ of spherical harmonics of order k . For $f \in L^2(S^{n-1})$, if we use the expansion $f=\frac{a_0}{\sqrt{|S^{n-1}|}}+\sum_{k=1}^{\infty}\sum_{j=1}^{N(n,k)}a_{k,j}Y_{k,j}(n;x)$ then the space $\mathcal{H}_K$ takes the form

$$\mathcal{H}_K=\{f \in L^2(S^{n-1}): ||f||_K^2=\frac{a_0^2}{|S^{n-1}|}+\sum_{k=1}^{\infty}[k(k+n-2)]^m\sum_{j=1}^{N(n,k)}a_{k,j}^2<\infty\}$$

and thus the corresponding kernel is

$$K(x,t) = 1 + \sum_{k=1}^{\infty} \frac{1}{[k(k+n-2)]^m} \sum_{j=1}^{N(n,k)} Y_{k,j}(n;x)Y_{k,j}(n;t) \tag{21}$$

which is well-defined iff $m > \frac{n-1}{2}$. Let $P_k(n;t)$ denote the Legendre polynomial of degree k in dimension n, then (21) takes the form

$$K(x,t) = 1 + \frac{1}{|S^{n-1}|} \sum_{k=1}^{\infty} \frac{N(n,k)}{[k(k+n-2)]^m} P_k(n;\langle x,t\rangle) \tag{22}$$

which does not have a closed form in general - for the case $n = 3$, see [11].

Remark 6. Clearly m can be replaced by any real number $s > \frac{n-1}{2}$.

5.2 Smoothing Properties of Polynomial and Gaussian Kernels

Let ∇^*_{n-1} denote the gradient operator on S^{n-1} (also called the first-order Beltrami operator, see [6] page 79 for a definition). Let $Y_k \in \mathcal{Y}_k(n)$, $k \geq 0$, then

$$||\nabla^*_{n-1} Y_k||^2_{L^2(S^{n-1})} = \int_{S^{n-1}} |\nabla^*_{n-1} Y_k(x)|^2 dS^{n-1}(x) = k(k+n-2) \tag{23}$$

This shows that spherical harmonics of higher-order are less smooth. This is particularly straightforward in the case $n = 2$ with the Fourier basis functions $\{1, \cos k\theta, \sin k\theta\}_{k\in\mathbb{N}}$ - as k increases, the functions oscillate more rapidly.

It follows that any regularization term $||f||^2_K$ in problems such as (20), where K possesses a decreasing spectrum λ_k - k corresponds to the order of the spherical harmonics - will have a smoothing effect. That is, the higher-order spherical harmonics, which are less smooth, will be penalized more. The decreasing spectrum property is true for the spline kernels, the polynomial kernel $(1 + \langle x,t\rangle)^d$, and the Gaussian kernel for $\sigma \geq (\frac{2}{n})^{1/2}$, as we showed in Theorems 2 and 3. Hence all these kernels possess smoothing properties on S^{n-1}.

Furthermore, Theorem 2 shows that for the Gaussian kernel, for all $k \geq 1$

$$(\frac{2e}{\sigma^2})^k \frac{A_1}{(2k+n-2)^{k+\frac{n-1}{2}}} < \lambda_k < (\frac{2e}{\sigma^2})^k \frac{A_2}{(2k+n-2)^{k+\frac{n-1}{2}}}$$

and Theorem 3 shows that for the polynomial kernel $(1 + \langle x,t\rangle)^d$

$$\frac{B_1}{(k+d+n-2)^{2d+n-\frac{3}{2}}} < \lambda_k < \frac{B_2}{(k+d+n-2)^{d+n-\frac{3}{2}}}$$

for $0 \leq k \leq d$. Compare these with the eigenvalues of the spline kernels

$$\lambda_k = \frac{1}{[k(k+n-2)]^m}$$

for $k \geq 1$, we see that the Gaussian kernel has the sharpest smoothing property, as can be seen from the exponential decay of the eigenvalues.

For $K(x,t) = (1+\langle x,t\rangle)^d$, if $d > 2m - n + \frac{3}{2}$, then K has sharper smoothing property than a spline kernel of order m. Moreover, all spherical harmonics of order greater than d are filtered out, hence choosing K amounts to choosing a hypothesis space of bandlimited functions on S^{n-1}.

A Proofs of Results

The proofs for results on S^{n-1} all make use of properties of spherical harmonics on S^{n-1}, which can be found in [6]. We will prove Theorem 2 (Theorem 3 is similar) and Theorem 5.

A.1 Proof of Theorem 2

Let $f : [-1,1] \to \mathbb{R}$ be a continuous function. Let $Y_k \in \mathcal{Y}_k(n)$ for $k \geq 0$. Then the Funk-Hecke formula ([6], page 30) states that for any $x \in S^{n-1}$:

$$\int_{S^{n-1}} f(\langle x,t\rangle) Y_k(t) dS^{n-1}(t) = \lambda_k Y_k(x) \tag{24}$$

where

$$\lambda_k = |S^{n-2}| \int_{-1}^{1} f(t) P_k(n;t)(1-t^2)^{\frac{n-3}{2}} dt \tag{25}$$

and $P_k(n;t)$ denotes the Legendre polynomial of degree k in dimension n. Since the spherical harmonics $\{\{Y_{k,j}(n;x)\}_{j=1}^{N(n,k)}\}_{k=0}^{\infty}$ form an orthonormal basis for $L^2(S^{n-1})$, an immediate consequence of the Funk-Hecke formula is that if K on $S^{n-1} \times S^{n-1}$ is defined by $K(x,t) = f(\langle x,t\rangle)$, and μ is the Lebesgue measure on S^{n-1}, then the eigenvalues of $L_K : L^2_\mu(S^{n-1}) \to L^2_\mu(S^{n-1})$ are given precisely by (25), with the corresponding orthonormal eigenfunctions of λ_k being $\{Y_{k,j}(n;x)\}_{j=1}^{N(n,k)}$. The multiplicity of λ_k is therefore $N(n,k) = \dim(\mathcal{Y}_k(n))$.

On S^{n-1} $e^{-\frac{||x-t||^2}{\sigma^2}} = e^{-\frac{2}{\sigma^2}} e^{\frac{2\langle x,t\rangle}{\sigma^2}}$. Thus

$\lambda_k = e^{-\frac{2}{\sigma^2}} |S^{n-2}| \int_{-1}^{1} e^{\frac{2t}{\sigma^2}} P_k(n;t)(1-t^2)^{\frac{n-3}{2}} dt$

$= e^{-\frac{2}{\sigma^2}} |S^{n-2}| \sqrt{\pi} \Gamma(\frac{n-1}{2})(\sigma^2)^{n/2-1} I_{k+n/2-1}(\frac{2}{\sigma^2})$ by Lemma 1

$= e^{-2/\sigma^2} \sigma^{n-2} I_{k+n/2-1}(\frac{2}{\sigma^2}) \Gamma(\frac{n}{2}) |S^{n-1}|$

Normalizing by setting $|S^{n-1}| = 1$ gives the required expression for λ_k as in (5).

Lemma 1. *Let $f(t) = e^{rt}$, then*

$$\int_{-1}^{1} f(t) P_k(n;t)(1-t^2)^{\frac{n-3}{2}} dt = \sqrt{\pi}\Gamma(\frac{n-1}{2}) \left(\frac{2}{r}\right)^{n/2-1} I_{k+n/2-1}(r) \tag{26}$$

Proof. We apply the following which follows from ([13], page 79, formula 9)

$$\int_{-1}^{1} e^{rt}(1-t^2)^{\nu-1}dt = \sqrt{\pi}\left(\frac{2}{r}\right)^{\nu-1/2}\Gamma(\nu)I_{\nu-1/2}(r) \tag{27}$$

and Rodrigues' rule ([6], page 23), which states that for $f \in C^k([-1,1])$

$$\int_{-1}^{1} f(t)P_k(n;t)(1-t^2)^{\frac{n-3}{2}}dt = R_k(n)\int_{-1}^{1} f^{(k)}(t)(1-t^2)^{k+\frac{n-3}{2}}dt \tag{28}$$

where $R_k(n) = \frac{1}{2^k}\frac{\Gamma(\frac{n-1}{2})}{\Gamma(k+\frac{n-1}{2})}$. For $f(t) = e^{rt}$, we have

$\int_{-1}^{1} e^{rt}P_k(n;t)(1-t^2)^{\frac{n-3}{2}}dt = R_k(n)r^k\int_{-1}^{1} e^{rt}(1-t^2)^{k+\frac{n-3}{2}}$

$= R_k(n)r^k\sqrt{\pi}\left(\frac{2}{r}\right)^{k+n/2-1}\Gamma(k+\frac{n-1}{2})I_{k+n/2-1}(r)$

Substituting in the values of $R_k(n)$ gives the desired answer. □

Lemma 2. *The sequence* $\{\lambda_k\}_{k=0}^{\infty}$ *is decreasing if* $\sigma \geq \left(\frac{2}{n}\right)^{1/2}$.

Proof. We will first prove that $\frac{\lambda_k}{\lambda_{k+1}} > (k+n/2)\sigma^2$. We have

$I_{k+n/2}(\frac{2}{\sigma^2}) = (\frac{1}{\sigma^2})^{k+n/2}\sum_{j=0}^{\infty}\frac{(\frac{1}{\sigma^2})^{2j}}{j!\Gamma(j+k+n/2+1)}$

$= (\frac{1}{\sigma^2})^{k+n/2}\sum_{j=0}^{\infty}\frac{(\frac{1}{\sigma^2})^{2j}}{j!(j+k+n/2)\Gamma(j+k+n/2)}$

$< (\frac{1}{\sigma^2})^{k+n/2}\frac{1}{k+n/2}\sum_{j=0}^{\infty}\frac{(\frac{1}{\sigma^2})^{2j}}{j!\Gamma(j+k+n/2)} = \frac{1}{\sigma^2(k+n/2)}I_{k+n/2-1}(\frac{2}{\sigma^2})$

which implies $\frac{\lambda_k}{\lambda_{k+1}} > (k+n/2)\sigma^2$. The inequality $\frac{\lambda_k}{\lambda_{k+1}} \geq 1$ thus is satisfied if $\sigma^2(k+n/2) \geq 1$ for all $k \geq 0$. It suffices to require that it holds for $k = 0$, that is $\sigma^2 n/2 \geq 1 \iff \sigma \geq \left(\frac{2}{n}\right)^{1/2}$. □

*Proof (**of (6)**).* By definition of $I_\nu(z)$, we have for $z > 0$

$$I_\nu(z) < \frac{(\frac{z}{2})^\nu}{\Gamma(\nu+1)}\sum_{j=0}^{\infty}\frac{(\frac{z}{2})^{2j}}{j!} = \frac{(\frac{z}{2})^\nu}{\Gamma(\nu+1)}e^{z^2/4}$$

Then for $\nu = k+\frac{n}{2}-1$ and $z = \frac{2}{\sigma^2}$: $I_{k+\frac{n}{2}-1}(\frac{2}{\sigma^2}) < \frac{1}{\Gamma(k+\frac{n}{2})}(\frac{1}{\sigma})^{2k+n-2}e^{1/\sigma^4}$. Consider Stirling's series for $a > 0$

$$\Gamma(a+1) = \sqrt{2\pi a}\left(\frac{a}{e}\right)^a\left[1+\frac{1}{12a}+\frac{1}{288a^2}-\frac{139}{51840a^3}+\ldots\right] \tag{29}$$

Thus for all $a > 0$ we can write $\Gamma(a+1) = e^{A(a)}\sqrt{2\pi e}\left(\frac{a}{e}\right)^{a+\frac{1}{2}}$ where $0 < A(a) < \frac{1}{12a}$. Hence for all $k \geq 1$

$$\Gamma(k+\frac{n}{2}) = e^{A(k,n)}\sqrt{2\pi e}(\frac{k+\frac{n}{2}-1}{e})^{k+\frac{n-1}{2}} = e^{A(k,n)}\sqrt{2\pi e}(\frac{2k+n-2}{2e})^{k+\frac{n-1}{2}}$$

where $0 < A(k,n) < \frac{1}{12(k+\frac{n}{2}-1)} \leq \frac{1}{12}$. Then

$$I_{k+\frac{n}{2}-1}(\frac{2}{\sigma^2}) < \frac{1}{\sqrt{\pi}}\frac{(2e)^{k+\frac{n}{2}-1}}{(2k+n-2)^{k+\frac{n-1}{2}}}(\frac{1}{\sigma})^{2k+n-2}e^{1/\sigma^4} \text{ implying (6).}$$

The other direction is obtained similarly. □

A.2 Proof of Theorem 5

We will first show an upper bound for $||Y_k||_\infty$, where Y_k is any $L^2(S^{n-1})$-normalized function in $\mathcal{Y}_k(n)$, then exhibit a one-dimensional subspace of functions in $\mathcal{Y}_k(n)$ that attain this upper bound. Observe that Y_k belongs to an orthonormal basis $\{Y_{k,j}(n;x)\}_{j=1}^{N(n,k)}$ of $\mathcal{Y}_k(n)$. The following highlights the crucial difference between the case $n = 2$ and $n \geq 3$.

Lemma 3. *For any $n \geq 2$, $k \geq 0$, for all $j \in \mathbb{N}$, $1 \leq j \leq N(n,k)$*

$$||Y_{k,j}(n;.)||_\infty \leq \sqrt{\frac{N(n,k)}{|S^{n-1}|}} \tag{30}$$

In particular, for $n = 2$ and all $k \geq 0$: $||Y_{k,j}(n;.)||_\infty \leq \frac{1}{\sqrt{\pi}}$.

Proof. The Addition Theorem for spherical harmonics ([6], page 18) states that for any $x, \alpha \in S^{n-1}$

$$\sum_{j=1}^{N(n,k)} Y_{k,j}(n;x)\overline{Y_{k,j}(n;\alpha)} = \frac{N(n,k)}{|S^{n-1}|} P_k(n;\langle x,\alpha\rangle)$$

which implies that for any $x \in S^{n-1}$

$$|Y_{k,j}(n;x)|^2 \leq \frac{N(n,k)}{|S^{n-1}|} P_k(n;\langle x,x\rangle) = \frac{N(n,k)}{|S^{n-1}|} P_k(n;1) = \frac{N(n,k)}{|S^{n-1}|}$$

giving us the first result. For $n = 2$, we have $N(n,k) = 1$ for $k = 0$, $N(n,k) = 2$ for $k \geq 1$, and $|S^1| = 2\pi$, giving us the second result. □

Definition 2. *Consider the group $O(n)$ of all orthogonal transformations in $\mathbb{R}^n$, that is $O(n) = \{A \in \mathbb{R}^{n\times n} : A^T A = AA^T = I\}$. A function $f : S^{n-1} \to \mathbb{R}$ is said to be* ***invariant*** *under a transformation $A \in O(n)$ if $f_A(x) = f(Ax) = f(x)$ for all $x \in S^{n-1}$. Let $\alpha \in S^{n-1}$. The* ***isotropy*** *group $J_{n,\alpha}$ is defined by $J_{n,\alpha} = \{A \in O(n) : A\alpha = \alpha\}$.*

Lemma 4. *Assume that $Y_k \in \mathcal{Y}_k(n)$ is invariant with respect to $J_{n,\alpha}$ and satisfies $\int_{S^{n-1}} |Y_k(x)|^2 dS^{n-1}(x) = 1$. Then Y_k is unique up to a multiplicative constant $C_{\alpha,n,k}$ with $|C_{\alpha,n,k}| = 1$ and*

$$||Y_k||_\infty = |Y_k(\alpha)| = \sqrt{\frac{N(n,k)}{|S^{n-1}|}} \tag{31}$$

Proof. If Y_k is invariant with respect to $J_{n,\alpha}$, then by ([6], Lemma 3, page 17), it must satisfy $Y_k(x) = Y_k(\alpha)P_k(n;\langle x,\alpha\rangle)$, showing that the subspace of $\mathcal{Y}_k(n)$ invariant with respect to $J_{n,\alpha}$ is one-dimensional. The Addition Theorem implies that for any $\alpha \in S^{n-1}$

$$\int_{S^{n-1}}|P_k(n;\langle x,\alpha\rangle)|^2dS^{n-1}(x)=\tfrac{|S^{n-1}|}{N(n,k)}$$

By assumption, we then have

$1=\int_{S^{n-1}}|Y_k(x)|^2dS^{n-1}(x)=|Y_k(\alpha)|^2\int_{S^{n-1}}|P_k(n;\langle x,\alpha\rangle)|^2dS^{n-1}(x)$

$=|Y_k(\alpha)|^2\frac{|S^{n-1}|}{N(n,k)}$, giving us $|Y_k(\alpha)|=\sqrt{\frac{N(n,k)}{|S^{n-1}|}}$. Thus we for all $x\in S^{n-1}$

$$|Y_k(x)|=\sqrt{\tfrac{N(n,k)}{|S^{n-1}|}}|P_k(n;\langle x,\alpha\rangle)|\le\sqrt{\tfrac{N(n,k)}{|S^{n-1}|}}$$

by the property $|P_k(n;t)|\le 1$ for $|t|\le 1$. Thus $||Y_k||_\infty=\sqrt{\frac{N(n,k)}{|S^{n-1}|}}$ as desired. □

Proposition 2 (Orthonormal basis of $\mathcal{Y}_k(n)$ [6]). *Let $n\ge 3$. Let $e_1,\dots,e_n$ be the canonical basis of $\mathbb{R}^n$. Let $x\in S^{n-1}$. We write $x=te_n+\sqrt{1-t^2}\begin{pmatrix}x_{(n-1)}\\0\end{pmatrix}$ where $t\in[-1,1]$ and $x_{(n-1)}\in S^{n-2}$, $(x_{(n-1)},0)^T\in span\{e_1,\dots,e_{n-1}\}$. Suppose that for $m=0,1,\dots,k$, the orthonormal bases $Y_{m,j}$, $j=1,\dots,N(n-1,m)$ of $\mathcal{Y}_m(n-1)$ are given, then an orthonormal basis for $\mathcal{Y}_k(n)$ is*

$$Y_{k,m,j}(n;x)=A_k^m(n;t)Y_{m,j}(n-1;x_{(n-1)}):j=1,2\dots,N(n-1,m)\qquad(32)$$

starting with the Fourier basis for $n=2$, where

$$A_k^m(n;t)=\frac{\sqrt{2^{2-n}(2k+n-2)(k-m)!(k+n+m-3)!}}{k!\Gamma(\frac{n-1}{2})}P_k^m(n;t)\qquad(33)$$

Proposition 3. *Let $n\in\mathbb{N}$, $n\ge 3$. Let μ be the Lebesgue measure on S^{n-1}. For each $k\ge 0$, any orthonormal basis of the space $\mathcal{Y}_k(n)$ of spherical harmonics of order k contains an L^2_μ-normalized spherical harmonic Y_k such that*

$$||Y_k||_\infty=\sqrt{\frac{N(n,k)}{|S^{n-1}|}}=\sqrt{\frac{(2k+n-2)(k+n-3)!}{k!(n-2)!|S^{n-1}|}}\to\infty\qquad(34)$$

as $k\to\infty$, where $|S^{n-1}|=\frac{2\pi^{\frac{n}{2}}}{\Gamma(\frac{n}{2})}$ is the surface area of S^{n-1}.

Proof. Let $x=te_n+\sqrt{1-t^2}\begin{pmatrix}x_{(n-1)}\\0\end{pmatrix}$, $-1\le t\le 1$. For each $k\ge 0$, the orthonormal basis for $\mathcal{Y}_k(n)$ in Proposition 2 contains the function

$$Y_{k,0,1}(n;x)=A_k^0(t)Y_{0,1}(n-1;x_{(n-1)})=A_k^0(t)\frac{1}{\sqrt{|S^{n-2}|}}\qquad(35)$$

$$Y_{k,0,1}(n;x)=\tfrac{1}{\Gamma(\frac{n-1}{2})}\sqrt{\tfrac{(2k+n-2)(k+n-3)!}{2^{n-2}k!|S^{n-2}|}}P_k(n;t)=\sqrt{\tfrac{N(n,k)}{|S^{n-1}|}}P_k(n;t)$$

Then $Y_{k,0,1}(n;x)$ is invariant with respect to $J_{n,\alpha}$ where $\alpha=(0,\dots,0,1)$. Thus $Y_k=Y_{k,0,1}$ is the desired function for the current orthonormal basis. For any orthonormal basis of $\mathcal{Y}_k(n)$, the result follows by Lemma 4 and rotational symmetry on the sphere. □

*Proof (**of Theorem 5**).* By the Funk-Hecke formula, all spherical harmonics of order k are eigenfunctions corresponding to the eigenvalue λ_k as given by (25). If infinitely many of the λ_k's are nonzero, then the corresponding set of $L^2(S^{n-1})$-orthonormal eigenfunctions $\{\phi_k\}$, being an orthonormal basis of $L^2(S^{n-1})$, contains a spherical harmonic Y_k satisfying (34), for infinitely many k. It follows from Proposition 3 then that $\sup_k ||\phi_k||_\infty = \infty$. □

References

1. N. Aronszajn. Theory of Reproducing Kernels. *Transactions of the American Mathematical Society*, vol. 68, pages 337-404, 1950.
2. M. Belkin, P. Niyogi, and V. Sindwani. Manifold Regularization: a Geometric Framework for Learning from Examples. University of Chicago Computer Science Technical Report TR-2004-06, 2004, accepted for publication.
3. M. Belkin and P. Niyogi. Semi-supervised Learning on Riemannian Manifolds. *Machine Learning*, Special Issue on Clustering, vol. 56, pages 209-239, 2004.
4. E. De Vito, A. Caponnetto, and L. Rosasco. Model Selection for Regularized Least-Squares Algorithm in Learning Theory. *Foundations of Computational Mathematics*, vol. 5, no. 1, pages 59-85, 2005.
5. J. Lafferty and G. Lebanon. Diffusion Kernels on Statistical Manifolds. *Journal of Machine Learning Research*, vol. 6, pages 129-163, 2005.
6. C. Müller. *Analysis of Spherical Symmetries in Euclidean Spaces.* Applied Mathematical Sciences 129, Springer, New York, 1997.
7. B. Schölkopf and A.J. Smola. *Learning with Kernels.* The MIT Press, Cambridge, Massachusetts, 2002.
8. S. Smale and D.X. Zhou. Learning Theory Estimates via Integral Operators and Their Approximations, 2005, to appear.
9. A.J. Smola, Z.L. Ovari and R.C. Williamson. Regularization with Dot-Product Kernels. *Advances in Information Processing Systems*, 2000.
10. I. Steinwart, D. Hush, and C. Scovel. An Explicit Description of the Reproducing Kernel Hilbert Spaces of Gaussian RBF Kernels Kernels. Los Alamos National Laboratory Technical Report LA-UR-04-8274, December 2005.
11. G. Wahba. Spline Interpolation and Smoothing on the Sphere. *SIAM Journal of Scientific and Statistical Computing*, vol. 2, pages 5-16, 1981.
12. G. Wahba. *Spline Models for Observational Data.* CBMS-NSF Regional Conference Series in Applied Mathematics 59, Society for Industrial and Applied Mathematics, Philadelphia, 1990.
13. G.N. Watson. *A Treatise on the Theory of Bessel Functions*, 2nd edition, Cambridge University Press, Cambridge, England, 1944.
14. D.X. Zhou. The Covering Number in Learning Theory. *Journal of Complexity*, vol. 18, pages 739-767, 2002.