An Explicit Description of the Extended Gaussian Kernel

Yong Liu and Shizhong Liao

School of Computer Science and Technology
Tianjin University,
Tianjin, China
szliao@tju.edu.cn

Abstract. Kernel methods play an important role in machine learning, pattern recognition and data mining. Although the kernel functions are the central part of the kernel methods, little is known about the structure of its reproducing kernel Hilbert spaces (RKHS) and the eigenvalues of the integral operator. In this paper, we first give the definition of the extended Gaussian kernel which includes the Gaussian kernel as its special case. Then, through a generalization form of the Weyl inner product, we present an explicit description of the RKHS of the extended Gaussian kernel. Furthermore, using the Funk-Hecke formula, we get the eigenvalues and eigenfunctions of the integral operator on the unit sphere.

Keywords: Integral operator, Reproducing kernel Hilbert space, Extended Gaussian kernel, Eigenvalues.

1 Introduction

The reproducing kernel Hilbert space (RKHS) and the eigenvalues of the integral operator recently have attracted more and more attentions in machine learning and data mining (comprehensive treatments are found in [15,18,9,16,12]). It is thus of crucial importance, for both practical and theoretical purposes, to have a deep understanding of the RKHS and the eigenvalues of the integral operator. Steinwart et al [13] first studied the structure of the RKHS induced by the popular Gaussian kernel, and they presented an orthonormal basis for this space. Minh [6] also discussed the RKHS of the Gaussian kernel and its orthonormal basis. Scovel et al [11] developed a general theory regarding mixtures of kernels, and analyzed the RKHS of the mixture in terms of the RKHSs of the mixture components. Sun and Zhou [14] explored the RKHS associated with the translation-invariant Mercer kernels, and derived some estimates for the covering numbers which form an essential part for the analysis of some algorithms in the learning theory. Kadri et al [5] explored the potential of adopting an operator-valued kernel feature space perspective for the analysis of functional data. Ferreira and Manegatto [3,4] analyzed the reproducing kernel Hilbert spaces of positive definite kernels on a topological space.

T. Washio and J. Luo (Eds.): PAKDD 2012 Workshops, LNAI 7769, pp. 88–99, 2013.

In this paper, we generalize the results associated with the Gaussian kernel [13,6] to general kernel, namely as the extended Gaussian kernel. Compared to the Gaussian kernel, the extended Gaussian kernel can be used to solve the problems where the input data need to be scaled. In addition, we also present an explicit description for the eigenvalues and the eigenfunctions of the integral operator on the unit sphere, which can be used in the theoretical analysis of kernel principal component analysis [8] and other methods that need eigenvalue and eigenfunction.

The contribution of our paper mainly consists of two aspects:

- An explicit description of the RKHS with its orthonormal basis induced by the extended Gaussian kernel.
- An explicit description of the eigenvalues and the eigenfunctions of the integral operator associated with the extended Gaussian kernel on the unit sphere.

The rest of the paper is organized as following. In Section 2, we introduce the basic facts on an RKHS, In Section 3, we define the extended Gaussian kernel and present our main results, i.e., the explicit description of the RKHS and the eigenvalues of the extended Gaussian kernel. We conclude this paper in Section 4.

2 Preliminaries

Let $\mathcal{X}$ be a nonempty set. A function K is called a kernel on $\mathcal{X}$ if there exists a Hilbert space $\mathcal{H}$ and a map $\Phi : \mathcal{X} \to \mathcal{H}$ such that for all $\boldsymbol{x}, \boldsymbol{x}' \in \mathcal{X}$ we have

$$K(\boldsymbol{x}, \boldsymbol{x}') = \langle \Phi(\boldsymbol{x}'), \Phi(\boldsymbol{x}) \rangle.$$

We call Φ a feature map and $\mathcal{H}$ a feature space of K. For any finite set of points $\{\boldsymbol{x}_i\}_{i=1}^N$ in $\mathcal{X}$ and $\{a_i \in \mathbb{R}\}_{i=1}^N$, if

$$\sum_{i,j=1}^{N} a_i a_j K(\boldsymbol{x}_i, \boldsymbol{x}_j) \geq 0,$$

then the function K is said to be positive definite kernel on $\mathcal{X}$.

For a given kernel, neither the feature map nor the feature space is uniquely determined. However, one can always construct a canonical feature space, namely, the reproducing kernel Hilbert space (RKHS). Let us now recall the basic theory of this space [1].

Definition 1. *Let $\mathcal{X}$ be a nonempty set and $\mathcal{H}$ be a Hilbert function space over $\mathcal{X}$, i.e., a Hilbert space that consists of functions mapping from $\mathcal{X}$ into $\mathbb{R}$.*

1. *The space* $\mathcal{H}$ *is called a reproducing kernel Hilbert space (RKHS) if for all* $\boldsymbol{x} \in \mathcal{X}$ *the Dirac functional* $\delta_{\boldsymbol{x}} : \mathcal{H} \to \mathbb{R}$ *defined by* $\delta_{\boldsymbol{x}}(f) := f(\boldsymbol{x}), f \in \mathcal{H}$, *is continuous.*
2. *A function* $K : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ *is called a reproducing kernel of* $\mathcal{H}$ *if we have* $K(\cdot, \boldsymbol{x}) \in \mathcal{H}$ *for all* $\boldsymbol{x} \in \mathcal{X}$ *and the reproducing property*

$$f(\boldsymbol{x}) = \langle f, K(\cdot, \boldsymbol{x}) \rangle$$

holds for all $f \in \mathcal{H}$ *and all* $\boldsymbol{x} \in \mathcal{X}$.

A Hilbert function space $\mathcal{H}$ that has a reproducing kernel K is always an RKHS. Vice versa, i.e., every RKHS has a (unique) reproducing kernel (see [10]).

3 Main Results

In this section, we will first give the definition of the extended Gaussian kernel, and then we will present an explicit description of the RKHS and the eigenvalues of the integral operator associated with the extended Gaussian kernel.

3.1 Extended Gaussian Kernel

For a multi-index $\boldsymbol{x} = (x_1, \ldots, x_d)^{\mathrm{T}} \in \mathbb{R}^d$, if $\boldsymbol{b} = \{b_1, \ldots, b_d\}^{\mathrm{T}} \in \mathbb{R}^d$, we write $\boldsymbol{x}^{(\boldsymbol{b})} = (x_1^{b_1}, \ldots, x_d^{b_d})^{\mathrm{T}}$, if $b \in \mathbb{R}$, we write $\boldsymbol{x}^{[b]} = (x_1^b, \ldots, x_d^b)^{\mathrm{T}}$.

Definition 2 (Extended Gaussian Kernel). *Let* $\mathcal{X} \subset \mathbb{R}^d$ *be a nonempty set. For* $\boldsymbol{b} \in \mathbb{R}^d$, *the extended Gaussian kernel* $K_{\boldsymbol{b}} : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ *is written as*

$$K_{\boldsymbol{b}}(\boldsymbol{x}, \boldsymbol{z}) := \exp\left(-\frac{\|\boldsymbol{x}^{(\boldsymbol{b})} - \boldsymbol{z}^{(\boldsymbol{b})}\|^2}{\sigma^2}\right).$$

For $b \in \mathbb{R}$, *the extended Gaussian kernel* $K_b : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ *is written as*

$$K_b(\boldsymbol{x}, \boldsymbol{z}) := \exp\left(-\frac{\|\boldsymbol{x}^{[b]} - \boldsymbol{z}^{[b]}\|^2}{\sigma^2}\right),$$

where $\sigma > 0$.

Remark 1. According to the definition of the extended Gaussian kernel, we know that the popular Gaussian kernel is a special case of the extended Gaussian kernel (when $\boldsymbol{b} = \{1, \ldots, 1\}^{\mathrm{T}}$ for $\boldsymbol{b} \in \mathbb{R}^d$ or $b = 1$ for $b \in \mathbb{R}$), thus the results associated with the extended Gaussian kernels can be easily applied to the Gaussian kernel. Moreover, in practice, the input data need to be scaled, so the extended Gaussian kernel with an advisable value of $\boldsymbol{b}$ may be more useful than the Gaussian kernel.

3.2 RKHS of Extended Gaussian Kernel

Let

$$\begin{aligned}\boldsymbol{b} &= (b_1, \ldots, b_d)^{\mathrm{T}} \in \mathbb{R}^d, d \in \mathbb{N};\\ \boldsymbol{\alpha} &= (\alpha_1, \ldots, \alpha_d)^{\mathrm{T}} \in (\mathbb{N} \cup \{0\})^d;\\ |\boldsymbol{\alpha}| &= \sum_{i=1}^{d} \alpha_i;\\ \boldsymbol{x}^{\boldsymbol{\alpha}} &= \prod_{i=1}^{d} x_i^{\alpha_i};\\ \boldsymbol{x}^{\boldsymbol{b},\boldsymbol{\alpha}} &= \prod_{i=1}^{d} x_i^{\alpha_i b_i}.\end{aligned}$$

We show the RKHS $\mathcal{H}_{\boldsymbol{b}}$ of the extended Gaussian kernel $K_{\boldsymbol{b}}$ in the following theorem.

Theorem 1. *Let $\mathcal{X} \subset \mathbb{R}^d$ be a nonempty set, for every $\sigma > 0$, $\boldsymbol{b} \in \mathbb{R}^d$. Then the extended Gaussian kernel $K_{\boldsymbol{b}}(\boldsymbol{x}, \boldsymbol{z}) = \exp\left(-\frac{\|\boldsymbol{x}^{(\boldsymbol{b})} - \boldsymbol{z}^{(\boldsymbol{b})}\|^2}{\sigma^2}\right)$ is the reproducing kernel of the space*

$$\mathcal{H}_{\boldsymbol{b}} = \left\{ f = e^{-\frac{\|\boldsymbol{x}^{(\boldsymbol{b})}\|^2}{\sigma^2}} \sum_{|\boldsymbol{\alpha}|=0}^{\infty} w_{\boldsymbol{\alpha}} \boldsymbol{x}^{\boldsymbol{b},\boldsymbol{\alpha}} : \|f\|_K^2 < \infty \right\}, \tag{1}$$

where the inner product $\langle \cdot, \cdot \rangle_K$ on $\mathcal{H}_{\boldsymbol{b}}$ is given by

$$\langle f, g \rangle_K = \sum_{k=0}^{\infty} \frac{k!}{(2/\sigma^2)^k} \sum_{|\boldsymbol{\alpha}|=k} \frac{w_{\boldsymbol{\alpha}} \nu_{\boldsymbol{\alpha}}}{C_{\boldsymbol{\alpha}}^k}$$

for

$$\begin{aligned} f &= e^{-\frac{\|\boldsymbol{x}^{(\boldsymbol{b})}\|^2}{\sigma^2}} \sum_{|\boldsymbol{\alpha}|=0}^{\infty} w_{\boldsymbol{\alpha}} \boldsymbol{x}^{\boldsymbol{b},\boldsymbol{\alpha}},\\ g &= e^{-\frac{\|\boldsymbol{x}^{(\boldsymbol{b})}\|^2}{\sigma^2}} \sum_{|\boldsymbol{\alpha}|=0}^{\infty} \nu_{\boldsymbol{\alpha}} \boldsymbol{x}^{\boldsymbol{b},\boldsymbol{\alpha}},\\ & f, g \in \mathcal{H}_{\boldsymbol{b}} \wedge f, g : \mathbb{R}^d \to \mathbb{R}.\end{aligned}$$

An orthonormal basis for $\mathcal{H}_{\boldsymbol{b}}$ is

$$\left\{ e_k(\boldsymbol{x}) = e^{-\frac{\|\boldsymbol{x}^{(\boldsymbol{b})}\|^2}{\sigma^2}} \sum_{|\boldsymbol{\alpha}|=k} \sqrt{\frac{(2/\sigma^2)^k C_{\boldsymbol{\alpha}}^k}{k!}} \boldsymbol{x}^{\boldsymbol{b},\boldsymbol{\alpha}} \right\}_{k=0}^{\infty}. \tag{2}$$

Proof. See in Appendix.A.

Remark 2. Obviously, $\mathcal{H}_{\boldsymbol{b}}$ is a function space with Hilbert norm $\|\cdot\|_K$, and the inner product $\langle\cdot,\cdot\rangle_K$ in $\mathcal{H}_{\boldsymbol{b}}$ is a simple generalization of the Weyl inner product for the homogeneous polynomial space $\mathcal{H}_d(\mathbb{R}^d)$.

Remark 3. An orthonormal basis for the RKHS induced by the Gaussian kernel $K(\boldsymbol{x},\boldsymbol{z}) = \exp\left(-\frac{\|\boldsymbol{x}-\boldsymbol{z}\|^2}{\sigma^2}\right)$ has been known in the literature ([13] and references therein). We generalize this result to the extended Gaussian kernels $K(\boldsymbol{x},\boldsymbol{z}) = \exp\left(-\frac{\|\boldsymbol{x}^{(\boldsymbol{b})}-\boldsymbol{z}^{(\boldsymbol{b})}\|^2}{\sigma^2}\right)$. In addition, our approach using the Weyl inner product leads to a much shorter proof.

Remark 4. In [13], Steinwart et al discussed how to use the explicit description of RKHS to analyze support vector machines. Thus, we can use the above results to analyze support vector machines with the extended Gaussian kernels.

3.3 Eigenvalues and Eigenfunctions of Integral Operator

In the theoretical analysis of a broad variety of methods for machine learning and data analysis, such as kernel principal component analysis [8] and spectral clustering [17], the eigenvalues and the eigenfunctions of the integral operator play a crucial role. For this reason, we will study the eigenvalues and the eigenfunctions of L_{K_b} associated with the extended Gaussian kernel.

To state our results, we need the following connection between the theory of the reproducing kernels and the theory of the integral operators, which is manifested via Mercer's theorem. Let $\mathcal{X}$ be a complete, separable metric space, equipped with a finite Borel measure μ, that is $\mu(\mathcal{X}) < \infty$. Let $K : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ be a positive definite kernel on $\mathcal{X}$ satisfying

$$\kappa = \sup_{\boldsymbol{x}\in\mathcal{X}} \sqrt{K(\boldsymbol{x},\boldsymbol{x})} < \infty.$$

We consider the integral operator $L_K : L^2_\mu(\mathcal{X}) \to L^2_\mu(\mathcal{X})$,

$$(L_K f)(\boldsymbol{x}) = \int_{\mathcal{X}} K(\boldsymbol{x},\boldsymbol{t}) f(\boldsymbol{t}) d\mu(\boldsymbol{t}).$$

This is a self-adjoint, compact operator that has eigenvalues

$$\lambda_1 \geq \lambda_2 \geq \ldots, \lambda_i, \ldots \geq 0,$$

with the corresponding L^2_μ-normalized eigenfunctions $\{\phi_k\}_{k=1}^\infty$ forming an orthonormal basis for $L^2_\mu(\mathcal{X})$. Mercer's theorem (we refer to [2] for more detail) states that

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

where the series converges absolutely for each $(\boldsymbol{x}, \boldsymbol{t}) \in \mathcal{X} \times \mathcal{X}$ and uniformly on compact subsets of $\mathcal{X} \times \mathcal{X}$.

Let $S^{d-1} = \{\boldsymbol{x} \in \mathbb{R}^d : \|\boldsymbol{x}\| = 1\}$ be the d-dimensional unit sphere, with surface area $|S^{d-1}| = \frac{2\pi^{\frac{d}{2}}}{\Gamma(\frac{d}{2})}$, where Γ is the gamma function defined by $\Gamma(k) = \int_0^\infty e^{-u} u^{k-1} du$. We review the concept of spherical harmonics which is defined in [7].

Definition 3 (Spherical Harmonics). *Let* $\Delta_d = -\left[\frac{\partial^2}{\partial x_1^2} + \cdots + \frac{\partial^2}{\partial x_d^2}\right]$ *denote the Laplacian operator on* $\mathbb{R}^d$*. A homogeneous polynomial of degree* k *in* $\mathbb{R}^d$ *is called a homogeneous harmonic of order* k *when its Laplacian vanishes. Let* $\mathcal{Y}_k(d)$ *denote the subspace of all homogeneous harmonics of order* k *on the unit sphere* S^{d-1} *in* $\mathbb{R}^d$*. The functions in* $\mathcal{Y}_k(d)$ *are called spherical harmonics of order* k*. We denote by* $\{Y_{k,j}(d; \boldsymbol{x})\}_{j=1}^{N(d,k)}$ *any fixed orthonormal basis for* $\mathcal{Y}_k(d)$ *where* $N(d,k) = \dim \mathcal{Y}_k(d) = \frac{(2k+d-2)(k+d-3)!}{k!(d-2)!}$, $k \geq 0$.

Theorem 2. *Let* $b \in \mathbb{R}, d \in \mathbb{N}, d \geq 2$*, be fixed. Let* $\mathcal{X} = S^{d-1}$ *and* μ *be the uniform probability distribution on* S^{d-1}*. If* $\langle \boldsymbol{x}^{[b]}, \boldsymbol{z}^{[b]} \rangle = \langle \boldsymbol{x}, \boldsymbol{z} \rangle^b$ *for all* $\boldsymbol{x}, \boldsymbol{z} \in \mathcal{X}$*, for the extended Gaussian kernel*

$$K_b(\boldsymbol{x}, \boldsymbol{z}) = \exp\left(-\frac{\|\boldsymbol{x}^{[b]} - \boldsymbol{z}^{[b]}\|^2}{\sigma^2}\right), \sigma > 0,$$

the eigenvalues of $L_{K_b} : L^2_\mu(\mathcal{X}) \to L^2_\mu(\mathcal{X})$ *are*

$$\lambda_k = |S^{d-2}| \int_{-1}^{1} \exp\left(-\frac{2 - 2t^b}{\sigma^2}\right) P_k(d; t)(1 - t^2)^{\frac{d-3}{2}} dt,$$

for all $k \in \mathbb{N} \cup \{0\}$*. Each* λ_k *occurs with multiplicity* $N(d,k)$*, and the corresponding eigenfunctions are the spherical harmonics of order* k *on* S^{d-1}*.*

Proof. See in Appendix.B.

Remark 5. Note that if $b = 1$ or $d = 1$, the assumption $\langle \boldsymbol{x}^{[b]}, \boldsymbol{z}^{[b]} \rangle = \langle \boldsymbol{x}, \boldsymbol{z} \rangle^b$ for all $\boldsymbol{x}, \boldsymbol{z} \in \mathcal{X}$ in the above theorem is satisfied. Thus, when we let $b = 1$, we can obtain the eigenvalues and the eigenfunctions of the integral operator induced by the Gaussian kernel.

Corollary 1. *Let* $d \in \mathbb{N}, d \geq 2$, $\mathcal{X} = S^{d-1}$*, and* μ *be the uniform probability distribution on* S^{d-1}*. For the Gaussian kernel*

$$K(\boldsymbol{x}, \boldsymbol{z}) = \exp\left(-\frac{\|\boldsymbol{x} - \boldsymbol{z}\|^2}{\sigma^2}\right), \sigma > 0,$$

the eigenvalues of $L_K : L^2_\mu(\mathcal{X}) \to L^2_\mu(\mathcal{X})$ *are*

$$\lambda_k = |S^{d-2}| \int_{-1}^{1} \exp\left(-\frac{2 - 2t}{\sigma^2}\right) P_k(d; t)(1 - t^2)^{\frac{d-3}{2}} dt,$$

for all $k \in \mathbb{N} \cup \{0\}$*. Each* λ_k *occurs with multiplicity* $N(d,k)$ *with the corresponding eigenfunctions being spherical harmonics of order* k *on* S^{d-1}*.*

Proof. Since the Gaussian kernel $K(\boldsymbol{x},\boldsymbol{z}) = \exp\left(-\frac{\|\boldsymbol{x}-\boldsymbol{z}\|^2}{\sigma^2}\right)$ is a special case of extended Gaussian kernel when $b = 1$, we can prove the corollary by using the result of Theorem 2.

The radial kernel $K(\boldsymbol{x},\boldsymbol{z}) = \exp\left(-\frac{\|\boldsymbol{x}-\boldsymbol{z}\|}{\sigma^2}\right), \sigma > 0$ is another popular kernel in machine learning and data mining. For both theoretical and practical purposes, we need to study the eigenvalues and the eigenfunctions of the integral operator associated with this radial kernel.

Theorem 3. *Let $d \in \mathbb{N}, d \geq 2$, $\mathcal{X} = S^{d-1}$, and μ be the uniform probability distribution on S^{d-1}. For the radial kernel*

$$K(\boldsymbol{x},\boldsymbol{z}) = \exp\left(-\frac{\|\boldsymbol{x}-\boldsymbol{z}\|}{\sigma^2}\right), \sigma > 0,$$

the eigenvalues of $L_K : L^2_\mu(\mathcal{X}) \to L^2_\mu(\mathcal{X})$ are

$$\lambda_k = |S^{d-2}| \int_{-1}^{1} \exp\left(-\frac{\sqrt{2-2t}}{\sigma^2}\right) P_k(d;t)(1-t^2)^{\frac{n-3}{2}} dt,$$

for all $k \in \mathbb{N} \cup \{0\}$. Each λ_k occurs with multiplicity $N(d,k)$ with the corresponding eigenfunctions being spherical harmonics of order k on S^{d-1}.

Proof. See in Appendix.C.

4 Conclusion

In this paper, we have generalized the results of the explicit description of the reproducing kernel Hilbert space (RKHS) associated with the Gaussian kernel [13,6] to the extended Gaussian kernel. In addition, we have presented the explicit description for the eigenvalues and eigenfunctions of the integral operator. These results can be used in the theoretical analysis of the kernel principal component analysis and other methods which need analysis of the eigenvalues and the eigenfunctions.

We will apply the results of this paper to analyze the learning performance of SVM or other kernel-based methods, and to explore a new criterion for model selection of kernel methods.

Acknowledgments. The work is supported in part by the Natural Science Foundation of China under grant No. 61170019, and the Natural Science Foundation of Tianjin under grant No. 11JCYBJC00700.

Appendix

This section gives the proofs for the theorems in the main text.

Appendix A

In order to prove Theorem 1, we first introduce the following lemma.

Lemma 1 (Aronszajn [1]). *Let $\mathcal{H}$ be a separable Hilbert space of functions over $\mathcal{X}$ with orthonormal basis $\{e_k\}_{k=0}^{\infty}$. $\mathcal{H}$ is a reproducing kernel Hilbert space iff*

$$\sum_{k=0}^{\infty} |e_k(\boldsymbol{x})|^2 < \infty$$

for all $\boldsymbol{x} \in \mathcal{X}$. The unique kernel K is defined by

$$K(\boldsymbol{x}, \boldsymbol{z}) = \sum_{k=0}^{\infty} e_k(\boldsymbol{x}) e_k(\boldsymbol{z}).$$

Proof (of Theorem 1). Note that for any vector $\boldsymbol{x}, \boldsymbol{z}$,

$$e^{\langle \boldsymbol{x}, \boldsymbol{z} \rangle} = \sum_{k=0}^{\infty} \frac{1}{k!} \sum_{|\boldsymbol{\alpha}|=k} C_{\boldsymbol{\alpha}}^k \boldsymbol{x}^{\boldsymbol{\alpha}} \boldsymbol{z}^{\boldsymbol{\alpha}},$$

thus we can obtain that

$$\begin{aligned} K_{\boldsymbol{b}}(\boldsymbol{x}, \boldsymbol{z}) &= \exp\left(-\frac{\|\boldsymbol{x}^{(\boldsymbol{b})} - \boldsymbol{z}^{(\boldsymbol{b})}\|^2}{\sigma^2}\right) \\ &= \exp\left(-\frac{\|\boldsymbol{x}^{(\boldsymbol{b})}\|^2}{\sigma^2}\right) \exp\left(-\frac{\|\boldsymbol{z}^{(\boldsymbol{b})}\|^2}{\sigma^2}\right) \exp\left(\frac{2\langle \boldsymbol{x}^{(\boldsymbol{b})}, \boldsymbol{z}^{(\boldsymbol{b})}\rangle}{\sigma^2}\right) \\ &= \exp\left(-\frac{\|\boldsymbol{x}^{(\boldsymbol{b})}\|^2}{\sigma^2}\right) \exp\left(-\frac{\|\boldsymbol{z}^{(\boldsymbol{b})}\|^2}{\sigma^2}\right) \sum_{k=0}^{\infty} \frac{(2/\sigma^2)^k}{k!} \sum_{|\boldsymbol{\alpha}|=k} C_{\boldsymbol{\alpha}}^k \boldsymbol{x}^{\boldsymbol{b},\boldsymbol{\alpha}} \boldsymbol{z}^{\boldsymbol{b},\boldsymbol{\alpha}}. \end{aligned}$$

Let $\mathcal{H}_0 = \left\{ f = e^{-\frac{\|\boldsymbol{x}^{(\boldsymbol{b})}\|^2}{\sigma^2}} \sum_{|\boldsymbol{\alpha}|=0}^{\infty} w_{\boldsymbol{\alpha}} \boldsymbol{x}^{\boldsymbol{b},\boldsymbol{\alpha}} : \sum_{k=0}^{\infty} \frac{k!}{(2/\sigma^2)^k} \sum_{|\boldsymbol{\alpha}|=k} \frac{w_{\boldsymbol{\alpha}}^2}{C_{\boldsymbol{\alpha}}^k} < \infty \right\}$. For

$$f = e^{-\frac{\|\boldsymbol{x}^{(\boldsymbol{b})}\|}{\sigma^2}} \sum_{|\boldsymbol{\alpha}|=0}^{\infty} w_{\boldsymbol{\alpha}} \boldsymbol{x}^{\boldsymbol{b},\boldsymbol{\alpha}} \in \mathcal{H}_0,$$

$$g = e^{-\frac{\|\boldsymbol{x}^{(\boldsymbol{b})}\|}{\sigma^2}} \sum_{|\boldsymbol{\alpha}|=0}^{\infty} \nu_{\boldsymbol{\alpha}} \boldsymbol{x}^{\boldsymbol{b},\boldsymbol{\alpha}} \in \mathcal{H}_0,$$

we define the inner product

$$\langle f, g \rangle_{K,0} = \sum_{k=0}^{\infty} \frac{k!}{(2/\sigma^2)^k} \sum_{|\boldsymbol{\alpha}|=k} \frac{w_{\boldsymbol{\alpha}} \nu_{\boldsymbol{\alpha}}}{C_{\boldsymbol{\alpha}}^k}.$$

We will show that $\mathcal{H}_0$ is itself a separable Hilbert space under $\langle\cdot,\cdot\rangle_{K,0}$. For simplicity, let $d=1$. Then

$$\mathcal{H}_0 = \left\{ f = e^{-\frac{x^{2b}}{\sigma^2}} \sum_{k=0}^{\infty} w_k x^{bk} : \ \sum_{k=0}^{\infty} \frac{k!}{(2/\sigma^2)^k} w_k^2 < \infty \right\}.$$

It is clear that $\mathcal{H}_0$ is an inner product space under $\langle\cdot,\cdot\rangle_{K,0}$. Its completeness under the induced norm $\|\cdot\|_{K,0}$ is equivalent to the completeness of the weighted ℓ^2 sequence space

$$\ell^2_\sigma = \left\{ (w_k)_{k=0}^{\infty} : \|(w_k)_{k=0}^{\infty}\|_{\ell^2_\sigma} = \left(\sum_{k=0}^{\infty} \frac{k!}{(2/\sigma^2)^k} w_k^2 \right)^{1/2} \right\},$$

which is itself a separable Hilbert space. Thus $(H_0, \|\cdot\|_{K,0})$ is a separable Hilbert space.

If $\mathcal{X} \subset \mathbb{R}^d$ has non-empty interior, then the mononomials $\boldsymbol{x}^{\boldsymbol{b},\boldsymbol{\alpha}}$ are all distinct. From the definition of the inner product $\langle\cdot,\cdot\rangle_{K,0}$, it is easy to obtain that

$$\langle e_i, e_j \rangle_K = \begin{cases} 0, & \text{if } i \neq j; \\ 1, & \text{otherwise;} \end{cases}$$

where e_k are given in (2). So $\{e_k\}_{k=0}^{\infty}$ are orthonormal under $\langle\cdot,\cdot\rangle_{K,0}$. Moreover, $\mathcal{H}_0 = \text{span}\{e_k, k=0,1,\ldots\}$, thus, $\{e_k\}_{k=0}^{\infty}$ forms an orthonormal basis for $(\mathcal{H}_0, \|\cdot\|_{K,0})$. By Lemma 1 and the following equation

$$\sum_{k=0}^{\infty} |e_k(\boldsymbol{x})|^2 = K(\boldsymbol{x},\boldsymbol{x}) = 1 < \infty,$$

we can obtain that $\mathcal{H}_{\boldsymbol{b}}$ is a reproducing kernel Hilbert space. Note that

$$\sum_{k=0}^{\infty} e_k(\boldsymbol{x}) e_k(\boldsymbol{z}) = K_b(\boldsymbol{x},\boldsymbol{z}),$$

and since the RKHS induced by a kernel on a set $\mathcal{X}$ is unique, thus $(\mathcal{H}_0, \|\cdot\|_{K,0})$ is the reproducing kernel Hilbert space of functions on $\mathcal{X}$ with the extended Guassian kernel $K_{\boldsymbol{b}}(\boldsymbol{x},\boldsymbol{z})$.

Appendix B

In order to obtain the eigenvalues and eigenfunctions of the integral operator associated with the extended Gaussian kernel, we first give the following lemma.

Lemma 2. *Let $d \in \mathbb{N}, d \geq 2$ be fixed. Let $K : [-1,1] \to \mathbb{R}$ be a continuous function, giving rise to a continuous, positive definite kernel $K(\boldsymbol{x},\boldsymbol{t}) = K(\langle \boldsymbol{x},\boldsymbol{t}\rangle)$ on $S^{d-1} \times S^{d-1}$. Let μ be the Lebesgue measure on S^{d-1}. The eigenvalues λ_k of*

$$L_K : L^2_\mu(S^{d-1}) \to L^2_\mu(S^{d-1})$$

are given by

$$\lambda_k = |S^{d-2}| \int_{-1}^{1} K(t) P_k(d;t)(1-t^2)^{\frac{d-3}{2}} dt,$$

each with multiplicity $N(d,k)$*, for* $k \in \mathbb{Z}, k \geq 0$*, where* $P_k(d;t)$ *is Legendre polynomial of degree* k *in dimension* d*,*

$$P_k(d;t) = k!\Gamma\left(\frac{d-1}{2}\right) \sum_{l=0}^{\lceil \frac{k}{2} \rceil} \left(\frac{-1}{4}\right)^l \frac{(1-t^2)^l t^{k-2l}}{l!(k-2l)!\Gamma(l+\frac{d-1}{2})}.$$

The corresponding eigenfunctions for each λ_k *are the spherical harmonics* $\{Y_{k,j}(d;\boldsymbol{x})\}_{j=1}^{N(d,k)}$ *of the order* k*.*

Proof. Let $f : [-1,1] \to \mathbb{R}$ be a continuous function. Let $Y_k \in \mathcal{Y}_k(d)$ for $k \geq 0$. Then Funk-Hecke formula ([7], p 30) states that for any $\boldsymbol{x} \in S^{d-1}$:

$$\int_{S^{d-1}} f(\langle \boldsymbol{x}, \boldsymbol{t} \rangle) Y_k(t) dS^{d-1}(\boldsymbol{t}) = \lambda_k Y_k(\boldsymbol{x}), \tag{3}$$

where

$$\lambda_k = |S^{d-2}| \int_{-1}^{1} f(t) P_k(d;\boldsymbol{t})(1-t^2)^{\frac{d-3}{2}} dt \tag{4}$$

and $P_k(d;\boldsymbol{t})$ denotes the Legendre polynomial of degree k in dimension d. The spherical harmonics $\left\{\{Y_{k,j}(d;\boldsymbol{x})\}_{j=1}^{N(d,k)}\right\}_{k=0}^{\infty}$ form an orthonormal basis for $L^2(S^{d-1})$. So if the kernel K on $S^{d-1} \times S^{d-1}$ is defined by $K(\boldsymbol{x},\boldsymbol{t}) = f(\langle \boldsymbol{x}, \boldsymbol{t} \rangle)$, via the Funk-Hecke formula, it is easy to verify that the eigenvalues of

$$L_K : L^2_\mu(S^{d-1}) \to L^2_\mu(S^{d-1})$$

are given precisely by (4), with the corresponding orthonormal eigenfunctions of $\{Y_{k,j}(d;\boldsymbol{x})\}_{j=1}^{N(d,k)}$. The multiplicity of λ_k is therefore $N(d,k) = \dim(\mathcal{Y}_k(d))$.

Proof (of Theorem 2). Note that

$$\exp\left(-\frac{\|\boldsymbol{x}^{[b]} - \boldsymbol{z}^{[b]}\|^2}{\sigma^2}\right) = \exp\left(-\frac{\|\boldsymbol{x}^{[b]}\|^2 + \|\boldsymbol{z}^{[b]}\|^2 - 2\langle \boldsymbol{x}^{[b]}, \boldsymbol{z}^{[b]} \rangle}{\sigma^2}\right),$$

since $\boldsymbol{x}, \boldsymbol{z} \in S^{d-1}$ and $\langle \boldsymbol{x}^{[b]}, \boldsymbol{z}^{[b]} \rangle = \langle \boldsymbol{x}, \boldsymbol{z} \rangle^b$, so it is easy to obtain that

$$\exp\left(-\frac{\|\boldsymbol{x}^{[b]} - \boldsymbol{z}^{[b]}\|^2}{\sigma^2}\right) = \exp\left(-\frac{2 - 2\langle \boldsymbol{x}, \boldsymbol{z} \rangle^b}{\sigma^2}\right).$$

Thus, using the Lemma 2, we know that the eigenvalues of

$$L_{K_b} : L^2_\mu(\mathcal{X}) \to L^2_\mu(\mathcal{X})$$

are

$$\lambda_k = |S^{d-2}| \int_{-1}^{1} \exp\left(-\frac{2-2t^b}{\sigma^2}\right) P_k(d;t)(1-t^2)^{\frac{d-3}{2}}\,dt,$$

and each λ_k occurs with multiplicity $N(d,k)$ with the corresponding eigenfunctions being spherical harmonics of order k on S^{d-1}.

Appendix C

Proof (of Theorem 3). On S^{d-1}, it is easy to verify that

$$\exp\left(-\frac{\|\boldsymbol{x}-\boldsymbol{z}\|}{\sigma^2}\right) = \exp\left(-\frac{\sqrt{(2-2\langle \boldsymbol{x},\boldsymbol{z}\rangle)}}{\sigma^2}\right).$$

Thus, using the Lemma 2, we know that the eigenvalues of

$$L_K : L^2_\mu(\mathcal{X}) \to L^2_\mu(\mathcal{X})$$

are

$$\lambda_k = |S^{d-2}| \int_{-1}^{1} \exp\left(-\frac{\sqrt{2-2t}}{\sigma^2}\right) P_k(d;t)(1-t^2)^{\frac{d-3}{2}}\,dt,$$

and each λ_k occurs with multiplicity $N(d,k)$ with the corresponding eigenfunctions being spherical harmonics of order k on S^{d-1}.

References

1. Aronszajn, N.: Theory of reproducing kernels. Transactions of the American Mathematical Society 68, 337–404 (1950)
2. Cucker, F., Smale, S.: On the mathematical foundations of learning. Bulletin of the American Mathematical Society 39(1), 1–49 (2001)
3. Ferreira, J.C., Manegatto, V.A.: Reproducing kernel hilbert spaces associated with kernels on topological spaces. Functional Analysis and Its Applications 46(2), 152–154 (2012)
4. Ferreira, J.C., Manegatto, V.A.: Reproducing properties of differentiable mercer-like kernels. Mathematische Nachrichten 285(8-9), 959–973 (2012)
5. Kadri, H., Rabaoui, A., Preux, P., Duflos, E., Rakotomamonjy, A.: Functional regularized least squares classification with operator-valued kernels. In: Proceeding of the 28th International Conference on Machine Learning (ICML 2011), pp. 993–1000 (2011)
6. Minh, H.Q., Niyogi, P., Yao, Y.: Mercer's Theorem, Feature Maps, and Smoothing. In: Lugosi, G., Simon, H.U. (eds.) COLT 2006. LNCS (LNAI), vol. 4005, pp. 154–168. Springer, Heidelberg (2006)
7. Müller, C.: Analysis of spherical symmetries in Euclidean spaces. Applied Mathematical Sciences, vol. 129. Springer, New York (1998)
8. Schölkopf, B., Smola, A.J., Müller, K.R.: Nonlinear component analysis as a kernel eigenvalue problem. Neural Computation 10(5), 1299–1319 (1998)

9. Schölkopf, B., Smola, A.J.: Learning with kernels: Support vector machines, regularization, optimization, and beyond. The MIT Press (2002)
10. Schölkopf, B., Smola, A.J., Müller, K.-R.: Kernel Principal Component Analysis. In: Gerstner, W., Hasler, M., Germond, A., Nicoud, J.-D. (eds.) ICANN 1997. LNCS, vol. 1327, pp. 583–588. Springer, Heidelberg (1997)
11. Scovel, C., Hush, D., Steinwart, I., Theiler, J.: Radial kernels and their reproducing kernel Hilbert spaces. Journal of Complexity 26(6), 641–660 (2010)
12. Smale, S., Zhou, D.X.: Learning theory estimates via integral operators and their approximations. Constructive Approximation 26(2), 153–172 (2007)
13. Steinwart, I., Hush, D., Scovel, C.: An explicit description of the reproducing kernel Hilbert spaces of Gaussian RBF kernels. IEEE Transactions on Information Theory 52(10), 4635–4643 (2006)
14. Sun, H.W., Zhou, D.X.: Reproducing kernel Hilbert spaces associated with analytic translation-invariant Mercer kernels. Journal of Fourier Analysis and Applications 14(1), 89–101 (2008)
15. Vapnik, V.: The nature of statistical learning theory. Springer (2000)
16. Vito, E.D., Caponnetto, A., Rosasco, L.: Model selection for regularized least-squares algorithm in learning theory. Foundations of Computational Mathematics 5(1), 59–85 (2005)
17. Von Luxburg, U., Bousquet, O., Belkin, M.: On the Convergence of Spectral Clustering on Random Samples: The Normalized Case. In: Shawe-Taylor, J., Singer, Y. (eds.) COLT 2004. LNCS (LNAI), vol. 3120, pp. 457–471. Springer, Heidelberg (2004)
18. Wahba, G.: Spline models for observational data. SIAM (1990)