Journal of Machine Learning Research 8 (2007) 2083-2120 Submitted 5/07; Revised 8/07; Published 9/07

Refinable Kernels*

Yesheng Xu YXU06@SYR.EDU
Haizhang Zhang HZHANG12@SYR.EDU
Department of Mathematics
Syracuse University
Syracuse, NY 13244, USA

Editor: Bernhard Schölkopf

Abstract

Motivated by mathematical learning from training data, we introduce the notion of *refinable kernels*. Various characterizations of refinable kernels are presented. The concept of refinable kernels leads to the introduction of *wavelet-like reproducing kernels*. We also investigate a refinable kernel that forms a Riesz basis. In particular, we characterize refinable translation invariant kernels, and refinable kernels defined by refinable functions. This study leads to multiresolution analysis of reproducing kernel Hilbert spaces.

Keywords: refinable kernels, refinable feature maps, wavelet-like reproducing kernels, dual kernels, learning with kernels, reproducing kernel Hilbert spaces, Riesz bases

1. Introduction

The main purpose of this paper is to introduce the notion of refinable kernels, wavelet-like reproducing kernels and multiresolution analysis of a reproducing kernel Hilbert space. Before proceeding to the motivation, it is worthwhile to know that there has been a large body of literature on similar notions such as refinable functions (Cavaretta et al., 1991; Daubechies, 1992), multiresolution analysis of $L^2(\mathbb{R})$ (Mallat, 1989; Meyer, 1992) and kernels constructed by wavelet functions (Amato et al., 2006; Rakotomamonjy and Canu, 2005; Rakotomamonjy et al., 2005). The connection of these well-known notions with those to be presented will become clear as we proceed this study.

We first motivate the concept of refinable kernels by learning via a kernel. Let X be a prescribed set which is called in the theory of learning an input space and is associated with an output space $Y \subseteq \mathbb{C}$. A typical learning task aims at inferring from a finite set of training data $\mathbf{z} := \{(x_j, y_j) : j \in \mathbb{N}_m\}$, where $\mathbb{N}_m := \{1, 2, \ldots, m\}$, a function f from X to Y so that $f(x)$ gives a satisfactory output of an input $x \in X$. A popular choice of f is a minimizer of a certain error functional. Specifically, we let $\mathcal{H}$ be a given class of functions on X, $Q : \mathbb{C} \times \mathbb{C} \to \mathbb{R}_+$ be a *loss function* (Schölkopf and Smola, 2002) measuring how well g fits the training data $\mathbf{z}$, $\mathcal{N} : \mathcal{H} \to \mathbb{R}_+$ be a controller of the set of functions in $\mathcal{H}$ from which we choose f, and μ be a positive regularization parameter. The function f may be chosen as

$$\arg\min_{g \in \mathcal{H}} \sum_{j \in \mathbb{N}_m} Q(g(x_j), y_j) + \mu \mathcal{N}(g). \tag{1}$$

*. Dedicated to Dr. Charles Micchelli's 65th birthday for friendship and esteem.

The selection of the function class $\mathcal{H}$ is critical for the behavior of f and thus it deserves special attention. In practice, $\mathcal{H}$ may be chosen through a *kernel* K on X, a function from $X \times X$ to $\mathbb{C}$ such that for all finite sets $\mathbf{t} := \{t_j : j \in \mathbb{N}_n\} \subseteq X$ the matrix

$$K[\mathbf{t}] := [K(t_j, t_k) : j, k \in \mathbb{N}_n] \tag{2}$$

is hermitian and positive semi-definite (see, for example, Cucker and Smale, 2002; Schölkopf and Smola, 2002; Shawe-Taylor and Cristianini, 2004; Vapnik, 1998). The importance of kernels in learning is that the function evaluation $K(x,y)$ is able to measure the similarity of $x, y \in X$. A kernel K on X corresponds to a Hilbert space

$$\mathcal{H}_K := \overline{\operatorname{span}}\{K(\cdot, y) : y \in X\} \tag{3}$$

of functions on X with an inner product determined by

$$(K(\cdot, y), K(\cdot, x))_{\mathcal{H}_K} = K(x, y), \quad x, y \in X. \tag{4}$$

The space $\mathcal{H}_K$ is a *reproducing kernel Hilbert space* (RKHS), that is, point evaluations are continuous linear functionals on $\mathcal{H}_K$ (Aronszajn, 1950). Moreover, $\mathcal{H}_K$ is the only Hilbert space of functions on X such that for all $x \in X$, we have that $K(\cdot, x) \in \mathcal{H}_K$ and

$$f(x) = (f, K(\cdot, x))_{\mathcal{H}_K}, \quad f \in \mathcal{H}_K. \tag{5}$$

Due to Equation (5), K is often interpreted as the *reproducing kernel* of $\mathcal{H}_K$. A RKHS has exactly one reproducing kernel (Aronszajn, 1950). To construct a learning function $f : X \to Y$ from the training data $\mathbf{z}$, we start with a kernel K on X. Choose in (1) $\mathcal{H} := \mathcal{H}_K$ and $\mathcal{N} := \|\cdot\|^2_{\mathcal{H}_K}$, the square of the norm on $\mathcal{H}_K$. In this case, the minimization problem in (1) reduces to a regularization in the RKHS, which has received much attention in the literature (see, for example, Bousquet and Elisseeff, 2002; Cucker and Smale, 2002; Micchelli and Pontil, 2005a,b; Mukherjee et al., 2006; Schölkopf and Smola, 2002; Smale and Zhou, 2003; Steinwart and Scovel, 2005; Vapnik, 1998; Wahba, 1999; Walder et al., 2006; Ying and Zhou, 2007; Zhang, 2004, and the references cited therein). In this setting, the *representer theorem* in learning (see, for example, Kimeldorf and Wahba, 1971; Micchelli and Pontil, 2004; Schölkopf et al., 2001; Schölkopf and Smola, 2002; Shawe-Taylor and Cristianini, 2004; Walder et al., 2006) asserts that there exists $\mathbf{c} := [c_j : j \in \mathbb{N}_m]^T \subseteq \mathbb{C}^m$, depending on the training data $\mathbf{z}$ and the kernel K, such that the minimizer (1) is

$$f = \sum_{j \in \mathbb{N}_m} c_j K(\cdot, x_j). \tag{6}$$

The choice of kernels K is certainly one of the most important issues in the above learning scheme via regularization. It is often based on the training data $\mathbf{z}$ currently available to us. However, the old training data may be updated to $\mathbf{z}' := \{(x'_j, y'_j) : j \in \mathbb{N}_{m'}\}$ by adding to $\mathbf{z}$ more new samples from $X \times Y$. The kernels that we use should offer us a convenient way to update the kernel. In other words, we are looking for kernels with the ability of learning dynamically expanding training data. Specifically, we demand kernels K having the feature that there is a cheap way of updating K to a new kernel K' such that $\mathcal{H}_K \preceq \mathcal{H}_{K'}$. Here and throughout the paper, we make the convention that whenever we write $\mathcal{W}_1 \preceq \mathcal{W}_2$ for Hilbert spaces $\mathcal{W}_1, \mathcal{W}_2$, the inclusion is in the sense that $\mathcal{W}_1 \subseteq \mathcal{W}_2$

and for all $u,v \in \mathcal{W}_1$, $(u,v)_{\mathcal{W}_1} = (u,v)_{\mathcal{W}_2}$. The inclusion $\mathcal{H}_K \preceq \mathcal{H}_{K'}$ has a natural interpretation. When we have a larger training data set $\mathbf{z}'$, we may expect that the minimization

$$\min_{g \in \mathcal{H}_{K'}} \sum_{j \in \mathbb{N}_{m'}} Q(g(x'_j), y'_j) + \mu' \|g\|^2_{\mathcal{H}_{K'}}, \tag{7}$$

yields a better predictor f' than f. This becomes possible only if the new space $\mathcal{H}_{K'}$ includes $\mathcal{H}_K$ as a subspace. Moreover, we require the updating from K to K' to have the feature that computing minimizer f' from (7) should be able to make use of the previously computed minimizer f from (1). Moreover, when the representation (6) for f is available, we want to process it efficiently.

This motivates us to introduce the concept of refinable kernels. We shall study characterizations of a refinable kernel, fundamental properties of refinable kernels and wavelet-like reproducing kernels, and multiscale structures of a RKHS induced by refinable kernels. It is important to note that the concept of wavelet-like reproducing kernels differs from that of "wavelet kernels" in Amato et al. (2006), Rakotomamonjy and Canu (2005), and Rakotomamonjy et al. (2005). The earlier means the kernels defined by the difference of kernels at two consecutive scales while the latter means the kernels defined by a linear combination of dilations and translations of wavelet functions. This paper is organized in seven sections. We present in Section 2 two characterizations of refinable kernels. Section 3 is devoted to wavelet-like reproducing kernels and a multiscale decomposition of the RKHS of a refinable kernel. In Section 4, we investigate refinable kernels of a Riesz type and we also introduce the notion of a multiresolution analysis for a RKHS. As concrete examples of refinable kernels, we formulate in Sections 5 and 6, respectively, conditions for translation invariant kernels and kernels defined by a refinable function to be refinable. In Section 7, we have a brief discussion of potential applications of refinable kernels and make a conclusion.

2. Characterizations of γ-Refinable Kernels

We define in this section γ-refinable kernels and present their characterizations. Let $\gamma : X \to X$ be a given bijective mapping and let ι denote the identity mapping from X to itself. We introduce a sequence of mappings from X to itself by recursions

$$\gamma_{-1} := \gamma^{-1}, \;\; \gamma_{-n-1} := \gamma^{-1} \circ \gamma_{-n}, \;\; \text{and} \;\; \gamma_0 := \iota, \;\; \gamma_n := \gamma \circ \gamma_{n-1}, \;\; n \in \mathbb{N}.$$

A kernel K on X is called γ-*refinable* if there exists a positive constant λ depending only on K and γ such that

$$\mathcal{H}_K \preceq \mathcal{H}_{K_1},$$

where

$$K_1(x,y) := \lambda K(\gamma(x), \gamma(y)), \;\; x,y \in X.$$

The constant λ is a normalization factor which ensures the inner product on $\mathcal{H}_K$ identical to that on $\mathcal{H}_{K_1}$. Examples of the mappings γ include the dilation mapping $x \to 2x$ in $\mathbb{R}^d$ and in general, the dilation mapping $x \to Ax$ where A is an expanding matrix. We simply call the γ-refinable kernels in this case *refinable kernels*.

Let K be a kernel on the input space X, and λ a positive constant. For any bijective mapping γ and any $n \in \mathbb{Z}$, we let

$$K_n(x,y) := \lambda^n K(\gamma_n(x), \gamma_n(y)), \;\; x,y \in X. \tag{8}$$

We next identify the RKHS defined by K_n.

Theorem 1 *If K is a kernel on X, then for each $n \in \mathbb{Z}$, the RKHS of kernel K_n is*

$$\mathcal{H}_{K_n} = \{f \circ \gamma_n : f \in \mathcal{H}_K\} \tag{9}$$

with inner product

$$(f,g)_{\mathcal{H}_{K_n}} := \lambda^{-n}(f \circ \gamma_{-n}, g \circ \gamma_{-n})_{\mathcal{H}_K}, \quad f,g \in \mathcal{H}_{K_n}. \tag{10}$$

Proof It can be shown that K_n is a kernel on X. We set $\mathbb{H}_n := \{f \circ \gamma_n : f \in \mathcal{H}_K\}$ and introduce an inner product on $\mathbb{H}_n$ by

$$(f,g)_{\mathbb{H}_n} := \lambda^{-n}(f \circ \gamma_{-n}, g \circ \gamma_{-n})_{\mathcal{H}_K}, \quad f,g \in \mathbb{H}_n.$$

It is clear that $\mathbb{H}_n$ is a Hilbert space with this inner product. Note that for any $x \in X$, $K_n(\cdot,x) \in \mathbb{H}_n$ and for any $f \in \mathbb{H}_n$, $f \circ \gamma_{-n} \in \mathcal{H}_K$. Hence, for $f \in \mathbb{H}_n$, we obtain by (5) and the definition of the inner product $(\cdot,\cdot)_{\mathcal{H}_{K_n}}$ that for $x \in X$

$$f(x) = (f \circ \gamma_{-n})(\gamma_n(x)) = (f \circ \gamma_{-n}, K(\cdot,\gamma_n(x)))_{\mathcal{H}_K} = \lambda^n (f, K(\gamma_n(\cdot),\gamma_n(x))_{\mathbb{H}_n}.$$

Combining this equation with the definition of the kernel K_n leads to $f(x) = (f, K_n(\cdot,x))_{\mathbb{H}_n}$, $x \in X$. This implies that K_n is the reproducing kernel for $\mathbb{H}_n$. By the unique correspondence between a RKHS and its reproducing kernel, we conclude that $\mathcal{H}_{K_n} = \mathbb{H}_n$. ∎

A direct consequence of Theorem 1 is that if K is a γ-refinable kernel then for each $n \in \mathbb{Z}$, the kernel K_n is γ-refinable. This result justifies the usage of the same mapping γ to update K_n to K_{n+1} in (8) for each $n \in \mathbb{Z}$.

Proposition 2 *If the kernel K is γ-refinable, then for each $n \in \mathbb{Z}$, K_n is γ-refinable. Conversely, if for some $n \in \mathbb{Z}$, K_n is γ-refinable, then K is γ-refinable.*

Proof Suppose that K is γ-refinable. Then, we have that $\mathcal{H}_K \preceq \mathcal{H}_{K_1}$. Let $f \in \mathcal{H}_{K_n}$. By Theorem 1, we deduce that $f \circ \gamma_{-n} \in \mathcal{H}_K \preceq \mathcal{H}_{K_1}$, which implies that $f \in \mathcal{H}_{K_{n+1}}$. In addition, Theorem 1, Equation (10) and the γ-refinability of K ensure that for $f,g \in \mathcal{H}_{K_n}$,

$$\begin{aligned}(f,g)_{\mathcal{H}_{K_n}} &= \lambda^{-n}(f \circ \gamma_{-n}, g \circ \gamma_{-n})_{\mathcal{H}_K} = \lambda^{-n}(f \circ \gamma_{-n}, g \circ \gamma_{-n})_{\mathcal{H}_{K_1}} \\ &= \lambda^{-n-1}(f \circ \gamma_{-n-1}, g \circ \gamma_{-n-1})_{\mathcal{H}_K} = (f,g)_{\mathcal{H}_{K_{n+1}}}.\end{aligned}$$

This confirms that K_n is γ-refinable.

Conversely, we suppose that K_n is γ-refinable for some $n \in \mathbb{Z}$. Since

$$K(x,y) = \lambda^{-n} K_n(\gamma_{-n}(x), \gamma_{-n}(y)), \quad x,y \in X,$$

the arguments in the proof of the first statement of this proposition show that K is γ-refinable. ∎

The next result follows immediately from the last proposition.

Corollary 3 *If K is γ-refinable, then for all $f,g \in \mathcal{H}_K$ and $n \in \mathbb{N}$, $f,g \in \mathcal{H}_{K_n}$ and $(f,g)_{\mathcal{H}_{K_n}} = (f,g)_{\mathcal{H}_K}$.*

We now present our first characterization of γ-refinable kernels.

Theorem 4 *A kernel K is γ-refinable if and only if*

$$K(\gamma_{-1}(\cdot),x) \in \mathcal{H}_K, \ \ \textit{for all } x \in X \tag{11}$$

and

$$(K(\gamma_{-1}(\cdot),y),K(\gamma_{-1}(\cdot),x))_{\mathcal{H}_K} = \lambda K(x,y), \ \ \textit{for all } x,y \in X, \tag{12}$$

where λ is the same constant in the definition of a γ-refinable kernel.

Proof By Proposition 2, K is γ-refinable if and only if

$$\mathcal{H}_{K_{-1}} \preceq \mathcal{H}_K. \tag{13}$$

Suppose that K is γ-refinable. The definition of kernel K_{-1} leads to

$$K(\gamma_{-1}(\cdot),x) = \lambda K_{-1}(\cdot,\gamma(x)) \in \mathcal{H}_{K_{-1}}, \ \ \text{for} \ \ x \in X,$$

for some constant λ. This combined with relation (13) ensures the validity of (11). By (13), (10) and (4), we obtain for all $x,y \in X$ that

$$\begin{aligned}(K(\gamma_{-1}(\cdot),y),K(\gamma_{-1}(\cdot),x))_{\mathcal{H}_K} &= (K(\gamma_{-1}(\cdot),y),K(\gamma_{-1}(\cdot),x))_{\mathcal{H}_{K_{-1}}} \\ &= \lambda(K(\cdot,y),K(\cdot,x))_{\mathcal{H}_K} = \lambda K(x,y),\end{aligned}$$

which is Equation (12).

Conversely, we suppose that (11) holds and (12) is satisfied with some constant λ, and we prove that inclusion relation (13) is valid. Let $f \in \mathcal{H}_K$. By (3), there exists a sequence

$$f_n \in \operatorname{span}\{K(\cdot,y) : y \in X\}, \ \ n \in \mathbb{N}$$

that converges to f in $\mathcal{H}_K$. Equation (11) implies that $f_n \circ \gamma_{-1} \in \mathcal{H}_K$, $n \in \mathbb{N}$, and Equation (12) implies for all $m,n \in \mathbb{N}$ that

$$\| f_m \circ \gamma_{-1} - f_n \circ \gamma_{-1} \|_{\mathcal{H}_K} = \lambda^{1/2} \| f_m - f_n \|_{\mathcal{H}_K}.$$

Therefore, $f_n \circ \gamma_{-1}$ is a Cauchy sequence in $\mathcal{H}_K$, whose limit is denoted by f_{-1}. Recall that point evaluations are continuous linear functionals on $\mathcal{H}_K$. An application of this fact yields that

$$f_{-1}(x) = \lim_{n\to\infty} (f_n \circ \gamma_{-1})(x), \ \ x \in X. \tag{14}$$

In the same manner, since f_n converges to f in $\mathcal{H}_K$, we have for each $x \in X$ that

$$(f \circ \gamma_{-1})(x) = \lim_{n\to\infty} (f_n \circ \gamma_{-1})(x). \tag{15}$$

We observe from (14) and (15) that $f \circ \gamma_{-1} = f_{-1}$. It is hence proved that $f \circ \gamma_{-1} \in \mathcal{H}_K$ for each $f \in \mathcal{H}_K$. This combined with (9) shows that the elements of $\mathcal{H}_{K_{-1}}$ are contained in $\mathcal{H}_K$. Finally, we verify by (12) for each $f,g \in \mathcal{H}_K$ that

$$(f \circ \gamma_{-1}, g \circ \gamma_{-1})_{\mathcal{H}_K} = (f_{-1}, g_{-1})_{\mathcal{H}_K} = \lim_{n\to\infty} (f_n \circ \gamma_{-1}, g_n \circ \gamma_{-1})_{\mathcal{H}_K} = \lambda \lim_{n\to\infty} (f_n, g_n)_{\mathcal{H}_K} = \lambda(f,g)_{\mathcal{H}_K}.$$

By the above equation and (10), the inner product on $\mathcal{H}_{K_{-1}}$ coincides with the one on $\mathcal{H}_K$. We conclude that (13) holds true and complete the proof. ∎

Another characterization of γ-refinable kernels K is in terms of a *feature map* for K. A function Φ from X to a Hilbert space $\mathcal{W}$ is called a feature map for the kernel K if

$$K(x,y) = (\Phi(x),\Phi(y))_{\mathcal{W}}, \ \ x,y \in X. \tag{16}$$

We call the Hilbert space $\mathcal{W}$ the feature space of K. It is known (Aronszajn, 1950) that K is a kernel on X if and only if there exists a map $\Phi : X \to \mathcal{W}$ satisfying (16). In the next result, we identify the RKHS $\mathcal{H}_K$ in terms of a feature map Φ for K. To state the result, we denote by $\Phi(X)$ the image of X under Φ, $\overline{\operatorname{span}}\,\Phi(X)$ the closure of $\operatorname{span}\Phi(X)$ in $\mathcal{W}$, and P_Φ the orthogonal projection from $\mathcal{W}$ onto $\overline{\operatorname{span}}\,\Phi(X)$.

Lemma 5 *Let K be a kernel having a representation (16) in terms of a feature map Φ from X to $\mathcal{W}$. Then $\mathcal{H}_K = \{(\Phi(\cdot),u)_{\mathcal{W}} : u \in \mathcal{W}\}$ with inner product*

$$((\Phi(\cdot),u)_{\mathcal{W}},(\Phi(\cdot),v)_{\mathcal{W}})_{\mathcal{H}_K} = (P_\Phi v, P_\Phi u)_{\mathcal{W}}, \ \ u,v \in \mathcal{W}. \tag{17}$$

A proof of this result in the special case that K is a *Hilbert-Schmidt kernel* was provided in Opfer (2006) (see Lemma 3.4, Theorem 3.5 therein). The proof works for the general case described in Lemma 5. The result can also be found in Micchelli and Pontil (2005a). In the application of Lemma 5, it is always convenient to assume that there holds

$$\overline{\operatorname{span}}\,\Phi(X) = \mathcal{W} \tag{18}$$

since otherwise $\mathcal{W}$ can be replaced by $\overline{\operatorname{span}}\,\Phi(X)$. If (18) holds then for each $f \in \mathcal{H}_K$ there exists a unique $u_f \in \mathcal{W}$ such that $f = (\Phi(\cdot),u_f)_{\mathcal{W}}$. Moreover, one can see by Lemma 5 that the linear transformation Γ from $\mathcal{H}_K$ to $\mathcal{W}$ defined by

$$\Gamma f := u_f \tag{19}$$

is an *isomorphism*, that is, it is one-to-one, onto and satisfies $\|\Gamma f\|_{\mathcal{W}} = \|f\|_{\mathcal{H}_K}$, for $f \in \mathcal{H}_K$.

We call a feature map Φ from X to $\mathcal{W}$ γ-*refinable* provided that there is a bounded linear operator T on $\mathcal{W}$ such that

$$\lambda^{-1/2}\Phi \circ \gamma_{-1} = T\Phi, \tag{20}$$

where $\lambda^{-1/2}$ plays the role of a normalization parameter. Throughout this paper, we mean that T is a function from a Hilbert space $\mathcal{W}$ to itself whenever we say that T is an operator on $\mathcal{W}$. Recall that a linear operator A on $\mathcal{W}$ is *isometric* if for all $u \in \mathcal{W}$, $\|Au\|_{\mathcal{W}} = \|u\|_{\mathcal{W}}$. One can see that A is isometric if and only if A^*A is equal to the identity operator on $\mathcal{W}$, where A^* denotes the *adjoint operator* of A.

We characterize a refinable kernel in terms of its feature map.

Theorem 6 *Suppose that K is a kernel on X with a feature map $\Phi : X \to \mathcal{W}$ satisfying (18). Then K is γ-refinable if and only if Φ is γ-refinable and the adjoint operator T^* of T in (20) is isometric.*

Proof Suppose that Φ is γ-refinable, that is, it satisfies (20) for some bounded operator T on $\mathcal{W}$, and suppose that T^* is isometric. We first observe by (16) and (20) for each $x \in X$ that

$$K(\gamma_{-1}(\cdot),x) = (\Phi \circ \gamma_{-1}(\cdot), \Phi(x))_{\mathcal{W}} = \lambda^{1/2}(T\Phi(\cdot), \Phi(x))_{\mathcal{W}} = \lambda^{1/2}(\Phi(\cdot), T^*\Phi(x))_{\mathcal{W}}. \tag{21}$$

Lemma 5 with the equation above yields that for each $x \in X$, $K(\gamma_{-1}(\cdot),x) \in \mathcal{H}_K$. Moreover, (18) implies that P_Φ is the identity operator on $\mathcal{W}$. This fact, together with Equations (21) and (17) ensures for all $x,y \in X$ that

$$(K(\gamma_{-1}(\cdot),y), K(\gamma_{-1}(\cdot),x))_{\mathcal{H}_K} = \lambda((\Phi(\cdot), T^*\Phi(y))_{\mathcal{W}}, (\Phi(\cdot), T^*\Phi(x))_{\mathcal{W}})_{\mathcal{H}_K} = \lambda(T^*\Phi(x), T^*\Phi(y))_{\mathcal{W}}.$$

By hypothesis, TT^* is the identity. Hence, the right hand side of the above equation becomes $\lambda(\Phi(x),\Phi(y))_{\mathcal{W}}$, which is equal to $\lambda K(x,y)$, since Φ is a feature map for K. That is, (12) holds. We conclude by Theorem 4 that K is γ-refinable.

Conversely, suppose that K is γ-refinable, that is, $\mathcal{H}_{K_{-1}} \preceq \mathcal{H}_K$. We shall choose a bounded linear operator T on $\mathcal{W}$ such that Φ satisfies (20) and T^* is isometric. By Theorem 1 and Lemma 5, functions in $\mathcal{H}_{K_{-1}}$ have the form $(\Phi \circ \gamma_{-1}(\cdot), u)_{\mathcal{W}}$, $u \in \mathcal{W}$. The inclusion $\mathcal{H}_{K_{-1}} \preceq \mathcal{H}_K$ implies that for each $u \in \mathcal{W}$ there exists $v_u \in \mathcal{W}$ such that

$$\lambda^{-1/2}(\Phi \circ \gamma_{-1}(\cdot), u)_{\mathcal{W}} = (\Phi(\cdot), v_u)_{\mathcal{W}}. \tag{22}$$

Equation (18) ensures that for each $u \in \mathcal{W}$ there is a unique $v_u \in \mathcal{W}$ satisfying (22). Let A denote the map $u \to v_u$ and observe that A is a linear operator on $\mathcal{W}$. We shall prove that it is isometric. Since by (16) for all $x,y \in X$

$$K_{-1}(x,y) = \lambda^{-1}K(\gamma_{-1}(x),\gamma_{-1}(y)) = \lambda^{-1}(\Phi(\gamma_{-1}(x)), \Phi(\gamma_{-1}(y)))_{\mathcal{W}},$$

the map $\Phi_{-1} := \lambda^{-1/2}\Phi \circ \gamma_{-1} : X \to \mathcal{W}$ is a feature map for K_{-1}. Since γ is a bijective map from X to itself, $P_{\Phi_{-1}}$ is also equal to the identity operator on $\mathcal{W}$. Therefore, by Lemma 5, we have for all $u \in \mathcal{W}$ that

$$\left\|\lambda^{-1/2}(\Phi \circ \gamma_{-1}(\cdot), u)_{\mathcal{W}}\right\|_{\mathcal{H}_{K_{-1}}} = \|u\|_{\mathcal{W}}. \tag{23}$$

Likewise, condition (18) and Lemma 5 imply that

$$\|(\Phi(\cdot), v_u)_{\mathcal{W}}\|_{\mathcal{H}_K} = \|v_u\|_{\mathcal{W}}. \tag{24}$$

In addition, by the relation $\mathcal{H}_{K_{-1}} \preceq \mathcal{H}_K$ and (22), there holds

$$\left\|\lambda^{-1/2}(\Phi \circ \gamma_{-1}(\cdot), u)_{\mathcal{W}}\right\|_{\mathcal{H}_{K_{-1}}} = \|(\Phi(\cdot), v_u)_{\mathcal{W}}\|_{\mathcal{H}_K}. \tag{25}$$

Combining Equations (23), (24) and (25) shows that A is isometric. By (22) we conclude for all $u \in \mathcal{W}$ that

$$\lambda^{-1/2}(\Phi \circ \gamma_{-1}(\cdot), u)_{\mathcal{W}} = (\Phi(\cdot), Au)_{\mathcal{W}} = (A^*\Phi(\cdot), u)_{\mathcal{W}}.$$

We choose $T := A^*$ and observe from the above equation that (20) holds. Thus, Φ is γ-refinable and T^* is isometric. ∎

3. Wavelet-like Reproducing Kernels

This section is devoted to developing a multiscale decomposition of the RKHS $\mathcal{H}_K$ of a γ-refinable kernel K. Specifically, we construct the nontrivial orthogonal complement of $\mathcal{H}_{K_n}$ in $\mathcal{H}_{K_{n+1}}$. In this regard, an issue important to us is when $\mathcal{H}_{K_n}$ is a proper subspace of $\mathcal{H}_{K_{n+1}}$. Our first result concerns this proper inclusion question. Let $\mathcal{R}(A)$ and $\mathcal{N}(A)$ denote the range and null space of an operator A on $\mathcal{W}$, respectively. We also denote for every $\mathcal{V} \subseteq \mathcal{W}$ by $\mathcal{V}^\perp$ the set of all elements in $\mathcal{W}$ that are orthogonal to $\mathcal{V}$.

Theorem 7 *Suppose that K defined by (16) is γ-refinable and the feature map Φ satisfies (18). Then $\mathcal{H}_{K_{-1}}$ is a proper subspace of $\mathcal{H}_K$ if and only if the operator T in (20) is not injective. Moreover, if $\mathcal{H}_{K_{-1}}$ is a proper subspace of $\mathcal{H}_K$, then for all $n \in \mathbb{Z}$, $\mathcal{H}_{K_n}$ is a proper subspace of $\mathcal{H}_{K_{n+1}}$.*

Proof Since K is γ-refinable, by Theorem 6, the feature map Φ is γ-refinable and T^* is isometric. Hence, by (20), functions in $\mathcal{H}_{K_{-1}}$ are of the form

$$\lambda^{-1/2}(\Phi \circ \gamma_{-1}(\cdot), u)_{\mathcal{W}} = (T\Phi(\cdot), u)_{\mathcal{W}} = (\Phi(\cdot), T^*u)_{\mathcal{W}}, \;\; u \in \mathcal{W}. \tag{26}$$

On the other hand, Lemma 5 ensures that functions in $\mathcal{H}_K$ have the form

$$(\Phi(\cdot), u)_{\mathcal{W}}, \;\; u \in \mathcal{W}. \tag{27}$$

The isometry of T^* guarantees that $\mathcal{R}(T^*)$ is a closed subspace of $\mathcal{W}$. This fact, together with Equations (26), (27) and the isomorphism Γ introduced in (19), implies that $\mathcal{H}_{K_{-1}}$ is a proper subspace of $\mathcal{H}_K$ if and only if $\mathcal{R}(T^*)$ is a proper subspace of $\mathcal{W}$. The relation $\mathcal{R}(T^*)^\perp = \mathcal{N}(T)$ (see, Conway, 1990, page 35) proves that $\mathcal{H}_{K_{-1}}$ is a proper subspace of $\mathcal{H}_K$ if and only if $\mathcal{N}(T) \neq \{0\}$. Hence, the first claim of this theorem is valid. The proof of the second statement is straightforward. ∎

Theorem 7 allows us to construct the nontrivial orthogonal complement of $\mathcal{H}_{K_n}$ in $\mathcal{H}_{K_{n+1}}$. For this purpose, we define

$$G := K_1 - K, \;\text{ and }\; G_n(x,y) := \lambda^n G(\gamma_n(x), \gamma_n(y)), \;\; x,y \in X, \;\; n \in \mathbb{Z}.$$

Theorem 8 *Suppose that K is a γ-refinable kernel on X. Then the following statements hold:*

(1) For each $n \in \mathbb{Z}$, G_n is a kernel on X.

(2) There holds $\mathcal{H}_{G_n} \preceq \mathcal{H}_{K_{n+1}}$, and $\mathcal{H}_{G_n}$ is the orthogonal complement of $\mathcal{H}_{K_n}$ in $\mathcal{H}_{K_{n+1}}$.

(3) For each $n \in \mathbb{Z}$ that

$$\mathcal{H}_{G_n} = \{f \circ \gamma_n : f \in \mathcal{H}_G\} \tag{28}$$

and the inner product on $\mathcal{H}_{G_n}$ satisfies

$$(f,g)_{\mathcal{H}_{G_n}} = \lambda^{-n}(f \circ \gamma_{-n}, g \circ \gamma_{-n})_{\mathcal{H}_G}, \;\; f,g \in \mathcal{H}_{G_n}. \tag{29}$$

Proof Since K is γ-refinable, we have by Proposition 2 that K_n is γ-refinable, namely, $\mathcal{H}_{K_n} \preceq \mathcal{H}_{K_{n+1}}$. Let $\mathcal{W}_n$ be the orthogonal complement of $\mathcal{H}_{K_n}$ in $\mathcal{H}_{K_{n+1}}$. It is clear that $\mathcal{W}_n$ is a RKHS with the inner product of $\mathcal{H}_{K_{n+1}}$. By a property of reproducing kernels (see, Aronszajn, 1950, page 345), the sum of K_n and the kernel of $\mathcal{W}_n$ is equal to K_{n+1}. Therefore, G is the kernel of $\mathcal{W}_0$, and it is observed by (8) that the kernel of $\mathcal{W}_n$ is G_n. This proves (1) and (2). The result (3) follows directly from Theorem 1. ■

A direct consequence of Theorem 8 (2) is that for all $n \in \mathbb{Z}$ and for all $f, g \in \mathcal{H}_{G_n}$,

$$(f,g)_{\mathcal{H}_{G_n}} = (f,g)_{\mathcal{H}_{K_{n+1}}}.$$

Theorem 8 leads to the decomposition

$$\mathcal{H}_{K_{n+1}} = \mathcal{H}_{K_n} \oplus \mathcal{H}_{G_n},$$

where the notation $A \oplus B$ denotes the orthogonal direct sum of A and B. We call G_n the *wavelet-like reproducing kernels* and in particular, G the initial wavelet-like kernel. It is clear that the initial wavelet-like kernel G is nontrivial if and only if $\mathcal{H}_{K_{-1}}$ is a proper subspace of $\mathcal{H}_K$. Repeatedly using the above decomposition with $n = -1$, we have the decomposition for the RKHS

$$\mathcal{H}_K = \mathcal{H}_{G_{-1}} \oplus \cdots \oplus \mathcal{H}_{G_{-m}} \oplus \mathcal{H}_{K_{-m}}, \quad m \geq 1. \tag{30}$$

One should notice the difference between wavelet-like reproducing kernels that we introduce here and the "wavelet kernels" studied in Amato et al. (2006), Rakotomamonjy and Canu (2005), and Rakotomamonjy et al. (2005). The latter are a class of Hilbert-Schmidt kernels defined as a superposition of dilations and translations of a wavelet function.

We now consider the decomposition (30) when $m \to \infty$. To this end, we define the space

$$\mathcal{H}_{-\infty} := \bigcap_{n \in \mathbb{Z}} \mathcal{H}_{K_n}$$

and we describe the space in the next theorem.

Theorem 9 *Suppose that K defined by (16) is γ-refinable and the feature map Φ satisfies (18). Then the closed subspace $\mathcal{H}_{-\infty}$ of $\mathcal{H}_K$ has the form*

$$\mathcal{H}_{-\infty} = \left\{ (\Phi(\cdot), u)_{\mathcal{W}} : u \in \bigcap_{n \in \mathbb{N}} \mathcal{R}((T^*)^n) \right\}. \tag{31}$$

Moreover, $\mathcal{H}_{-\infty} = \{0\}$ if and only if

$$\lim_{n \to \infty} \|T^n u\|_{\mathcal{W}} = 0, \quad \textit{for all } u \in \mathcal{W}. \tag{32}$$

Proof Since $\mathcal{H}_{K_n}$, $n \leq 0$, are closed subspaces of $\mathcal{H}_K$, $\mathcal{H}_{-\infty}$ is a closed subspace of $\mathcal{H}_K$. By (20), we use induction to conclude for each $n \in \mathbb{N}$ that

$$\lambda^{-n/2}(\Phi \circ \gamma_{-n}(\cdot), u)_{\mathcal{W}} = (T^n \Phi(\cdot), u)_{\mathcal{W}} = (\Phi(\cdot), (T^*)^n u)_{\mathcal{W}}, \quad u \in \mathcal{W}. \tag{33}$$

By Theorem 1 and Equation (27), functions in $\mathcal{H}_{K_{-n}}$ are of the form $\lambda^{-n/2}(\Phi \circ \gamma_{-n}(\cdot), u)_{\mathcal{W}}$. This combined with (33) proves formula (31).

It remains to prove the second statement. Since for each $n \in \mathbb{N}$, $(T^*)^n$ is isometric, $\mathcal{R}((T^*)^n)$ is a closed subspace of $\mathcal{W}$. Therefore,

$$\mathcal{W}_{-\infty} := \bigcap_{n\in\mathbb{N}} \mathcal{R}((T^*)^n)$$

is a closed subspace of $\mathcal{W}$. It suffices to show that $\mathcal{W}_{-\infty} = \{0\}$ if and only if (32) holds. Suppose that (32) is satisfied. Let $v \in \mathcal{W}_{-\infty}$ and by the definition of $\mathcal{W}_{-\infty}$, there exists for each $n \in \mathbb{N}$ a $v_n \in \mathcal{W}$ such that $(T^*)^n v_n = v$. Since T^* is isometric, $\|v_n\|_{\mathcal{W}} = \|v\|_{\mathcal{W}}$, which ensures that for each $u \in \mathcal{W}$ and $n \in \mathbb{N}$,

$$|(u,v)_{\mathcal{W}}| = |(u,(T^*)^n v_n)_{\mathcal{W}}| = |(T^n u, v_n)_{\mathcal{W}}| \le \|T^n u\|_{\mathcal{W}} \|v_n\|_{\mathcal{W}} = \|T^n u\|_{\mathcal{W}} \|v\|_{\mathcal{W}}.$$

Let $n \to \infty$ in the above inequality and by condition (32) we conclude that each $u \in \mathcal{W}$ is orthogonal to $\mathcal{W}_{-\infty}$. Consequently, $\mathcal{W}_{-\infty}$ contains only the zero element.

Conversely, we suppose that $\mathcal{W}_{-\infty} = \{0\}$. By the relation that

$$\bigcap_{n\in\mathbb{N}} \mathcal{R}((T^*)^n) = \left(\bigcup_{n\in\mathbb{N}} \mathcal{N}(T^n)\right)^{\perp}, \tag{34}$$

the union of $\mathcal{N}(T^n)$, $n \in \mathbb{N}$, is dense in $\mathcal{W}$. Let $u \in \mathcal{W}$. For each $\varepsilon > 0$ there exists an $m \in \mathbb{N}$ and $v \in \mathcal{N}(T^m)$ such that $\|u - v\|_{\mathcal{W}} \le \varepsilon$. For each operator A on $\mathcal{W}$, its norm $\|A\|$ is defined as

$$\|A\| := \sup\{\|Aw\|_{\mathcal{W}} : w \in \mathcal{W},\ \|w\|_{\mathcal{W}} = 1\}.$$

An operator on $\mathcal{W}$ has the same norm as its adjoint (Conway, 1990). Since T^* is isometric, we have $\|T\| = \|T^*\| = 1$. By the definition of the norm of an operator on $\mathcal{W}$, there holds

$$\|Tw\|_{\mathcal{W}} \le \|w\|_{\mathcal{W}},\ \ w \in \mathcal{W},$$

We get from the above equation for all $n \ge m$ that $T^n v = 0$ and

$$\|T^n u\|_{\mathcal{W}} = \|T^n u - T^n v\|_{\mathcal{W}} \le \|u - v\|_{\mathcal{W}} \le \varepsilon.$$

This verifies (32) and completes the proof. ∎

The decomposition (30) can now be extended to the decomposition

$$\mathcal{H}_K = (\mathcal{H}_{-\infty}) \bigoplus_{n\in\mathbb{N}} \mathcal{H}_{G_{-n}}. \tag{35}$$

This decomposition gives a multiresolution analysis (Mallat, 1989; Meyer, 1992) of the RKHS $\mathcal{H}_K$, in terms of a sequence of orthogonal subspaces, each of which is a RKHS corresponding to the wavelet-like kernels.

In passing, we make an additional remark on condition (32). It has a close relation with the translation invariant subspaces in Hardy spaces (Beurling, 1949), which in turn has an important application to the Bedrosian identity (Yu and Zhang, 2006). Under the assumption that the linear span of the eigenelements of T is dense in $\mathcal{W}$, the condition is equivalent to that all the eigenvalues of T have the absolute value less than one (see Beurling, 1949, and the references therein).

To close this section, we prove a corollary of Theorem 7, which concerns the finite dimensional feature space and presents an example of *trivial* refinable kernels, in the sense that its wavelet-like kernel is the zero kernel.

Corollary 10 *If K defined by (16) is γ-refinable, the feature map Φ satisfies (18), and the feature space $\mathcal{W}$ is finite dimensional, then $\mathcal{H}_{K_{-1}} = \mathcal{H}_K$.*

Proof Since K is γ-refinable, by Theorem 6, T^* is isometric, or equivalently, TT^* is equal to the identity operator on $\mathcal{W}$. It follows that for every $w \in \mathcal{W}$, there holds $w = T(T^*w)$. Therefore, T is a surjective operator on $\mathcal{W}$. Since $\mathcal{W}$ is finite dimensional, T must be injective as well. By Theorem 7, there holds $\mathcal{H}_{K_{-1}} = \mathcal{H}_K$. ■

As an application of Corollary 10, we investigate the finite *dot-product kernel* (FitzGerald et al., 1995). Set $\mathbb{Z}_+ := \mathbb{N} \cup \{0\}$, and $\mathbb{A}_n := \{\alpha := (\alpha_j : j \in \mathbb{N}_d) \in \mathbb{Z}_+^d : \sum_{j \in \mathbb{N}_d} \alpha_j = n\}$, $n \in \mathbb{Z}_+$. It can be seen that the kernel

$$K(x,y) := \sum_{\alpha \in \mathbb{A}_n} c_\alpha x^\alpha y^\alpha, \ \ x,y \in \mathbb{R}^d,$$

is refinable on $\mathbb{R}^d$, where c_α, $\alpha \in \mathbb{A}_n$, are positive constants. A feature map Φ for this kernel is

$$\Phi(x) := [\sqrt{c_\alpha} x^\alpha : \alpha \in \mathbb{A}_n] \in \ell^2(\mathbb{A}_n), \ \ x \in \mathbb{R}^d$$

and the feature space $\ell^2(\mathbb{A}_n)$ is of finite dimension. By Corollary 10, K is a trivial refinable kernel on $\mathbb{R}^d$ in the sense that $\mathcal{H}_{K_{-1}} = \mathcal{H}_K$. Examples of nontrivial refinable kernels on $\mathbb{R}^d$ will be given in Sections 5 and 6.

4. γ-Refinable Kernels of a Riesz Type

The study in this section is motivated by the representation (6), which indicates a need of a countable subset $\mathcal{X}$ of X such that the linear span of $K_\mathcal{X} := \{K(\cdot,x) : x \in \mathcal{X}\}$ is dense in $\mathcal{H}_K$. In practical computations, it is also desirable to have a convenient and stable way of finding an approximation $\tilde{f} \in \operatorname{span} K_\mathcal{X}$ of a function $f \in \mathcal{H}_K$. This leads to consideration of requiring $K_\mathcal{X}$ to be a frame or a Riesz basis for $\mathcal{H}_K$.

We recall the basic concept of frames and Riesz bases (cf., Daubechies, 1992; Duffin and Schaeffer, 1952; Mallat, 1998; Young, 1980). Let J be an index set. A family of elements $\{\varphi_j : j \in J\}$ in a Hilbert space $\mathcal{W}$ forms a *frame* if there exist $0 < \alpha \le \beta < \infty$ such that for all $f \in \mathcal{W}$

$$\alpha \|f\|_\mathcal{W}^2 \le \sum_{j \in J} |(f,\varphi_j)_\mathcal{W}|^2 \le \beta \|f\|_\mathcal{W}^2. \tag{36}$$

The constants α, β are called the *frame bounds* for $\{\varphi_j : j \in J\}$. We call a frame $\{\varphi_j : j \in J\}$ a *tight frame* when its two frame bounds are equal, that is, $\alpha = \beta$. If, in addition to (36), $\{\varphi_j : j \in J\}$ is a linearly independent set, we call it a *Riesz basis* for $\mathcal{W}$. A Riesz basis $\{\varphi_j : j \in J\}$ is *equivalent* to an orthonormal basis $\{\psi_j : j \in J\}$ for $\mathcal{W}$, namely, there exists a bounded linear operator L on $\mathcal{W}$ having a bounded inverse such that $L(\psi_j) = \varphi_j$, $j \in J$.

An arbitrary element in $\mathcal{W}$ can be expressed as a linear combination of a frame $\{\varphi_j : j \in J\}$ for $\mathcal{W}$. We denote by $\ell^2(J)$ the Hilbert space of the square summable sequences on J with the inner product $(c,d)_{\ell^2(J)} := \sum_{j \in J} c_j \overline{d_j}$. Define the *frame operator* F from $\mathcal{W}$ to $\ell^2(J)$ by setting for all $f \in \mathcal{W}$, $(Ff)_j := (f,\varphi_j)_\mathcal{W}$, for all $j \in J$. Then F^*F is a bounded positive self-adjoint operator on $\mathcal{W}$ with a bounded inverse and

$$\left\{\tilde{\varphi}_j := (F^*F)^{-1}\varphi_j : j \in J\right\}$$

is a frame for $\mathcal{W}$. We shall refer to it as the *dual frame* of $\{\varphi_j : j \in J\}$ since we have for all $f \in \mathcal{W}$

$$f = \sum_{j\in J}(f,\varphi_j)_{\mathcal{W}}\tilde{\varphi}_j = \sum_{j\in J}(f,\tilde{\varphi}_j)_{\mathcal{W}}\varphi_j. \tag{37}$$

When $\{\varphi_j : j \in J\}$ is a Riesz basis, one can see from (37) that $(\varphi_j,\tilde{\varphi}_k)_{\mathcal{W}} = \delta_{j,k}$, $j,k \in J$. We hence say in this case that $\{\varphi_j : j \in J\}$ and $\{\tilde{\varphi}_j : j \in J\}$ constitute a pair of *biorthogonal bases* for $\mathcal{W}$.

Let $\mathcal{Z} := \{z_j : j \in J\} \subseteq X$ be a countable set. For a kernel K on X, we are interested in the conditions for which the set $K_{\mathcal{Z}} := \{K(\cdot,z_j) : j \in J\}$ is a Riesz basis for $\mathcal{H}_K$. Let us begin with a simple necessary and sufficient condition which follows directly from the reproducing property (5) and the definition of a Riesz basis.

Proposition 11 *The family $K_{\mathcal{Z}}$ is a Riesz basis for $\mathcal{H}_K$ with frame bounds $0 < \alpha \le \beta < \infty$ if and only if for every finite subset $\mathcal{X} \subseteq \mathcal{Z}$, $K[\mathcal{X}]$ is nonsingular, and for all $f \in \mathcal{H}_K$*

$$\alpha\|f\|^2_{\mathcal{H}_K} \le \sum_{j\in J}|f(z_j)|^2 \le \beta\|f\|^2_{\mathcal{H}_K}.$$

We remark that Proposition 11 is closely related to the concept of universal kernels. If X is a topological space, we call a kernel K on X a *universal kernel* if for all compact $\mathcal{X} \subseteq X$, the linear span of $\{K(\cdot,y) : y \in \mathcal{X}\}$ is dense in $C(\mathcal{X})$, the Banach space of continuous functions on $\mathcal{X}$. Various characterizations of universal kernels are studied in Micchelli et al. (2006). By a result of Zhou (2003), universal kernels have the property that for all finite subsets $\mathbf{x}$ of X, $K[\mathbf{x}]$ is nonsingular.

We next present a characterization for $K_{\mathcal{Z}}$ to be a Riesz basis for $\mathcal{H}_K$ in terms of a uniqueness set $\mathcal{Z}$. We call $\mathcal{X} \subseteq X$ a *uniqueness set* for $\mathcal{H}_K$ if there is not a nontrivial $f \in \mathcal{H}_K$ that vanishes on $\mathcal{X}$ (Micchelli et al., 2003, 2006). We shall also need the matrix

$$\Lambda := [K(z_j,z_k) : j,k \in J]. \tag{38}$$

Note that each bounded operator $\mathcal{A} : \ell^2(J) \to \ell^2(J)$ can be represented via a unique matrix $[\mathcal{A}_{j,k} : j,k \in J]$ as

$$(\mathcal{A}c)_j := \sum_{k\in J}\mathcal{A}_{j,k}c_k,\ \ c \in \ell^2(J),\ \ j \in J.$$

For simplicity, we shall not distinguish a linear operator on $\ell^2(J)$ from its corresponding representation matrix. The matrix associated with the adjoint operator $\mathcal{A}^*$ of $\mathcal{A}$ is

$$(\mathcal{A}^*)_{j,k} := \overline{\mathcal{A}_{k,j}},\ \ j,k \in J.$$

We also denote by $\mathcal{A}^T$ and $\overline{\mathcal{A}}$ the transpose and conjugate of a matrix $\mathcal{A}$, respectively, namely,

$$(\mathcal{A}^T)_{j,k} = \mathcal{A}_{k,j},\ \ (\overline{\mathcal{A}})_{j,k} = \overline{\mathcal{A}_{j,k}},\ \ j,k \in J.$$

Proposition 12 *The family $K_{\mathcal{Z}}$ forms a Riesz basis for $\mathcal{H}_K$ if and only if $\mathcal{Z}$ is a uniqueness set for $\mathcal{H}_K$ and there exist $0 < \alpha \le \beta < \infty$ such that*

$$\alpha\|c\|^2_{\ell^2(J)} \le (\Lambda c,c)_{\ell^2(J)} \le \beta\|c\|^2_{\ell^2(J)}. \tag{39}$$

Proof The proof follows directly from the fact (see, for example, Young, 1980, page 32) that for $\{\varphi_j : j \in J\}$ to be a Riesz basis for a Hilbert space $\mathcal{W}$ with frame bounds $0 < \alpha \le \beta < \infty$, it is necessary and sufficient that its linear span is dense in $\mathcal{W}$ and for all $c \in \ell^2(J)$

$$\alpha\|c\|^2_{\ell^2(J)} \le \left\|\sum_{j\in J} c_j\varphi_j\right\|^2_{\mathcal{W}} \le \beta\|c\|^2_{\ell^2(J)}.$$

■

The third characterization is in terms of a feature map for the kernel K.

Theorem 13 *Let K be given by (16). Then $K_{\mathcal{Z}}$ is a Riesz basis for $\mathcal{H}_K$ if and only if $\Phi(\mathcal{Z})$ is a Riesz basis for $\overline{\mathrm{span}}\,\Phi(X)$.*

Proof We only discuss the case when (18) holds true, for the other can be handled in a similar way. Since the operator Γ defined by (19) is an isomorphism from $\mathcal{H}_K$ onto $\mathcal{W}$, $K_{\mathcal{Z}}$ is a Riesz basis if and only if $\Gamma(K_{\mathcal{Z}})$ is a Riesz basis for $\Gamma(\mathcal{H}_K) = \mathcal{W}$. By (16) and the definition of Γ, $\Gamma(K_{\mathcal{Z}}) = \Phi(\mathcal{Z})$. This completes the proof. ■

The following corollary is a direct consequence of Theorem 13.

Corollary 14 *Let K be given by (16). If (18) holds then $K_{\mathcal{Z}}$ is a Riesz basis for $\mathcal{H}_K$ if and only if the features $\Phi(\mathcal{Z})$ of $\mathcal{Z}$ is a Riesz basis for the feature space $\mathcal{W}$.*

For the simplicity of notations, we set

$$\phi_{n,j} := \lambda^{n/2}K(\gamma_n(\cdot), z_j), \ \ (n, j) \in \mathbb{Z}\times J.$$

Note that $\phi_{0,j} = K(\cdot,z_j)$, $j \in J$. In the following presentation, we shall adopt the convention that we use m,n to denote integers and j,k,l to denote indices in J. The next result regards the sequence $\phi_{n,j}$ being a Riesz basis for $\mathcal{H}_{K_n}$. Since the proof is standard (cf., Daubechies, 1992), we omit the details.

Proposition 15 *Suppose that $K_{\mathcal{Z}}$ is a Riesz basis for $\mathcal{H}_K$ with frame bounds $0 < \alpha \le \beta < \infty$. Then for each $n \in \mathbb{Z}$, $\{\phi_{n,j} : j \in J\}$ is a Riesz basis for $\mathcal{H}_{K_n}$ with the same frame bounds α, β.*

For the rest of this section we assume that $K_{\mathcal{Z}}$ is a Riesz basis for $\mathcal{H}_K$. This assumption implies that the linear operator S on $\mathcal{H}_K$ defined by

$$Sf := \sum_{j\in J} f(z_j)K(\cdot,z_j), \ \ f \in \mathcal{H}_K \tag{40}$$

is bounded positive self-adjoint and so is its inverse operator S^{-1} on $\mathcal{H}_K$ (see, for example, Daubechies, 1992, pages 58–59). Note that S is a special case of the operator F^*F introduced at the beginning of this section. A particular example of the operator S was studied in Smale and Zhou (2007) to approximate integral operators.

Proposition 16 *Suppose that $K_{\mathcal{Z}}$ is a Riesz basis for $\mathcal{H}_K$ and let S be the operator on $\mathcal{H}_K$ defined by (40). Then the function*

$$\tilde{K}(x,y) := (S^{-1}K(\cdot,y))(x), \ \ x,y \in X \tag{41}$$

is a kernel on X and the corresponding RKHS is

$$\mathcal{H}_{\tilde{K}} = \{f : f \in \mathcal{H}_K\} \tag{42}$$

with inner product $(f,g)_{\mathcal{H}_{\tilde{K}}} := (Sf,g)_{\mathcal{H}_K}$.

Proof Since for all $x,y \in X$

$$(S^{-1}K(\cdot,y))(x) = (S^{-1}(K(\cdot,y)),K(\cdot,x))_{\mathcal{H}_K}$$

and S^{-1} is positive self-adjoint, the function $\tilde{K}$ defined by (41) is a kernel on X. Clearly, $\mathcal{W} := \{f : f \in \mathcal{H}_K\}$ with inner product $(f,g)_{\mathcal{W}} := (Sf,g)_{\mathcal{H}_K}$ is a RKHS. We also observe by (5) and the fact that $S = S^*$ for each $f \in \mathcal{H}_K$ and $x \in X$ that

$$f(x) = (f,K(\cdot,x))_{\mathcal{H}_K} = (Sf,S^{-1}(K(\cdot,x)))_{\mathcal{H}_K} = (f,S^{-1}(K(\cdot,x)))_{\mathcal{W}} = (f,\tilde{K}(\cdot,x))_{\mathcal{W}},$$

which implies that $\tilde{K}$ is the kernel of $\mathcal{W}$. Consequently, we have (42). ∎

We call $\tilde{K}$ defined by (41) the *dual kernel* of K. By the general theory of frames introduced at the beginning of this section, the dual Riesz basis of $K_{\mathcal{Z}}$ is

$$\tilde{K}_{\mathcal{Z}} := \{\tilde{K}(\cdot,z_j) : j \in J\}. \tag{43}$$

To obtain an explicit expression of this dual basis, we need to understand the operator S^{-1}. We shall use notation $\tilde{\Lambda}$ for the inverse Λ^{-1} of matrix Λ defined by (38).

Theorem 17 *If $K_{\mathcal{Z}}$ is a Riesz basis for $\mathcal{H}_K$, then for each $n \in \mathbb{Z}$, the dual Riesz basis $\{\tilde{\phi}_{n,j} : j \in J\}$ of $\{\phi_{n,j} : j \in J\}$ for $\mathcal{H}_{K_n}$ has the form*

$$\tilde{\phi}_{n,j} = \sum_{k \in J} \tilde{\Lambda}_{k,j}\phi_{n,k}, \ \ j \in J \tag{44}$$

and

$$S^{-1}f = \sum_{j \in J}\left(\sum_{k \in J}\tilde{\Lambda}_{j,k}f(z_k)\right)\left(\sum_{l \in J}\tilde{\Lambda}_{l,j}K(\cdot,z_l)\right), \ \ f \in \mathcal{H}_K. \tag{45}$$

Proof Since $K_{\mathcal{Z}}$ is a Riesz basis for $\mathcal{H}_K$, by Proposition 12, (39) holds for some positive constants α,β. We hence have for all $c \in \ell^2(J)$ that (see, for example, Daubechies, 1992, page 58)

$$\beta^{-1}\|c\|^2_{\ell^2(J)} \leq (\tilde{\Lambda}c,c)_{\ell^2(J)} \leq \alpha^{-1}\|c\|^2_{\ell^2(J)}. \tag{46}$$

Moreover, by Proposition 15, $\{\phi_{n,j} : j \in J\}$ is a Riesz basis for $\mathcal{H}_{K_n}$. This combined with (46) shows that $\tilde{\phi}_{n,j}$ defined by (44) belong to $\mathcal{H}_{K_n}$. By (10) and (5), we have for all $j,k \in J$ that

$$(\tilde{\phi}_{n,j},\phi_{n,k})_{\mathcal{H}_{K_n}} = (\tilde{\phi}_{0,j},K(\cdot,z_k))_{\mathcal{H}_K} = \tilde{\phi}_{0,j}(z_k) = \sum_{l \in J}\tilde{\Lambda}_{l,j}K(z_k,z_l) = \sum_{l \in J}\Lambda_{k,l}\tilde{\Lambda}_{l,j} = \delta_{j,k},$$

which shows that $\{\tilde{\phi}_{n,j} : j \in J\}$ is the dual Riesz basis of $\{\phi_{n,j} : j \in J\}$ in $\mathcal{H}_{K_n}$.

We next establish the representation (45) of S^{-1}. It follows that for each $f \in \mathcal{H}_K$ the function (denoted by g) in the right hand side of (45) is in $\mathcal{H}_K$. Recalling the definition of matrix $\tilde{\Lambda}$, the function g satisfies for $j \in J$ that

$$g(z_j) = \sum_{l \in J} \tilde{\Lambda}_{j,l} f(z_l).$$

As a consequence, we have by the definition (40) of S for all $k \in J$ that

$$(Sg)(z_k) = \sum_{j \in J} \left(\sum_{l \in J} \tilde{\Lambda}_{j,l} f(z_l) \right) K(z_k, z_j) = \sum_{l \in J} f(z_l) \sum_{j \in J} \tilde{\Lambda}_{j,l} \Lambda_{k,j} = \sum_{l \in J} f(z_l) \delta_{k,l} = f(z_k).$$

Since, by Proposition 12, $\mathcal{Z}$ is a uniqueness set for $\mathcal{H}_K$, we have $Sg = f$. ■

We remark that the functions defined by Equation (44) satisfy

$$\tilde{\phi}_{n,j} = \lambda^{n/2} \tilde{\phi}_{0,j} \circ \gamma_n, \quad j \in J, \ n \in \mathbb{Z}.$$

Another implication of Theorem 17 is that $\tilde{\phi}_{0,j}$, $j \in J$, are the *interpolating functions* on $\mathcal{Z}$, that is,

$$\tilde{\phi}_{0,j}(z_k) = \delta_{j,k}, \quad j,k \in J.$$

For each $n \in \mathbb{Z}$ we introduce the sampling operator $I_{n,J}$ by

$$(I_{n,J} f)_j := \lambda^{-n/2} f(\gamma_{-n}(z_j)), \quad j \in J, \quad f \in \mathcal{H}_{K_n}.$$

It is pointed out that a special case of $I_{0,J}$ has been introduced in Smale and Zhou (2007). A function $f \in \mathcal{H}_{K_n}$ can be completely recovered from its sample $I_{n,J} f$, that is,

$$f = \sum_{j \in J} (\tilde{\Lambda} I_{n,J} f)_j \phi_{n,j} = \sum_{j \in J} (I_{n,J} f)_j \tilde{\phi}_{n,j}. \tag{47}$$

In particular, we have the representation for our original kernel K

$$K(x,y) = \sum_{j,k \in J} K(x, z_j) \tilde{\Lambda}_{j,k} K(z_k, y), \quad x,y \in X. \tag{48}$$

The Riesz basis provides a characterization of γ-refinable kernels in terms of the sampling operator, which we present next.

Theorem 18 *Suppose that $K_{\mathcal{Z}}$ is a Riesz basis for $\mathcal{H}_K$. Then K is γ-refinable if and only if*

$$\{I_{-1,J} \phi_{0,j} : j \in J\} \subseteq \ell^2(J), \tag{49}$$

$$(\tilde{\Lambda} I_{-1,J} \phi_{0,j}, I_{-1,J} \phi_{0,k})_{\ell^2(J)} = K(z_k, z_j), \quad j,k \in J, \tag{50}$$

and

$$\phi_{-1,j} = \sum_{k \in J} (\tilde{\Lambda} I_{-1,J} \phi_{0,j})_k K(\cdot, z_k), \quad j \in J. \tag{51}$$

Proof Suppose that conditions (49), (50) and (51) hold true. For $j,k\in J$, we set $\mathcal{C}_{j,k}:=(\tilde{\Lambda}I_{-1,J}\phi_{0,j})_k$. By (49) and (46), $[\mathcal{C}_{j,k}:k\in J]\in\ell^2(J)$. Since $K_{\mathcal{Z}}$ is a Riesz basis for $\mathcal{H}_K$, it follows from (51) that $\phi_{-1,j}\in\mathcal{H}_K$, $j\in J$. Equations (50), (47) and (10) imply for all $j,k\in J$ that

$$(\phi_{-1,j},\phi_{-1,k})_{\mathcal{H}_K}=K(z_k,z_j)=(\phi_{-1,j},\phi_{-1,k})_{\mathcal{H}_{K_{-1}}}. \tag{52}$$

For each $f\in\mathcal{H}_{K_{-1}}$, we have by Proposition 15 a sequence $f_n\in\operatorname{span}\{\phi_{-1,j}:j\in J\}$, $n\in\mathbb{N}$ that converges to f in $\mathcal{H}_{K_{-1}}$. By (52), there holds for each $n\in\mathbb{N}$ that $f_n\in\mathcal{H}_K$ and

$$(f_n,f_n)_{\mathcal{H}_K}=(f_n,f_n)_{\mathcal{H}_{K_{-1}}}. \tag{53}$$

This means that f_n is a Cauchy sequence in $\mathcal{H}_K$. There hence exists $g\in\mathcal{H}_K$ that is the limit of f_n in $\mathcal{H}_K$. We get by (5) for each $x\in X$ that

$$g(x)=(g,K(\cdot,x))_{\mathcal{H}_K}=\lim_{n\to\infty}(f_n,K(\cdot,x))_{\mathcal{H}_K}=\lim_{n\to\infty}f_n(x)=\lim_{n\to\infty}(f_n,K_{-1}(\cdot,x))_{\mathcal{H}_{K_{-1}}}=f(x).$$

Therefore, $f\in\mathcal{H}_K$, and by (53)

$$(f,f)_{\mathcal{H}_{K_{-1}}}=\lim_{n\to\infty}(f_n,f_n)_{\mathcal{H}_{K_{-1}}}=\lim_{n\to\infty}(f_n,f_n)_{\mathcal{H}_K}=(g,g)_{\mathcal{H}_K}=(f,f)_{\mathcal{H}_K}.$$

We conclude that $\mathcal{H}_{K_{-1}}\preceq\mathcal{H}_K$, that is, K is γ-refinable.

Conversely, suppose that $\mathcal{H}_{K_{-1}}\preceq\mathcal{H}_K$. Then since $K_{\mathcal{Z}}$ is a Riesz basis for $\mathcal{H}_K$ with the dual basis $\tilde{K}_{\mathcal{Z}}$ defined by (43), we let $n=0$ and $f=\phi_{-1,j}$ in (47) to get that (49), (51) hold true. The inclusion $\mathcal{H}_{K_{-1}}\preceq\mathcal{H}_K$ also implies (52). Through a calculation, we notice that (50) is a consequence of (52). The proof is complete. ■

In the rest of this section, we construct a frame for the RKHS $\mathcal{H}_G$ of the wavelet-like kernel $G:=K_1-K$.

Lemma 19 *Suppose that K is γ-refinable and $K_{\mathcal{Z}}$ is a Riesz basis for $\mathcal{H}_K$ with frame bounds $0<\alpha\le\beta<\infty$. Then*

$$\psi_{0,j}:=\lambda^{-1/2}G(\cdot,\gamma_{-1}(z_j)),\quad j\in J$$

form a frame for $\mathcal{H}_G$ with the same frame bounds α,β.

Proof By the definition $G=K_1-K$, we have that

$$\psi_{0,j}=\phi_{1,j}-\lambda^{-1/2}K(\cdot,\gamma_{-1}(z_j)),\quad j\in J.$$

Let $f\in\mathcal{H}_G$. Since f is orthogonal to $\mathcal{H}_K$ in $\mathcal{H}_{K_1}$, we have for each $j\in J$ that

$$(f,\psi_{0,j})_{\mathcal{H}_G}=(f,\psi_{0,j})_{\mathcal{H}_{K_1}}=(f,\phi_{1,j})_{\mathcal{H}_{K_1}},$$

where the first equality holds because of Theorem 8 (2). Applying Proposition 15 and $\|f\|_{\mathcal{H}_G}=\|f\|_{\mathcal{H}_{K_1}}$ yields that $\{\psi_{0,j}:j\in J\}$ is a frame for $\mathcal{H}_G$ with frame bounds α,β. ■

The frame for the RKHS $\mathcal{H}_G$ is now translated to a frame for the RKHSs $\mathcal{H}_{G_n}$.

Proposition 20 *Suppose that K is γ-refinable and $K_{\mathcal{Z}}$ is a Riesz basis for $\mathcal{H}_K$ with frame bounds $0<\alpha\le\beta<\infty$. Then for each $n\in\mathbb{Z}$, $\psi_{n,j}:=\lambda^{n/2}\psi_{0,j}\circ\gamma_n$, $j\in J$ form a frame for $\mathcal{H}_{G_n}$ with the same frame bounds α,β. Furthermore, there holds*

$$\psi_{n,j}=\sum_{k\in J}\mathcal{D}_{j,k}\phi_{n+1,k}, \tag{54}$$

where

$$\mathcal{D}_{j,k}:=\delta_{j,k}-\lambda^{-1}\sum_{l\in J}\tilde{\Lambda}_{k,l}K(\gamma_{-1}(z_l),\gamma_{-1}(z_j)). \tag{55}$$

Proof Arguments similar to those in the proof of Proposition 15 with Lemma 19, equations (28) and (29) prove the first claim of this proposition. For each $n\in\mathbb{Z}$, by Theorem 17, $\{\tilde{\phi}_{n+1,k}:k\in J\}$ is the dual Riesz basis of $\{\phi_{n+1,k}:k\in J\}$ for $\mathcal{H}_{K_{n+1}}$. Since $\mathcal{H}_{G_n}\preceq\mathcal{H}_{K_{n+1}}$, we obtain for each $j\in J$

$$\psi_{n,j}=\sum_{k\in J}(\psi_{n,j},\tilde{\phi}_{n+1,k})_{\mathcal{H}_{K_{n+1}}}\phi_{n+1,k}.$$

By (44) we confirm that $(\psi_{n,j},\tilde{\phi}_{n+1,k})_{\mathcal{H}_{K_{n+1}}}$ is equal to $\mathcal{D}_{j,k}$ defined by (55), proving the result. ∎

We next present the reconstruction of a function $f\in\mathcal{H}_{G_n}$ from its samples $(f,\psi_{n,j})_{\mathcal{H}_{G_n}}$, $j\in J$. To describe the reconstruction, we remark that for each $(n,j)\in\mathbb{Z}\times J$ there holds

$$\phi_{n,j}=\sum_{k\in J}\mathcal{C}_{j,k}\phi_{n+1,k} \tag{56}$$

and

$$\tilde{\phi}_{n,j}=\sum_{k\in J}\tilde{\mathcal{C}}_{j,k}\tilde{\phi}_{n+1,k}, \tag{57}$$

where

$$\tilde{\mathcal{C}}_{j,k}:=\lambda^{-1/2}\sum_{l\in J}\tilde{\Lambda}_{l,j}K(\gamma_{-1}(z_k),z_l).$$

For $j,k\in J$, we let

$$\tilde{\mathcal{D}}_{j,k}:=\delta_{j,k}-\sum_{l\in J}\overline{\mathcal{C}_{l,j}}\tilde{\mathcal{C}}_{l,k}.$$

Theorem 21 *Suppose that K is γ-refinable and $K_{\mathcal{Z}}$ is a Riesz basis for $\mathcal{H}_K$. For each $n\in\mathbb{Z}$, the functions*

$$\tilde{\psi}_{n,j}:=\tilde{\phi}_{n+1,j}-\sum_{k\in J}\overline{\mathcal{C}_{k,j}}\tilde{\phi}_{n,k},\quad j\in J \tag{58}$$

constitute a frame for $\mathcal{H}_{G_n}$ and have the representation

$$\tilde{\psi}_{n,j}=\sum_{k\in J}\tilde{\mathcal{D}}_{j,k}\tilde{\phi}_{n+1,k}. \tag{59}$$

Moreover, there holds for all $f\in\mathcal{H}_{G_n}$ that

$$f=\sum_{j\in J}(f,\tilde{\psi}_{n,j})_{\mathcal{H}_{G_n}}\psi_{n,j}=\sum_{j\in J}(f,\psi_{n,j})_{\mathcal{H}_{G_n}}\tilde{\psi}_{n,j}. \tag{60}$$

Proof Since K is γ-refinable, by Proposition 2, $\mathcal{H}_{K_n} \preceq \mathcal{H}_{K_{n+1}}$. This together with Theorem 17 follows that the functions $\tilde{\psi}_{n,j}$ defined by (58) are in $\mathcal{H}_{K_{n+1}}$. Moreover, it can be verified by (56) that they are orthogonal to $\mathcal{H}_{K_n}$. Therefore, $\{\tilde{\psi}_{n,j} : j \in J\} \subseteq \mathcal{H}_{G_n}$. Arguments similar to those in the proof of Proposition 20 yield that $\tilde{\psi}_{n,j}$, $j \in J$ form a frame for $\mathcal{H}_{G_n}$ with the same frame bounds as those of $\{\tilde{\phi}_{0,j} : j \in J\}$ for $\mathcal{H}_K$. Equation (59) is obtained by substituting (57) into (58).

We next prove the first equality of (60). To this end, for any fixed $f \in \mathcal{H}_{G_n}$ we define

$$g := \sum_{j\in J} (f, \tilde{\psi}_{n,j})_{\mathcal{H}_{G_n}} \psi_{n,j}$$

and observe that $g \in \mathcal{H}_{G_n}$. It can be verified that

$$\begin{aligned} g &= \sum_{j\in J} (f, \tilde{\phi}_{n+1,j})_{\mathcal{H}_{K_{n+1}}} \psi_{n,j} \\ &= \sum_{j\in J} (f, \tilde{\phi}_{n+1,j})_{\mathcal{H}_{K_{n+1}}} (\phi_{n+1,j} - \lambda^{-(n+1)/2} K_n(\cdot, \gamma_{-n-1}(z_j))) \\ &= f - \lambda^{-(n+1)/2} \sum_{j\in J} (f, \tilde{\phi}_{n+1,j})_{\mathcal{H}_{K_{n+1}}} K_n(\cdot, \gamma_{-n-1}(z_j)). \end{aligned}$$

The above equation ensures that $g - f \in \mathcal{H}_{G_n} \cap \mathcal{H}_{K_n}$, which implies that $g = f$. Likewise, we may prove the second equality of (60). ■

Suppose that K is a kernel on X and K_n, $n \in \mathbb{Z}$, are defined by (8). The RKHS $\mathcal{H}_K$ is said to have a multiresolution analysis if

$$\cdots \preceq \mathcal{H}_{K_{-2}} \preceq \mathcal{H}_{K_{-1}} \preceq \mathcal{H}_K, \quad \bigcap_{n\in\mathbb{N}} \mathcal{H}_{K_{-n}} = \{0\},$$

and $K_{\mathcal{Z}}$ with a countable set $\mathcal{Z} := \{z_j : j \in J\} \subseteq X$ is a Riesz basis for $\mathcal{H}_K$. The following theorem characterizes a RKHS that has a multiresolution analysis.

Theorem 22 *Let K be a kernel on an input space X with a feature map Φ from X to a Hilbert space $\mathcal{W}$ that satisfies (18). The RKHS $\mathcal{H}_K$ has a multiresolution analysis if and only if Φ is refinable, that is, it satisfies (20) for some bounded linear operator T on $\mathcal{W}$ whose adjoint T^* is isometric, T has the property (32) and there exists a countable subset $\mathcal{Z}$ of X such that $\Phi(\mathcal{Z})$ is a Riesz basis for $\mathcal{W}$.*

Proof The result of this theorem follows directly from Theorems 6, 9 and 13. ■

Suppose that $\mathcal{H}_K$ has a multiresolution analysis. By (35), Lemma 19 and Proposition 20, we have that

$$\mathcal{H}_K = \bigoplus_{n\in\mathbb{N}} \mathcal{H}_{G_{-n}}$$

and $\lambda^{(n-1)/2} G(\gamma_n(\cdot), \gamma_{-1}(z_j))$, $j \in J$ form a frame for $\mathcal{H}_{G_n}$, $n \leq -1$. The multiresolution analysis on $\mathcal{H}_K$ is hence generated by the wavelet-like kernel G.

To close this section, we present the decomposition and reconstruction algorithms. These algorithms are analogues to the Mallat algorithms (cf., Mallat, 1989) in wavelet analysis. They are important for fast computation. With Equations (56), (57), (54) and (59), we now establish a recursive scheme for the decomposition (30). For

$$f \in \bigcup_{m\geq 0} \mathcal{H}_{K_m},$$

we denote by $P_n f$ the orthogonal projection of f onto $\mathcal{H}_{K_n}$, for $n \leq 0$. We have that

$$P_{n+1} f = P_n f + \sum_{j\in J} (f, \tilde{\psi}_{n,j})_{\mathcal{H}_{G_n}} \psi_{n,j} = P_n f + \sum_{j\in J} (f, \psi_{n,j})_{\mathcal{H}_{G_n}} \tilde{\psi}_{n,j}, \ \ n \leq -1.$$

We define four vectors

$$\alpha_n := [(f, \phi_{n,j})_{\mathcal{H}_{K_n}} : j \in J], \ \ \tilde{\alpha}_n := [(f, \tilde{\phi}_{n,j})_{\mathcal{H}_{K_n}} : j \in J],$$

$$\beta_n := [(f, \psi_{n,j})_{\mathcal{H}_{G_n}} : j \in J] \ \text{ and } \ \tilde{\beta}_n := [(f, \tilde{\psi}_{n,j})_{\mathcal{H}_{G_n}} : j \in J].$$

By (47), the projection $P_n f$ is completely described by α_n or $\tilde{\alpha}_n$. Likewise, by (60), the *difference* $P_{n+1} f - P_n f$ between two levels of consecutive projections is completely determined by β_n or $\tilde{\beta}_n$. We introduce the matrix notations:

$$\mathcal{C} := [\mathcal{C}_{j,k} : j,k \in J], \ \tilde{\mathcal{C}} := [\tilde{\mathcal{C}}_{j,k} : j,k \in J], \ \mathcal{D} := [\mathcal{D}_{j,k} : j,k \in J], \ \tilde{\mathcal{D}} := [\tilde{\mathcal{D}}_{j,k} : j,k \in J].$$

Suppose that α_0 is given and for each $n \leq -1$ we then use $\mathcal{C}$ and $\mathcal{D}$ to decompose α_n, $n \leq 0$ recursively to obtain α_n, β_n

$$\alpha_n = \overline{\mathcal{C}} \alpha_{n+1}, \ \ \beta_n = \overline{\mathcal{D}} \alpha_{n+1}. \tag{61}$$

Conversely, α_{n+1} can be reconstructed from α_n and β_n by using $\tilde{\mathcal{C}}$ and $\tilde{\mathcal{D}}$

$$\alpha_{n+1} = \tilde{\mathcal{C}}^T \alpha_n + \tilde{\mathcal{D}}^T \beta_n, \ \ n \leq -1. \tag{62}$$

Alternatively, the decomposition and reconstruction process can start from $\tilde{\alpha}_0$

$$\tilde{\alpha}_n = \overline{\tilde{\mathcal{C}}} \tilde{\alpha}_{n+1}, \ \ \tilde{\beta}_n = \overline{\tilde{\mathcal{D}}} \tilde{\alpha}_{n+1}, \ \ \tilde{\alpha}_{n+1} = \mathcal{C}^T \tilde{\alpha}_n + \mathcal{D}^T \tilde{\beta}_n, \ \ n \leq -1. \tag{63}$$

Since $\{\phi_{0,j} : j \in J\}$ and $\{\tilde{\phi}_{0,j} : j \in J\}$ are Riesz bases for $\mathcal{H}_K$ and Riesz bases are equivalent to orthonormal bases,

$$\{[(f, \phi_{0,j})_{\mathcal{H}_K} : j \in J] : f \in \mathcal{H}_K\} = \{[(f, \tilde{\phi}_{0,j})_{\mathcal{H}_K} : j \in J] : f \in \mathcal{H}_K\} = \ell^2(J),$$

we have by (61), (62) and (63) that

$$\tilde{\mathcal{C}}^T \overline{\mathcal{C}} + \tilde{\mathcal{D}}^T \overline{\mathcal{D}} = \mathcal{C}^T \overline{\tilde{\mathcal{C}}} + \mathcal{D}^T \overline{\tilde{\mathcal{D}}} = I, \tag{64}$$

where I is the identity matrix. We say that $\mathcal{C}, \tilde{\mathcal{C}}, \mathcal{D}, \tilde{\mathcal{D}}$ form a *perfect reconstruction system* if they satisfy (64).

5. Refinable Translation Invariant Kernels

In this section, we consider refinable kernels with specializing our input space to $\mathbb{R}^d$, $d \in \mathbb{N}$, the mapping γ to the dilation mapping $x \to 2x$ in $\mathbb{R}^d$ and kernels to *translation invariant* kernels K on $\mathbb{R}^d$, that is, for all $x, y, a \in \mathbb{R}^d$

$$K(x-a, y-a) = K(x,y).$$

In other words, the main purpose of this section is to characterize refinable translation invariant kernels on $\mathbb{R}^d$.

We need the notation of *Fourier transform* defined for $f \in L^1(\mathbb{R}^d)$ as

$$\hat{f}(t) := \frac{1}{(2\pi)^d}\int_{\mathbb{R}^d} f(x)e^{-i(x,t)}dx,\ \ t \in \mathbb{R}^d,$$

where (x,t) denotes the inner product of x,t in $\mathbb{R}^d$. The Fourier transform $\hat{f}$ of $f \in L^p(\mathbb{R}^d)$, $1 < p \leq \infty$, is defined in the weak sense (Grafakos, 2004). If both f, $\hat{f}$ belong to $L^1(\mathbb{R}^d)$ then there holds

$$f(x) = \int_{\mathbb{R}^d} \hat{f}(t)e^{i(x,t)}dt,\ \ x \in \mathbb{R}^d.$$

Clearly, K is translation invariant if and only if there exists a function $k : \mathbb{R}^d \to \mathbb{C}$ such that

$$K(x,y) = k(x-y),\ \ x,y \in \mathbb{R}^d. \tag{65}$$

Note that in this section k will always denote a function. It was established by Bochner (1959) that if k is continuous on $\mathbb{R}^d$ then (65) defines a kernel if and only if there exists a finite positive Borel measure μ on $\mathbb{R}^d$ such that

$$k(x) = \int_{\mathbb{R}^d} e^{i(x,t)}d\mu(t),\ \ x \in \mathbb{R}^d. \tag{66}$$

We shall consider only measures μ that are *absolutely continuous* with respect to the Lebesgue measure. This means that $\mu(A) = 0$ whenever the Lebesgue measure $|A|$ of a Borel subset $A \subseteq \mathbb{R}^d$ is zero. By the Radon-Nikodym theorem (Rudin, 1987), μ in (66) is absolutely continuous with respect to the Lebesgue measure if and only if $\hat{k}$ is a nonnegative Lebesgue integrable function on $\mathbb{R}^d$.

Specifically, we shall characterize nonnegative $\hat{k} \in L^1(\mathbb{R}^d)$ for which the kernel K given by

$$K(x,y) = k(x-y) = \int_{\mathbb{R}^d} e^{i(x-y,t)}\hat{k}(t)dt,\ \ x,y \in \mathbb{R}^d \tag{67}$$

is refinable, that is, there holds

$$\mathcal{H}_{K_{-1}} \preceq \mathcal{H}_K. \tag{68}$$

Note that in this section K_j, $j \in \mathbb{Z}$ are defined through a positive constant λ as

$$K_j(x,y) = \lambda^j K(2^j x, 2^j y),\ \ x,y \in \mathbb{R}^d. \tag{69}$$

We shall use j to denote integers while m,n,l to denote elements in $\mathbb{Z}^d$. We shall also discuss conditions for $K_{\mathbb{Z}^d} := \{K(\cdot,n) : n \in \mathbb{Z}^d\}$ to be a Riesz basis for $\mathcal{H}_K$.

We next identify a feature map for the kernel K. Let $L^2(\mathbb{R}^d,\hat{k}dt)$ be the space of Borel measurable functions $f:\mathbb{R}^d\to\mathbb{C}$ such that $\int_{\mathbb{R}^d}|f(t)|^2\hat{k}(t)dt<\infty$. It is a Hilbert space with the inner product

$$(f,g)_{L^2(\mathbb{R}^d,\hat{k}dt)}:=\int_{\mathbb{R}^d}f(t)\overline{g(t)}\hat{k}(t)dt.$$

We observe from (67) that

$$K(x,y)=(\Phi(x),\Phi(y))_{L^2(\mathbb{R}^d,\hat{k}dt)},\ \ x,y\in\mathbb{R}^d,$$

where the feature map $\Phi:\mathbb{R}^d\to L^2(\mathbb{R}^d,\hat{k}dt)$ is defined by $\Phi(x)(t):=e^{i(x,t)},\ \ t\in\mathbb{R}^d$. It is clear that $\operatorname{span}\Phi(\mathbb{R}^d)$ is dense in $L^2(\mathbb{R}^d,\hat{k}dt)$. By Lemma 5, the functions in $\mathcal{H}_K$ are of the form

$$f_\Phi:=(\Phi(\cdot),f)_{L^2(\mathbb{R}^d,\hat{k}dt)},\ \ f\in L^2(\mathbb{R}^d,\hat{k}dt) \tag{70}$$

and the inner product on $\mathcal{H}_K$ is given by

$$(f_\Phi,g_\Phi)_{\mathcal{H}_K}=(g,f)_{L^2(\mathbb{R}^d,\hat{k}dt)},\ \ f,g\in L^2(\mathbb{R}^d,\hat{k}dt). \tag{71}$$

We now present a special result on characterization of refinable translation invariant kernels. For this purpose, we set

$$\Omega:=\{t\in\mathbb{R}^d:\hat{k}(t)>0\} \tag{72}$$

and denote by χ_Ω the characteristic function of Ω. We remark that Ω is a Lebesgue measurable subset of $\mathbb{R}^d$.

Theorem 23 *Let K be the translation invariant kernel given in (67) through a nonnegative $\hat{k}\in L^1(\mathbb{R}^d)$ and let Ω be defined by (72). Then K is a refinable kernel on $\mathbb{R}^d$ if and only if*

$$\Omega\subseteq 2\Omega \tag{73}$$

and

$$\hat{k}=\frac{\lambda}{2^d}\chi_\Omega\hat{k}\left(\frac{\cdot}{2}\right). \tag{74}$$

Proof This proof is based on Theorem 4. Through a change of variables, we have for each $y\in\mathbb{R}^d$ that

$$K\left(\frac{x}{2},y\right)=2^d\int_{\mathbb{R}^d}e^{i(x,t)}e^{-i2(y,t)}\hat{k}(2t)dt,\ \ x\in\mathbb{R}^d.$$

It follows by (70) that $K(\frac{\cdot}{2},y)\in\mathcal{H}_K$ for each $y\in\mathbb{R}^d$ if and only if there exists a nonnegative $g\in L^2(\mathbb{R}^d,\hat{k}dt)$ such that

$$\hat{k}(2\cdot)=g\hat{k}. \tag{75}$$

Suppose that (75) is valid. Then it can be verified by the uniqueness of Fourier transforms, and (71) that

$$\left(K\left(\frac{\cdot}{2},y\right),K\left(\frac{\cdot}{2},x\right)\right)_{\mathcal{H}_K}=\lambda K(x,y),\quad \text{for all } x,y\in\mathbb{R}^d$$

if and only if

$$2^dg^2\,\hat{k}=\lambda\hat{k}(2\cdot). \tag{76}$$

Suppose that (73) and (74) are true. We then obtain by hypothesis (74) that equations (75) and (76) hold true for $g = \frac{\lambda}{2^d}\chi_{\frac{\Omega}{2}}$, which, by (73), is contained in $L^2(\mathbb{R}^d, \hat{k}dt)$. Conversely, if (75) and (76) are valid then (73) follows from (75), and (74) is a consequence of (73), (75) and (76). ■

In the next corollary, we prove special properties of a refinable translation invariant kernel.

Corollary 24 *Let K be a refinable translation invariant kernel defined by (67). If $\hat{k}$ is nontrivial, then $\lambda \geq 1$ and if $\Omega = \frac{\Omega}{2} \neq \emptyset$, then $\hat{k}(t)$ is not continuous at $t = 0$.*

Proof Since K is refinable, by Theorem 23, Equations (73) and (74) hold. We then observe by (74) that

$$\int_{\Omega} \hat{k}(t)dt = \lambda \int_{\frac{\Omega}{2}} \hat{k}(t)dt. \tag{77}$$

The inclusion (73) and Equation (77) imply that there must hold $\lambda \geq 1$ since $\hat{k}$ is nontrivial.

We next prove that $\hat{k}(t)$ is not continuous at $t = 0$. Assume to the contrary that $\hat{k}$ is continuous at $t = 0$ and $\Omega = \frac{\Omega}{2} \neq \emptyset$. In this case, we first see from (77) that $\lambda = 1$ and then by (74) for each $t \in \Omega$ that

$$\hat{k}(t) = \lim_{j\to\infty} \frac{\hat{k}(2^{-j}t)}{2^{dj}} = 0$$

because by hypothesis $\hat{k}(t)$ is continuous at $t = 0$. This contradicts the assumption that $\Omega \neq \emptyset$. ■

As a consequence of this corollary, we have the following interesting observation. For an inclusion relation of the RKHSs of Gaussian kernels, see Walder et al. (2007).

Corollary 25 *The Gaussian kernels*

$$G_\sigma(x,y) := \exp\left(-\sigma\|x-y\|^2\right), \ \ x,y \in \mathbb{R}^d, \ \ \sigma > 0,$$

where $\|x\| := (x,x)^{1/2}$, are not refinable.

Proof Since the Gaussian kernels can be represented as

$$G_\sigma(x,y) = \frac{1}{(2\pi)^d}\int_{\mathbb{R}^d}\left(\frac{\pi}{\sigma}\right)^{d/2} e^{i(x-y,t)} e^{-\frac{\|t\|^2}{4\sigma}} dt,$$

and $e^{-\frac{\|\cdot\|^2}{4\sigma}}$, supported on the whole $\mathbb{R}^d$, is clearly continuous at $t = 0$, by Corollary 24, they are not refinable. ■

We now present a nontrivial refinable translation invariant kernel.

Corollary 26 *For $a, b, \sigma \geq 0$, let $\hat{k} := \|\cdot\|^\sigma \chi_{[-a,b]^d}$. Then the kernel K defined by (67) with $\hat{k}$ is refinable with $\Omega = [-a,b]^d \setminus \{0\}$ and $\lambda = 2^{\sigma+d}$.*

Proof It can be verified directly that $\hat{k}$ satisfies condition (73) and (74) with $\Omega=[-a,b]^d\setminus\{0\}$ and $\lambda=2^{\sigma+d}$. Hence, by Theorem 23, K is refinable. ■

We next characterize when $\mathcal{H}_{K_{-1}}$ is a proper subspace of $\mathcal{H}_K$ if K is a refinable translation invariant kernel. For this purpose, we identify the feature map for such a K with

$$\lambda^{-1/2}\Phi\left(\frac{\cdot}{2}\right)=T\Phi,$$

where

$$Tf:=\lambda^{-1/2}f\left(\frac{\cdot}{2}\right)\chi_\Omega,\quad f\in L^2(\mathbb{R}^d,\hat{k}dt). \tag{78}$$

Theorem 27 *Suppose that a kernel K defined by (67) is refinable on $\mathbb{R}^d$ and Ω is defined by (72). Then $\mathcal{H}_{K_{-1}}$ is a proper subspace of $\mathcal{H}_K$ if and only if*

$$|2\Omega-\Omega|>0. \tag{79}$$

If (79) holds true then $\lambda>1$ and

$$\bigcap_{j\in\mathbb{Z}}\mathcal{H}_{K_j}=\{0\}. \tag{80}$$

Proof Set $f\in L^2(\mathbb{R}^d,\hat{k}dt)$. We observe by (78) that $Tf=0$ if and only if $f(x)=0$, a.e. $x\in\frac{\Omega}{2}$. Therefore, $\mathcal{N}(T)$ is nontrivial if and only if (79) is true. The first statement of this theorem hence follows from Theorem 7.

To prove the second statement, we suppose that (79) holds. We prove (80) by verifying condition (32) in Theorem 9. To this end, we note by (78) for each $j\in\mathbb{N}$ that

$$T^jf=\lambda^{-\frac{j}{2}}f\left(\frac{\cdot}{2^j}\right)\chi_\Omega$$

and

$$\hat{k}(2^j\cdot)=\left(\frac{\lambda}{2^d}\right)^j\hat{k}\chi_{\frac{\Omega}{2^j}}. \tag{81}$$

The above two equations imply that

$$\|T^jf\|^2_{L^2(\mathbb{R}^d,\hat{k}dt)}=\int_{\frac{\Omega}{2^j}}|f(t)|^2\hat{k}(t)dt. \tag{82}$$

We integrate both sides of (81) over $\mathbb{R}^d$ to get that

$$\frac{1}{\lambda^j}\int_\Omega\hat{k}(t)dt=\int_{\frac{\Omega}{2^j}}\hat{k}(t)dt.$$

This with (79) implies that $\lambda>1$. Hence,

$$\lim_{j\to\infty}\int_{\frac{\Omega}{2^j}}\hat{k}(t)dt=0,$$

which with (82) ensures that

$$\lim_{j\to\infty}\|T^jf\|_{L^2(\mathbb{R}^d,\hat{k}dt)}=0,$$

proving the result. ∎

Let us turn to establishing conditions for $K_{\mathbb{Z}^d}$ to be a Riesz basis for $\mathcal{H}_K$. We begin with a technical lemma.

Lemma 28 *The family $K_{\mathbb{Z}^d}$ is a Riesz basis for $\mathcal{H}_K$ if and only if $\mathcal{E}_{\mathbb{Z}^d} := \{e^{i(n,t)} : n \in \mathbb{Z}^d\}$ is a Riesz basis for $L^2(\mathbb{R}^d, \hat{k}dt)$.*

Proof This lemma follows from Theorem 13 and the density of $\operatorname{span}\Phi(\mathbb{R}^d)$ in $L^2(\mathbb{R}^d, \hat{k}dt)$. ∎

Our next approach is based on the characterization of Riesz bases mentioned at the beginning of the proof for Proposition 12. We shall use the property of Ω that for all Lebesgue measurable sets $A \subseteq \Omega$ with $|A| > 0$ there exists a $\sigma > 0$ such that $\left|\{t \in A : \hat{k}(t) \geq \sigma\}\right| > 0$.

Lemma 29 *The linear span of $\mathcal{E}_{\mathbb{Z}^d}$ is dense in $L^2(\mathbb{R}^d, \hat{k}dt)$ if and only if*

$$|\Omega \cap (\Omega + 2n\pi)| = 0, \quad n \in \mathbb{Z}^d \setminus \{0\}. \tag{83}$$

Proof Suppose that there exists a nonzero $n \in \mathbb{Z}^d$ such that $|\Omega \cap (\Omega + 2n\pi)| > 0$. For $\sigma_1 > 0$, $a \in \mathbb{R}$, $0 < \delta < \pi$ we set

$$A_1 := \{t : t \in \Omega \cap (\Omega + 2n\pi), \hat{k}(t) \geq \sigma_1\} \cap [a, a+\delta]^d.$$

We can choose some $\sigma_1 > 0$, $a \in \mathbb{R}$ such that A_1 has nonzero Lebesgue measure. Since the set $A_1 - 2n\pi$ is contained in Ω with $|A_1 - 2n\pi| = |A_1| > 0$, we can find a $\sigma_2 > 0$ such that $|A_2| > 0$, where $A_2 := \{t : t \in A_1 - 2n\pi, \hat{k}(t) \geq \sigma_2\}$. Set $A := A_2 + 2n\pi$, $\sigma := \min\{\sigma_1, \sigma_2\}$. The set A so constructed has the properties that $|A| > 0$, $A \cap (A - 2n\pi) = \emptyset$ and $\hat{k}(t) \geq \sigma$, for $t \in A \cup (A - 2n\pi)$. Using this set, we define a function $f \in L^2(\mathbb{R}^d, \hat{k}dt)$ as

$$f(t) := \begin{cases} -1, & t \in A, \\ 1, & t \in A - 2n\pi, \\ 0, & \text{otherwise}. \end{cases}$$

For an arbitrary $g \in \mathcal{E} := \operatorname{span}\mathcal{E}_{\mathbb{Z}^d}$, we have that

$$\begin{aligned} \int_{\mathbb{R}^d} |g(t) - f(t)|^2 \hat{k}(t)dt &\geq \int_A |g(t) - f(t)|^2 \hat{k}(t)dt + \int_{A-2n\pi} |g(t) - f(t)|^2 \hat{k}(t)dt \\ &\geq \sigma \int_A (|g(t)+1|^2 + |g(t)-1|^2)dt \geq \sigma|A|. \end{aligned}$$

This shows that $\mathcal{E}$ would not be dense in $L^2(\mathbb{R}^d, \hat{k}dt)$ if (83) were invalid.

On the other hand, suppose that (83) is true. For $n \in \mathbb{Z}^d$, we define $\tilde{\Omega}_n := (\Omega - 2n\pi) \cap [0, 2\pi]^d$ and observe that these sets satisfy the condition

$$\bigcup\{\tilde{\Omega}_n : n \in \mathbb{Z}^d\} \subseteq [0, 2\pi]^d, \quad |\tilde{\Omega}_n \cap \tilde{\Omega}_m| = 0, \quad n \neq m.$$

Let

$$\rho(t) := \begin{cases} \hat{k}(t + 2n\pi), & t \in \tilde{\Omega}_n, \; n \in \mathbb{Z}^d, \\ 0, & \text{otherwise} \end{cases}$$

and for $f \in L^2(\mathbb{R}^d, \hat{k}dt)$ we introduce a new function $g \in L^2([0,2\pi]^d, \rho dt)$ by setting

$$g(t) := \begin{cases} f(t+2n\pi), & t \in \tilde{\Omega}_n,\ n \in \mathbb{Z}^d, \\ 0, & \text{otherwise.} \end{cases}$$

Since $\mathcal{E}$ is dense in $L^2([0,2\pi]^d, \rho dt)$, for each $\varepsilon > 0$ there exists $\tilde{f} \in \mathcal{E}$ such that

$$\|g - \tilde{f}\|_{L^2([0,2\pi]^d, \rho dt)} < \varepsilon.$$

Note that

$$\|f - \tilde{f}\|_{L^2(\mathbb{R}^d, \hat{k}dt)} = \|g - \tilde{f}\|_{L^2([0,2\pi]^d, \rho dt)}.$$

Combining the two relations above proves the lemma. ■

We next present a necessary and sufficient condition for $K_{\mathbb{Z}^d}$ to be a Riesz basis for $\mathcal{H}_K$ if K is a translation invariant kernel defined by (67) through a nonnegative $\hat{k} \in L^1(\mathbb{R}^d)$.

Theorem 30 *Let K be a translation invariant kernel defined by (67) through a nonnegative $\hat{k} \in L^1(\mathbb{R}^d)$ and Ω be defined by (72). Then $K_{\mathbb{Z}^d}$ is a Riesz basis for $\mathcal{H}_K$ if and only if*

$$\sum_{n \in \mathbb{Z}^d} \chi_\Omega(\cdot + 2n\pi) = 1, \ \ a.e. \tag{84}$$

and there exist $0 < \alpha \le \beta < \infty$ such that

$$\alpha \le \hat{k}(t) \le \beta, \quad a.e.\ t \in \Omega. \tag{85}$$

Proof By Lemma 28, $K_{\mathbb{Z}^d}$ is a Riesz basis for $\mathcal{H}_K$ if and only if the linear span of $\mathcal{E}_{\mathbb{Z}^d}$ is dense in $L^2(\mathbb{R}^d, \hat{k}dt)$, which is equivalent to (83) by Lemma 29,

$$\sum_{n \in \mathbb{Z}^d} c_n e^{i(n,t)} \in L^2(\mathbb{R}^d, \hat{k}dt), \ \ c \in \ell^2(\mathbb{Z}^d) \tag{86}$$

and for some constants $0 < \alpha \le \beta < \infty$

$$\alpha(2\pi)^d \|c\|^2_{\ell^2(\mathbb{Z}^d)} \le \left\| \sum_{n \in \mathbb{Z}^d} c_n e^{i(n,t)} \right\|^2_{L^2(\mathbb{R}^d, \hat{k}dt)} \le \beta(2\pi)^d \|c\|^2_{\ell^2(\mathbb{Z}^d)}, \ \ c \in \ell^2(\mathbb{Z}^d). \tag{87}$$

One can use arguments similar to those on pages 139–140 of Daubechies (1992) to show that relation (86) and inequality (87) hold true if and only if there exist $0 < \alpha \le \beta < \infty$ such that

$$\alpha \le \sum_{n \in \mathbb{Z}^d} \hat{k}(\cdot + 2n\pi) \le \beta, \quad \text{a.e.} \tag{88}$$

The proof is completed by noting that (83) and (88) hold true if and only if (84) and (85) are satisfied. ■

Inequality (88) was established for a different purpose in Smale and Zhou (2004), where it was proved that $\Lambda := [K(m,n) : m, n \in \mathbb{Z}^d]$ satisfies (39) for $J := \mathbb{Z}^d$ if and only if (88) holds true.

In the next theorem, we construct $\hat{k}$ that satisfy conditions (73), (74), (84) and (85) to obtain a refinable kernel K on $\mathbb{R}^d$ with $K_{\mathbb{Z}^d}$ being a Riesz basis for $\mathcal{H}_K$.

Theorem 31 *Let K be defined by (67) where nonnegative $\hat{k} \in L^1(\mathbb{R}^d)$ is continuous at* 0. *Then K is refinable and $K_{\mathbb{Z}^d}$ is a Riesz basis for $\mathcal{H}_K$ if and only if $\lambda = 2^d$ and*

$$\hat{k} = \eta\chi_\Omega, \quad a.e. \tag{89}$$

where η is a positive constant, and Ω satisfies (73) and (84). Moreover, if (84) and (89) hold true then functions

$$\frac{1}{\sqrt{\eta}(2\pi)^{d/2}} K(\cdot, n), \quad n \in \mathbb{Z}^d \tag{90}$$

form an orthonormal basis for $\mathcal{H}_K$.

Proof Suppose that $\lambda = 2^d$, $\hat{k}$ is given by (89) with a positive constant η, and Ω satisfies (73) and (84). By Theorem 23, K defined by (67) is refinable, and by Theorem 30, $K_{\mathbb{Z}^d}$ is a Riesz basis for $\mathcal{H}_K$.

Conversely, we suppose that $\hat{k} \in L^1(\mathbb{R}^d)$ is continuous at 0, K is refinable and $K_{\mathbb{Z}^d}$ is a Riesz basis for $\mathcal{H}_K$. These hypotheses imply that there hold (73), (74), (84) and (85). Choose $\eta = \hat{k}(0)$. Repeatedly using Equation (74) with iterations, we have for all $t \in \Omega$ that

$$\hat{k}(t) = \left(\frac{\lambda}{2^d}\right)^j \hat{k}\left(\frac{t}{2^j}\right), \quad j \in \mathbb{N}.$$

This formula with condition (85) ensures that $\lambda = 2^d$. Equation (89) follows from the formula above, (85) and the continuity of $\hat{k}$ at 0. This completes the proof of the necessary and sufficient condition.

It remains to show that the functions defined by (90) form an orthonormal basis for $\mathcal{H}_K$. Since (84) and (89) are true, we obtain by direct computation for all $m, n \in \mathbb{Z}^d$ that

$$\begin{aligned}
\frac{1}{\eta(2\pi)^d}(K(\cdot,n),K(\cdot,m))_{\mathcal{H}_K} &= \frac{1}{(2\pi)^d}\int_\Omega e^{i(m-n,t)}dt \\
&= \frac{1}{(2\pi)^d}\int_{\mathbb{R}^d} e^{i(m-n,t)}\chi_\Omega(t)dt \\
&= \frac{1}{(2\pi)^d}\int_{[0,2\pi]^d} e^{i(m-n,t)}\sum_{l\in\mathbb{Z}^d}\chi_\Omega(\cdot+2l\pi)dt \\
&= \frac{1}{(2\pi)^d}\int_{[0,2\pi]^d} e^{i(m-n,t)}dt = \delta_{m,n}.
\end{aligned}$$

This proves that functions defined by (90) constitute an orthonormal basis for $\mathcal{H}_K$ and completes the proof. ■

By noting that (84) can be interpreted as that $\Omega + 2n\pi$, $n \in \mathbb{Z}^d$, form a *tiling* of $\mathbb{R}^d$, we construct examples of refinable kernels K such that $K_{\mathbb{Z}^d}$ are Riesz bases for $\mathcal{H}_K$. The readers are referred to Grünbaum and Shephard (1989) for the subject of tiling. We describe our examples in the following two corollaries.

Corollary 32 *For $a \in [0, 2\pi]$, let $\Omega := [-a, 2\pi - a]^d$. Then the kernel K defined by (67) with $\hat{k}$ having the form (89) is a refinable kernel such that $K_{\mathbb{Z}^d}$ is a Riesz basis for $\mathcal{H}_K$.*

We remark that when $a = \pi$, k is the well-known sinc function. The result in the last corollary can be extended.

Corollary 33 *Suppose that* α *and* β *are constants satisfying either* $\frac{\pi}{2} \le \alpha \le \beta \le \frac{2\pi}{3}$ *or* $\frac{2\pi}{3} \le \alpha \le \beta \le \pi$*. Let*

$$\Omega := ([-2\pi+\alpha,-2\pi+\beta]\cup[-\pi,-\beta]\cup[-\alpha,\alpha]\cup[\beta,\pi]\cup[2\pi-\beta,2\pi-\alpha])^d. \tag{91}$$

Then the kernel K *defined by (67) with* $\hat{k}$ *having the form (89) with* $\eta = 1$ *is a refinable kernel such that* $K_{\mathbb{Z}^d}$ *is a Riesz basis for* $\mathcal{H}_K$*. Moreover,*

$$k(x) = \prod_{j\in\mathbb{N}_d} \frac{2\sin\alpha x_j}{x_j} + \prod_{j\in\mathbb{N}_d} \frac{4\sin\frac{\pi-\beta}{2}x_j}{x_j}\cos\frac{\pi+\beta}{2}x_j + \prod_{j\in\mathbb{N}_d} \frac{4\sin\frac{\beta-\alpha}{2}x_j}{x_j}\cos\frac{\alpha+\beta}{2}x_j,\ \ x\in\mathbb{R}^d,$$

where x_j *denotes the* j*-th component of* x*.*

When both α and β in (91) are chosen as π, k is also reduced to the sinc function.

6. Refinable Kernels Defined by Refinable Functions

We present in this section a construction of refinable kernels via refinable functions. For a complete reference of refinable functions, the readers are referred to Cavaretta et al. (1991) and Daubechies (1992). As in the last section, we assume that the mapping γ has the form $x \to 2x$ throughout this section.

Let φ be a compactly supported continuous function on $\mathbb{R}^d$ that is *refinable*, namely, there exists $h := [h_n : n \in \mathbb{Z}^d]$ such that

$$\varphi\left(\frac{\cdot}{2}\right) = \sum_{n\in\mathbb{Z}^d} h_n \varphi(\cdot - n). \tag{92}$$

We always assume that φ is nontrivial, and the cardinality of $\{n : h_n \ne 0,\ n \in \mathbb{Z}^d\}$ is finite by the compact support of φ. Suppose further that we have an infinite matrix A satisfying for some positive constants α, β that

$$\alpha\|c\|^2_{\ell^2(\mathbb{Z}^d)} \le (Ac,c)_{\ell^2(\mathbb{Z}^d)} \le \beta\|c\|^2_{\ell^2(\mathbb{Z}^d)},\ \ c\in\ell^2(\mathbb{Z}^d). \tag{93}$$

The above inequality implies that A is a bounded positive self-adjoint operator on $\ell^2(\mathbb{Z}^d)$ and its inverse A^{-1} is also bounded positive self-adjoint (see, for example, Daubechies, 1992, page 58).

Motivated by (48), associated with the matrix A we define our kernel K by

$$K(x,y) := (A\Psi(x),\Psi(y))_{\ell^2(\mathbb{Z}^d)},\ \ x,y\in\mathbb{R}^d, \tag{94}$$

where Ψ is a mapping from $\mathbb{R}^d$ to $\ell^2(\mathbb{Z}^d)$ given by $\Psi(x) := [\varphi(x-n) : n\in\mathbb{Z}^d],\ \ x\in\mathbb{R}^d$. Assuming that $A := [A_{m,n} : m,n\in\mathbb{Z}^d]$, we have that

$$K(x,y) = \sum_{m\in\mathbb{Z}^d}\sum_{n\in\mathbb{Z}^d} A_{m,n}\varphi(x-n)\overline{\varphi(y-m)},\ \ x,y\in\mathbb{R}^d. \tag{95}$$

Kernels in the form

$$\sum_{n\in\mathbb{Z}^d} \psi(x-n)\overline{\psi(y-n)} \tag{96}$$

constructed by a refinable function ψ were considered in Opfer (2006), and kernels defined as a superposition of frame elements in RKHS were discussed in Gao et al. (2001), Opfer (2006), and

Rakotomamonjy and Canu (2005). When A is the identity infinite matrix I, we see that K defined by (95) has the form (96) with $\psi = \varphi$. We are interested in the necessary and sufficient condition for K to degenerate to the form (96) for some function ψ on $\mathbb{R}^d$ such that the series converges for all $x, y \in \mathbb{R}^d$. We need the following technical lemma, whose proof is standard and thus is omitted.

Lemma 34 *The linear span of $\Psi(\mathbb{R}^d)$ is dense in $\ell^2(\mathbb{Z}^d)$, that is, $\Psi(\mathbb{R}^d)^\perp = \{0\}$.*

The next proposition shows that kernels in the form (95) are more general than those in the degenerate form (96) and in general cannot be written in the degenerate form.

Proposition 35 *Let A be an infinite matrix satisfying (93) and K be defined by (95) through a compactly supported continuous function φ. Then K can be represented as (96) if and only if*

$$A_{m,n} = A_{m-n,0}, \quad m, n \in \mathbb{Z}^d. \tag{97}$$

Proof Suppose that the kernel K defined by (95) has the form (96). It follows that for all $x, y \in \mathbb{R}^d$ and for $l \in \mathbb{Z}^d$, $K(x-l, y-l) = K(x, y)$. Using (95), we rewrite the above equation as

$$\sum_{m\in\mathbb{Z}^d}\sum_{n\in\mathbb{Z}^d} A_{m-l,n-l}\varphi(x-n)\overline{\varphi(y-n)} = \sum_{m\in\mathbb{Z}^d}\sum_{n\in\mathbb{Z}^d} A_{m,n}\varphi(x-n)\overline{\varphi(y-n)}.$$

By Lemma 34, we have for all $m, n, l \in \mathbb{Z}^d$ that $A_{m-l,n-l} = A_{m,n}$. In this equation, letting $l = n$ yields (97).

Conversely, we suppose that (97) is satisfied. For $n \in \mathbb{Z}^d$, we set $a_n := A_{n,0}$ and observe that for all $c \in \ell^2(\mathbb{Z}^d)$

$$(Ac, c)_{\ell^2(\mathbb{Z}^d)} = \frac{1}{(2\pi)^d}\int_{[0,2\pi]^d}\left(\sum_{n\in\mathbb{Z}^d} a_n e^{i(n,t)}\right)\left|\sum_{n\in\mathbb{Z}^d} c_n e^{i(n,t)}\right|^2 dt.$$

This with (93) implies that

$$\alpha \le \sum_{n\in\mathbb{Z}^d} a_n e^{i(n,t)} \le \beta, \quad \text{a.e. } t \in [0, 2\pi]^d,$$

where the constants α and β are the lower and upper bound in (93). Therefore, there exists $b \in \ell^2(\mathbb{Z}^d)$ such that

$$\sum_{n\in\mathbb{Z}^d} b_n e^{i(n,t)} = \left(\sum_{n\in\mathbb{Z}^d} a_n e^{i(n,t)}\right)^{1/2}, \quad \text{a.e. } t \in [0, 2\pi]^d.$$

We then define the matrix B by setting $B_{m,n} := b_{m-n,0}$, $m, n \in \mathbb{Z}^d$. Clearly, we have $B = B^*$ and $A = B^2$, which ensures that

$$K(x, y) = (B\Psi(x), B\Psi(y))_{\ell^2(\mathbb{Z}^d)}, \quad x, y \in \mathbb{R}^d.$$

One can see that K can be rewritten as (96) with $\psi := \sum_{m\in\mathbb{Z}^d} B_{0,m}\varphi(\cdot - m)$. ■

The main purpose of this section is to formulate conditions on h and A so that the kernel K in the form (95) is refinable. Our discussions will be based on the following result concerning the RKHS $\mathcal{H}_K$.

Proposition 36 *The RKHS of the kernel K defined by (94) is*

$$\mathcal{H}_K := \{c_\Psi := (\Psi(\cdot), c)_{\ell^2(\mathbb{Z}^d)} : c \in \ell^2(\mathbb{Z}^d)\}$$

with inner product

$$(c_\Psi, d_\Psi)_{\mathcal{H}_K} = (A^{-1}d, c)_{\ell^2(\mathbb{Z}^d)}, \ \ c, d \in \ell^2(\mathbb{Z}^d).$$

Proof Since the operator A satisfies (93), there exists a bounded positive self-adjoint operator $A^{1/2}$ on $\ell^2(\mathbb{Z}^d)$ such that $A^{1/2}A^{1/2} = A$ (see, Conway, 1990, ,page 240). It is observed that K has the following feature map representation $K(x,y) = (\Phi(x), \Phi(y))_{\ell^2(\mathbb{Z}^d)}$, $x, y \in \mathbb{R}^d$, where

$$\Phi := A^{1/2}\Psi. \tag{98}$$

The proposition now follows immediately from Lemmas 5 and 34. ■

Let λ be a fixed positive number and K_j, $j \in \mathbb{Z}$ be defined as in (69). We shall give a characterization for K to be refinable, that is, (68) holds. To this end, we introduce the infinite matrix H associated with h as

$$H_{m,n} := h_{n-2m}, \ \ m, n \in \mathbb{Z}^d. \tag{99}$$

It can be seen by the generalized Minkowski inequality that the matrix H induces a bounded operator on $\ell^2(\mathbb{Z}^d)$. In fact, we have for each $c \in \ell^2(\mathbb{Z}^d)$ that

$$\|Hc\|_{\ell^2(\mathbb{Z}^d)} \leq \left(\sum_{n\in\mathbb{Z}^d} |h_n|\right) \|c\|_{\ell^2(\mathbb{Z}^d)}.$$

We next characterize refinable kernels in terms of matrices A and H.

Theorem 37 *Suppose that φ is a nontrivial compactly supported refinable function satisfying (92). Then the kernel K defined by (94) is refinable if and only if*

$$HA^{-1}H^*A = \lambda I. \tag{100}$$

Proof The function $\Phi : \mathbb{R}^d \to \ell^2(\mathbb{Z}^d)$ defined by (98) is a feature map for K. We observe by (92) that it satisfies a refinement equation

$$\lambda^{-1/2}\Phi\left(\frac{\cdot}{2}\right) = \lambda^{-1/2}A^{1/2}\Psi\left(\frac{\cdot}{2}\right) = \lambda^{-1/2}A^{1/2}H\Psi = \lambda^{-1/2}A^{1/2}HA^{-1/2}\Phi,$$

where $A^{-1/2}$ denotes the inverse of $A^{1/2}$. Setting

$$T := \lambda^{-1/2}A^{1/2}HA^{-1/2}, \tag{101}$$

by Theorem 6, K is refinable if and only if T^* is isometric, or equivalently, $TT^* = I$. The proof is complete by noting that equation $TT^* = I$ has the form (100). ■

We need the following lemma to study the proper inclusion of $\mathcal{H}_{K_{-1}}$ in $\mathcal{H}_K$.

Lemma 38 *Let $[a_n : n \in \mathbb{Z}^d]$ be a nontrivial vector in $\ell^2(\mathbb{Z}^d)$ with a finite number of nonzero components. Then the linear span of $\{\tilde{a}_m := [a_{m-n} : n \in \mathbb{Z}^d] : m \in \mathbb{Z}^d\}$ is dense in $\ell^2(\mathbb{Z}^d)$, and $\tilde{a}_m$, $m \in \mathbb{Z}^d$ are linearly independent.*

Proposition 39 *Let φ be a nontrivial compactly supported continuous refinable function on $\mathbb{R}^d$, A, h satisfy (93) and (100), and K be defined by (94). Then $\mathcal{H}_{K_{-1}}$ is a proper subspace of $\mathcal{H}_K$.*

Proof By Theorem 7, $\mathcal{H}_{K_{-1}}$ is a proper subspace of $\mathcal{H}_K$ if and only if the null space $\mathcal{N}(T)$ of operator T defined by (101) contains nonzero elements in $\ell^2(\mathbb{Z}^d)$, which is equivalent to that

$$\mathcal{N}(H) \neq \{0\}. \tag{102}$$

Set $\tilde{b}_m := [h_{n-m} : n \in \mathbb{Z}^d]$, $m \in \mathbb{Z}^d$. Assume that $\mathcal{N}(H) = \{0\}$. This implies that span$\{\tilde{b}_{2m} : m \in \mathbb{Z}^d\}$ is dense in $\ell^2(\mathbb{Z}^d)$. Choose $l \in \mathbb{Z}^d \setminus 2\mathbb{Z}^d$. Since $[h_n : n \in \mathbb{Z}^d]$ has a finite number of nonzero components, $\tilde{b}_l$ can be represented as a finite linear combination of $\tilde{b}_{2m}$, $m \in \mathbb{Z}^d$. However, Lemma 38 ensures that $\tilde{b}_m$, $m \in \mathbb{Z}^d$ are linearly independent. This contradiction implies the validity of (102). ∎

When A and H commute, Equation (100) reduces to

$$HH^* = \lambda I. \tag{103}$$

Through a scaling of the matrix H, one may consider $HH^* = I$. This equation arose also in the construction of orthonormal wavelets and it has been well understood in the one-dimensional case (cf., Daubechies, 1992). For a special class of solutions h of (103) in the multidimensional case, see Chen et al. (2003) and Chen et al. (2007). These h can be used to construct A and H satisfying (93) and (100).

Proposition 40 *Let h be a solution of (103). Then for each real number $a \in \mathbb{R} \setminus \{\pm\lambda^{-1/2}\}$ the matrix A defined by*

$$A := ((I + aH)(I + aH^*))^{-1} \tag{104}$$

satisfies (93) with $\alpha := (1 + |a|\sqrt{\lambda})^{-2}$ and $\beta := (1 - |a|\sqrt{\lambda})^{-2}$, and (100).

Proof For $a \in \mathbb{R} \setminus \{\pm\lambda^{-1/2}\}$ we set $B := (I + aH)(I + aH^*)$. By direct computation, we obtain for each $c \in \ell^2(\mathbb{Z}^d)$ that

$$(Bc, c)_{\ell^2(\mathbb{Z}^d)} = (1 + a^2\lambda)\|c\|^2_{\ell^2(\mathbb{Z}^d)} + a((H + H^*)c, c)_{\ell^2(\mathbb{Z}^d)}. \tag{105}$$

Equation (103) leads to the estimates

$$\|Hc\|_{\ell^2(\mathbb{Z}^d)} \leq \sqrt{\lambda}\|c\|_{\ell^2(\mathbb{Z}^d)}, \text{ and } \|H^*c\|_{\ell^2(\mathbb{Z}^d)} \leq \sqrt{\lambda}\|c\|_{\ell^2(\mathbb{Z}^d)}.$$

Equation (105) with these two estimates implies that

$$(1 - |a|\sqrt{\lambda})^2\|c\|^2_{\ell^2(\mathbb{Z}^d)} \leq (Bc, c)_{\ell^2(\mathbb{Z}^d)} \leq (1 + |a|\sqrt{\lambda})^2\|c\|^2_{\ell^2(\mathbb{Z}^d)}.$$

The inverse operator A of B hence satisfies (93) with $\alpha := (1 + |a|\sqrt{\lambda})^{-2}$ and $\beta := (1 - |a|\sqrt{\lambda})^{-2}$.

It remains to show that A satisfies Equation (100). It can be verified by (103) that the following equation holds

$$H(I+aH)(I+aH^*)H^* = \lambda(I+aH)(I+aH^*).$$

By the definition (104) of A, this is equivalent to the equation $HA^{-1}H^* = \lambda A^{-1}$, which confirms that A satisfies (100). ■

The last proposition provides a class of refinable kernels given by (95) that never degenerate to the form (96).

Proposition 41 *Let h satisfy (103), where matrix H is defined by (99), and matrix A be of the form (104) for some $a \in \mathbb{R} \setminus \{\pm\lambda^{-1/2}\}$. If there exists at least one $n \in \mathbb{Z}^d$ such that the real part $\operatorname{Re}(h_n)$ of h_n is not zero then A satisfies (97) if and only if $a = 0$.*

Proof If $a = 0$ then $A = I$ satisfies (97). Let $a \in \mathbb{R} \setminus \{0, \pm\lambda^{-1/2}\}$. One can see that A satisfies (97) if and only if A^{-1} does. Since $A^{-1} = (1+a^2\lambda)I + a(H+H^*)$, it satisfies (97) only if $H+H^*$ does. Choose $n \in \mathbb{Z}^d$ such that $\operatorname{Re}(h_n) \neq 0$. There exists an $m \in \mathbb{Z}^d$ such that $\operatorname{Re}(h_m) \neq \operatorname{Re}(h_n)$ since $h \in \ell^2(\mathbb{Z}^d)$. As a consequence, we have

$$(H+H^*)_{-m,-m} = 2\operatorname{Re}(h_m) \neq 2\operatorname{Re}(h_n) = (H+H^*)_{-n,-n}.$$

This shows that $H+H^*$ does not satisfy (97). The proof is complete. ■

By Propositions 35 and 41, if h is a real vector in $\ell^2(\mathbb{Z}^d)$ satisfying (103) then for all $a \in \mathbb{R} \setminus \{0, \pm\lambda^{-1/2}\}$, the refinable kernel K defined by (95) through A given in (104) can not be rewritten as (96).

We now turn to an investigation of the intersection of $\mathcal{H}_{K_j}$, for $j \in \mathbb{Z}$, where K_j are kernels defined by (69).

Theorem 42 *Suppose that φ is a nontrivial compactly supported continuous refinable function on $\mathbb{R}^d$, h satisfies (103), A satisfies (93) and (100), and K is defined by (94). Then (80) holds true if and only if*

$$h_n \neq 0, \ \text{ for at least one } n \in \mathbb{Z}^d \setminus 2\mathbb{Z}^d. \tag{106}$$

If (106) does not hold then

$$\bigcap_{j\in\mathbb{Z}} \mathcal{H}_{K_j} = \{(\Psi(\cdot), c)_{\ell^2(\mathbb{Z}^d)} : c \in \mathcal{N}(H)^\perp\}. \tag{107}$$

Proof Let $\tilde{\mathcal{F}}$ be the function from $\ell^2(\mathbb{Z}^d)$ to $L^2([0,2\pi]^d)$ defined for $c \in \ell^2(\mathbb{Z}^d)$ by $(\tilde{\mathcal{F}}c)(t) := \sum_{n\in\mathbb{Z}^d} c_n e^{i(n,t)}$, $t \in [0,2\pi]^d$. Set $B := \lambda^{-1/2}H$, $m_0 := \tilde{\mathcal{F}}(\lambda^{-1/2}h)$, and let $\{\nu_j : j \in \mathbb{N}_{2^d}\} \subseteq \mathbb{Z}^d$ denote the set of extreme points of cube $[0,1]^d$. By condition (32) in Theorem 9, (80) holds true if and only if

$$\lim_{j\to\infty} \|T^j c\|_{\ell^2(\mathbb{Z}^d)} = 0, \quad \text{for all } c \in \ell^2(\mathbb{Z}^d), \tag{108}$$

where T is the operator defined by (101). Noting that $T^j = A^{1/2}B^jA^{-1/2}$, Equation (108) is equivalent to the condition

$$\lim_{j\to\infty} (2\pi)^{-d/2} \|\tilde{\mathcal{F}}(B^j c)\|_{L^2([0,2\pi]^d)} = \lim_{j\to\infty} \|B^j c\|_{\ell^2(\mathbb{Z}^d)} = 0, \quad \text{for all } c \in \ell^2(\mathbb{Z}^d). \tag{109}$$

For each $c \in \ell^2(\mathbb{Z}^d)$, we define $m_1 := \tilde{\mathcal{F}} c$. It can be verified by direct calculation that

$$\tilde{\mathcal{F}}(B^j c)(t) = \left(\frac{1}{2^d} \sum_{j \in \mathbb{N}_{2^d}} m_0(-t - \pi\nu_j) m_1(t + \pi\nu_j)\right) \left(\frac{1}{2^d} \sum_{j \in \mathbb{N}_{2^d}} m_0(-t - \pi\nu_j)\right)^{j-1}, \ t \in [0, 2\pi]^d.$$

Equation (103) can be rewritten in terms of m_0 as

$$\sum_{j \in \mathbb{N}_{2^d}} |m_0(\cdot + \pi\nu_j)|^2 = 2^d. \tag{110}$$

The Cauchy-Schwartz inequality with (110) ensures for all $t \in [0, 2\pi]^d$ that

$$\left|\sum_{j \in \mathbb{N}_{2^d}} m_0(t + \pi\nu_j)\right| \leq 2^d, \tag{111}$$

where the equality holds at a point $t_0 \in [0, 2\pi]^d$ if and only if

$$m_0(t_0 + \pi\nu_j) = a, \ \ j \in \mathbb{N}_{2^d} \ \text{ for some } \ a \in \mathbb{C} \ \text{ with } \ |a| = 1. \tag{112}$$

If $h_n = 0$ for each $n \in \mathbb{Z}^d \setminus 2\mathbb{Z}^d$ then $m_0(\cdot + \pi\nu_j) = m_0$ for all $j \in \mathbb{N}_{2^d}$. This together with (110) implies that (112) holds for all $t_0 \in [0, 2\pi]^d$. Thus, the equality in (111) holds for all $t \in [0, 2\pi]^d$. We now choose c such that $m_1 = \overline{m_0(-\cdot)}$ and clearly for such a c (109) does not hold. Consequently, (80) does not hold and this is equivalent to saying that (80) implies (106).

Conversely, suppose that (106) holds. By the fact that the zeros of a nontrivial real-analytic function on $\mathbb{R}^d$ form a set of Lebesgue measure zero, the set of points $t \in [0, 2\pi]^d$ for which the equality in (111) holds has zero Lebesgue measure. Therefore, for a fixed $c \in \ell^2(\mathbb{Z}^d)$, $\tilde{\mathcal{F}}(B^j c)$ goes to zero almost everywhere on $[0, 2\pi]^d$. Since

$$|\tilde{\mathcal{F}}(B^j c)(t)| \leq \left(\frac{1}{2^d} \sum_{j \in \mathbb{N}_{2^d}} |m_1(t + \pi\nu_j)|^2\right)^{1/2}, \ \ t \in [0, 2\pi]^d,$$

Equation (109) holds true by the Lebesgue dominated convergence theorem. Thus, we conclude that (80) holds.

Now, suppose that (106) does not hold. Note that $c \in \ell^2(\mathbb{Z}^d)$ is in the union of $\mathcal{N}(T^j)$, $j \in \mathbb{N}$ if and only if $A^{-1/2} c \in \mathcal{N}(H)$. We use Theorem 7 and (34) to get that

$$\bigcap_{j \in \mathbb{Z}} \mathcal{H}_{K_j} = \{(A^{1/2}\Psi(\cdot), c)_{\ell^2(\mathbb{Z}^d)} : c \in (A^{1/2}\mathcal{N}(H))^\perp\}.$$

The above equation can be rewritten as (107). ■

We next present a characterization for $K_{\mathbb{Z}^d}$ to be a Riesz basis for $\mathcal{H}_K$.

Theorem 43 *Let K be defined by (94) through a matrix A satisfying (93) and a compactly supported continuous function φ on $\mathbb{R}^d$. Then $K_{\mathbb{Z}^d}$ is a Riesz basis for $\mathcal{H}_K$ if and only if the polynomial*

$$q(t) := \sum_{n \in \mathbb{Z}^d} \varphi(n) e^{i(n,t)}, \ \ t \in \mathbb{R}^d \tag{113}$$

has no zeros.

Proof By Theorem 13 and condition (93), $K_{\mathbb{Z}^d}$ is a Riesz basis for $\mathcal{H}_K$ if and only if $\tilde{\varphi}_m := [\varphi(m-n) : n \in \mathbb{Z}^d]$, for $m \in \mathbb{Z}^d$ form a Riesz basis for $\ell^2(\mathbb{Z}^d)$. Lemma 38 states that $\tilde{\varphi}_m$, $m \in \mathbb{Z}^d$ are linearly independent if there exists $t \in \mathbb{Z}^d$ such that $\varphi(t) \neq 0$. Therefore, $\{\tilde{\varphi}_m : m \in \mathbb{Z}^d\}$ is Riesz basis for $\ell^2(\mathbb{Z}^d)$ if and only if there exist $0 < \alpha \leq \beta < \infty$ such that for every $c \in \ell^2(\mathbb{Z}^d)$ there holds

$$\alpha\|c\|^2_{\ell^2(\mathbb{Z}^d)} \leq \sum_{m\in\mathbb{Z}^d} |(\tilde{\varphi}_m, c)_{\ell^2(\mathbb{Z}^d)}|^2 \leq \beta\|c\|^2_{\ell^2(\mathbb{Z}^d)}. \tag{114}$$

Since

$$\sum_{m\in\mathbb{Z}^d} |(\tilde{\varphi}_m, c)_{\ell^2(\mathbb{Z}^d)}|^2 = \frac{1}{(2\pi)^d}\int_{[0,2\pi]^d} \left|\sum_{n\in\mathbb{Z}^d} \varphi(n)e^{i(n,t)}\right|^2 \left|\sum_{n\in\mathbb{Z}^d} \overline{c_n}e^{i(n,t)}\right|^2 dt,$$

Equation (114) holds for all $c \in \ell^2(\mathbb{Z}^d)$ if and only if $\alpha \leq |q(t)|^2 \leq \beta$, $t \in [0,2\pi]^d$. The theorem hence follows from the continuity of q on $\mathbb{R}^d$. ■

We conclude this section by a result regarding a multiresolution analysis for $\mathcal{H}_K$.

Theorem 44 *Let φ be a nontrivial compactly supported continuous refinable function on $\mathbb{R}^d$, h satisfy (103), A satisfy (93), and K be defined by (94). Then $\mathcal{H}_K$ has a multiresolution analysis with $K_{\mathbb{Z}^d}$ being a Riesz basis for $\mathcal{H}_K$ if and only if there holds (100), (106), and the polynomial (113) has no zeros.*

Proof This result is a direct consequence of Theorems 37, 42 and 43. ■

7. A Discussion of Applications and Conclusions

For the completeness of the paper, in this section we discuss how refinable kernels can be used to efficiently update kernels for learning from increasing training data. Here we only use a simple learning example to illustrate the main points. The general case requires further substantial research and it will be reported on a different occasion.

In this special example, we assume that the input space X is $\mathbb{R}$ and that the *initial* training data set is given by $\mathbf{z} := \{(j, y_j) : j \in \mathbb{B}_m\} \subseteq X \times Y$, where we have set for each $n \in \mathbb{N}$, $\mathbb{B}_n := \{-n, \ldots, -1, 0, 1, \ldots, n\}$. Let K be a kernel on X and consider the loss function $Q(p,q) := |p-q|^2$, for $p, q \in \mathbb{C}$. This loss function is important in practice (for example, in regularization networks Evgeniou et al., 2000; Schölkopf and Smola, 2002; Vapnik, 1998). The predictor f from the training data $\mathbf{z}$ is hence the minimizer of

$$\min_{g\in\mathcal{H}_K} \sum_{j\in\mathbb{B}_m} |g(j) - y_j|^2 + \mu\|g\|^2_{\mathcal{H}_K},$$

where μ is a positive regularization parameter. The representer theorem (Kimeldorf and Wahba, 1971; Schölkopf et al., 2001; Schölkopf and Smola, 2002) ensures in this case that

$$f = \sum_{j\in\mathbb{B}_m} c_j K(\cdot, j). \tag{115}$$

Here, the vector $\mathbf{c} := [c_j : j \in \mathbb{B}_m]^T$ satisfies the linear system

$$(\mu I_{2m+1} + K[\mathbb{B}_m])\mathbf{c} = \mathbf{y}, \tag{116}$$

where I_n denotes the $n \times n$ identity matrix and $\mathbf{y} := [y_j : j \in \mathbb{B}_m]^T$.

Suppose that the initial training data set is updated to a new data set $\mathbf{z}' := \{(j/2, y'_j) : j \in \mathbb{B}_{2m}\}$ where $y'_{2j} = y_j$ for $j \in \mathbb{B}_m$. We divide $\mathbf{x}' := \mathbb{B}_{2m}/2$ into two disjoint subsets $\mathbf{x}'_1 := \mathbb{B}_m/2$ and $\mathbf{x}'_2 = \mathbb{B}_{2m}/2 \setminus \mathbb{B}_m/2$, and $\mathbf{y}' := \{y'_j : j \in \mathbb{B}_{2m}\}$ into $\mathbf{y}'_1$ and $\mathbf{y}'_2$, accordingly. For convenience, we set $\mathbf{x}'_2 := \{x'_{2,j} : j \in \mathbb{N}_{2m}\}$. If K is refinable on $X = \mathbb{R}$ then we update the kernel K to a new kernel $K_1 := \lambda K(2\cdot, 2\cdot)$. A new predictor f' is then obtained as the minimizer of

$$\min_{g \in \mathcal{H}_{K_1}} \sum_{j \in \mathbb{B}_{2m}} |g(j/2) - y'_j|^2 + \mu' \|g\|^2_{\mathcal{H}_{K_1}},$$

where μ' is an updated regularization parameter. By the representer theorem, we have that

$$f' = \sum_{j \in \mathbb{B}_m} c'_{1,j} K(\cdot, j/2) + \sum_{j \in \mathbb{N}_{2m}} c'_{2,j} K(\cdot, x'_{2,j}). \tag{117}$$

The above vectors $\mathbf{c}'_1 := [c'_{1,j} : j \in \mathbb{B}_m]^T$ and $\mathbf{c}'_2 := [c'_{2,j} : j \in \mathbb{N}_{2m}]^T$ satisfy the linear system

$$\begin{bmatrix} \mu' I_{2m+1} + K_1[\mathbf{x}'_1] & K_1[\mathbf{x}'_1, \mathbf{x}'_2] \\ K_1[\mathbf{x}'_2, \mathbf{x}'_1] & \mu' I_{2m} + K_1[\mathbf{x}'_2] \end{bmatrix} \begin{bmatrix} \mathbf{c}'_1 \\ \mathbf{c}'_2 \end{bmatrix} = \begin{bmatrix} \mathbf{y}'_1 \\ \mathbf{y}'_2 \end{bmatrix}, \tag{118}$$

where $K_1[\mathbf{x}'_1, \mathbf{x}'_2] := [K_1(p,q) : p \in \mathbf{x}'_1, q \in \mathbf{x}'_2]$.

It can be easily seen that the computational advantages offered by the refinability of the kernel include:

- **Efficient updating the kernel**. Kernels K_n in all scales can be efficiently updated from a refinable kernel K.

- **Improvement of the predictor**. Since $\mathcal{H}_K \preceq \mathcal{H}_{K_1}$, the class of candidate functions for the predictor is enlarged and the initial predictor f is in $\mathcal{H}_{K_1}$. Consequently, we can expect an improvement in approximation quality from the initial predictor to the new predictor f'.

- **Efficiency in setting up the coefficient matrix**. By refinability, we observe that the block matrix $K_1[\mathbf{x}'_1]$ in (118) satisfies the relation

  $$K_1[\mathbf{x}'_1] = K_1[\mathbb{B}_m/2] = \lambda K[\mathbb{B}_m].$$

  Therefore, the coefficient matrix of system (118) is an augmentation of that of system (116). As a result, we do not need to recompute the entries in the block $K_1[\mathbf{x}'_1]$.

- **Fast solving the linear system for the updated data set**. Because of the special structure in its coefficient matrix that results from the refinability of the kernel, the linear system can be solved efficiently by a fast algorithm analogous to the multi-level (wavelet) method (cf., Chen et al., 2005, 2006a,b).

- **Fast algorithms for processing the predictor**. Since the predictor (115) or (117) is expressed as a linear combination of the kernel, when the kernel is refinable we can use the (Mallat-type) decomposition and reconstruction algorithms developed in Section 4 to process the predictor.

Finally, we close this paper with the conclusion: Motivated by efficient mathematical learning, we introduce the notion of refinable kernels and characterize various types of refinable kernels. Examples of refinable kernels are presented. A special learning example illustrates that refinable kernels should provide computational advantages for solving various learning problems. Important examples of refinable kernels and their applications deserve further investigation.

Acknowledgments

Yuesheng Xu is supported in part by the US National Science Foundation under grants CCR-0407476 and DMS-0712827, by National Aeronautics and Space Administration under Cooperative Agreement NNX07AC37A, by the Natural Science Foundation of China under grants 10371122 and 10631080, and by the Education Ministry of the People's Republic of China under the Changjiang Scholar Chair Professorship Program through Sun Yat-sen University. He is also with the Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100080, P. R. China. Haizhang Zhang is supported by Syracuse University under a Syracuse University Graduate Fellowship.

References

U. Amato, A. Antoniadis and M. Pensky. Wavelet kernel penalized estimation for non-equispaced design regression. *Stat. Comput.*, 16: 37–55, 2006.

N. Aronszajn. Theory of reproducing kernels. *Trans. Amer. Math. Soc.*, 68: 337–404, 1950.

A. Beurling. On two problems concerning linear transformations in Hilbert space. *Acta Math.*, 81: 239–255, 1949.

S. Bochner. *Lectures on Fourier Integrals with an Author's Supplement on Monotonic Functions, Stieltjes Integrals, and Harmonic Analysis*. Annals of Mathematics Studies 42, Princeton University Press, New Jersey, 1959.

O. Bousquet and A. Elisseeff. Stability and generalization. *Journal of Machine Learning Research*, 2: 499–526, 2002.

A. S. Cavaretta, W. Dahmen and C. A. Micchelli. Stationary subdivision. *Mem. Amer. Math. Soc.*, 93, no. 453, 1991.

Q. Chen, C. A. Micchelli, S. Peng and Y. Xu. Multivariate filters banks having a matrix factorization. *SIAM J. Matrix Anal. Appl.*, 25: 517-531, 2003.

Q. Chen, C. A. Micchelli and Y. Xu. On the matrix completion problem for multivariate filter bank construction. *Adv. Comput. Math.*, 26: 173–204, 2007.

Z. Chen, B. Wu and Y. Xu. Multilevel augmentation methods for solving operator equations. *Numer. Math. J. Chinese Univ.*, 14: 31–55, 2005.

Z. Chen, B. Wu and Y. Xu. Multilevel augmentation methods for differential equations. *Adv. Comput. Math.*, 24: 213–238, 2006.

Z. Chen, Y. Xu and H. Yang. A multilevel augmentation method for solving ill-posed operator equations. *Inverse Problems*, 22: 155–174, 2006.

J. B. Conway. *A Course in Functional Analysis*. 2nd Edition, Springer-Verlag, New York, 1990.

F. Cucker and S. Smale. On the mathematical foundations of learning. *Bull. Amer. Math. Soc.*, 39: 1–49, 2002.

I. Daubechies. *Ten Lectures on Wavelets*. CBMS-NSF Regional Conference Series in Applied Mathematics 61, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1992.

R. J. Duffin and A. C. Schaeffer. A class of nonharmonic Fourier series. *Trans. Amer. Math. Soc.*, 72: 341–366, 1952.

T. Evgeniou, M. Pontil and T. Poggio. Regularization networks and support vector machines. *Adv. Comput. Math.*, 13: 1–50, 2000.

C. H. FitzGerald, C. A. Micchelli and A. Pinkus. Functions that preserve families of positive semidefinite matrices. *Linear Algebra Appl.*, 221: 83–102, 1995.

J. B. Gao, C. J. Harris and S. R. Gunn. On a class of support vector kernels based on frames in function Hilbert spaces. *Neural Comput.*, 13: 1975–1994, 2001.

L. Grafakos. *Classical and Modern Fourier Analysis*. Prentice Hall, New Jersey, 2004.

B. Grünbaum and G. C. Shephard. *Tilings and Patterns*. W. H. Freeman and Company, New York, 1989.

G. Kimeldorf and G. Wahba. Some results on Tchebycheffian spline functions. *J. Math. Anal. Appl.*, 33: 82–95, 1971.

S. Mallat. Multiresolution approximations and wavelet orthonormal bases of $L^2(R)$. *Trans. Amer. Math. Soc.*, 315: 69–87, 1989.

S. Mallat. *A Wavelet Tour of Signal Processing*. 2nd Edition, Academic Press, San Diego, CA, 1998.

Y. Meyer. *Wavelets and Operators*. Cambridge University Press, Cambridge, 1992.

C. A. Micchelli and M. Pontil. A function representation for learning in Banach spaces. In *Proceeding of the 17th Annual Conference on Learning Theory* (COLT 04), pages 255–269, Banff, Alberta, 2004.

C. A. Micchelli and M. Pontil. Learning the kernel function via regularization. *Journal of Machine Learning Research*, 6: 1099–1125, 2005.

C. A. Micchelli and M. Pontil. On learning vector-valued functions. *Neural Comput.*, 17: 177–204, 2005.

C. A. Micchelli, Y. Xu and P. Ye. Cucker Smale learning theory in Besov spaces. In *Advances in Learning Theory: Methods, Models and Applications*, pages 47–68, IOS Press, Amsterdam, The Netherlands, 2003.

C. A. Micchelli, Y. Xu and H. Zhang. Universal kernels. *Journal of Machine Learning Research*, 7: 2651–2667, 2006.

S. Mukherjee, P. Niyogi, T. Poggio and R. Rifkin. Learning theory: stability is sufficient for generalization and necessary and sufficient for empirical risk minimization. *Adv. Comput. Math.*, 25: 161–193, 2006.

R. Opfer. Multiscale kernels. *Adv. Comput. Math.*, 25: 357–380, 2006.

R. Opfer. Tight frame expansions of multiscale reproducing kernels in Sobolev spaces. *Appl. Comput. Harmon. Anal.*, 20: 357–374, 2006.

A. Rakotomamonjy and S. Canu. Frames, reproducing kernels, regularization and learning. *Journal of Machine Learning Research,* 6: 1485–1515, 2005.

A. Rakotomamonjy, X. Mary and S. Canu. Non-parametric regression with wavelet kernels. *Appl. Stoch. Models Bus. Ind.*, 21: 153–163, 2005.

W. Rudin. *Real and Complex Analysis*. 3rd Edition, McGraw-Hill, New York, 1987.

B. Schölkopf, R. Herbrich and A. J. Smola. A generalized representer theorem. In *Proceeding of the 14th Annual Conference on Computational Learning Theory and the 5th European Conference on Computational Learning Theory,* pages 416–426, Springer-Verlag, London, UK, 2001.

B. Schölkopf and A. J. Smola. *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. MIT Press, Cambridge, Mass, 2002.

J. Shawe-Taylor and N. Cristianini. *Kernel Methods for Pattern Analysis*. Cambridge University Press, Cambridge, 2004.

S. Smale and D. X. Zhou. Estimating the approximation error in learning theory. *Anal. Appl.*, 1: 17–41, 2003.

S. Smale and D. X. Zhou. Shannon sampling and function reconstruction from point values. *Bull. Amer. Math. Soc.*, 41: 279–305, 2004.

S. Smale and D. X. Zhou. Learning theory estimates via integral operators and their approximations. *Constr. Approx.*, 26: 153–172, 2007.

I. Steinwart and C. Scovel. Fast rates for support vector machines using Gaussian kernels. In *Proceeding of the 18th Annual Conference on Learning Theory* (COLT 05), pages 279–294, Bertinoro, 2005.

V. N. Vapnik. *Statistical Learning Theory*. Wiley, New York, 1998.

G. Wahba. Support vector machines, reproducing kernel Hilbert spaces and the randomized GACV. In *Advances in Kernel Methods–Support Vector Learning,* pages 69–86, MIT Press, Cambridge, Mass, 1999.

C. Walder, K. I. Kim and B. Schölkopf. Sparse multiscale Gaussian process regression. Technical Report No. TR-162, Max Planck Institute for Biological Cybernetics, 2007.

C. Walder, B. Schölkopf and O. Chapelle. Implicit surface modelling with a globally regularised basis of compact support. *Computer Graphics Forum*, 25: 635–644, 2006.

Y. Ying and D. X. Zhou. Learnability of Gaussians with flexible variances. *Journal of Machine Learning Research*, 8: 249–276, 2007.

R. M. Young. *An Introduction to Nonharmonic Fourier Series*. Academic Press, New York, 1980.

B. Yu and H. Zhang. The Bedrosian identity and homogeneous semi-convolution equations. *J. Integral Equations Appl.*, accepted, 2006.

T. Zhang. Statistical behavior and consistency of classification methods based on convex risk minimization. *Ann. Statis.*, 32: 56–85, 2004.

D. X. Zhou. Density problem and approximation error in learning theory. *Preprint*, 2003.