Solving Support Vector Machines in Reproducing Kernel Banach Spaces with Positive Definite Functions

Gregory E. Fasshauer[a], Fred J. Hickernell[a], Qi Ye[b,*]

[a]*Department of Applied Mathematics, Illinois Institute of Technology, Chicago, Illinois 60616*
[b]*Department of Mathematics, Syracuse University, Syracuse, NY 13244*

Abstract

In this paper we solve support vector machines in reproducing kernel Banach spaces with reproducing kernels defined on nonsymmetric domains instead of the traditional methods in reproducing kernel Hilbert spaces. Using the orthogonality of semi-inner-products, we can obtain the explicit representations of the dual (normalized-duality-mapping) elements of support vector machine solutions. In addition, we can introduce the reproduction property in a generalized native space by Fourier transform techniques such that it becomes a reproducing kernel Banach space, which can be even embedded into Sobolev spaces, and its reproducing kernel is set up by the related positive definite function. The representations of the optimal solutions of support vector machines (regularized empirical risks) in these reproducing kernel Banach spaces are formulated explicitly in terms of positive definite functions, and their finite numbers of coefficients can be computed by fixed point iteration. We also give some typical examples of reproducing kernel Banach spaces induced by Matérn functions (Sobolev splines) so that their support vector machine solutions are well computable as the classical algorithms. Moreover, each of their reproducing bases includes information from multiple training data points. The concept of reproducing kernel Banach spaces offers us a new numerical tool for solving support vector machines.

Keywords: support vector machine, regularized empirical risk, reproducing kernel, reproducing kernel Banach space, positive definite function, Matérn function, Sobolev spline.

---

*Corresponding author
Email addresses: `fasshauer@iit.edu` (Gregory E. Fasshauer), `hickernell@iit.edu` (Fred J. Hickernell), `qiye@syr.edu` (Qi Ye)

1. Introduction

The theory and practice of kernel-based methods is a fast growing research area. They have been used for both scattered data approximation and machine learning. Applications come from such different fields as physics, biology, geology, meteorology and finance. The books [4, 7, 20, 21] show how to use (conditionally) positive definite kernels to construct interpolants for observation data sampled from some unknown functions in the native spaces induced by the kernel functions. In the books [2, 18], the optimal support vector machine solutions are obtained in reproducing kernel Hilbert spaces (RKHSs), and these solutions are formulated in terms of the related reproducing kernels and given data values. Actually, as long as the same inner product is used, the concepts of native spaces and RKHSs are interchangeable. It is just that researchers in numerical analysis and statistical learning use different terminology and techniques to introduce those spaces. Moreover, the recent contributions [9, 10, 22] develop a clear and detailed framework for generalized Sobolev spaces and RKHSs by establishing a connection between Green functions and reproducing kernels.

Related to the current research work, [5, 6, 23] all generalize classical native spaces (RKHSs) to Banach spaces in different ways. However, the reproducing property in generalized native spaces is not discussed in [5, 6], and [23] does not mention how to use reproducing kernels to introduce the explicit forms of their reproducing kernel Banach spaces (RKBSs) analogous to the typical cases of RKHSs induced by Gaussian kernels and Sobolev-spline kernels, etc. Using [23] it is therefore difficult to obtain explicit and simple support vector machine (SVM) solutions and perform practical computations. Following the results of these earlier authors, [22, Section 6] tries to combine both of these ideas, and uses Fourier transform techniques to construct RKBSs.

In this paper we want to complete and extend the theoretical results in [22, Section 6]. In addition, the RKBS given in Definition 4.1 is different from that of [23]. Our RKBS can be *one-sided* or *two-sided* and its reproducing kernel K can be defined on nonsymmetric domains, i.e., $K : \Omega_2 \times \Omega_1 \to \mathbb{C}$, where Ω_1 and Ω_2 can be various subsets of $\mathbb{R}^{d_1}$ and $\mathbb{R}^{d_2}$, respectively (see Definition 4.1). Our RKBS is an extension of the RKHS and it does not require the reflexivity condition. The RKBS defined in [23] can be seen as a special case of the RKBS defined in this paper. According to Lemma 4.1, we can still obtain the optimal solution in the one-sided RKBS using the techniques of semi-inner-products.

It is well known that for given training data $D := \left\{(x_j, y_j)\right\}_{j=1}^N$ the classical

SVM (regularized empirical risk) in the RKHS $\mathcal{H}$ has the form

$$\min_{f \in \mathcal{H}} \sum_{j=1}^{N} L\left(\boldsymbol{x}_j, y_j, f(\boldsymbol{x}_j)\right) + R\left(\|f\|_{\mathcal{H}}\right),$$

where L is a *loss function* and R is a *regularization function* (see Theorem 3.1). In the same way we are able to apply an optimal recovery of RKBSs to solve SVMs in RKBSs. Theorem 4.2 establishes that the SVM in the right-sided RKBS $\mathcal{B}$ with the reproducing kernel $K : \Omega_2 \times \Omega_1 \to \mathbb{C}$ based on the training data $D \subseteq \Omega_1 \times \mathbb{C}$ satisfies

$$\min_{f \in \mathcal{B}} \sum_{j=1}^{N} L\left(\boldsymbol{x}_j, y_j, f(\boldsymbol{x}_j)\right) + R\left(\|f\|_{\mathcal{B}}\right).$$

Moreover, this problem has a unique optimal solution $s_{D,L,R}$ and its dual (normalized-duality-mapping) element $s_{D,L,R}^*$ is a linear combination of the reproducing kernel centered at the training data points $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N\} \subseteq \Omega_1$, i.e.,

$$s_{D,L,R}^*(\boldsymbol{x}) = \sum_{k=1}^{N} c_k K(\boldsymbol{x}, \boldsymbol{x}_k), \quad \boldsymbol{x} \in \Omega_2.$$

According to Corollary 4.3, the coefficient vector $\boldsymbol{c} := (c_1, \cdots, c_N)^T$ of $s_{D,L,R}^*$ is a fixed point of the function $F_{D,L,R}^* : \mathbb{R}^N \to \mathbb{R}^N$ dependent of the differential loss function L and the differential regularization function R, i.e., $F_{D,L,R}^*(\boldsymbol{c}) = \boldsymbol{c}$. From this it is obvious that the SVM in the RKBS is the generalization of the classical method in the RKHS.

In Section 5, we show how to use a positive definite function Φ to set up different RKBSs $\mathcal{B}_\Phi^p(\mathbb{R}^d)$ and $\mathcal{B}_\Phi^p(\Omega)$ with $p > 1$ whose two-sided reproducing kernel is given by $K(\boldsymbol{x}, \boldsymbol{y}) = \Phi(\boldsymbol{x} - \boldsymbol{y})$ (see Theorems 5.1 and 5.6). We can observe that $\mathcal{B}_\Phi^p(\mathbb{R}^d)$ is a kind of generalized native space. Furthermore, $\mathcal{B}_\Phi^p(\mathbb{R}^d)$ and $\mathcal{B}_\Phi^p(\Omega)$ coincide with the definition of RKBSs given in [23]. The SVM solution $s_{D,L,R}$ in $\mathcal{B}_\Phi^p(\mathbb{R}^d)$ can be represented by the positive definite function Φ, which means that we can obtain an explicit formula for the SVM solution $s_{D,L,R}$ in $\mathcal{B}_\Phi^p(\mathbb{R}^d)$ (see Theorem 5.4). Corollary 5.5 shows that the finite dimensional coefficients of the SVM solution $s_{D,L,R}$ can even be obtained by solving a fixed point iteration problem for differentiable loss functions and regularization functions. Theorem 5.6 and Corollary 5.7 give some examples of reproducing kernels defined on nonsymmetric domains. Corollary 5.3 and 5.8 provide that RKBSs can be embedded into Sobolev spaces for some special reproducing kernels, e.g., Sobolev-spline kernels (Matérn functions).

The Matérn functions represent a fast growing research area which has frequent applications in approximation theory and statistical learning, and moreover, they are positive definite functions and (full-space) Green functions (see [7, 9, 14, 22]). In Section 6, we solve the SVMs in the RKBSs of Matérn functions. If $G_{\theta,n}$ is the Matérn function with parameter $\theta > 0$ and degree $n > 3d/2$ then, according to our theoretical results, $\mathcal{B}^2_{G_{\theta,n}}(\mathbb{R}^d)$ is an RKHS, while $\mathcal{B}^4_{G_{\theta,n}}(\mathbb{R}^d)$ is only an RKBS. Their reproducing kernels, however, are the same Sobolev-spline kernel $K_{\theta,n}(\boldsymbol{x}, \boldsymbol{y}) := G_{\theta,n}(\boldsymbol{x} - \boldsymbol{y})$. It is well known that the SVM solution in $\mathcal{B}^2_{G_{\theta,n}}(\mathbb{R}^d) \equiv \mathcal{H}_{G_{\theta,n}}(\mathbb{R}^d)$ has the explicit expression

$$s_{D,L,R}(\boldsymbol{x}) := \sum_{k=1}^{N} c_k K_{\theta,n}(\boldsymbol{x}, \boldsymbol{x}_k), \quad \boldsymbol{x} \in \mathbb{R}^d,$$

(see Theorem 3.1). In this paper we discover a new fact that the SVM solution in $\mathcal{B}^4_{G_{\theta,n}}(\mathbb{R}^d)$ also has an explicit form, namely

$$s_{D,L,R}(\boldsymbol{x}) = \sum_{k_1,k_2,k_3=1}^{N,N,N} c_{k_1} \overline{c_{k_2}} c_{k_3} \mathcal{K}_{\theta,3n}\left(\boldsymbol{x}, \boldsymbol{x}_{k_1}, \boldsymbol{x}_{k_2}, \boldsymbol{x}_{k_3}\right), \quad \boldsymbol{x} \in \mathbb{R}^d,$$

where $\mathcal{K}_{\theta,3n}(\boldsymbol{x}, \boldsymbol{y}_1, \boldsymbol{y}_2, \boldsymbol{y}_3) := G_{\theta,3n}(\boldsymbol{x} - \boldsymbol{y}_1 + \boldsymbol{y}_2 - \boldsymbol{y}_3)$. Section 6 shows that several other explicit representations of SVM solutions in the RKBS $\mathcal{B}^p_{G_{\theta,n}}(\mathbb{R}^d)$ are easily computable when p is an even number. This discovery could lead to a new numerical tool for SVMs.

For the binary classification problems, it is well-known that the classical hinge loss is designed to maximize the 2-norm margins by using the linear functions. However, we can not employ the hinge loss to set up the SVMs in order to maximize other p-norm margins. We guess that for applications to the problems that arise in current practice it will be necessary to construct loss functions depending on different kinds of RKBSs.

Remark 1.1. In this paper, the third author hopes to correct a mistake concerning the optimal recovery of RKBS $\mathcal{B}^p_\Phi(\mathbb{R}^d)$ mentioned in [22, Section 6.2]. Theorem 5.4 is the correction of [22, Theorem 6.5], which was the result of a misconception that the normalized duality mapping is linear. The main ideas and techniques used in the corrected version below are still the same as in [22]. An updated version of [22] has been posted on Ye's webpage.

2. Banach Spaces

In this section, we review some classical theoretical results for Banach spaces from [11, 13, 15, 16]. We denote the *dual space* (the collection of all bounded linear functionals) of a Banach space $\mathcal{B}$ by $\mathcal{B}'$ and its *dual bilinear product* as $\langle \cdot, \cdot \rangle_{\mathcal{B}}$, i.e.,

$$\langle f, T \rangle_{\mathcal{B}} := T(f), \quad \text{for all } T \in \mathcal{B}' \text{ and all } f \in \mathcal{B}.$$

[16, Theorem 1.10.7] states that $\mathcal{B}'$ is also a Banach space.

If the Banach spaces $\mathcal{B}_1$ and $\mathcal{B}_2$ are *isometrically isomorphic* (equivalent), i.e., $\mathcal{B}_1 \equiv \mathcal{B}_2$, then we can think of both spaces as being identical in the sense that their norms and their elements can be seen to be the same in both spaces (see [16, Definition 1.4.13]). We say that $\mathcal{B}_1$ is *embedded into* $\mathcal{B}_2$ if there exists a positive constant C such that $\|f\|_{\mathcal{B}_2} \le C \|f\|_{\mathcal{B}_1}$ for all $f \in \mathcal{B}_1 \subseteq \mathcal{B}_2$ (see [1, Section 1.25]).

If the Banach space $\mathcal{B}$ is *reflexive* (see [16, Definition 1.11.6]), then we have $\mathcal{B}'' \equiv \mathcal{B}$ and $\langle f, g \rangle_{\mathcal{B}} = \langle g, f \rangle_{\mathcal{B}'}$ for all $f \in \mathcal{B}$ and all $g \in \mathcal{B}'$. For example, the function space $L_p(\Omega; \mu)$ defined on the positive measure space $(\Omega, \mathscr{B}_\Omega, \mu)$ is a reflexive Banach space and its dual space is isometrically equivalent to $L_q(\Omega; \mu)$ where $p, q > 1$ and $p^{-1} + q^{-1} = 1$ (see [16, Example 1.10.2 and Theorem 1.11.10]). For the complex situation, the isometric isomorphism from $L_p(\Omega; \mu)'$ onto $L_q(\Omega; \mu)$ is antilinear.

We say that $\mathcal{B}$ is *uniformly convex* if, for every $\epsilon > 0$, there is $\delta > 0$ such that

$$\left\| \frac{f + g}{2} \right\|_{\mathcal{B}} \le 1 - \delta, \text{ whenever } \|f\|_{\mathcal{B}} = \|g\|_{\mathcal{B}} = 1 \text{ and } \|f - g\|_{\mathcal{B}} \ge \epsilon$$

(see [16, Definition 5.2.1]). According to [16, Definition 5.4.1, 5.4.15 and Corollary 5.4.18], $\mathcal{B}$ is said to be *smooth* or *Gâteaux differentiable* if

$$\lim_{\lambda \to 0} \frac{\|f + \lambda g\|_{\mathcal{B}} - \|f\|_{\mathcal{B}}}{\lambda} \text{ exists }, \quad \text{for all } f, g \in \mathcal{B}.$$

A typical case is that $L_p(\Omega; \mu)$ is uniformly convex and smooth if $1 < p < \infty$.

It is well known that we can discuss the orthogonality in Banach spaces with a more general axiom system than that in Hilbert spaces. The papers [11, 13, 15] show that every Banach space can be represented as a semi-inner-product space in order that the theories of Banach space can be penetrated by Hilbert space type arguments. A *semi-inner-product* $[\cdot, \cdot]_{\mathcal{B}} : \mathcal{B} \times \mathcal{B} \to \mathbb{C}$ defined on a Banach space $\mathcal{B}$ is given by

(i) $[f + g, h]_{\mathcal{B}} = [f, h]_{\mathcal{B}} + [g, h]_{\mathcal{B}}$, (ii) $[f, f]_{\mathcal{B}} = \|f\|_{\mathcal{B}}^2$,

(iii) $[\lambda f, g]_{\mathcal{B}} = \lambda[f, g]_{\mathcal{B}}$, $[f, \lambda g]_{\mathcal{B}} = \overline{\lambda}[f, g]_{\mathcal{B}}$, (iv) $|[f, g]_{\mathcal{B}}| \le [f, f]_{\mathcal{B}}[g, g]_{\mathcal{B}}$,

5

for all $f, g, h \in \mathcal{B}$ and all $\lambda \in \mathbb{C}$. However, Hermitian symmetry of the semi-inner-product may not hold, i.e., $[f, g]_{\mathcal{B}} \neq \overline{[g, f]_{\mathcal{B}}}$. This indicates that the generality of the semi-inner-product in Banach space is a serious limitation for any extensive development that parallels the inner product of Hilbert space.

For example, a semi-inner-product of $L_p(\Omega; \mu)$ with $1 < p < \infty$ is given by

$$[g, f]_{L_p(\Omega;\mu)} = \frac{1}{\|f\|_{L_p(\Omega;\mu)}^{p-2}} \int_\Omega g(\boldsymbol{x})\overline{f(\boldsymbol{x})}\,|f(\boldsymbol{x})|^{p-2}\,\mathrm{d}\mu(\boldsymbol{x}), \quad \text{for all } f, g \in L_p(\Omega; \mu),$$

(see examples in [11, 13]).

We say that f is *orthogonal* to g in a Banach space $\mathcal{B}$ if

$$\|f + \lambda g\|_{\mathcal{B}} \geq \|f\|_{\mathcal{B}}, \quad \text{for all } \lambda \in \mathbb{C},$$

(see the definitions in [11, 13]). Suppose that the Banach space $\mathcal{B}$ is smooth. Using [11, Theorem 2], we can determine that f is orthogonal to g if and only if f is *normal to g*, i.e.,

$$[g, f]_{\mathcal{B}} = 0.$$

We can also obtain a representation theorem in Banach space by an adaptation of the representation theorem in Hilbert space. Suppose that the Banach space $\mathcal{B}$ is uniformly convex and smooth. According to [11, Theorem 3 and 6], for every bounded linear functional $T \in \mathcal{B}'$, there exists a unique $f \in \mathcal{B}$ such that

$$T(g) = \langle g, T \rangle_{\mathcal{B}} = [g, f]_{\mathcal{B}}, \quad \text{for all } g \in \mathcal{B},$$

and $\|T\|_{\mathcal{B}'} = \|f\|_{\mathcal{B}}$. This mapping is also surjective. We call T the *normalized-duality-mapping element* of f and rewrite it as $f^* := T$. For convenience we simplify normalized-duality-mapping element to dual element in this paper. The normalized duality mapping is a one-to-one and norm-preserving mapping from $\mathcal{B}$ onto $\mathcal{B}'$. Note that this mapping is usually nonlinear. According to [11, Theorem 7], the semi-inner-product of $\mathcal{B}'$ has the form $[f^*, g^*]_{\mathcal{B}'} = [g, f]_{\mathcal{B}}$ for all $f^*, g^* \in \mathcal{B}'$. For example, the dual element of $f \in L_p(\Omega; \mu)$ with $1 < p < \infty$ is given by

$$f^* = \frac{f(\boldsymbol{x})\,|f(\boldsymbol{x})|^{p-2}}{\|f\|_{L_p(\Omega;\mu)}^{p-2}} \in L_q(\Omega; \mu),$$

where q is the conjugate exponent of p. Let $\mathcal{N}$ be a subset of $\mathcal{B}$. We can check that f is orthogonal to $\mathcal{N}$ if and only if its dual element $f^* \in \mathcal{N}^\perp = \{\eta \in \mathcal{B}' : \langle h, \eta \rangle_{\mathcal{B}} = 0, \text{ for all } h \in \mathcal{N}\}$, i.e.,

$$[h, f]_{\mathcal{B}} = \langle h, f^* \rangle_{\mathcal{B}} = 0, \quad \text{for all } h \in \mathcal{N}.$$

3. Reproducing Kernels and Reproducing Kernel Hilbert Spaces

Most of the material presented in this section can be found in the monographs [7, 18, 21]. For the reader's convenience we repeat here what is essential to our discussion later on.

Definition 3.1 ([21, Definition 10.1])**.** Let $\Omega \subseteq \mathbb{R}^d$ and $\mathcal{H}$ be a Hilbert space consisting of functions $f : \Omega \to \mathbb{C}$. $\mathcal{H}$ is called a *reproducing kernel Hilbert space* (RKHS) and a kernel function $K : \Omega \times \Omega \to \mathbb{C}$ is called a *reproducing kernel* for $\mathcal{H}$ if

(i) $K(\cdot, \mathbf{y}) \in \mathcal{H}$ and (ii) $f(\mathbf{y}) = (f, K(\cdot, \mathbf{y}))_{\mathcal{H}}$, for all $f \in \mathcal{H}$ and all $\mathbf{y} \in \Omega$,

where $(\cdot, \cdot)_{\mathcal{H}}$ is used to denote the inner product of $\mathcal{H}$.

Remark 3.1. In order to simplify our discussion and proofs, we let all kernel functions be complex-valued and all function spaces be composed of complex-valued functions in this paper. According to [16, Proposition 1.9.3], it is not difficult for us to restrict the theoretical results to real kernel functions and function spaces.

3.1. Optimal Recovery in Reproducing Kernel Hilbert Spaces

Theorem 3.1 (Representer theorem [18, Theorem 5.5])**.** *Let $\mathcal{H}$ be a reproducing kernel Hilbert space with a reproducing kernel K defined on $\Omega \subseteq \mathbb{R}^d$, and a regularization function $R : [0, \infty) \to [0, \infty)$ be convex and strictly increasing. We choose the loss function $L : \Omega \times \mathbb{C} \times \mathbb{C} \to [0, \infty)$ such that $L(\mathbf{x}, y, \cdot)$ is a convex map for any fixed $\mathbf{x} \in \Omega$ and any fixed $y \in \mathbb{C}$. Given the data $D := \{(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_N, y_N)\}$, with pairwise distinct data points $X = \{\mathbf{x}_1, \dots, \mathbf{x}_N\} \subseteq \Omega$ and associated data values $Y = \{y_1, \dots, y_N\} \subset \mathbb{C}$, the optimal solution (support vector machine solution) $s_{D,L,R}$ of*

$$\min_{f \in \mathcal{H}} \sum_{j=1}^{N} L\left(\mathbf{x}_j, y_j, f(\mathbf{x}_j)\right) + R\left(\|f\|_{\mathcal{H}}\right),$$

has the explicit representation

$$s_{D,L,R}(\mathbf{x}) = \sum_{k=1}^{N} c_k K(\mathbf{x}, \mathbf{x}_k), \quad \mathbf{x} \in \Omega,$$

for some coefficients $c_1, \dots, c_N \in \mathbb{C}$.

7

3.2. Constructing Reproducing Kernel Hilbert Spaces by Positive Definite Functions

Definition 3.2 ([21, Definition 6.1]). A continuous even function $\Phi : \mathbb{R}^d \to \mathbb{C}$ is called *positive definite* if, for all $N \in \mathbb{N}$ and all sets of pairwise distinct centers $X = \{x_1, \ldots, x_N\} \subset \mathbb{R}^d$, the quadratic form

$$\sum_{j=1}^{N} \sum_{k=1}^{N} c_j \overline{c_k} \Phi(x_j - x_k) = c^* A_{\Phi,X} c > 0, \quad \text{for all } c \in \mathbb{C}^N \backslash \{0\}.$$

Here the interpolation matrix $A_{\Phi,X} := \left(\Phi(x_j - x_k) \right)_{j,k=1}^{N,N} \in \mathbb{C}^{N \times N}$ and $c^* = \overline{c}^T$.

We say Φ is even if $\Phi(x) = \overline{\Phi(-x)}$. This shows that Φ is a positive definite function if and only if $A_{\Phi,X}$ is a positive definite matrix for any pairwise distinct finite set X of data points in $\mathbb{R}^d$. The application and history of positive definite functions can be seen in the review paper [8]. [21, Section 10.2] shows how to use positive definite functions to construct RKHSs.

Theorem 3.2 ([21, Theorem 6.11]). *Suppose that $\Phi \in C(\mathbb{R}^d) \cap L_1(\mathbb{R}^d)$. Then Φ is positive definite if and only if Φ is bounded and its Fourier transform $\hat{\Phi}$ is nonnegative and nonvanishing (nonzero everywhere).*

Remark 3.2. In this paper, the Fourier transform of $f \in L_1(\mathbb{R}^d)$ is defined by

$$\hat{f}(x) := (2\pi)^{-d/2} \int_{\mathbb{R}^d} f(y) e^{-ix^T y} dy,$$

where i is the imaginary unit, i.e., $i^2 = -1$.

Theorem 3.3 ([21, Theorem 10.12]). *Suppose that $\Phi \in C(\mathbb{R}^d) \cap L_1(\mathbb{R}^d)$ is a positive definite function. Then the space*

$$\mathcal{H}_\Phi(\mathbb{R}^d) := \left\{ f \in L_2(\mathbb{R}^d) \cap C(\mathbb{R}^d) : \hat{f}/\hat{\Phi}^{1/2} \in L_2(\mathbb{R}^d) \right\},$$

equipped with the norm

$$\|f\|_{\mathcal{H}_\Phi(\mathbb{R}^d)} := \left((2\pi)^{-d/2} \int_{\mathbb{R}^d} \frac{\left| \hat{f}(x) \right|^2}{\hat{\Phi}(x)} dx \right)^{1/2}$$

is a reproducing kernel Hilbert space (native space) with reproducing kernel given by

$$K(x, y) := \Phi(x - y), \quad x, y \in \mathbb{R}^d,$$

8

where $\hat{\Phi}$ and $\hat{f}$ are the Fourier transforms of Φ and f, respectively. The inner product in $\mathcal{H}_\Phi(\mathbb{R}^d)$ has the form

$$(f, g)_{\mathcal{H}} = (2\pi)^{-d/2} \int_{\mathbb{R}} \frac{\hat{f}(\boldsymbol{x})\overline{\hat{g}(\boldsymbol{x})}}{\hat{\Phi}(\boldsymbol{x})} \mathrm{d}\boldsymbol{x}, \quad f, g \in \mathcal{H}_\Phi(\mathbb{R}^d).$$

Using Fourier transform techniques similar to those in Theorem 3.3, we can employ positive definite functions to set up RKBSs (see Section 5).

4. Reproducing Kernels and Reproducing Kernel Banach Spaces

Now we give the definition of RKBSs as a natural generalization of RKHSs by viewing the inner product as a dual bilinear product.

Definition 4.1. Let Ω_1 and Ω_2 be two subsets of $\mathbb{R}^{d_1}$ and $\mathbb{R}^{d_2}$ respectively, and $\mathcal{B}$ be a Banach space composed of functions $f : \Omega_1 \to \mathbb{C}$, whose dual space $\mathcal{B}'$ is isometrically equivalent to a function space $\mathcal{F}$ with $g : \Omega_2 \to \mathbb{C}$. Denote that $K : \Omega_2 \times \Omega_1 \to \mathbb{C}$ is a kernel function.

We call $\mathcal{B}$ a *reproducing kernel Banach space* (RKBS) and K its *right-sided reproducing kernel* if

(i) $K(\cdot, \boldsymbol{y}) \in \mathcal{F} \equiv \mathcal{B}'$ and (ii) $f(\boldsymbol{y}) = \langle f, K(\cdot, \boldsymbol{y}) \rangle_{\mathcal{B}}$, for all $f \in \mathcal{B}$ and all $\boldsymbol{y} \in \Omega_1$.

If the Banach space $\mathcal{B}$ reproduces from the other side, i.e.,

(iii) $\overline{K(\boldsymbol{x}, \cdot)} \in \mathcal{B}$ and (iv) $\overline{g(\boldsymbol{x})} = \langle \overline{K(\boldsymbol{x}, \cdot)}, g \rangle_{\mathcal{B}}$, for all $g \in \mathcal{F} \equiv \mathcal{B}'$ and all $\boldsymbol{x} \in \Omega_2$,

then $\mathcal{B}$ is called a *reproducing kernel Banach space* and K its *left-sided reproducing kernel*.

For two-sided reproduction as above we say that $\mathcal{B}$ is a *reproducing kernel Banach space* with the *two-sided reproducing kernel K*.

Remark 4.1. We know that the Riesz representer map on complex Hilbert space $\mathcal{H}$ is antilinear, i.e.,

$$T_{\lambda g}(f) = \langle f, \lambda g \rangle_{\mathcal{H}} = \overline{\lambda}(f, g)_{\mathcal{H}} = \overline{\lambda}\langle f, g \rangle_{\mathcal{H}} = \overline{\lambda} T_g(f),$$

for all $f, g \in \mathcal{H}$ and all $\lambda \in \mathbb{C}$. Here we also let the isometrical isomorphism from the dual space $\mathcal{B}'$ onto the related function space $\mathcal{F}$ be antilinear. Thus, the format of two-sided RKBSs coincides with complex RKHSs, i.e.,

$$\langle \overline{K(\boldsymbol{y}, \cdot)}, f \rangle_{\mathcal{H}} = (\overline{K(\boldsymbol{y}, \cdot)}, f)_{\mathcal{H}} = \overline{(f, K(\cdot, \boldsymbol{y}))}_{\mathcal{H}} = \overline{f(\boldsymbol{y})}, \quad \text{for all } f \in \mathcal{H} \text{ and all } \boldsymbol{y} \in \Omega,$$

which indicates that the RKHS is a special case of a two-sided RKBS.

Why do we define our RKBSs differently from [23, Definition 1]? The reason is that we can show the optimal recovery in an RKBS even if it is only one-sided. We do not require a reflexivity condition for the definition of our RKBS. Moreover, since the dual space of a Hilbert space is isometrically equivalent to itself, we can choose the equivalent function space $\mathcal{F} \equiv \mathcal{H}$ such that the domain of the reproducing kernel K is symmetric, i.e., $\Omega_2 = \Omega_1$. Actually, the Banach space $\mathcal{B}$ is usually not equal to any equivalent function space $\mathcal{F}$ of its dual $\mathcal{B}'$ even though we only require them to be isomorphic. We naturally do not need any symmetry conditions in the Banach space. Therefore the nonsymmetric domain is used to define the RKBS $\mathcal{B}$ and its reproducing kernel K, i.e., $\Omega_2 \neq \Omega_1$. The domain of K is related to both $\mathcal{B}$ and $\mathcal{F} \equiv \mathcal{B}'$. If we choose a different $\mathcal{F}$ which is isometrically equivalent to the dual $\mathcal{B}'$, then we can obtain a different reproducing kernel K of the RKBS $\mathcal{B}$ dependent on its equivalent dual space $\mathcal{F}$.

The functional $K(\cdot, \mathbf{y})$ can be seen as a point evaluation function $\delta_{\mathbf{y}}$ defined on $\mathcal{B}$. This implies that $\delta_{\mathbf{y}}$ is a bounded linear functional on $\mathcal{B}$, i.e., $\delta_{\mathbf{y}} \in \mathcal{B}'$. If the Banach space $\mathcal{B}$ is further uniformly convex and smooth, then its semi-inner-product and its normalized duality mapping are well-defined, which can be used to set up the equivalent conditions of right-sided RKBSs, i.e.,

$$\delta_{\mathbf{y}} \in \mathcal{B}' \equiv \mathcal{F} \text{ which indicates that } f(\mathbf{y}) = \langle f, \delta_{\mathbf{y}} \rangle_{\mathcal{B}} = [f, \delta_{\mathbf{y}}^*]_{\mathcal{B}},$$

for all $f \in \mathcal{B}$ and all $\mathbf{y} \in \Omega_1$ (see the discussions of the semi-inner products in Section 2).

If $\mathcal{B}$ is a reflexive two-sided RKBS, then the equivalent dual space $\mathcal{F}$ of $\mathcal{B}$ is also a reflexive two-sided RKBS. All RKBSs and reproducing kernels set up in Section 5 satisfy the two-sided definition but their domains can be symmetric or nonsymmetric.

If a sequence $\{f_n\}_{n=1}^{\infty} \subset \mathcal{B}$ and $f \in \mathcal{B}$ such that $\|f - f_n\|_{\mathcal{B}} \to 0$ when $n \to \infty$, then

$$|f(\mathbf{y}) - f_n(\mathbf{y})| = |\langle f - f_n, K(\cdot, \mathbf{y}) \rangle_{\mathcal{B}}| \leq \|K(\cdot, \mathbf{y})\|_{\mathcal{B}'} \|f - f_n\|_{\mathcal{B}} \to 0, \quad \mathbf{y} \in \Omega_1,$$

when $n \to \infty$. This means that convergence in the right-sided RKBS $\mathcal{B}$ implies pointwise convergence.

Suppose that $\mathcal{B}$ is a reflexive right-sided RKBS. We show that $\{K(\cdot, \mathbf{y}) : \mathbf{y} \in \Omega_1\}$ is a linear vector space basis of $\mathcal{F}$ and $\text{span}\{K(\cdot, \mathbf{y}) : \mathbf{y} \in \Omega_1\}$ is dense in $\mathcal{F}$. Let $\mathcal{N}$ be a completion (closure) of $\text{span}\{K(\cdot, \mathbf{y}) : \mathbf{y} \in \Omega_1\} \subseteq \mathcal{F} \equiv \mathcal{B}'$ with its dual norm. Now we prove that $\mathcal{N} \equiv \mathcal{F} \equiv \mathcal{B}'$. Since [16, Theorem 1.10.7] provides that $\mathcal{F}$ is also a Banach space, we have $\mathcal{N} \subseteq \mathcal{F}$. Assume that $\mathcal{N} \subsetneq \mathcal{F}$. According

to [16, Corollary 1.9.7] (application of Hahn-Banach extension theorems) there is an element $f \in \mathcal{B} \equiv \mathcal{B}'' \equiv \mathcal{F}'$ such that $\|f\|_{\mathcal{B}} = 1$ and $f(\mathbf{y}) = \langle f, K(\cdot, \mathbf{y}) \rangle_{\mathcal{B}} = 0$ for all $\mathbf{y} \in \Omega_1$. We find the contradiction between $\|f\|_{\mathcal{B}} = 1$ and $f = 0$. Thus the first assumption is not true and then we can conclude that $\mathcal{N} \equiv \mathcal{F} \equiv \mathcal{B}'$, which indicates that $\{K(\cdot, \mathbf{y}) : \mathbf{y} \in \Omega_1\}$ is a linear vector space basis of $\mathcal{F}$ and $\{\delta_{\mathbf{y}} : \mathbf{y} \in \Omega_1\}$ is a linear vector space basis of $\mathcal{B}'$.

Example 4.1. We give a simple example of a two-sided RKBS. Let $\Omega_2 = \Omega_1 := \{1, \cdots, n\}$ and $\mathsf{A} \in \mathbb{C}^{n \times n}$ be a symmetric positive definite matrix. It can be decomposed into $\mathsf{A} = \mathsf{V}\mathsf{D}\mathsf{V}^*$, where D is a positive diagonal matrix and V is an orthogonal matrix. We choose $p, q > 1$ such that $p^{-1} + q^{-1} = 1$. Define $\mathcal{B} := \{f : \Omega_1 \to \mathbb{C}\}$ equipped with the norm

$$\|f\|_{\mathcal{B}} := \left\|\mathsf{D}^{-1/q}\mathsf{V}^*\boldsymbol{f}\right\|_q, \quad \text{where } \boldsymbol{f} := (f(1), \cdots, f(n))^T.$$

We can check that $\mathcal{B}$ is a Banach space and its dual space $\mathcal{B}'$ is isometrically equivalent to $\mathcal{F} := \{g : \Omega_2 \to \mathbb{C}\}$ equipped with the norm

$$\|g\|_{\mathcal{B}'} := \left\|\mathsf{D}^{-1/p}\mathsf{V}^*\boldsymbol{g}\right\|_p, \quad \text{where } \boldsymbol{g} := (g(1), \cdots, g(n))^T.$$

Moreover, its dual bilinear form is given by

$$\langle f, g \rangle_{\mathcal{B}} = \boldsymbol{g}^*\mathsf{A}^{-1}\boldsymbol{f}, \quad \text{for all } f \in \mathcal{B} \text{ and all } g \in \mathcal{B}'.$$

If the kernel function is defined by

$$K(j, k) := \mathsf{A}_{jk}, \quad j \in \Omega_2, \ k \in \Omega_1,$$

then the reproduction can easily be verified, i.e.,

$$\langle f, K(\cdot, k) \rangle_{\mathcal{B}} = f(k), \quad k \in \Omega_1, \quad \text{and} \quad \overline{\langle K(j, \cdot), g \rangle_{\mathcal{B}}} = \overline{g(j)}, \quad j \in \Omega_2.$$

Therefore $\mathcal{B}$ is indeed a two-sided RKBS.

(In the same way, we can also employ the singular value decomposition of a nonsymmetric and nonsingular square matrix A to introduce the two-sided RKBS.)

4.1. Optimal Recovery in Reproducing Kernel Banach Spaces

It is well-known that any Hilbert space is uniformly convex and smooth. It is natural for us to assume the right-sided RKBS is further uniformly convex and smooth to discuss optimal recovery in it. The definitions of uniform convexity and smoothness of Banach spaces are given in Section 2.

Given the pairwise distinct data points $X = \{x_1, \ldots, x_N\} \subseteq \Omega_1$ and the associated data values $Y = \{y_1, \ldots, y_N\} \subset \mathbb{C}$, we define a subset of the right-sided RKBS $\mathcal{B}$ by

$$\mathcal{N}_\mathcal{B}(X, Y) := \left\{ f \in \mathcal{B} : \ f(x_j) = y_j, \ \text{for all } j = 1, \ldots, N \right\}.$$

If $\mathcal{N}_\mathcal{B}(X, Y)$ is the null set, then there is no meaning for the SVMs. So we need to assume that $\mathcal{N}_\mathcal{B}(X, Y)$ is always non-null for the given data sites. Actually we can show that $\mathcal{N}_\mathcal{B}(X, Y)$ is non-null for any data values Y if and only if $\delta_{x_1}, \ldots, \delta_{x_N}$ are linearly independent on $\mathcal{B}$ because $\sum_{k=1}^N c_k \delta_{x_k} = 0$ if and only if $\sum_{k=1}^N c_k f(x_k) = 0$ for all $f \in \mathcal{B}$, and moreover, $c = (c_1, \cdots, c_N)^T = 0$ if and only if $b^* c = 0$ for all $b \in \mathbb{C}^N$.

In this section, we suppose that $\delta_{x_1}, \ldots, \delta_{x_N}$ are always *linearly independent* on $\mathcal{B}$ for the given pairwise distinct data points X, which is equivalent to the fact that $K(\cdot, x_1), \ldots, K(\cdot, x_N)$ are linearly independent. We use the techniques of [23, Theorem 19] to verify the following lemma.

Lemma 4.1. *Let $\mathcal{B}$ be a reproducing kernel Banach space with a right-sided reproducing kernel K defined on $\Omega_2 \times \Omega_1 \subseteq \mathbb{R}^{d_2} \times \mathbb{R}^{d_1}$. Suppose that $\mathcal{B}$ is uniformly convex and smooth. Given the data $D := \{(x_1, y_1), \ldots, (x_N, y_N)\}$ with pairwise distinct data points $X = \{x_1, \ldots, x_N\} \subseteq \Omega_1$ and associated data values $Y = \{y_1, \ldots, y_N\} \subset \mathbb{C}$, the dual element s_D^* of the unique optimal solution*

$$s_D := \underset{f \in \mathcal{B}}{argmin} \left\{ \|f\|_\mathcal{B} : \ f(x_j) = y_j, \text{ for all } j = 1, \ldots, N \right\}, \tag{4.1}$$

is the linear combination of $K(\cdot, x_1), \ldots, K(\cdot, x_N)$, i.e.,

$$s_D^*(x) = \sum_{k=1}^N c_k K(x, x_k), \quad x \in \Omega_2.$$

Proof. We first prove the uniqueness of the optimal solution of the minimization problem (4.1). Let us assume that the minimization problem (4.1) has two optimal solutions $s_1, s_2 \in \mathcal{B}$ with $s_1 \neq s_2$. Since $\mathcal{B}$ is uniformly convex, [16, Corollary 5.1.12] provides that $\left\| \frac{1}{2}(s_1 + s_2) \right\|_\mathcal{B} < \frac{1}{2}\|s_1\|_\mathcal{B} + \frac{1}{2}\|s_2\|_\mathcal{B}$. Then $\|s_1\|_\mathcal{B} = \|s_2\|_\mathcal{B}$ shows for $s_3 := \frac{1}{2}(s_1 + s_2)$ that $\|s_3\|_\mathcal{B} < \|s_1\|_\mathcal{B}$ and $s_3 \in \mathcal{N}_\mathcal{B}(X, Y)$, i.e., s_1 is not an optimal solution of the minimization problem (4.1). The assumption that there are two minimizers is false.

Next we show the existence of the minimizer. The minimization problem (4.1) is equivalent to $\min_{f \in \mathcal{N}_\mathcal{B}(X,Y)} \|f\|_\mathcal{B}$. Since convergence in a one-sided RKBS $\mathcal{B}$ implies pointwise convergence, we can check that $\mathcal{N}_\mathcal{B}(X, Y)$ is a closed convex subset of $\mathcal{B}$. Combining this with the uniform convexity of $\mathcal{B}$, [16, Corollary 5.2.17]

shows that $\mathcal{N}_{\mathcal{B}}(X, Y)$ is a Chebyshev set (see [16, Definition 5.1.17]). Thus an optimal solution $\min_{f \in \mathcal{N}_{\mathcal{B}}(X,Y)} \|f\|_{\mathcal{B}}$ exists.

Because $\mathcal{N}_{\mathcal{B}}(X, Y) + \mathcal{N}_{\mathcal{B}}(X, \{0\}) = \mathcal{N}_{\mathcal{B}}(X, Y)$ and $\mathcal{N}_{\mathcal{B}}(X, \{0\})$ is a closed subspace of $\mathcal{B}$ we can determine that the optimal solution s_D is orthogonal to $\mathcal{N}_{\mathcal{B}}(X, \{0\})$, i.e., $\|s_D + h\|_{\mathcal{B}} \geq \|s_D\|_{\mathcal{B}}$ for all $h \in \mathcal{N}_{\mathcal{B}}(X, \{0\})$. Since $\mathcal{B}$ is uniformly convex and smooth, the dual element s_D^* of s_D is well-defined and

$$[h, s_D]_{\mathcal{B}} = \langle h, s_D^* \rangle_{\mathcal{B}} = 0, \text{ for all } h \in \mathcal{N}_{\mathcal{B}}(X, \{0\}),$$

which implies that

$$s_D^* \in \mathcal{N}_{\mathcal{B}}(X, \{0\})^\perp = \{g \in \mathcal{F} \equiv \mathcal{B}' : \langle h, g \rangle_{\mathcal{B}} = 0, \text{ for all } h \in \mathcal{N}_{\mathcal{B}}(X, \{0\})\}.$$

It is obvious that

$$\mathcal{N}_{\mathcal{B}}(X, \{0\}) = \left\{f \in \mathcal{B} : f(\boldsymbol{x}_j) = \langle f, K(\cdot, \boldsymbol{x}_j) \rangle_{\mathcal{B}} = 0, \ j = 1, \ldots, N\right\}$$
$$= \left\{f \in \mathcal{B} : \langle f, h \rangle_{\mathcal{B}} = 0, \text{ for all } h \in \text{span}\{K(\cdot, \boldsymbol{x}_k)\}_{k=1}^N\right\} = {}^\perp \text{span}\{K(\cdot, \boldsymbol{x}_k)\}_{k=1}^N.$$

According to [16, Proposition 1.10.15], we have

$$s_D^* \in \left({}^\perp \text{span}\{K(\cdot, \boldsymbol{x}_k)\}_{k=1}^N\right)^\perp = \text{span}\{K(\cdot, \boldsymbol{x}_1), \ldots, K(\cdot, \boldsymbol{x}_N)\}.$$

Here $\mathcal{N}_1^\perp$ and ${}^\perp \mathcal{N}_2$ denote the annihilator of $\mathcal{N}_1$ in $\mathcal{B}'$ and the annihilator of $\mathcal{N}_2$ in $\mathcal{B}$, respectively, where $\mathcal{N}_1 \subseteq \mathcal{B}$ and $\mathcal{N}_2 \subseteq \mathcal{B}'$ (see [16, Definition 1.10.14]). □

Now we verify the representer theorem for SVMs in a right-sided RKBS.

Theorem 4.2. *Let $\mathcal{B}$ be a reproducing kernel Banach space with a right-sided reproducing kernel K defined on $\Omega_2 \times \Omega_1 \subseteq \mathbb{R}^{d_2} \times \mathbb{R}^{d_1}$, and a regularization function $R : [0, \infty) \rightarrow [0, \infty)$ be convex and strictly increasing. Suppose that $\mathcal{B}$ is uniformly convex and smooth. We choose the loss function $L : \Omega_1 \times \mathbb{C} \times \mathbb{C} \rightarrow [0, \infty)$ such that $L(\boldsymbol{x}, y, \cdot)$ is a convex map for any fixed $\boldsymbol{x} \in \Omega_1$ and any fixed $y \in \mathbb{C}$. Given the data $D := \{(\boldsymbol{x}_1, y_1), \ldots, (\boldsymbol{x}_N, y_N)\}$ with pairwise distinct data points $X = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N\} \subseteq \Omega_1$ and associated data values $Y = \{y_1, \ldots, y_N\} \subset \mathbb{C}$, the dual element of the unique optimal solution (support vector machine solution) $s_{D,L,R}$ of*

$$\min_{f \in \mathcal{B}} \sum_{j=1}^N L\left(\boldsymbol{x}_j, y_j, f(\boldsymbol{x}_j)\right) + R\left(\|f\|_{\mathcal{B}}\right), \tag{4.2}$$

13

has the explicit representation

$$s^*_{D,L,R}(\boldsymbol{x}) = \sum_{k=1}^{N} c_k K(\boldsymbol{x}, \boldsymbol{x}_k), \quad \boldsymbol{x} \in \Omega_2,$$

for some coefficients $c_1, \ldots, c_N \in \mathbb{C}$.

Proof. Let

$$T_{D,L,R}(f) := \sum_{j=1}^{N} L\left(\boldsymbol{x}_j, y_j, f(\boldsymbol{x}_j)\right) + R\left(\|f\|_{\mathcal{B}}\right), \quad f \in \mathcal{B}.$$

The minimization problem (4.2) is equivalent to $\min_{f \in \mathcal{B}} T_{D,L,R}(f)$. Since $\mathcal{B}$ is uniformly convex and R is convex and strictly increasing, the regularization $f \mapsto R\left(\|f\|_{\mathcal{B}}\right)$ is continuous and strictly convex. Because the $\mathcal{B}$-norm convergence implies the pointwise convergence and $L\left(\boldsymbol{x}_j, y_j, \cdot\right)$ is convex for all $j = 1, \ldots, N$, the mapping $f \mapsto \sum_{j=1}^{N} L\left(\boldsymbol{x}_j, y_j, f(\boldsymbol{x}_j)\right)$ is also continuous and convex. This indicates the continuity and strict convexity of $T_{D,L,R}$. Using the increasing property of R, we can check that the set $\{f \in \mathcal{B} : T_{D,L,R}(f) \leq T_{D,L,R}(0)\}$ is nonempty and bounded. Moreover, the uniformly convex norm implies its reflexivity by the Milman-Pettis Theorem [16, Theorem 5.2.15]. Thus the existence of minimizers theorem [18, Theorem A.6.9] gives the existence of the unique solution $s_{D,L,R}$ to minimize $T_{D,L,R}$ over $\mathcal{B}$.

We fix any $f \in \mathcal{B}$ and let $D_f := \{(\boldsymbol{x}_k, f(\boldsymbol{x}_k))\}_{k=1}^{N}$. According to Lemma 4.1, there exists an element s_{D_f} whose dual element $s^*_{D_f} \in \text{span}\{K(\cdot, \boldsymbol{x}_k)\}_{k=1}^{N}$ such that s_{D_f} interpolates the data values $\{f(\boldsymbol{x}_k)\}_{k=1}^{N}$ at the centers points $X = \{\boldsymbol{x}_k\}_{k=1}^{N}$ and $\left\|s_{D_f}\right\|_{\mathcal{B}} \leq \|f\|_{\mathcal{B}}$. This indicates that

$$T_{D,L,R}(s_{D_f}) \leq T_{D,L,R}(f).$$

Therefore the dual element $s^*_{D,L,R}$ of the optimal solution $s_{D,L,R}$ of the minimization problem (4.2) belongs to $\text{span}\{K(\cdot, \boldsymbol{x}_k)\}_{k=1}^{N}$.

$\square$

Remark 4.2. Since $K(\cdot, \boldsymbol{x}_j)$ can be seen as a point evaluation functional $\delta_{\boldsymbol{x}_j}$ defined on $\mathcal{B}$, it indicates that the dual element of $s_{D,L,R}$ can be also written as a linear combination of $\delta_{\boldsymbol{x}_1}, \ldots, \delta_{\boldsymbol{x}_N}$, i.e., $s^*_{D,L,R} = \sum_{j=1}^{N} c_j \delta_{\boldsymbol{x}_j}$.

The uniform convexity and smoothness of $\mathcal{B}$ imply the uniform convexity and smoothness of its dual $\mathcal{B}' \equiv \mathcal{F}$. If $\mathcal{B}$ is a left-sided RKBS satisfying uniform

convexity and smoothness conditions, then we can further perform optimal recovery in $\mathcal{F}$ in the same way, i.e., the dual element of the optimal solution (SVM solution) of

$$\min_{g \in \mathcal{F} \equiv \mathcal{B}'} \sum_{j=1}^{N} \tilde{L}\left(\boldsymbol{x}_j, y_j, \overline{g(\boldsymbol{x}_j)}\right) + R\left(\|g\|_{\mathcal{B}'}\right),$$

is a linear combination of $\overline{K(\boldsymbol{x}_1, \cdot)}, \ldots, \overline{K(\boldsymbol{x}_N, \cdot)}$, where $X = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N\} \subseteq \Omega_2$ and $\tilde{L} : \Omega_2 \times \mathbb{C} \times \mathbb{C} \to [0, \infty)$.

Moreover, since the normalized duality mapping is an identity mapping on the Hilbert space and the reproducing kernel of an RKHS is symmetric, optimal recovery in RKBSs as in Theorem 4.2 can be seen as a generalization of optimal recovery in RKHSs as in Theorem 3.1.

Since the normalized duality mapping is one-to-one, for any fixed $\boldsymbol{c} \in \mathbb{C}^N$, there exists an unique $s_{\boldsymbol{c}} \in \mathcal{B}$ such that its dual element has the form $s_{\boldsymbol{c}}^* = \sum_{k=1}^{N} c_k K(\cdot, \boldsymbol{x}_k) = \boldsymbol{k}_X^T \boldsymbol{c}$, where $\boldsymbol{k}_X := (K(\cdot, \boldsymbol{x}_1), \cdots, K(\cdot, \boldsymbol{x}_N))^T$ and $\boldsymbol{c} := (c_1, \cdots, c_N)^T$. According to Theorem 4.2, the SVM (4.2) can be transformed to solve a finite-dimensional optimization problem, i.e.,

$$\boldsymbol{c}_{opt} := \operatorname*{argmin}_{\boldsymbol{c} \in \mathbb{C}^N} \sum_{j=1}^{N} L\left(\boldsymbol{x}_j, y_j, s_{\boldsymbol{c}}(\boldsymbol{x}_j)\right) + R\left(\|s_{\boldsymbol{c}}\|_{\mathcal{B}}\right),$$

and the dual element of the SVM solution has the form $s_{D,L,R}^* = \boldsymbol{k}_X^T \boldsymbol{c}_{opt}$.

Now we want to show that these optimal coefficients $\boldsymbol{c}_{opt}$ can be computed by a fixed point iteration method similar as in [17]. Suppose that $L(\boldsymbol{x}, y, \cdot) \in \mathrm{C}^1(\mathbb{C})$ for all $\boldsymbol{x} \in \Omega_1$ and all $y \in \mathbb{C}$, and $R \in \mathrm{C}^1([0, \infty))$. Let

$$\phi_j^*(\boldsymbol{c}) := [K(\cdot, \boldsymbol{x}_j), \boldsymbol{k}_X^T \boldsymbol{c}]_{\mathcal{B}'} = [K(\cdot, \boldsymbol{x}_j), s_{\boldsymbol{c}}^*]_{\mathcal{B}'}, \quad \boldsymbol{c} \in \mathbb{C}^N, \quad j = 1, \ldots, N,$$

and

$$L'(\boldsymbol{x}, y, t) := \frac{\mathrm{d}}{\mathrm{d}t} L(\boldsymbol{x}, y, t), \quad \boldsymbol{x} \in \Omega_1, \quad y \in \mathbb{C},$$

where $\frac{\mathrm{d}}{\mathrm{d}t}$ represents the Wirtinger derivative defined by

$$\frac{\mathrm{d}}{\mathrm{d}t} := \frac{1}{2}\left(\frac{\mathrm{d}}{\mathrm{d}u} - i\frac{\mathrm{d}}{\mathrm{d}v}\right), \quad \text{where } t = u + iv \text{ with } i^2 = -1 \text{ and } u, v \in \mathbb{R}.$$

Thus we have

$$s_{\boldsymbol{c}}(\boldsymbol{x}_j) = \langle s_{\boldsymbol{c}}, K(\cdot, \boldsymbol{x}_j) \rangle_{\mathcal{B}} = [s_{\boldsymbol{c}}, K(\cdot, \boldsymbol{x}_j)^*]_{\mathcal{B}} = [K(\cdot, \boldsymbol{x}_j), s_{\boldsymbol{c}}^*]_{\mathcal{B}'} = \phi_j^*(\boldsymbol{c}), \quad j = 1, \ldots, N,$$

and

$$\|s_{\boldsymbol{c}}\|_{\mathcal{B}}^2 = [s_{\boldsymbol{c}}, s_{\boldsymbol{c}}]_{\mathcal{B}} = \langle s_{\boldsymbol{c}}, s_{\boldsymbol{c}}^*\rangle_{\mathcal{B}} = \sum_{j=1}^{N} \overline{c_j} \langle s_{\boldsymbol{c}}, K(\cdot, \boldsymbol{x}_j)\rangle_{\mathcal{B}} = \sum_{j=1}^{N} \overline{c_j} \phi_j^*(\boldsymbol{c}) = \boldsymbol{c}^* \boldsymbol{\phi}^*(\boldsymbol{c}),$$

where $\boldsymbol{\phi}^* := \left(\phi_1^*, \cdots, \phi_N^*\right)^T$. Denote that

$$\mathcal{T}_{D,L,R}^*(\boldsymbol{c}) := \sum_{j=1}^{N} L\left(\boldsymbol{x}_j, y_j, \phi_j^*(\boldsymbol{c})\right) + R\left(\sqrt{\boldsymbol{c}^* \boldsymbol{\phi}^*(\boldsymbol{c})}\right) = \sum_{j=1}^{N} L\left(\boldsymbol{x}_j, y_j, s_{\boldsymbol{c}}(\boldsymbol{x}_j)\right) + R\left(\|s_{\boldsymbol{c}}\|_{\mathcal{B}}\right).$$

Since $\boldsymbol{c}_{opt}$ is the global minimizer of $\mathcal{T}_{D,L,R}^*$ over $\mathbb{C}^N$, $\boldsymbol{c}_{opt}$ is a stationary point of $\mathcal{T}_{D,L,R}^*$, i.e., $\nabla \mathcal{T}_{D,L,R}^*(\boldsymbol{c}_{opt}) = 0$. We compute the gradient of $\mathcal{T}_{D,L,R}^*$ by Wirtinger partial derivatives, i.e.,

$$\nabla \mathcal{T}_{D,L,R}^*(\boldsymbol{c})^T = \boldsymbol{l}_D'\left(\boldsymbol{\phi}^*(\boldsymbol{c})\right)^T \nabla \boldsymbol{\phi}^*(\boldsymbol{c}) + \frac{R'\left(\sqrt{\boldsymbol{c}^* \boldsymbol{\phi}^*(\boldsymbol{c})}\right)}{4\sqrt{\boldsymbol{c}^* \boldsymbol{\phi}^*(\boldsymbol{c})}} \boldsymbol{c}^* \nabla \boldsymbol{\phi}^*(\boldsymbol{c}),$$

where $\boldsymbol{l}_D'\left(\boldsymbol{\phi}^*\right) := \left(L'(\boldsymbol{x}_1, y_1, \phi_1^*), \cdots, L'(\boldsymbol{x}_N, y_N, \phi_N^*)\right)^T$ and $\nabla \boldsymbol{\phi}^* := \left(\frac{\partial}{\partial c_k} \phi_j^*\right)_{j,k=1}^{N,N}$ is the Jacobian (gradient) matrix of $\boldsymbol{\phi}^*$ by Wirtinger partial derivatives. The optimal solution $\boldsymbol{c}_{opt}$ is also a fixed point of the function $F_{D,L,R}^*$, i.e.,

$$F_{D,L,R}^*(\boldsymbol{c}_{opt}) = \boldsymbol{c}_{opt},$$

where

$$F_{D,L,R}^*(\boldsymbol{c}) := \boldsymbol{c} + \nabla \mathcal{T}_{D,L,R}^*(\boldsymbol{c}), \quad \boldsymbol{c} \in \mathbb{C}^N \backslash \{\boldsymbol{0}\}. \tag{4.3}$$

Corollary 4.3. *Suppose that the loss function $L(\boldsymbol{x}, y, \cdot) \in \mathrm{C}^1(\mathbb{C})$ for all $\boldsymbol{x} \in \Omega_1$ and all $y \in \mathbb{C}$, and the regularization function $R \in \mathrm{C}^1([0, \infty))$. Then the coefficients $\boldsymbol{c}$ of the dual element $s_{D,L,R}^*$ of the support vector machine solution $s_{D,L,R}$ given in Theorem 4.2 is a fixed point of the function $F_{D,L,R}^*$ defined in Equation (4.3), i.e., $F_{D,L,R}^*(\boldsymbol{c}) = \boldsymbol{c}$.*

Remark 4.3. Even though we can obtain the coefficients of $s_{D,L,R}^*$ by the fixed point iteration method, it is still difficult for us to recover the explicit form $s_{D,L,R}$ in many cases. In Section 5 we discuss how to obtain the SVM solutions in RKBSs induced by positive definite functions (see Theorem 5.4). In that setting the coefficients of the explicit form are also computable by a fixed point iteration method for differentiable loss functions and regularization functions.

5. Constructing Reproducing Kernel Banach Spaces by Positive Definite Functions

Now we construct RKBSs based on positive definite functions in a way similar to the construction of RKHSs in Theorem 3.3. Let $1 < p, q < \infty$ and $p^{-1} + q^{-1} = 1$. Suppose that $\Phi \in C(\mathbb{R}^d) \cap L_1(\mathbb{R}^d)$ is a positive definite function. According to Theorem 3.2, we know that $\hat{\Phi} \in L_1(\mathbb{R}^d) \cap C(\mathbb{R}^d)$ is nonnegative and nonvanishing. We define

$$\mathcal{B}_\Phi^p(\mathbb{R}^d) := \Big\{ f \in C(\mathbb{R}^d) \cap \mathcal{SI} : \text{ the distributional Fourier transform } \hat{f} \text{ of } f \text{ is}$$
$$\text{a measurable function defined on } \mathbb{R}^d \text{ such that } \hat{f}/\hat{\Phi}^{1/q} \in L_q(\mathbb{R}^d) \Big\},$$
$$\tag{5.1}$$

equipped with the norm

$$\|f\|_{\mathcal{B}_\Phi^p(\mathbb{R}^d)} := \left((2\pi)^{-d/2} \int_{\mathbb{R}^d} \frac{\left|\hat{f}(\boldsymbol{x})\right|^q}{\hat{\Phi}(\boldsymbol{x})} \mathrm{d}\boldsymbol{x} \right)^{1/q},$$

where $\mathcal{SI}$ is the collection of all slowly increasing functions (see [21, Definition 5.19]). We define $\mathcal{B}_\Phi^q(\mathbb{R}^d)$ in an analogous way as above.

Remark 5.1. Following the theoretical results of [12, Section 7.1] and [19, Section 1.3] we can define the distributional Fourier transform $\hat{T} \in \mathscr{S}'$ of the tempered distribution $T \in \mathscr{S}'$ by

$$\langle \gamma, \hat{T} \rangle_{\mathscr{S}} := \langle \hat{\gamma}, T \rangle_{\mathscr{S}}, \quad \text{for all } \gamma \in \mathscr{S},$$

where $\mathscr{S}$ is the Schwartz space (see [21, Definition 5.17]) and $\mathscr{S}'$ is its dual space with the dual bilinear form $\langle \cdot, \cdot \rangle_{\mathscr{S}}$. We can also verify that $C(\mathbb{R}^d) \cap \mathcal{SI} \subset L_1^{loc}(\mathbb{R}^d) \cap \mathcal{SI}$ is embedded into $\mathscr{S}'$.

When $p \geq q$, then $\hat{\Phi} \in L_1(\mathbb{R}^d) \cap C(\mathbb{R}^d)$ implies that $\hat{\Phi}^{p/q} \in L_1(\mathbb{R}^d)$ which will be used in the proof of the following theorem. We also need to impose an additional symmetry condition on $\hat{\Phi}^{q/p} \in L_1(\mathbb{R}^d)$ which is needed in the proof. Since $p/q = p - 1$ and $q/p = q - 1$, this condition can be represented as $\hat{\Phi}^{\min\{p,q\}-1} \in L_1(\mathbb{R}^d)$.

Since we can denote the positive measure μ on $\mathbb{R}^d$ as

$$\mu(A) := (2\pi)^{-d/2} \int_A \frac{\mathrm{d}\boldsymbol{x}}{\hat{\Phi}(\boldsymbol{x})}, \quad \text{for any open set } A \text{ of } \mathbb{R}^d.$$

17

[16, Example 1.2.6] provides that the space $L_q(\mathbb{R}^d; \mu)$ is well-defined on the positive measure space $(\mathbb{R}^d, \mathscr{B}_{\mathbb{R}^d}, \mu)$, i.e.,

$$L_q(\mathbb{R}^d; \mu) := \left\{ f : \mathbb{R}^d \to \mathbb{C} : f \text{ is measurable and } \int_{\mathbb{R}^d} |f(\boldsymbol{x})|^q \, d\mu(\boldsymbol{x}) < \infty \right\},$$

equipped with the norm

$$\|f\|_{L_q(\mathbb{R}^d; \mu)} := \left(\int_{\mathbb{R}^d} |f(\boldsymbol{x})|^q \, d\mu(\boldsymbol{x}) \right)^{1/q}.$$

$L_p(\mathbb{R}^d; \mu)$ is also defined in an analogous way. [16, Example 1.10.2 and Theorem 1.10.7] show that $L_q(\mathbb{R}^d; \mu)$ is a Banach space and its dual space $L_q(\mathbb{R}^d; \mu)'$ is isometrically equivalent to $L_p(\mathbb{R}^d; \mu)$. In analogy to the representation theorem on Hilbert space, the bounded linear functional $T_g \in L_q(\mathbb{R}^d; \mu)'$ associated with $g \in L_p(\mathbb{R}^d; \mu)$ is given by

$$T_g(f) := \int_{\mathbb{R}^d} f(\boldsymbol{x})\overline{g(\boldsymbol{x})} d\mu(\boldsymbol{x}), \quad \text{for all } f \in L_q(\mathbb{R}^d; \mu).$$

Here, this isometric isomorphism from $L_q(\mathbb{R}^d; \mu)'$ onto $L_p(\mathbb{R}^d; \mu)$ is antilinear, just as the dual of complex Hilbert spaces, i.e.,

$$T_{\lambda g}(f) = \int_{\mathbb{R}^d} f(\boldsymbol{x})\overline{\lambda g(\boldsymbol{x})} d\mu(\boldsymbol{x}) = \overline{\lambda} T_g(f), \quad \text{for all } f \in L_q(\mathbb{R}^d; \mu) \text{ and all } \lambda \in \mathbb{C}.$$

If we can show that $\mathcal{B}_\Phi^p(\mathbb{R}^d)$ and $L_q(\mathbb{R}^d; \mu)$ are isometrically isomorphic, then $\mathcal{B}_\Phi^p(\mathbb{R}^d)$ is a Banach space and its dual space $\mathcal{B}_\Phi^p(\mathbb{R}^d)'$ is isometrically equivalent to $L_p(\mathbb{R}^d; \mu)$. One can argue analogously for $\mathcal{B}_\Phi^q(\mathbb{R}^d) \equiv L_p(\mathbb{R}^d; \mu)$. If we can further verify the two-sided reproduction of $\mathcal{B}_\Phi^p(\mathbb{R}^d)$, then $\mathcal{B}_\Phi^p(\mathbb{R}^d)$ is a two-sided RKBS.

Theorem 5.1. *Let $1 < p, q < \infty$ and $p^{-1} + q^{-1} = 1$. Suppose that $\Phi \in L_1(\mathbb{R}^d) \cap C(\mathbb{R}^d)$ is a positive definite function on $\mathbb{R}^d$ and that $\hat{\Phi}^{\min\{p,q\}-1} \in L_1(\mathbb{R}^d)$. Then $\mathcal{B}_\Phi^p(\mathbb{R}^d)$ given in Equation (5.1) is a reproducing kernel Banach space with the two-sided reproducing kernel*

$$K(\boldsymbol{x}, \boldsymbol{y}) := \Phi(\boldsymbol{x} - \boldsymbol{y}), \quad \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^d.$$

Its dual space $\mathcal{B}_\Phi^p(\mathbb{R}^d)'$ and $\mathcal{B}_\Phi^q(\mathbb{R}^d)$ are isometrically isomorphic. Moreover, $\mathcal{B}_\Phi^p(\mathbb{R}^d)$ is uniformly convex and smooth.

In particular, when $p = 2$ then $\mathcal{B}_\Phi^2(\mathbb{R}^d) = \mathcal{H}_\Phi(\mathbb{R}^d)$ is a reproducing kernel Hilbert space as in Theorem 3.3.

18

Proof. For convenience, we assume that $p \geq q$. We first prove that $\mathcal{B}^p_\Phi(\mathbb{R}^d)$ and $L_q(\mathbb{R}^d; \mu)$ are isometrically isomorphic. The Fourier transform map can be seen as a one-to-one map from $\mathcal{B}^p_\Phi(\mathbb{R}^d)$ into $L_q(\mathbb{R}^d; \mu)$. We can check the equality of their norm

$$\|f\|_{\mathcal{B}^p_\Phi(\mathbb{R}^d)} = \left((2\pi)^{-d/2} \int_{\mathbb{R}^d} \frac{|\hat{f}(\boldsymbol{x})|^q}{\hat{\Phi}(\boldsymbol{x})} \mathrm{d}\boldsymbol{x}\right)^{1/q} = \left(\int_{\mathbb{R}^d} |\hat{f}(\boldsymbol{x})|^q \mathrm{d}\mu(\boldsymbol{x})\right)^{1/q} = \|\hat{f}\|_{L_q(\mathbb{R}^d;\mu)}.$$

So the Fourier transform map is an isometric isomorphism. Now we verify that the Fourier transform map is surjective. Fix any $h \in L_q(\mathbb{R}^d; \mu)$. We want to find an element in $\mathcal{B}^p_\Phi(\mathbb{R}^d)$ whose Fourier transform is equal to h. We conclude that $h \in L_1(\mathbb{R}^d)$ because

$$\int_{\mathbb{R}^d} |h(\boldsymbol{x})| \, \mathrm{d}\boldsymbol{x} \leq \left(\int_{\mathbb{R}^d} \frac{|h(\boldsymbol{x})|^q}{\hat{\Phi}(\boldsymbol{x})} \mathrm{d}\boldsymbol{x}\right)^{1/q} \left(\int_{\mathbb{R}^d} \hat{\Phi}(\boldsymbol{x})^{p/q} \mathrm{d}\boldsymbol{x}\right)^{1/p} < \infty.$$

Thus, the inverse Fourier transform of h given as $\check{h}(\boldsymbol{x}) = (2\pi)^{-d/2} \int_{\mathbb{R}^d} h(\boldsymbol{y}) e^{i\boldsymbol{x}^T\boldsymbol{y}} \mathrm{d}\boldsymbol{y}$ is well-defined and an element of $C(\mathbb{R}^d) \cap \mathcal{SI}$. This indicates that $\hat{\check{h}} = h$ and $\check{h} \in \mathcal{B}^p_\Phi(\mathbb{R}^d)$ because $\langle \hat{\check{h}}, \gamma \rangle_{\mathscr{S}} = \langle h, \hat{\check{\gamma}} \rangle_{\mathscr{S}} = \langle h, \gamma \rangle_{\mathscr{S}}$ for all $\gamma \in \mathscr{S}$. Therefore $\mathcal{B}^p_\Phi(\mathbb{R}^d)$ is isometrically equivalent to $L_q(\mathbb{R}^d; \mu)$.

Using $\hat{\Phi}^{q/p} = \hat{\Phi}^{q-1} \in L_1(\mathbb{R}^d)$ we can also prove that $\mathcal{B}^q_\Phi(\mathbb{R}^d) \equiv L_p(\mathbb{R}^d; \mu)$ in an analogous way. Therefore $\mathcal{B}^q_\Phi(\mathbb{R}^d)$ is isometrically equivalent to the dual space of $\mathcal{B}^p_\Phi(\mathbb{R}^d)$.

We fix any $\boldsymbol{y} \in \mathbb{R}^d$. The Fourier transform of $K(\cdot, \boldsymbol{y})$ is equal to $\hat{k}_{\boldsymbol{y}}(\boldsymbol{x}) := \hat{\Phi}(\boldsymbol{x}) e^{-i\boldsymbol{x}^T\boldsymbol{y}}$. Since $\hat{\Phi}^{p-1} \in L_1(\mathbb{R}^d)$ we have $\hat{k}_{\boldsymbol{y}} \in L_p(\mathbb{R}^d; \mu)$. Thus $K(\cdot, \boldsymbol{y})$ can be seen as an element of $\mathcal{B}^q_\Phi(\mathbb{R}^d) \equiv \mathcal{B}^p_\Phi(\mathbb{R}^d)'$. In addition, $\overline{K(\boldsymbol{x}, \cdot)} \in \mathcal{B}^p_\Phi(\mathbb{R}^d)$ for any $\boldsymbol{x} \in \mathbb{R}^d$ because $\hat{\Phi}^{q-1} \in L_1(\mathbb{R}^d)$ and $\left(\overline{K(\boldsymbol{x}, \cdot)}\right)\hat{} = \hat{k}_{\boldsymbol{x}} \in L_q(\mathbb{R}^d; \mu)$ by $\Phi = \overline{\Phi(-\cdot)}$.

Finally, we verify the right-sided reproduction. Fix any $f \in \mathcal{B}^p_\Phi(\mathbb{R}^d)$ and $\boldsymbol{y} \in \mathbb{R}^d$. We can verify that $\hat{f} \in L_1(\mathbb{R}^d)$ as in the above proof. Moreover, the continuity of f and $\check{\hat{f}}$ allows us to recover f pointwise from its Fourier transform via

$$f(\boldsymbol{x}) = \check{\hat{f}}(\boldsymbol{x}) = (2\pi)^{-d/2} \int_{\mathbb{R}^d} \hat{f}(\boldsymbol{y}) e^{i\boldsymbol{x}^T\boldsymbol{y}} \mathrm{d}\boldsymbol{y}.$$

Thus, we have

$$\langle f, K(\cdot, \boldsymbol{y}) \rangle_{\mathcal{B}^p_\Phi(\mathbb{R}^d)} = \langle \hat{f}, \hat{k}_{\boldsymbol{y}} \rangle_{L_q(\mathbb{R}^d;\mu)} = \int_{\mathbb{R}^d} \hat{f}(\boldsymbol{x}) \overline{\hat{k}_{\boldsymbol{y}}(\boldsymbol{x})} \mathrm{d}\mu(\boldsymbol{x})$$

$$= (2\pi)^{-d/2} \int_{\mathbb{R}^d} \frac{\hat{f}(\boldsymbol{x}) \overline{\hat{\Phi}(\boldsymbol{x}) e^{-i\boldsymbol{x}^T\boldsymbol{y}}}}{\hat{\Phi}(\boldsymbol{x})} \mathrm{d}\boldsymbol{x} = (2\pi)^{-d/2} \int_{\mathbb{R}^d} \hat{f}(\boldsymbol{x}) e^{i\boldsymbol{x}^T\boldsymbol{y}} \mathrm{d}\boldsymbol{x} = f(\boldsymbol{y}).$$

In the same way, we can also verify that $\mathcal{B}_\Phi^p(\mathbb{R}^d)$ satisfies the left-sided reproduction property, i.e.,

$$\langle \overline{K(\boldsymbol{x}, \cdot)}, g \rangle_{\mathcal{B}_\Phi^p(\mathbb{R}^d)} = \langle \hat{k}_{\boldsymbol{x}}, \hat{g} \rangle_{\mathrm{L}_q(\mathbb{R}^d;\mu)} = \overline{\int_{\mathbb{R}^d} \hat{g}(\boldsymbol{y}) \overline{\hat{k}_{\boldsymbol{x}}(\boldsymbol{y})} \mathrm{d}\mu(\boldsymbol{y})} = \overline{g(\boldsymbol{x})},$$

for all $g \in \mathcal{B}_\Phi^q(\mathbb{R}^d) \equiv \mathcal{B}_\Phi^p(\mathbb{R}^d)'$ and all $\boldsymbol{x} \in \mathbb{R}^d$. Therefore $\mathcal{B}_\Phi^p(\mathbb{R}^d)$ is an RKBS with the two-sided reproducing kernel K.

Since $\mathcal{B}_\Phi^p(\mathbb{R}^d) \equiv \mathrm{L}_q(\mathbb{R}^d; \mu)$ is reflexive and K is even, the dual space $\mathcal{B}_\Phi^p(\mathbb{R}^d)' \equiv \mathcal{B}_\Phi^q(\mathbb{R}^d)$ is also an RKBS with the two-sided reproducing kernel K.

Because $\mathrm{L}_q(\mathbb{R}^d; \mu)$ and $\mathrm{L}_p(\mathbb{R}^d; \mu)$ are uniformly convex and smooth by [16, Theorem 5.2.11 and Example 5.4.8]. $\mathcal{B}_\Phi^p(\mathbb{R}^d)$ and $\mathcal{B}_\Phi^q(\mathbb{R}^d)$ are also uniformly convex and smooth. □

Remark 5.2. We can combine our result with [16, Proposition 1.9.3] to conclude that the restriction of $\mathcal{B}_\Phi^p(\mathbb{R}^d)$ to the reals is also an RKBS with the two-sided reproducing kernel K and its dual is isometrically equivalent to the restriction of $\mathcal{B}_\Phi^q(\mathbb{R}^d)$ to the reals. It is well-known that the RKHS of a given reproducing kernel is unique. Theorem 5.1, however, shows that different RKBSs may have the same reproducing kernel. We will provide an example for this in Section 6. Moreover, the proof of Theorem 5.1 provides that $\mathcal{B}_\Phi^p(\mathbb{R}^d)$ with $p \geq 2$ is still a right-sided RKBS without the additional condition $\hat{\Phi}^{q-1} \in \mathrm{L}_1(\mathbb{R}^d)$.

According to [21, Theorem 10.10] any positive definite kernel can be used to construct an RKHS. We may extend the positive definite kernel into an RKBS.

Corollary 5.2. *Let* $\mathcal{B}_\Phi^p(\mathbb{R}^d)$ *with* $p \geq 2$ *be defined in Theorem 5.1. Then* $\mathcal{B}_\Phi^p(\mathbb{R}^d) \subseteq \mathrm{L}_p(\mathbb{R}^d)$.

Proof. We fix any $f \in \mathcal{B}_\Phi^p(\mathbb{R}^d)$. According to the proof of Theorem 5.1, we have $\hat{f} \in \mathrm{L}_q(\mathbb{R}^d)$ because

$$\int_{\mathbb{R}^d} \left| \hat{f}(\boldsymbol{x}) \right|^q \mathrm{d}\boldsymbol{x} \leq (2\pi)^{qd/2} \left(\int_{\mathbb{R}^d} \frac{\left| \hat{f}(\boldsymbol{x}) \right|^q}{\hat{\Phi}(\boldsymbol{x})} \mathrm{d}\boldsymbol{x} \right) \left(\sup_{\boldsymbol{x} \in \mathbb{R}^d} \hat{\Phi}(\boldsymbol{x}) \right) < \infty.$$

The Hausdorff-Young inequality [12, Theorem 7.1.13] provides that $f = \check{\hat{f}} \in \mathrm{L}_p(\mathbb{R}^d)$ because $1 < q \leq 2$. □

Remark 5.3. The RKBS $\mathcal{B}_\Phi^p(\mathbb{R}^d)$ with $p \geq 2$ can be precisely written as

$$\mathcal{B}_\Phi^p(\mathbb{R}^d) := \Big\{ f \in \mathrm{L}_p(\mathbb{R}^d) \cap \mathrm{C}(\mathbb{R}^d) : \text{ the distributional Fourier transform } \hat{f} \text{ of } f$$
$$\text{is a measurable function defined on } \mathbb{R}^d \text{ such that } \hat{f}/\hat{\Phi}^{1/q} \in \mathrm{L}_q(\mathbb{R}^d) \Big\}.$$

20

However, $\mathcal{B}_\Phi^p(\mathbb{R}^d) \not\subseteq L_p(\mathbb{R}^d)$ with $1 < p < 2$ because the Hausdorff-Young inequality does not work for $q > 2$.

We fix any positive number $m > d/2$. According to [21, Corollary 10.13], if there are two positive constants C_1, C_2 such that

$$C_1 \left(1 + \|x\|_2^2\right)^{-m/2} \le \hat{\Phi}(x)^{1/2} \le C_2 \left(1 + \|x\|_2^2\right)^{-m/2}, \quad x \in \mathbb{R}^d,$$

then the RKHS $\mathcal{B}_\Phi^2(\mathbb{R}^d) \equiv \mathcal{H}_\Phi(\mathbb{R}^d)$ and the classical L_2-based Sobolev space $W_2^m(\mathbb{R}^d) \equiv \mathcal{H}^m(\mathbb{R}^d)$ of order m are isomorphic, i.e., $\mathcal{H}_\Phi(\mathbb{R}^d) \cong \mathcal{H}^m(\mathbb{R}^d)$.

Following the ideas of RKHSs, we can also find a relationship between RKBSs and Sobolev spaces. Let $f_m(x) := \left(1 + \|x\|_2^2\right)^{m/2} \hat{f}(x)$ with $p \ge 2$. The theory of singular integrals then shows that f belongs to the classical L_p-based Sobolev space $W_p^m(\mathbb{R}^d)$ of order m if any only if the function f_m is the Fourier transform of some function in $L_p(\mathbb{R}^d)$, and the L_p-norm of the inverse Fourier transform f_m is equivalent to the W_p^m-norm of f (much more detail is mentioned in [1, Section 7.63] and [12, Section 7.9]). Using the Hausdorff-Young inequality, we can get $\|f\|_{W_p^m(\mathbb{R}^d)} \le C \left\|\check{f}_m\right\|_{L_p(\mathbb{R}^d)} \le C \|f_m\|_{L_q(\mathbb{R}^d)}$ for some positive constant C independent of f. Following these statements, we can introduce the following corollary.

Corollary 5.3. *Let the positive definite function Φ be as in Theorem 5.1 and $W_p^m(\mathbb{R}^d)$ be the classical L_p-based Sobolev space of order $m > pd/q - d/q$. Here q is the conjugate exponent of $p \ge 2$. If there are two positive constants C_1, C_2 such that*
$$C_1 \left(1 + \|x\|_2^2\right)^{-m/2} \le \hat{\Phi}(x)^{1/q} \le C_2 \left(1 + \|x\|_2^2\right)^{-m/2}, \quad x \in \mathbb{R}^d,$$

then $\mathcal{B}_\Phi^p(\mathbb{R}^d)$ is embedded into $W_p^m(\mathbb{R}^d)$, i.e.,

$$\|f\|_{W_p^m(\mathbb{R}^d)} \le C \|f\|_{\mathcal{B}_\Phi^p(\mathbb{R}^d)}, \quad f \in \mathcal{B}_\Phi^p(\mathbb{R}^d) \subseteq W_p^m(\mathbb{R}^d),$$

for some positive constant C independent on f.

Remark 5.4. Here the lower bound for m is induced by the condition that $\hat{\Phi}^{q/p} \in L_1(\mathbb{R}^d)$. According to Corollary 5.3, the dual space $W_q^{-m}(\mathbb{R}^d)$ of the Sobolev space $W_p^m(\mathbb{R}^d)$ is embedded into the dual space $\mathcal{B}_\Phi^p(\mathbb{R}^d)'$ of the RKBS $\mathcal{B}_\Phi^p(\mathbb{R}^d)$. It is well-known that the point evaluation functional δ_x belongs to $W_q^{-m}(\mathbb{R}^d)$ (see [1, Section 3.25]) which coincides with $\delta_x \in \mathcal{B}_\Phi^p(\mathbb{R}^d)'$.

Since $K(\cdot, x_1), \ldots, K(\cdot, x_N)$ are linearly independent in $\mathcal{B}_\Phi^q(\mathbb{R}^d) \equiv \mathcal{B}_\Phi^p(\mathbb{R}^d)'$ for any pairwise distinct data points $X = \{x_1, \ldots, x_N\} \subseteq \mathbb{R}^d$, $\delta_{x_1}, \ldots, \delta_{x_N}$ are linearly independent on $\mathcal{B}_\Phi^p(\mathbb{R}^d)$. Combining Theorems 4.2 and 5.1, we can solve the empirical SVM solution in $\mathcal{B}_\Phi^p(\mathbb{R}^d)$ with $p > 1$.

Theorem 5.4. *Let $\mathcal{B}_\Phi^p(\mathbb{R}^d)$ with $p > 1$ be defined as in Theorem 5.1 and the regularization function $R : [0, \infty) \to [0, \infty)$ be convex and strictly increasing. We choose the loss function $L : \mathbb{R}^d \times \mathbb{C} \times \mathbb{C} \to [0, \infty)$ such that $L(\boldsymbol{x}, y, \cdot)$ is a convex map for any fixed $\boldsymbol{x} \in \mathbb{R}^d$ and any fixed $y \in \mathbb{C}$. Given the data $D := \{(\boldsymbol{x}_1, y_1), \ldots, (\boldsymbol{x}_N, y_N)\}$ with pairwise distinct data points $X = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N\} \subseteq \mathbb{R}^d$ and associated data values $Y = \{y_1, \ldots, y_N\} \subset \mathbb{C}$, the unique optimal solution (support vector machine solution) $s_{D,L,R}$ of*

$$\min_{f \in \mathcal{B}_\Phi^p(\mathbb{R}^d)} \sum_{j=1}^N L\left(\boldsymbol{x}_j, y_j, f(\boldsymbol{x}_j)\right) + R\left(\|f\|_{\mathcal{B}_\Phi^p(\mathbb{R}^d)}\right), \tag{5.2}$$

has the explicit representation

$$s_{D,L,R}(\boldsymbol{x}) = (2\pi)^{-d/2} \int_{\mathbb{R}^d} \hat{\Phi}(\boldsymbol{y})^{p-1} \sum_{k=1}^N c_k e^{i(\boldsymbol{x}-\boldsymbol{x}_k)^T \boldsymbol{y}} \left| \sum_{l=1}^N c_l e^{-i\boldsymbol{x}_l^T \boldsymbol{y}} \right|^{p-2} \mathrm{d}\boldsymbol{y}, \quad \boldsymbol{x} \in \mathbb{R}^d, \tag{5.3}$$

for some coefficients $c_1, \ldots, c_N \in \mathbb{C}$ and $i^2 = -1$.

Proof. Using Theorems 4.2 and 5.1, the dual element of the SVM solution $s_{D,L,R}$ of the SVM (5.2) is a linear combination of $K(\cdot, \boldsymbol{x}_1), \ldots, K(\cdot, \boldsymbol{x}_N)$, i.e.,

$$s_{D,L,R}^*(\boldsymbol{x}) = \sum_{k=1}^N b_k K(\boldsymbol{x}, \boldsymbol{x}_k) = \sum_{k=1}^N b_k \Phi(\boldsymbol{x} - \boldsymbol{x}_k), \quad \boldsymbol{x} \in \mathbb{R}^d, \ \boldsymbol{b} := (b_1, \cdots, b_N)^T \in \mathbb{C}^N.$$

Suppose that $s_{D,L,R}$ is not trivial. According to the proof of Theorem 5.1, the identity element of $s_{D,L,R}^* \in \mathcal{B}_\Phi^q(\mathbb{R}^d)$ in $\mathrm{L}_p(\mathbb{R}^d; \mu)$ is the Fourier transform of $s_{D,L,R}^*$, i.e.,

$$f_s(\boldsymbol{x}) := \mathcal{F}\left(s_{D,L,R}^*\right)(\boldsymbol{x}) = \sum_{k=1}^N b_k \hat{\Phi}(\boldsymbol{x}) e^{-i\boldsymbol{x}^T \boldsymbol{x}_k}, \quad \boldsymbol{x} \in \mathbb{R}^d.$$

The dual element of $f_s \in \mathrm{L}_p(\mathbb{R}^d; \mu)$ in $\mathrm{L}_q(\mathbb{R}^d; \mu)$ has the form

$$f_s^*(\boldsymbol{x}) = \frac{f_s(\boldsymbol{x}) |f_s(\boldsymbol{x})|^{p-2}}{\|f_s\|_{\mathrm{L}_p(\mathbb{R}^d; \mu)}^{p-2}}, \quad \boldsymbol{x} \in \mathbb{R}^d.$$

Because the dual element of $s_{D,L,R}^*$ in $\mathcal{B}_\Phi^p(\mathbb{R}^d)$ is equal to the identity element of $f_s^* \in \mathrm{L}_q(\mathbb{R}^d; \mu)$ in $\mathcal{B}_\Phi^p(\mathbb{R}^d)$, which is the inverse Fourier transfer of f_s^*, we can determine that

$$s_{D,L,R}(\boldsymbol{x}) = \mathcal{F}^{-1}\left(f_s^*\right)(\boldsymbol{x}) = (2\pi)^{-d/2} \int_{\mathbb{R}^d} \hat{\Phi}(\boldsymbol{y})^{p-1} \sum_{k=1}^N c_k e^{i(\boldsymbol{x}-\boldsymbol{x}_k)^T \boldsymbol{y}} \left| \sum_{l=1}^N c_l e^{-i\boldsymbol{x}_l^T \boldsymbol{y}} \right|^{p-2} \mathrm{d}\boldsymbol{y},$$

22

and the coefficients are given by $c_k := \|f_s\|_{L_p(\mathbb{R}^d;\mu)}^{\frac{2-p}{p-1}} b_k = \|s_{D,L,R}\|_{\mathcal{B}_\Phi^p(\mathbb{R}^d)}^{q-2} b_k$ for all $k = 1, \ldots, N$, where q is the conjugate exponent of p.

$\square$

Remark 5.5. In particular, if p is an even positive integer, then $s_{D,L,R}$ is also a linear combination of some kernel function translated to the data points X. For example, when $p = 4$, then

$$s_{D,L,R} = \sum_{k_1,k_2,k_3=1}^{N,N,N} c_{k_1}\overline{c_{k_2}}c_{k_3}\Phi_3\left(\cdot - \boldsymbol{x}_{k_1} + \boldsymbol{x}_{k_2} - \boldsymbol{x}_{k_3}\right) = \sum_{k_1,k_2,k_3=1}^{N,N,N} c_{k_1}\overline{c_{k_2}}c_{k_3}\mathcal{K}_3\left(\cdot, \boldsymbol{x}_{k_1}, \boldsymbol{x}_{k_2}, \boldsymbol{x}_{k_3}\right),$$

where the kernel function $\mathcal{K}_3(\boldsymbol{x}, \boldsymbol{y}_1, \boldsymbol{y}_2, \boldsymbol{y}_3) := \Phi_3(\boldsymbol{x} - \boldsymbol{y}_1 + \boldsymbol{y}_2 - \boldsymbol{y}_3)$ and Φ_3 is the inverse Fourier transform of $\hat{\Phi}^3$. Moreover,

$$\|s_{D,L,R}\|_{\mathcal{B}_\Phi^p(\mathbb{R}^d)}^{4/3} = \|s_{D,L,R}\|_{\mathcal{B}_\Phi^p(\mathbb{R}^d)}^{-2/3} [s_{D,L,R}, s_{D,L,R}]_{\mathcal{B}_\Phi^p(\mathbb{R}^d)} = \|s_{D,L,R}\|_{\mathcal{B}_\Phi^p(\mathbb{R}^d)}^{-2/3} \langle s_{D,L,R}, s_{D,L,R}^*\rangle_{\mathcal{B}_\Phi^p(\mathbb{R}^d)}$$

$$= \sum_{j=1}^{N} \overline{c_j}\langle s_{D,L,R}, K(\cdot, \boldsymbol{x}_j)\rangle_{\mathcal{B}_\Phi^p(\mathbb{R}^d)} = \sum_{j,k_1,k_2,k_3=1}^{N,N,N,N} \overline{c_j}c_{k_1}\overline{c_{k_2}}c_{k_3}\mathcal{K}_3\left(\boldsymbol{x}_j, \boldsymbol{x}_{k_1}, \boldsymbol{x}_{k_2}, \boldsymbol{x}_{k_3}\right).$$

We can observe that the coefficients of the SVM solution $s_{D,L,R}$ given in Theorem 5.4 differ from the coefficients of its dual element $s_{D,L,R}^*$ only by a constant factor. As in Corollary 4.3, the coefficients of $s_{D,L,R}$ can also be computed by the fixed point iteration method. For any fixed $\boldsymbol{c} := (c_1, \cdots, c_N)^T \in \mathbb{C}^N$, we can define a unique function $s_{\boldsymbol{c}} \in \mathcal{B}_\Phi^p(\mathbb{R}^d)$ as in Equation (5.3). Let

$$\phi_j(\boldsymbol{c}) := s_{\boldsymbol{c}}(\boldsymbol{x}_j) = (2\pi)^{-d/2} \int_{\mathbb{R}^d} \hat{\Phi}(\boldsymbol{y})^{p-1} \sum_{k=1}^{N} c_k e^{i(\boldsymbol{x}_j - \boldsymbol{x}_k)^T \boldsymbol{y}} \left|\sum_{l=1}^{N} c_l e^{-i\boldsymbol{x}_l^T \boldsymbol{y}}\right|^{p-2} \mathrm{d}\boldsymbol{y}, \quad \boldsymbol{c} \in \mathbb{C}^N,$$

for all $j = 1, \ldots, N$, and $\boldsymbol{\phi} := (\phi_1, \cdots, \phi_N)^T$. Thus we have

$$\|s_{\boldsymbol{c}}\|_{\mathcal{B}_\Phi^p(\mathbb{R}^d)}^q = \|s_{\boldsymbol{c}}\|_{\mathcal{B}_\Phi^p(\mathbb{R}^d)}^{q-2} \langle s_{\boldsymbol{c}}, s_{\boldsymbol{c}}^*\rangle_{\mathcal{B}_\Phi^p(\mathbb{R}^d)} = \sum_{j=1}^{N} \overline{c_j}\langle s_{\boldsymbol{c}}, K(\cdot, \boldsymbol{x}_j)\rangle_{\mathcal{B}_\Phi^p(\mathbb{R}^d)} = \boldsymbol{c}^*\boldsymbol{\phi}(\boldsymbol{c}).$$

Here q is the conjugate exponent of p. Denote that

$$\mathcal{T}_{D,L,R}(\boldsymbol{c}) := \sum_{j=1}^{N} L\left(\boldsymbol{x}_j, y_j, \phi_j(\boldsymbol{c})\right) + R\left((\boldsymbol{c}^*\boldsymbol{\phi}(\boldsymbol{c}))^{1/q}\right) = \sum_{j=1}^{N} L\left(\boldsymbol{x}_j, y_j, s_{\boldsymbol{c}}(\boldsymbol{x}_j)\right) + R\left(\|s_{\boldsymbol{c}}\|_{\mathcal{B}_\Phi^p(\mathbb{R}^d)}\right).$$

23

It is easy to check that the coefficients of $s_{D,L,R}$ are the minimizers of $\mathcal{T}_{D,L,R}$ over $\mathbb{C}^N$, i.e.,

$$\boldsymbol{c}_{opt} := \operatorname*{argmin}_{\boldsymbol{c} \in \mathbb{C}^N} \mathcal{T}_{D,L,R}(\boldsymbol{c}) \text{ such that } s_{D,L,R} = s_{\boldsymbol{c}_{opt}}.$$

Suppose that $L(\boldsymbol{x}, y, \cdot) \in \mathrm{C}^1(\mathbb{C})$ for all $\boldsymbol{x} \in \mathbb{R}^d$ and all $y \in \mathbb{C}$, $R \in \mathrm{C}^1([0, \infty))$ and $p \geq 2$. We can compute the gradient of $\mathcal{T}_{D,L,R}$ by Wirtinger partial derivatives in the form

$$\nabla \mathcal{T}_{D,L,R}(\boldsymbol{c})^T = \boldsymbol{l}'_D(\boldsymbol{\phi}(\boldsymbol{c}))^T \nabla \boldsymbol{\phi}(\boldsymbol{c}) + \frac{R'\left((\boldsymbol{c}^*\boldsymbol{\phi}(\boldsymbol{c}))^{1/q}\right)}{2q\left(\boldsymbol{c}^*\boldsymbol{\phi}(\boldsymbol{c})\right)^{1/p}} \boldsymbol{c}^* \nabla \boldsymbol{\phi}(\boldsymbol{c}),$$

where $\boldsymbol{l}'_D(\boldsymbol{\phi}) := (L'(\boldsymbol{x}_1, y_1, \phi_1), \cdots, L'(\boldsymbol{x}_N, y_N, \phi_N))^T$ and the entries of the Jacobian (gradient) matrix $\nabla \boldsymbol{\phi} := \left(\frac{\partial}{\partial c_k} \phi_j\right)_{j,k=1}^{N,N}$ by Wirtinger partial derivatives have the forms

$$\frac{\partial}{\partial c_k} \phi_j(\boldsymbol{c}) = \frac{p}{2}(2\pi)^{-d/2} \int_{\mathbb{R}^d} \hat{\Phi}(\boldsymbol{y})^{p-1} e^{i(\boldsymbol{x}_j - \boldsymbol{x}_k)^T \boldsymbol{y}} \left| \sum_{l=1}^N c_l e^{-i\boldsymbol{x}_l^T \boldsymbol{y}} \right|^{p-2} d\boldsymbol{y}.$$

Moreover, $\boldsymbol{c}_{opt}$ is the stationary point of $\nabla \mathcal{T}_{D,L,R}$ which indicates that $\boldsymbol{c}_{opt}$ is a fixed point of the function

$$F_{D,L,R}(\boldsymbol{c}) := \boldsymbol{c} + \nabla \mathcal{T}_{D,L,R}(\boldsymbol{c}), \quad \boldsymbol{c} \in \mathbb{C}^N. \tag{5.4}$$

Therefore, we can introduce the following corollary.

Corollary 5.5. *Suppose that the loss function $L(\boldsymbol{x}, y, \cdot) \in \mathrm{C}^1(\mathbb{C})$ for all $\boldsymbol{x} \in \mathbb{R}^d$ and all $y \in \mathbb{C}$, the regularization function $R \in \mathrm{C}^1([0, \infty))$ and $p \geq 2$. Then the coefficient vector $\boldsymbol{c}$ of the support vector machine solution $s_{D,L,R}$ given in Theorem 5.4 is a fixed point of the function $F_{D,L,R}$ defined in Equation (5.4), i.e., $F_{D,L,R}(\boldsymbol{c}) = \boldsymbol{c}$.*

Remark 5.6. The coefficients $\boldsymbol{c} := (c_1, \cdots, c_N)^T$ of the SVM solution $s_{D,L,R}$ in $\mathcal{B}^p_\Phi(\mathbb{R}^d)$ differ from the coefficients $\boldsymbol{b} := (b_1, \cdots, b_N)^T$ of its dual element $s^*_{D,L,R}$ in $\mathcal{B}^q_\Phi(\mathbb{R}^d)$ only by a constant factor. Both coefficient vectors $\boldsymbol{b}$ and $\boldsymbol{c}$ are fixed points of the functions $F^*_{D,L,R}$ as in Equation (4.3) and $F_{D,L,R}$ as in Equation (5.4), respectively. Roughly speaking, $F^*_{D,L,R}$ can be seen as a conjugate of $F_{D,L,R}$. Much more contents of these fixed point iteration algorithms for the binary classification problems will be deeply discussed in our next papers.

We now use the techniques of [3, Theorem 6] to set up a two-sided RKBS defined on a subset Ω of $\mathbb{R}^d$.

Theorem 5.6. *Let the positive definite function Φ be as in Theorem 5.1 and $\Omega \subseteq \mathbb{R}^d$. Then the function space*

$$\mathcal{B}^p_\Phi(\Omega) := \left\{ h : \text{ there exists a function } h \in \mathcal{B}^p_\Phi(\mathbb{R}^d) \text{ such that } f|_\Omega = h \right\},$$

equipped with the norm

$$\|h\|_{\mathcal{B}^p_\Phi(\Omega)} := \inf_{f \in \mathcal{B}^p_\Phi(\mathbb{R}^d)} \|f\|_{\mathcal{B}^p_\Phi(\mathbb{R}^d)} \ \ s.t. \ f|_\Omega = h,$$

is a reproducing kernel Banach space with the two-sided reproducing kernel

$$K|_{\mathbb{R}^d \times \Omega}(\boldsymbol{x}, \boldsymbol{y}) := \Phi(\boldsymbol{x} - \boldsymbol{y}), \quad \boldsymbol{x} \in \mathbb{R}^d, \ \boldsymbol{y} \in \Omega,$$

where $f|_\Omega$ stands for the restriction of f to Ω. Its dual space $\mathcal{B}^p_\Phi(\Omega)'$ is isometrically equivalent to a closed subspace of $\mathcal{B}^q_\Phi(\mathbb{R}^d)$ (the annihilator of $\mathcal{N}_0$ in $\mathcal{B}^q_\Phi(\mathbb{R}^d)$)

$$\mathcal{N}_0^\perp = \left\{ g \in \mathcal{B}^q_\Phi(\mathbb{R}^d) \equiv \mathcal{B}^p_\Phi(\mathbb{R}^d)' : \ \langle f, g \rangle_{\mathcal{B}^p_\Phi(\mathbb{R}^d)} = 0, \text{ for all } f \in \mathcal{N}_0 \right\},$$

where q is the conjugate exponent of $p > 1$ and

$$\mathcal{N}_0 := \left\{ f \in \mathcal{B}^p_\Phi(\mathbb{R}^d) : \ f|_\Omega = 0 \right\}.$$

Moreover, $\mathcal{B}^p_\Phi(\Omega)$ is uniformly convex and smooth.

Proof. Since convergence in a two-sided RKBS $\mathcal{B}^p_\Phi(\mathbb{R}^d)$ implies pointwise convergence, we can determine that $\mathcal{N}_0$ is a closed subspace of $\mathcal{B}^p_\Phi(\mathbb{R}^d)$. According to the construction of $\mathcal{B}^p_\Phi(\Omega)$, $\mathcal{B}^p_\Phi(\Omega)$ is isometrically equivalent to the quotient space $\mathcal{B}^p_\Phi(\mathbb{R}^d)/\mathcal{N}_0$ (see [16, Definition 1.7.1 and 1.7.3]). Thus $\mathcal{B}^p_\Phi(\Omega)$ is a Banach space by [16, Theorem 1.7.9 and Corollary 1.11.19].

Next we use the identification of $\left(\mathcal{B}^p_\Phi(\mathbb{R}^d)/\mathcal{N}_0\right)' \equiv \mathcal{N}_0^\perp$ to verify the two-sided reproduction (see [16, Theorem 1.10.17]). Let K be the reproducing kernel of $\mathcal{B}^p_\Phi(\mathbb{R}^d)$ given in Theorem 5.1. We fix any $\boldsymbol{y} \in \Omega$. Since

$$\langle f, K(\cdot, \boldsymbol{y}) \rangle_{\mathcal{B}^p_\Phi(\mathbb{R}^d)} = f(\boldsymbol{y}) = 0, \quad \text{for all } f \in \mathcal{N}_0,$$

we have $K(\cdot, \boldsymbol{y}) \in \mathcal{N}_0^\perp \equiv \left(\mathcal{B}^p_\Phi(\mathbb{R}^d)/\mathcal{N}_0\right)' \equiv \mathcal{B}^p_\Phi(\Omega)'$. Combining this with the right-sided reproduction of $\mathcal{B}^p_\Phi(\mathbb{R}^d)$, we have

$$\langle h, K(\cdot, \boldsymbol{y}) \rangle_{\mathcal{B}^p_\Phi(\Omega)} = \langle Eh, K(\cdot, \boldsymbol{y}) \rangle_{\mathcal{B}^p_\Phi(\mathbb{R}^d)} = (Eh)(\boldsymbol{y}) = h(\boldsymbol{y}),$$

for all $h \in \mathcal{B}_\Phi^p(\Omega)$ and all $\boldsymbol{y} \in \Omega$, where E is the extension operator from $\mathcal{B}_\Phi^p(\Omega)$ into $\mathcal{B}_\Phi^p(\mathbb{R}^d)$ such that $Eh|_\Omega = h$ and $\|Eh\|_{\mathcal{B}_\Phi^p(\mathbb{R}^d)} = \|h\|_{\mathcal{B}_\Phi^p(\Omega)}$. Since $\overline{K(\boldsymbol{x}, \cdot)}|_\Omega \in \mathcal{B}_\Phi^p(\Omega)$ for all $\boldsymbol{x} \in \mathbb{R}^d$, we can also obtain the left-sided reproduction of $\mathcal{B}_\Phi^p(\Omega)$, i.e.,

$$\langle \overline{K(\boldsymbol{x}, \cdot)}|_\Omega, g \rangle_{\mathcal{B}_\Phi^p(\Omega)} = \langle \overline{K(\boldsymbol{x}, \cdot)}, g \rangle_{\mathcal{B}_\Phi^p(\mathbb{R}^d)} = g(\boldsymbol{x}),$$

for all $g \in \mathcal{N}_0^\perp \equiv \mathcal{B}_\Phi^p(\Omega)'$. Therefore $\mathcal{B}_\Phi^p(\Omega)$ is an RKBS with the two-sided reproducing kernel $K|_{\mathbb{R}^d \times \Omega}$.

Since $\mathcal{B}_\Phi^p(\mathbb{R}^d)$ is uniformly convex, [16, Theorem 5.2.24] provides that $\mathcal{B}_\Phi^p(\Omega) \equiv \mathcal{B}_\Phi^p(\mathbb{R}^d)/\mathcal{N}_0$ is uniformly convex. We also know that $\mathcal{B}_\Phi^p(\mathbb{R}^d)' \equiv L_q(\mathbb{R}^d; \mu)$ is uniformly convex and $\mathcal{N}_0^\perp$ is a closed subspace of $\mathcal{B}_\Phi^q(\mathbb{R}^d) \equiv \mathcal{B}_\Phi^p(\mathbb{R}^d)'$ by [16, Proposition 1.10.15]. Combining with [16, Proposition 5.1.20 and 5.4.5], we can also check that $\mathcal{B}_\Phi^p(\Omega)$ is smooth. $\qquad \square$

Remark 5.7. When $p = 2$, then we know that $\mathcal{B}_\Phi^2(\Omega)$ is a Hilbert space by Theorem 5.1. Thus the dual space and the space itself are isometrically isomorphic such that the reproducing kernel becomes $K|_{\Omega \times \Omega}$. Since $\mathcal{B}_\Phi^2(\mathbb{R}^d) = \mathcal{N}_0 \oplus \mathcal{N}_0^\perp$, we can determine that $\{g|_\Omega : g \in \mathcal{N}_0^\perp\} = \mathcal{B}_\Phi^2(\Omega)$ and $\|g\|_{\mathcal{B}_\Phi^2(\mathbb{R}^d)} = \|g|_\Omega\|_{\mathcal{B}_\Phi^2(\Omega)}$ for all $g \in \mathcal{N}_0^\perp$ which implies that $\mathcal{B}_\Phi^2(\Omega) \equiv \mathcal{N}_0^\perp \equiv \mathcal{B}_\Phi^2(\Omega)'$ and $\mathcal{B}_\Phi^2(\Omega)$ has the inner product

$$(h_1, h_2)_{\mathcal{B}_\Phi^2(\Omega)} = \langle h_1, h_2 \rangle_{\mathcal{B}_\Phi^2(\Omega)} = \langle Eh_1, Eh_2 \rangle_{\mathcal{B}_\Phi^2(\mathbb{R}^d)} = (Eh_1, Eh_2)_{\mathcal{B}_\Phi^2(\mathbb{R}^d)},$$

for all $h_1, h_2 \in \mathcal{B}_\Phi^2(\Omega)$. Therefore $\mathcal{B}_\Phi^2(\Omega)$ is an RKHS. Moreover, since $K(\cdot, \boldsymbol{y}) \in \mathcal{N}_0^\perp$ for any $\boldsymbol{y} \in \Omega$, we have $E(K(\cdot, \boldsymbol{y})|_\Omega) = K(\cdot, \boldsymbol{y})$. This shows that $K|_{\Omega \times \Omega}$ is a reproducing kernel of $\mathcal{B}_\Phi^2(\Omega)$. This conclusion is the same as in [3, Theorem 6].

If the RKBS is even a Hilbert space, then we can choose an equivalent function space of its dual as itself such that its reproducing kernel has symmetric domains. The difficulty to find an equivalent function space of the dual of RKBS, which is defined on the same domain of the RKBS, causes the domains of its reproducing kernel to be nonsymmetric. Theorems 5.1 and 5.6 provide us with examples of symmetric and nonsymmetric reproducing kernels of RKBSs, respectively.

Suppose that the positive definite function Φ given in Theorem 5.1 has a compact support Ω_Φ. Because of the positive definite properties of Φ, its support $\text{supp}(\Phi) = \Omega_\Phi$ with the origin is symmetric and bounded. Let Ω_1 and Ω_2 be two subsets of $\mathbb{R}^d$ such that the complement Ω_1^c includes $\Omega_2^c + \Omega_\Phi$. We fix any $\gamma \in \mathscr{S}$ so that its support $\text{supp}(\gamma) \subseteq \Omega_2^c$. Since the convolution function $\gamma * \Phi \in \mathcal{B}_\Phi^p(\mathbb{R}^d)$ and

its support $\text{supp}(\gamma * \Phi) \subseteq \text{supp}(\gamma) + \text{supp}(\Phi) \subseteq \Omega_2^c + \Omega_\Phi \subseteq \Omega_1^c$, we can determine that $\gamma * \Phi \in \mathcal{N}_0$ with $\Omega := \Omega_1$. For any $g \in \mathcal{N}_0^\perp$, we have

$$\int_{\mathbb{R}^d} \gamma(\boldsymbol{x})\overline{g(\boldsymbol{x})}\mathrm{d}\boldsymbol{x} = \int_{\mathbb{R}^d} \hat{\gamma}(\boldsymbol{x})\overline{\hat{g}(\boldsymbol{x})}\mathrm{d}\boldsymbol{x} = \int_{\mathbb{R}^d} \frac{\widehat{\gamma * \Phi}(\boldsymbol{x})\overline{\hat{g}(\boldsymbol{x})}}{\hat{\Phi}(\boldsymbol{x})}\mathrm{d}\boldsymbol{x} = \langle \gamma * \Phi, g \rangle_{\mathcal{B}_\Phi^p(\mathbb{R}^d)} = 0$$

which indicates that $g|_{\Omega_2^c} = 0$. According to this result we can deduce that $g = 0$ if and only if $g \in \mathcal{N}_0^\perp$ and $g|_{\Omega_2} = 0$. This means that the restriction map of $\mathcal{N}_0^\perp$ to Ω_2 is one-to-one. Thus the normed space

$$\mathcal{B}(\Omega_2) := \{\phi : \Omega_2 \to \mathbb{C} : \phi = g|_{\Omega_2} \text{ for some } g \in \mathcal{N}_0^\perp\}$$

equipped with the norm $\|\phi\|_{\mathcal{B}(\Omega_2)} := \|g\|_{\mathcal{B}_\Phi^q(\mathbb{R}^d)}$ is well-defined and it is obvious that $\mathcal{B}(\Omega_2) \equiv \mathcal{N}_0^\perp$. Under these additional conditions, the dual space of $\mathcal{B}_\Phi^p(\Omega_1)$ defined in Theorem 5.6 can be even isometrically equivalent to a space composed of functions defined on Ω_2, i.e., $\mathcal{B}_\Phi^p(\Omega_1)' \equiv \mathcal{N}_0^\perp \equiv \mathcal{B}(\Omega_2)$. In this case $\mathcal{B}_\Phi^p(\Omega_1)$ is also an RKBS with the two-sided reproducing kernel $K|_{\Omega_2 \times \Omega_1}$.

Corollary 5.7. *Suppose that the positive definite function Φ given in Theorem 5.1 has a compact support Ω_Φ in $\mathbb{R}^d$. Let Ω_1 and Ω_2 be two subsets of $\mathbb{R}^d$ such that the complement Ω_1^c includes $\Omega_2^c + \Omega_\Phi$. Then $\mathcal{B}_\Phi^p(\Omega_1)$ with $p > 1$ defined in Theorem 5.6 is a reproducing kernel Banach space with the two-sided reproducing kernel*

$$K|_{\Omega_2 \times \Omega_1}(\boldsymbol{x}, \boldsymbol{y}) := \Phi(\boldsymbol{x} - \boldsymbol{y}), \quad \boldsymbol{x} \in \Omega_2, \ \boldsymbol{y} \in \Omega_1.$$

If the subset Ω is a *regular* domain, then the definition of weak derivatives (see [1, Section 1.62]) provides that $f|_\Omega \in W_p^m(\Omega)$ and $\|f|_\Omega\|_{W_p^m(\Omega)} \leq \|f\|_{W_p^m(\mathbb{R}^d)}$ for all $f \in W_p^m(\mathbb{R}^d)$, where $W_p^m(\Omega)$ is the L_p-based Sobolev space of order m. Now we use the embeddings of $\mathcal{B}_\Phi^p(\mathbb{R}^d)$ given in Corollary 5.3 to derive the embeddings of $\mathcal{B}_\Phi^p(\Omega)$. We fix any $h \in \mathcal{B}_\Phi^p(\Omega)$. According to Corollary 5.3, we have

$$\|h\|_{W_p^m(\Omega)} \leq \|Eh\|_{W_p^m(\mathbb{R}^d)} \leq C \|Eh\|_{\mathcal{B}_\Phi^p(\mathbb{R}^d)} = C \|h\|_{\mathcal{B}_\Phi^p(\Omega)}, \quad h \in \mathcal{B}_\Phi^p(\Omega) \subseteq W_p^m(\Omega),$$

for some positive constant C independent on h.

Corollary 5.8. *Let Φ be a positive definite function and $m > pd/q - d/q$ be as in Corollary 5.3. Here q is the conjugate exponent of $p \geq 2$. Suppose that $\Omega \subseteq \mathbb{R}^d$ is regular. Then $\mathcal{B}_\Phi^p(\Omega)$ defined in Theorem 5.6 is embedded into the L_p-based Sobolev space of order m, $W_p^m(\Omega)$, i.e.,*

$$\|h\|_{W_p^m(\Omega)} \leq C \|h\|_{\mathcal{B}_\Phi^p(\Omega)}, \quad h \in \mathcal{B}_\Phi^p(\Omega) \subseteq W_p^m(\Omega),$$

for some positive constant C independent on h.

6. Examples for Matérn Functions

[9, Example 5.7] and [22, Example 4.4] show that *Matérn functions* (Sobolev splines) with shape parameter $\theta > 0$ and degree $n > d/2$

$$G_{\theta,n}(\boldsymbol{x}) := \frac{2^{1-n-d/2}}{\pi^{d/2}\Gamma(n)\theta^{2n-d}}(\theta\|\boldsymbol{x}\|_2)^{n-d/2}K_{d/2-n}(\theta\|\boldsymbol{x}\|_2), \quad \boldsymbol{x} \in \mathbb{R}^d,$$

are positive definite functions on $\mathbb{R}^d$, where $t \mapsto K_\nu(t)$ is the modified Bessel function of the second kind of order ν and $t \mapsto \Gamma(t)$ is the Gamma function. Moreover, $G_{\theta,n}$ is a full-space Green function of the differential operator $L_{\theta,n} := \left(\theta^2 I - \Delta\right)^n$, i.e., $L_{\theta,n}G_{\theta,n} = \delta_{\boldsymbol{0}}$. The Fourier transform of $G_{\theta,n}$ has the form

$$\hat{G}_{\theta,n}(\boldsymbol{x}) = \left(\theta^2 + \|\boldsymbol{x}\|_2^2\right)^{-n}, \quad \boldsymbol{x} \in \mathbb{R}^d.$$

Let $1 < q \leq 2 \leq p < \infty$ with $p^{-1} + q^{-1} = 1$ such that $nq/p > d/2$ and $m := 2n/q$. Since $\hat{G}_{\theta,n}^{\min\{p,q\}-1} \in L_1(\mathbb{R}^d)$, Theorem 5.1 provides that $\mathcal{B}_{G_{\theta,n}}^p(\mathbb{R}^d)$ is an RKBS on $\mathbb{R}^d$ with the two-sided reproducing kernel $K_{\theta,n}(\boldsymbol{x},\boldsymbol{y}) = G_{\theta,n}(\boldsymbol{x} - \boldsymbol{y})$. We can also check that there are two positive constants C_1, C_2 such that

$$C_1\left(1 + \|\boldsymbol{x}\|_2^2\right)^{-m/2} \leq \hat{G}_{\theta,n}(\boldsymbol{x})^{1/q} \leq C_2\left(1 + \|\boldsymbol{x}\|_2^2\right)^{-m/2}. \quad \boldsymbol{x} \in \mathbb{R}^d.$$

According to Corollary 5.3 and 5.8, the RKBS $\mathcal{B}_{G_{\theta,n}}^p(\mathbb{R}^d)$ is embedded into $W_p^m(\mathbb{R}^d)$ and the RKBS $\mathcal{B}_{G_{\theta,n}}^p(\Omega)$ is embedded into $W_p^m(\Omega)$ for any regular domain Ω of $\mathbb{R}^d$.

In particular, when $p := 4$, then $\hat{G}_{\theta,n}^3 = \hat{G}_{\theta,3n}$. According to the discussion of Theorem 5.4 and Remark 5.5, the optimal solution of the SVM

$$\min_{f \in \mathcal{B}_{G_{\theta,n}}^4(\mathbb{R}^d)} \sum_{j=1}^N L(\boldsymbol{x}_j, y_j, f(\boldsymbol{x}_j)) + R\left(\|f\|_{\mathcal{B}_{G_{\theta,n}}^4(\mathbb{R}^d)}\right),$$

has the explicit representation

$$s_{D,L,R}(\boldsymbol{x}) = \sum_{k_1,k_2,k_3=1}^{N,N,N} c_{k_1}\overline{c_{k_2}}c_{k_3}G_{\theta,3n}\left(\boldsymbol{x} - \boldsymbol{x}_{k_1} + \boldsymbol{x}_{k_2} - \boldsymbol{x}_{k_3}\right)$$

$$= \sum_{k_1,k_2,k_3=1}^{N,N,N} c_{k_1}\overline{c_{k_2}}c_{k_3}\mathcal{K}_{\theta,3n}\left(\boldsymbol{x}, \boldsymbol{x}_{k_1}, \boldsymbol{x}_{k_2}, \boldsymbol{x}_{k_3}\right), \quad \boldsymbol{x} \in \mathbb{R}^d,$$

and its coefficients $c = (c_1, \cdots, c_N)^T$ are obtained by solving the following minimization problem

$$\min_{c \in \mathbb{C}^N} \sum_{j=1}^{N} L\left(x_j, y_j, \sum_{k_1,k_2,k_3=1}^{N,N,N} c_{k_1} \overline{c_{k_2}} c_{k_3} \mathcal{K}_{\theta,3n}\left(x_j, x_{k_1}, x_{k_2}, x_{k_3} \right) \right)$$
$$+ R\left(\sum_{j,k_1,k_2,k_3=1}^{N,N,N,N} \overline{c_j} c_{k_1} \overline{c_{k_2}} c_{k_3} \mathcal{K}_{\theta,3n}\left(x_j, x_{k_1}, x_{k_2}, x_{k_3} \right) \right)^{3/4},$$

where $\mathcal{K}_{\theta,3n}(x, y_1, y_2, y_3) := G_{\theta,3n}(x - y_1 + y_2 - y_3)$, and the loss function L and the regularization function R are the same as in Theorem 5.4. More generally, when p is even, then the SVM solution $s_{D,L,R}$ in $\mathcal{B}^p_{G_{\theta,n}}(\mathbb{R}^d)$ is a linear combination of the product groups of the reproducing kernel bases, i.e.,

$$s_{D,L,R}(x) = \sum_{k \in \mathscr{G}^N_{p-1}} \prod_{j=1}^{p/2} c_{k_{2j-1}} \prod_{l=1}^{p/2-1} \overline{c_{k_{2l}}} \mathcal{K}_{\theta,(p-1)n}\left(x, x_{k_1}, \cdots, x_{k_{p-1}} \right), \quad x \in \mathbb{R}^d,$$

where $\mathcal{K}_{\theta,(p-1)n}\left(x, y_1, \cdots, y_{p-1} \right) := G_{\theta,(p-1)n}\left(x - y_1 + y_2 + \cdots + (-1)^{p-1} y_{p-1} \right)$ and $\mathscr{G}^N_{p-1} := \left\{ k := (k_1, \cdots, k_{p-1})^T \in \mathbb{N}^{p-1} : 1 \le k_j \le N, \ j = 1, \ldots, p-1 \right\}$.

According to some numerical experiments comparing $\mathcal{B}^2_{G_{\theta,n}}(\mathbb{R}^2)$ and $\mathcal{B}^4_{G_{\theta,n}}(\mathbb{R}^2)$, we find that the accuracy of the SVM solutions in $\mathcal{B}^4_{G_{\theta,n}}(\mathbb{R}^2)$ is better than in $\mathcal{B}^2_{G_{\theta,n}}(\mathbb{R}^2)$ for the same training data and testing data. The reason for this is that we use three data points to set up each reproducing kernel base for $p = 4$ but the reproducing kernel base for $p = 2$ only owns two data points. This means that the reproducing kernel base for $p = 4$ contains much more information than for $p = 2$. Many other numerical tests will appear in a future paper.

The Matérn functions have been applied in the field of statistical learning (see [14]). This new discovery about Matérn functions might help create new numerical tools for SVMs in RKBS.

Acknowledgments

References

[1] R. A. Adams and J. J. F. Fournier, *Sobolev Spaces*, Elservier/Academic Press, Amsterdam, 2003.

[2] E. Alpaydin, *Introduction to Machine Learning*, MIT Press, Cambridge, Massachusetts, 2010.

[3] A. Berlinet and C. Thomas-Agnan, *Reproducing Kernel Hilbert Spaces in Probability and Statistics*, Kluwer Academic Publishers, Dordrecht, 2004.

[4] M. D. Buhmann, *Radial Basis Functions: Theory and Implementations*, Cambridge University Press, Cambridge, 2003.

[5] J. F. Erickson, *Generalized native spaces*, Ph.D. thesis, Illinois Institute of Technology, Chicago, 2007.

[6] J. F. Erickson and G. E. Fasshauer, *Generalized native spaces*, Approximation Theory XII: San Antonio 2007, in: M. Neamtu, L. L. Schumaker (Eds.), Nashboro Press, Brentwood, TN, 2008, pp. 133–142.

[7] G. E. Fasshauer, *Meshfree Approximation Methods with* MATLAB. World Scientific Publishing Co. Pte. Ltd., Singapore, 2007.

[8] G. E. Fasshauer, *Positive definite kernels: past, present and future*. Dolomite Research Notes on Approximation, 4 (2011), pp. 21–63.

[9] G. E. Fasshauer and Q. Ye, *Reproducing kernels of generalized Sobolev spaces via a Green function approach with distributional operators*, Numer. Math., 119 (2011), pp. 585–611.

[10] G. E. Fasshauer and Q. Ye, *Reproducing kernels of Sobolev spaces via a Green kernel approach with differential operators and boundary operators*, Adv. Comput. Math., 38 (2013), pp. 891-921.

[11] J. R. Giles, *Classes of semi-inner-product spaces*, Trans. Amer. Math. Soc., 129 (1967), pp. 436–446.

[12] L. Hörmander, *The Analysis of Linear Partial Differential Operators I*, Classics in Mathematics, Springer, Berlin Heidelberg, 1983.

[13] R. C. James, *Orthogonality and linear functionals in normed linear spaces*, Trans. Amer. Math. Soc., 61 (1947), pp. 265–292.

[14] M. L. Stein, *Interpolation of Spatial Data: some theory for kriging*, Springer, New York, 1999.

[15] G. Lumer, *Semi-inner-product spaces*. Trans. Amer. Math. Soc., 100 (1961), pp. 29–43.

[16] R. E. Megginson, *An Introduction to Banach Space Theory*, Springer, New York, 1998.

[17] C. A. Micchelli, L. Shen and Y. Xu, *Proximity algorithms for image models: denoising*, Inverse Problems, 27 (2011), online at stacks.iop.org/IP/27/045009.

[18] I. Steinwart and A. Christmann, *Support Vector Machines*, Information Science and Statistics, Springer, New York, 2008.

[19] E. M. Stein and G. Weiss, *Introduction to Fourier Analysis on Euclidean Spaces*, Princeton University Press, Princeton, 1971.

[20] G. Wahba, *Spline Models for Observational Data*, SIAM, Philadelphia, 1990.

[21] H. Wendland, *Scattered Data Approximation*, Cambridge University Press, Cambridge, 2005.

[22] Q. Ye, *Analyzing reproducing kernel approximation methods via a Green function approach*. Ph.D. thesis, Illinois Institute of Technology, Chicago, 2012.

[23] H. Zhang, Y. Xu and J. Zhang, *Reproducing kernel Banach spaces for machine learning*. J. Mach. Learn. Res., 10 (2009), pp. 2741–2775.