

# Solving Neumann Boundary Problem with Kernel-Regularized Learning Approach

Xuexue Ran, Baohuai Sheng

Department of Mathematics, Shaoxing University, Shaoxing, China

Email: bhsheng@usx.edu.cn

**How to cite this paper:** Ran, X.X. and Sheng, B.H. (2024) Solving Neumann Boundary Problem with Kernel-Regularized Learning Approach. *Journal of Applied Mathematics and Physics*, 12, 1101-1125.  
<https://doi.org/10.4236/jamp.2024.124069>

**Received:** March 7, 2024

**Accepted:** April 16, 2024

**Published:** April 19, 2024

Copyright © 2024 by author(s) and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## Abstract

We provide a kernel-regularized method to give theory solutions for Neumann boundary value problem on the unit ball. We define the reproducing kernel Hilbert space with the spherical harmonics associated with an inner product defined on both the unit ball and the unit sphere, construct the kernel-regularized learning algorithm from the view of semi-supervised learning and bound the upper bounds for the learning rates. The theory analysis shows that the learning algorithm has better uniform convergence according to the number of samples. The research can be regarded as an application of kernel-regularized semi-supervised learning.

## Keywords

Neumann Boundary Value, Kernel-Regularized Approach, Reproducing Kernel Hilbert Space, The Unit Ball, The Unit Sphere

## 1. Introduction

It is known that approximation theory and skills have been used to give the analytic solution for PDE with boundary value problems and form the method of fundamental solutions (see e.g. [1]-[6]; **Appendices 1-3**). Recently, the kernel-based collocation method for solving several PDEs with boundary problems has been developed (see e.g. [7] [8] [9] [10] [11]) from the view of minimal norm interpolation of reproducing kernel Hilbert spaces (RKHSs), the existent theorem and the representation theorem for the numerical solutions are shown qualitatively. It is suggested by [12] that kernel-regularized gradient learning may be used to give numerical solution for PDE. Indeed, some kernel-regularized learning algorithms have been used to study the PDE with Dirichlet boundary value problem quantitatively (see e.g. [13] [14] [15] [16]). For a given domain  $D \subset R^d$ , the pairwise distinct collocation points chosen in the kernel-based collocation method are:

$$X_D = \{x_1, \dots, x_N\} \subset D,$$

$$X_{\partial D} = \{x_{N+1}, \dots, x_{N+M}\} \subset \partial D,$$

and the sample values from the PDE (where  $P$  and  $B$  are given differential operators):

$$\begin{cases} Pu = f^*, & \text{in } D \subset R^d, \\ Bu = g, & \text{on } \partial D \end{cases} \quad (1)$$

at the given collocation, points are (see [7] [11]):

$$y_j = f^*(x_j) + \eta_{x_j}, \quad j = 1, 2, \dots, N,$$

$$y_k = g(x_{N+k}), \quad k = 1, 2, \dots, M,$$

where the random variable  $\bar{\eta} := (\eta_{x_1}, \dots, \eta_{x_N})^T \sim \mathcal{N}(\bar{0}, \bar{\Psi}_{\partial D})$ .

There are two typical PDE problems (see Chapter 1 of [17]). When

$P = \nabla = \sum_{j=1}^d \frac{\partial^2}{\partial x_j^2}$  is the Laplace operator and  $B = I$  is the unit operator, we have the Dirichlet problem:

$$\begin{cases} \Delta u = f^*, & \text{in } D \subset R^d, \\ u = g, & \text{on } \partial D \end{cases} \quad (2)$$

When  $P = I$  is the unit operator and  $B = \frac{\partial}{\partial \bar{n}}$  is the directional derivative along the outward normal vector  $\bar{n}$ , i.e.  $Bu = \frac{\partial u}{\partial \bar{n}} = \nabla u \cdot \bar{n}$ , problem (1) become the Neumann boundary problem:

$$\begin{cases} u = f^*, & \text{in } D \subset R^d, \\ \frac{\partial u}{\partial \bar{n}} = g, & \text{on } \partial D. \end{cases} \quad (3)$$

The observations  $\{(x_j, y_j)\}_{j=1}^N$  can be regarded as the supervised labeled observations in the setting of semi-supervised learning and  $y_k (k = 1, 2, \dots, M)$  may be regarded as the unlabeled ones. This similarity encourages us to construct kernel-regularized learning algorithms to give numerical solution for problem (3) referring to the kernel semi-supervised learning frameworks (see e.g. [18] [19] [20] [21]). Along this line, we constructed in [15] a kind of kernel-regularized learning algorithm for solving problem (2) and showed the convergence rate. In the present paper, we shall construct a kind of kernel-regularized regression algorithm to solve problem (2) in case that  $D$  is the unit ball and  $\partial D$  is the unit sphere. To this aim, we restate problem (3) in the setting of Sobolev spaces.

Let  $\Omega \subset R^d$  be a given bounded closed domain with a smoothness surface (i.e. the outward normal derivative is continuous) and  $\rho_\Omega$  be a Borel measure on  $\Omega$ . Let  $C^{(s)}(\Omega)$  denote the set of functions such that  $\partial_x^\alpha f(x) \in C(\Omega)$  and  $|\alpha| \leq s$ , where for  $\alpha = (\alpha_1, \dots, \alpha_d) \in Z_+^d$  we define  $|\alpha| = \sum_{i=1}^d \alpha_i$  and

$$\partial^\alpha f(x) := \partial_x^\alpha f(x) = \partial_1^{\alpha_1} \cdots \partial_d^{\alpha_d} f(x) = \frac{\partial^{|\alpha|} f(x)}{\partial(x^1)^{\alpha_1} \cdots \partial(x^d)^{\alpha_d}}.$$

For a  $K \in C^{(s)}(\Omega \times \Omega)$ , we define:

$$\partial_x^\alpha \partial_y^\beta K(x, y) = \frac{\partial^{|\alpha|+|\beta|} K(x, y)}{\partial(x^1)^{\alpha_1} \cdots \partial(x^d)^{\alpha_d} \partial(y^1)^{\beta_1} \cdots \partial(y^d)^{\beta_d}}, x = (x^1, \dots, x^d), y = (y^1, \dots, y^d).$$

Denoted by  $W^1(\rho_\Omega)$ , the set of all functions whose 1-order partial derivatives are all in  $L^2(\rho_\Omega)$ , i.e.

$$W^1(\rho_\Omega) = \left\{ f : \|f\|_{W^1(\rho_\Omega)} = \left( \sum_{|\alpha| \leq 1} \|D^\alpha f\|_{2, \rho_\Omega}^2 \right)^{1/2} < +\infty \right\},$$

and  $L^2(\rho_\Omega) = \left\{ f(x) : \|f\|_{2, \rho_\Omega} = \left( \int_\Omega |f(x)|^2 d\rho_\Omega \right)^{1/2} < +\infty \right\}$ . Defined by  $\frac{\partial}{\partial \bar{n}}$ , the

outward normal derivative operator, i.e.  $\frac{\partial f(x)}{\partial \bar{n}} := \frac{\partial_x f(x)}{\partial \bar{n}} = \nabla f(x) \cdot \bar{n}$ , where

$$\nabla f(x) = \left( \frac{\partial f}{\partial x^1}, \dots, \frac{\partial f}{\partial x^d} \right) \text{ and } \bar{n} \text{ is the outward normal vector at}$$

$$x = \{x^1, \dots, x^d\} \in \partial\Omega.$$

To borrow the setting of learning theory (see [22]), we rewrite the problem (3). Let  $f^*$  be an unknown function and  $g$  be a given function. Then, the collocation points  $z = \{x_1, \dots, x_m\} \subset \text{int}\Omega$  and  $v = \{x_{m+1}, \dots, x_{m+u}\} \subset \partial\Omega$  for the Neumann boundary problem:

$$\begin{cases} f(x) = f^*(x), & x \in \Omega, \\ \frac{\partial_x f(x)}{\partial \bar{n}} = g(x), & x \in \partial\Omega \end{cases} \quad (4)$$

are scattered points with values  $y_{x_i} = f^*(x_i) + \eta_{x_i}$ ,  $i = 1, 2, \dots, m$ , and

$y_{x_i} = g(x_i)$ ,  $i = m+1, m+2, \dots, m+u$ , where for a given  $x_i = (x_i^1, \dots, x_i^d) \in \Omega$ ,

$\xi_{x_i}$  is a random variable subject to a condition distribution  $\rho(y | x_i)$  satisfy-

ing  $|\rho(y | x_i)| \leq B$ ,  $B$  is a given constant number,

$$E_{\rho(\cdot|x)}(\eta_x) = \int_{[-B, B]} \eta_x(y) d\rho(y | x) = 0 \text{ and } \sigma = \left( \sum_{i=1}^m \sigma_{x_i}^2 \right)^{1/2} < +\infty \text{ and}$$

$\sigma_x^2 = E_{\rho(\cdot|x)}(\eta_x^2)$ . The correspondence of (4) is:

$$\begin{cases} f(x_i) = y_{x_i}, & i = 1, 2, \dots, m, \\ \left. \frac{\partial_x f(x)}{\partial \bar{n}} \right|_{x=x_i} = g(x_i), & i = m+1, \dots, m+u \end{cases} \quad (5)$$

We shall give an investigation on the convergence analysis of problem (5) when  $\Omega$  is the unit ball  $B^d = \{x \in R^d : \|x\| \leq 1\}$  and  $\partial\Omega$  is the unit sphere  $S^{d-1} = \{x \in R^d : \|x\| = 1\}$ . The paper is organized as follows. In Section 2.1, we

shall provide some notions on the kernel-regularized regression learning, which contain the concept of reproducing kernel Hilbert spaces, the concept of kernel-regularized regression learning model and the kernel-regularized semi-supervised regression learning model. In Section 2.2, we shall provide some notions and results on spherical analysis. In particular, we shall define an RKHS with spherical harmonics, which have the reproducing property for the outward normal vector operator. The Neumann boundary value problem with the RKHS as the hypothesis space is defined. Based on these notions, we define in Section 2.3 a kind of kernel-regularized learning algorithm for solving the Neumann boundary value problem, show the representation theorem and give the error decomposition, with which we give the learning rates (*i.e.* the main Theorem 2.1) in Section 2.4. In Section 3, we shall give some lemmas, which will be used to show Theorem 2.1 in Section 4. Section 5 is the appendices containing some knowledge about the convex function, a kind of RKHS  $\mathcal{H}_K^{\bar{n}}(\rho_\Omega)$  defined in a Sobolev space  $H^{\bar{n}}(\rho_\Omega)$  associating a general domain  $\Omega$  and its boundary  $\partial\Omega$ , and a probability inequality defined on a general RKHS.

Throughout the paper, we denote by  $A = O(B)$  the fact that there is a constant  $C$  independent of  $A$  and  $B$  such that  $A \leq CB$ . We say  $a \sim B$  if both  $A = O(B)$  and  $B = O(A)$ .

## 2. Kernel-Regularized Regression and Error Analysis

We first provide some notions and results about the kernel-regularized regression learning problem.

### 2.1. Notions and Results on Kernel-Regularized Regression

Follow all the definitions and notions in Section 1. Let

$K_x(y) = K(x, y): \Omega \times \Omega \rightarrow \mathbb{R}$  be a Mercer kernel (*i.e.* it is continuous, symmetry and positive semi-definite, *i.e.* for any given integer  $l \geq 1$  and any given set  $\{x_1, x_2, \dots, x_l\} \subset \Omega$ , the matrix  $(K(x_i, x_j))_{i, j=1, 2, \dots, l}$  is positive definite) satisfy-

ing  $\left(\int_{\Omega \times \Omega} |K(x, y)|^2 d\rho_\Omega(x) d\rho_\Omega(y)\right)^{\frac{1}{2}} < +\infty$ . The reproducing kernel Hilbert space  $(H_K, \|\cdot\|_K)$  associated with  $K(x, y)$  is a Hilbert space consisting of all the real functions defined on  $\Omega$  such that:

$$f(x) = \langle f, K_x \rangle_K, \quad \forall x \in \Omega, \forall f \in H_K.$$

Define an operator  $\mathcal{L}_K(f, x)$  as:

$$\mathcal{L}_K(f, x) = \int_\Omega f(y) K_x(y) d\rho_\Omega, \quad x \in \Omega.$$

Then,  $\mathcal{L}: L^2(\rho_\Omega) \rightarrow L^2(\rho_\Omega)$ . Denoted by  $\lambda_k$  the  $k$ -th eigenvalue associated with eigenfunction  $\varphi_k$ . Then, we have by the Mercer theorem that:

$$K(x, y) = \sum_{l=0}^{+\infty} \lambda_l \varphi_l(x) \varphi_l(y), \quad x, y \in \Omega,$$

where we assume the convergence on the right side is absolute (for every  $x, y \in \Omega$ )

and uniform on  $x, y \in \Omega$ . If  $\{\varphi_k\}_{k=0}^{+\infty}$  forms an orthonormal system in  $L^2(\rho_\Omega)$ , then:

$$H_K = \mathcal{L}_K^{\frac{1}{2}}(L^2(\rho_\Omega)) = \left\{ f(x) = \sum_{l=0}^{+\infty} a_l(f) \varphi_l(x) : \|f\|_K = \left( \sum_{l=0}^{+\infty} \frac{|a_l(f)|^2}{\lambda_l} \right)^{\frac{1}{2}} < +\infty \right\},$$

where  $\mathcal{L}_K = \mathcal{L}_K^{\frac{1}{2}} \circ \mathcal{L}_K^{\frac{1}{2}}$ .

Let  $z = \{(x_k, y_k)\}_{k=1}^m$  be a set of observations about an unknown function  $f^*$ . Then, to obtain a good approximation of  $f^*$ , one usually borrow the kernel-regularized regression learning model:

$$f_z = \arg \min_{f \in H_K} (\mathcal{E}_z(f) + \lambda \|f\|_K^2), \tag{6}$$

where  $\mathcal{E}_z(f) = \frac{1}{m} \sum_{k=1}^m (y_k - f(x_k))^2$  is the empirical variance. In learning theory, the convergence analysis is sum up to bound the convergence rate for the error ([23] or [24]):

$$\|f_z - f^*\|_{L^2(\rho_\Omega)}. \tag{7}$$

When the observation set  $z$  has the form  $z = \{(x_k, y_k)\}_{k=1}^m \cup \{x_{m+l}\}_{l=1}^u$ , *i.e.* the nodes  $\{x_{m+l}\}_{l=1}^u$  have no labeled observation values  $\{y_{m+l}\}_{l=1}^u$ , we call this case the semi-supervised learning. In practical applications, most of the observations belong to semi-supervised samples since data is precious. Many mathematicians have paid their attentions to this field (see e.g. [18] [19] [21] [25] [26] [27] [28]). The main ideas of dealing with this problem are add a term to make the use of unlabeled data, *i.e.* we need to modify (8) as the form of:

$$f_{z,\lambda,\gamma} = \arg \min_{f \in H_K} (\mathcal{E}_z(f) + \gamma \Omega_{u,m}(f) + \lambda \|f\|_K^2), \quad \gamma > 0. \tag{8}$$

For example, in [19], one choose

$$\gamma \Omega_{u,m}(f) = \frac{\gamma}{(u+m)^2} \sum_{i,j=1}^{l+u} (f(x_i) - f(x_j))^2 W_{i,j},$$

where  $W_{i,j}$  are edge weights in the data adjacency graph. Also, in [21], one chooses:

$$\gamma \Omega_{u,m}(f) = \frac{\gamma}{(m+u)(m+u-1)} \sum_{i,j=1, i \neq j}^{m+u} w_{ij}^{(\sigma)} \times (f(x_i) - f(x_j))^2,$$

$$w_{ij}^{(\sigma)} = \exp \left\{ -\frac{d_K(x_i, x_j)}{\sigma} \right\},$$

where  $\sigma > 0$  and

$$d_K(x, y) = \|K_x - K_y\|_K = \sqrt{K(x, x) + K(y, y) - 2K(x, y)}.$$

These choices encourage us to choose suitable  $\gamma \Omega_{u,m}(f)$  to give good approximation solution for problem (5).

### 2.2. Some Results on Spherical Analysis

In this subsection, we shall define a kind of RKHS with spherical harmonics, with which define a kernel-regularized regression learning algorithm for solving problem (5) when  $\Omega$  is the unit ball  $B^d$  and show the learning rates.

Let  $B^d = \{x \in R^d : \|x\| \leq 1\}$  denote the unit ball in  $d$ -dimensional Euclidean space  $R^d$  with the usual inner product  $\langle x, y \rangle$ , and  $\|x\| = \sqrt{\langle x, x \rangle}$  is the usual Euclidean norm. For weight  $W_\mu(x) = (1 - \|x\|^2)^\mu$ ,  $\mu > -1$ , we denote by  $L_{p,\mu}(B^d) \equiv L_p(B^d, W_\mu)$ ,  $1 \leq p < +\infty$ , the space of measurable functions defined on  $B^d$  with:

$$\|f\|_{L_{p,\mu}(B^d)} = \left( \int_{B^d} |f(x)|^p W_\mu(x) dx \right)^{\frac{1}{p}} < +\infty, \quad 1 \leq p < +\infty$$

and for  $p = +\infty$ , we assume that  $L_{\infty,\mu}$  denotes the space  $C(B^d)$  of continuous functions on  $B^d$  with the uniform norm.

We denote by  $\Pi_n^d$  the space of all polynomials in  $d$  variables of degree at most  $n$ , and by  $\nu_n^d(W_\mu)$  the space of all polynomials of degree  $n$  which are orthogonal to polynomials of low degree in  $L_{2,\mu}(B^d)$ . The  $\nu_n^d(W_\mu)$  is mutually orthogonal in  $L_{2,\mu}(B^d)$  and (see [29]):

$$L_{2,\mu}(B^d) = \bigoplus_{n=0}^{\infty} \nu_n^d(W_\mu), \quad \Pi_n^d = \bigoplus_{k=0}^n \nu_k^d(W_\mu).$$

Let  $d\sigma$  denote the Lebesgue measure on  $S^{d-1} = \{x : \|x\| = 1\}$  and denote the area of  $S^{d-1}$  by  $\sigma_d$ ,  $\sigma_d = \int_{S^{d-1}} d\sigma = 2\pi^{d/2} / \Gamma(d/2)$ . Let  $\mathcal{H}_n^d$  denote the space of homogeneous harmonic polynomials of degree  $n$ , which are homogeneous polynomials of degree  $n$  satisfying equation  $\Delta p = \sum_{k=1}^d \frac{\partial^2 p}{\partial (x^k)^2} = 0$ . Also, we de-

note by  $\mathcal{P}_n^d$  the set of homogeneous polynomials of degree  $n$ . It is well known that:

$$\alpha_n^d := \dim H_n^d = \binom{n+d-1}{n-1} - \binom{n+d-3}{n-2}.$$

Let  $W^1(B^d)_\mu$  denote the set of functions whose 1-th derivatives are all in  $L_{2,\mu}(B^d)$ , i.e.

$$W^1(B^d)_\mu = \left\{ f : \|f\|_{W^2(B^d)_\mu} = \left( \sum_{|\alpha| \leq 1} \|\partial^\alpha f\|_{L_{2,\mu}(B^d)}^2 \right)^{1/2} < +\infty \right\}.$$

In this case,  $S^{d-1} = \left\{ x : \|x\|^2 = \sum_{i=1}^d (x^i)^2 = 1 \right\}$  and

$$\frac{\partial_x f}{\partial \vec{n}}(x) = \sum_{i=1}^d x^i \frac{\partial f}{\partial x^i}(x), \quad x = (x^1, \dots, x^d) \in S^{d-1}. \tag{9}$$

Define a subclass of  $W^1(B^d)_\mu$  as:

$$\mathcal{H}_\mu^{\bar{n}} = \left\{ f \in W^1(B^d)_\mu : \|f\|_{\mathcal{H}_\mu^{\bar{n}}} = \left( \int_{B^d} |f(x)|^2 W_\mu(x) dx + \int_{S^{d-1}} \left| \frac{\partial_\xi f}{\partial \bar{n}}(\xi) \right|^2 d\sigma(\xi) \right)^{\frac{1}{2}} < +\infty \right\}.$$

An inner product defined on  $\mathcal{H}_\mu^{\bar{n}}$  is:

$$\langle f, g \rangle_{\mathcal{H}_\mu^{\bar{n}}} = \int_{B^d} f(x)g(x)W_\mu(x)dx + \int_{S^{d-1}} \frac{\partial f}{\partial \bar{n}}(\xi) \frac{\partial g}{\partial \bar{n}}(\xi) d\sigma(\xi).$$

Denoted by  $\nu_n^d(W_\mu, S)$  the space of orthogonal polynomials with respect to  $\langle \cdot, \cdot \rangle_{\mathcal{H}_\mu^{\bar{n}}}$ . Then, by Theorem 3 of [30], we know  $\nu_n^d(W_\mu, S)$  contains a mutually orthonormal basis  $\{Q_{n,k} \equiv Q_{n,k}^d : k = 1, 2, \dots, a_n^d\}$  with respect to  $\langle \cdot, \cdot \rangle_{\mathcal{H}_\mu^{\bar{n}}}$ . Then, there holds the expansion:

$$f(x) \sim \sum_{k=0}^\infty \sum_{l=1}^{a_k^d} a_{k,l}(f) Q_{k,l}(x), \quad x \in B^d, f \in \mathcal{H}_\mu^{\bar{n}},$$

where  $a_{k,l}(f) = \langle f, Q_{k,l} \rangle_{\mathcal{H}_\mu^{\bar{n}}}$  and by the Bessel inequality, we have:

$$\left( \sum_{k=0}^\infty \sum_{l=1}^{a_k^d} |a_{k,l}(f)|^2 \right)^{\frac{1}{2}} \leq \|f\|_{\mathcal{H}_\mu^{\bar{n}}} < +\infty.$$

Let  $K^\mu(x, y) : B^d \times B^d \rightarrow R$  be a Mercer kernel with the form:

$$K_x^\mu(y) = K^\mu(x, y) := \sum_{k=0}^\infty \lambda_k P_k(x, y), \quad x \in B^d, y \in B^d, \tag{10}$$

where  $P_k(x, y) = \sum_{l=1}^{a_k^d} Q_{k,l}(x)Q_{k,l}(y)$  and  $\sum_{k=0}^\infty \lambda_k c_k < +\infty$  with  $\lambda_k > 0$  being defined as  $\sup_{x \in B^d} P_k(x, x) = c_k (k = 0, 1, 2, \dots)$ .

Define

$$\begin{aligned} \mathcal{L}_{K^\mu}(f)(x) &= \mathcal{L}_{K^\mu}(f, x) := \langle f, K_x(\cdot) \rangle_{\mathcal{H}_\mu^{\bar{n}}} \\ &= \sum_{k=0}^\infty \lambda_k \sum_{l=1}^{a_k^d} a_{k,l}(f) Q_{k,l}(x), \quad x \in B^d \end{aligned}$$

and  $\mathcal{H}_{K^\mu}^{\bar{n}} = \mathcal{L}_{K^\mu}^{\frac{1}{2}}(\mathcal{H}_\mu^{\bar{n}})$ . Then,

$$\mathcal{H}_{K^\mu}^{\bar{n}} = \left\{ \sum_{k=0}^\infty \sum_{l=1}^{a_k^d} a_{k,l}(f) Q_{k,l}(x) : \|f\|_{K^\mu} = \left( \sum_{k=0}^\infty \sum_{l=1}^{a_k^d} \frac{|a_{k,l}(f)|^2}{\lambda_k} \right)^{\frac{1}{2}} < +\infty \right\}.$$

For  $f(x) = \sum_{k=0}^\infty \sum_{l=1}^{a_k^d} a_{k,l}(f) Q_{k,l}(x) \in \mathcal{H}_{K^\mu}^{\bar{n}}$  and

$g(x) = \sum_{k=0}^\infty \sum_{l=1}^{a_k^d} a_{k,l}(g) Q_{k,l}(x) \in \mathcal{H}_{K^\mu}^{\bar{n}}$ , we define:

$$\langle f, g \rangle_{K^\mu} = \sum_{k=0}^\infty \sum_{l=1}^{a_k^d} \frac{a_{k,l}(f) a_{k,l}(g)}{\lambda_k}.$$

Then, we shall show in Proposition 2.1 that  $\mathcal{H}_{K^\mu}^{\bar{n}}$  is an RKHS associated with

kernel (10).

To give quantitatively description for the kernel  $K^\mu$ , we give two assumptions.

**Assumption A.** Assume  $K^\mu \in C^{(1)}(B^d \times B^d)$ .

**Assumption B.** Assume  $\sup_{x \in S^{d-1}} \left| \frac{\partial_x P_k(x, x)}{\partial \bar{n}} \right| = c'_k$  ( $k = 0, 1, 2, \dots$ ) and

$$\sum_{k=0}^{\infty} \lambda_k c'_k < +\infty. \tag{11}$$

Since  $\{Q_{n,k} \equiv Q_{n,k}^d : k = 1, 2, \dots, a_n^d\}$  are algebraic polynomials,  $c_k$  and  $c'_k$  must exist. The real numbers  $\lambda_k$  satisfy (11) are also existent, for example, we can take  $\lambda_k = e^{-(c_k + c'_k)}$  ( $k = 0, 1, 2, \dots$ ).

If (11) holds, then by Theorem 2.4 in [31], or Theorem 4.2 in [32] or Proposition 6.2 in [33] that:

$$\partial_x^\alpha (\partial_y^\beta K_x^\mu(y)) = K^\mu(x, y) := \sum_{k=0}^{\infty} \lambda_k \partial_x^\alpha (\partial_y^\beta P_k(x, y)), \quad x \in S^{d-1}, y \in S^{d-1}, \tag{12}$$

and the convergence in the right-side of (12) is absolute and uniform on  $S^{d-1} \times S^{d-1}$ .

**Proposition 2.1.** Assume above Assumptions A and B hold. Then,

1) There holds the reproducing property:

$$f(x) = \left\langle f, K_x^\mu(\cdot) \right\rangle_{K^\mu}, \quad f \in \mathcal{H}_{K^\mu}^{\bar{n}}, x \in B^d. \tag{13}$$

2) There holds the reproducing property for the outward normal vector operator, i.e.

$$\frac{\partial_x f(x)}{\partial \bar{n}} = \left\langle f, \frac{\partial_x K_x^\mu(\cdot)}{\partial \bar{n}} \right\rangle_{K^\mu}, \quad f \in \mathcal{H}_{K^\mu}^{\bar{n}}, x \in S^{d-1}. \tag{14}$$

3) Define

$$k = \sup_{x \in B^d} \|K_x^\mu(\cdot)\|_{K^\mu} + \sup_{x \in S^{d-1}} \left\| \frac{\partial_x K_x^\mu(\cdot)}{\partial \bar{n}} \right\|_{K^\mu}.$$

Then, for all  $x \in B^d$  and  $y \in S^{d-1}$ , we have:

$$|f(x)| \leq k \|f\|_{K^\mu}, \quad \left| \frac{\partial_x f(y)}{\partial \bar{n}} \right| \leq k \|f\|_{K^\mu}. \tag{15}$$

*Proof.* See the proofs in Section 4.

Let  $\{x_i\}_{i=1}^{m+1}$  be observations drawn i.i.d. according to  $\rho_{B^d}$ ,  $y_i = f^*(x_i) + \eta_{x_i}$ ,  $i = 1, 2, \dots, m$  and for a given  $x_i \in B^d$   $\xi_{x_i}$  is a random variable subject to a condition distribution  $\rho_{B^d}(y | x_i)$  satisfying  $|\rho(y | x_i)| \leq B$  ( $B$  is a given constant number),  $E_{\rho^{(\cdot|x)}}(\eta_x) = \int_{[-B, B]} \eta_x(y) d\rho_{B^d}(y | x) = 0$ ,  $\sigma = \left( \int_{B^d} \sigma_x^2 d\rho_{B^d} \right)^{\frac{1}{2}} < +\infty$  and  $\sigma_x^2 = E_{\rho_{B^d}^{(\cdot|x)}}(\eta_x^2)$ . Then,  $z = \{(x_i, y_i)\}_{i=1}^m$  can be regarded as observations drawn i.i.d. according to  $\rho(x, y) = \rho_{B^d}(x) \rho_{B^d}(y | x)$  and  $\{x_i\}_{i=m+1}^{m+1}$  be samples drawn i.i.d. according to  $\rho_{S^{d-1}}$ . The correspondence



of problem (5) then is:

$$\begin{cases} f(x_i) = y_{x_i}, & i = 1, 2, \dots, m, \\ \frac{\partial f(x_i)}{\partial \bar{n}} = \sum_{k=1}^d x_i^k \frac{\partial f(x_i)}{\partial x^k} = g(x_i), & i = m+1, \dots, m+l \end{cases} \quad (16)$$

We shall give an investigation on the numerical solutions of problem (16) with kernel-regularized approaches. A kernel learning algorithm with  $\mathcal{H}_{K^\mu}^{\bar{n}}$  being the hypothesis space will be defined in Section 2.2. The representation theorem for it is provided and an error decomposition for its error analysis is given, from which a learning rate for Algorithm (16) is shown.

### 2.3. Kernel-Regularized Regression

With above notions in hand, we now give following kernel-regularized learning algorithms for giving solutions for problem (16):

$$f_{z,\lambda} = \arg \min_{f \in \mathcal{H}_{K^\mu}^{\bar{n}}} \mathcal{E}_z(f) + \frac{1}{l} \sum_{i=m}^{m+l} \left( \frac{\partial_x f(x_i)}{\partial \bar{n}} - g(x_i) \right)^2 + \lambda \|f\|_{K^\mu}^2, \quad (17)$$

where  $l \sim m$ , i.e. there exist  $c_1 > 0, c_2 > 0$  such that  $c_1 \leq \frac{l}{m} \leq c_2$  and

$$\mathcal{E}_z(f) = \frac{1}{m} \sum_{i=1}^m (f(x_i) - y_{x_i})^2.$$

Corresponding to (17), we define a model for the observations without the disturbances  $\xi_{x_i}$  by:

$$f_{\bar{x},\lambda} = \arg \min_{f \in \mathcal{H}_{K^\mu}^{\bar{n}}} \mathcal{E}_{\bar{x}}(f) + \frac{1}{l} \sum_{i=m}^{m+l} \left( \frac{\partial_x f(x_i)}{\partial \bar{n}} - g(x_i) \right)^2 + \lambda \|f\|_{K^\mu}^2, \quad (18)$$

where

$$\mathcal{E}_{\bar{x}}(f) = \frac{1}{m} \sum_{i=1}^m (f(x_i) - f^*(x_i))^2.$$

The integral model for (18) is defined as:

$$f_\lambda = \arg \min_{f \in \mathcal{H}_{K^\mu}^{\bar{n}}} \mathcal{E}(f) + \int_{S^{d-1}} \left( \frac{\partial_x f(x)}{\partial \bar{n}} - g(x) \right)^2 d\rho_{S^{d-1}} + \lambda \|f\|_{K^\mu}^2, \quad (19)$$

where

$$\mathcal{E}(f) = \int_{B^d} (f(x) - f^*(x))^2 d\rho_{B^d}.$$

**Proposition 2.2.** (Representation theorem). *Assume above Assumptions A and B hold. Then,*

1) Algorithm (17) has unique solution  $f_{z,\lambda}$  and there are coefficients  $\{a_i\}_{i=1}^{m+l}$  depending upon  $\lambda, m, l, f_{z,\lambda}, z, f^*$  and  $g$  such that:

$$f_{z,\lambda}(\cdot) = \sum_{i=1}^m a_i K_{x_i}^\mu(\cdot) + \sum_{k=1}^l a_{m+k} \frac{\partial_x K_{x_{m+k}}^\mu(\cdot)}{\partial \bar{n}}. \quad (20)$$

2) Algorithm (18) has unique solution  $f_{\bar{x},\lambda}$  and there are coefficients  $\{b_i\}_{i=1}^{m+l}$

depending upon  $\lambda, m, l, f_{\bar{x},\lambda}, \bar{X}, f^*$  and  $g$  such that:

$$f_{\bar{x},\lambda}(\cdot) = \sum_{i=1}^m b_i K_{x_i}^\mu(\cdot) + \sum_{k=1}^l b_{m+k} \frac{\partial_x K_x^\mu(\cdot)}{\partial \bar{n}}. \tag{21}$$

3) Algorithm (19) has unique solution  $f_\lambda$  and there is a function  $G_{\lambda, f^*}(x)$  depending upon  $\lambda, f^*$  and a function  $\rho_{\lambda, g}(x)$  depending upon  $\lambda$  and  $g$  such that:

$$f_\lambda(\cdot) = \int_{B^d} G_{\lambda, f^*}(x) K_x^\mu(\cdot) d\rho_{B^d} + \int_{S^{d-1}} \rho_{\lambda, g}(x) \frac{\partial_x K_x^\mu(\cdot)}{\partial \bar{n}} d\rho_{S^{d-1}}. \tag{22}$$

*Proof.* See the proof in Section 4.

(20) shows that Algorithm (17) can be replaced by some coefficient regularized models and is a new topic, such kind of research can be found from literature [34] [35].

We give the following error decomposition:

$$\begin{aligned} & \|f_{z,\lambda} - f^*\|_{L^2(\rho_{B^d})} + \left\| \frac{\partial f_{z,\lambda}}{\partial \bar{n}} - g \right\|_{L^2(\rho_{S^{d-1}})} \\ & \leq \|f_{z,\lambda} - f_{\bar{x},\lambda}\|_{L^2(\rho_{B^d})} + \|f_{\bar{x},\lambda} - f^*\|_{L^2(\rho_{B^d})} \\ & \quad + \left\| \frac{\partial f_{z,\lambda}}{\partial \bar{n}} - \frac{\partial f_{\bar{x},\lambda}}{\partial \bar{n}} \right\|_{L^2(\rho_{S^{d-1}})} + \left\| \frac{\partial f_{\bar{x},\lambda}}{\partial \bar{n}} - g \right\|_{L^2(\rho_{S^{d-1}})} \\ & \leq \|f_{z,\lambda} - f_{\bar{x},\lambda}\|_{L^2(\rho_{B^d})} + \|f_{\bar{x},\lambda} - f_\lambda\|_{L^2(\rho_{B^d})} + \|f_\lambda - f^*\|_{L^2(\rho_{B^d})} \\ & \quad + \left\| \frac{\partial f_{z,\lambda}}{\partial \bar{n}} - \frac{\partial f_{\bar{x},\lambda}}{\partial \bar{n}} \right\|_{L^2(\rho_{S^{d-1}})} + \left\| \frac{\partial f_{\bar{x},\lambda}}{\partial \bar{n}} - \frac{\partial f_\lambda}{\partial \bar{n}} \right\|_{L^2(\rho_{S^{d-1}})} + \left\| \frac{\partial f_\lambda}{\partial \bar{n}} - g \right\|_{L^2(\rho_{S^{d-1}})} \\ & \leq \left( \|f_{z,\lambda} - f_{\bar{x},\lambda}\|_{L^2(\rho_{B^d})} + \left\| \frac{\partial f_{z,\lambda}}{\partial \bar{n}} - \frac{\partial f_{\bar{x},\lambda}}{\partial \bar{n}} \right\|_{L^2(\rho_{S^{d-1}})} \right) \\ & \quad + \left( \|f_{\bar{x},\lambda} - f_\lambda\|_{L^2(\rho_{B^d})} + \left\| \frac{\partial f_{\bar{x},\lambda}}{\partial \bar{n}} - \frac{\partial f_\lambda}{\partial \bar{n}} \right\|_{L^2(\rho_{S^{d-1}})} \right) + \sqrt{2} \sqrt{K(f^*, g, \lambda)} \tag{23} \\ & \leq k \left( \|f_{z,\lambda} - f_{\bar{x},\lambda}\|_{K^\mu} + \|f_{\bar{x},\lambda} - f_\lambda\|_{K^\mu} \right) + \sqrt{2} \sqrt{K(f^*, g, \lambda)}, \end{aligned}$$

where in the last derivation, we have used (15) and

$$K(f^*, g, \lambda) = \inf_{f \in \mathcal{H}_{K^\mu}^n} \left( \|f - f^*\|_{L^2(\rho_{B^d})}^2 + \int_{S^{d-1}} \left( \frac{\partial f(x)}{\partial \bar{n}} - g(x) \right)^2 d\rho_{S^{d-1}} + \lambda \|f\|_{K^\mu}^2 \right),$$

which controls the approximation errors. Then, to bound the error:

$$\mathcal{E}(f_{z,\lambda}, f^*, g) = \|f_{z,\lambda} - f^*\|_{L^2(\rho_{B^d})} + \left\| \frac{\partial f_{z,\lambda}}{\partial \bar{n}} - g \right\|_{L^2(\rho_{S^{d-1}})},$$

we need only to bound upper bounds for the sample errors  $\|f_{z,\lambda} - f_{\bar{x},\lambda}\|_{K^\mu}$  and  $\|f_{\bar{x},\lambda} - f_{\rho,\lambda}\|_{K^\mu}$  respectively.

### 2.4. Learning Rates

**Theorem 2.1.** Let the above Assumptions A and B hold and let  $f_{z,\lambda}$  be the solution of (17) and let  $f^* \in C(B^d)$  and  $g \in L^2(\rho_{S^{d-1}})$ . Then, for any  $\delta \in (0,1)$ , with confidence  $1 - \delta$ , holds:

$$\begin{aligned} & \|f_{z,\lambda} - f^*\|_{L^2(\rho_{B^d})} + \left\| \frac{\partial f_{z,\lambda}}{\partial \bar{n}} - g \right\|_{L^2(\rho_{S^{d-1}})} \\ &= O \left( \frac{\sigma}{\lambda\sqrt{m\delta}} + \left( \frac{\sqrt{K(f^*, g, \lambda)}}{\lambda^{\frac{3}{2}}\sqrt{m}} + \frac{\|f^*\|_{C(B^d)}}{\lambda\sqrt{m}} \right) \log \frac{4}{\delta} \right. \\ & \quad \left. + \frac{\sqrt{K(f^*, g, \lambda)}}{\lambda\sqrt{l\delta}} + \sqrt{K(f^*, g, \lambda)} \right). \end{aligned} \tag{24}$$

If  $g = \frac{\partial f^*(x)}{\partial \bar{n}}$  for  $x \in S^{d-1}$ , then

$$\mathcal{E}(f_{z,\lambda}, f^*, g) = \|f_{z,\lambda} - f^*\|_{H^{\bar{n}}(\rho_{B^d})}$$

and in this case,

$$K(f^*, g, \lambda) = K(f^*, \lambda) = \inf_{f \in \mathcal{H}_{K^\mu}^{\bar{n}}} \left( \|f - f^*\|_{H^{\bar{n}}(\rho_{B^d})}^2 + \lambda \|f\|_{K^\mu}^2 \right), \quad \lambda > 0,$$

where the norm  $\|\cdot\|_{H^{\bar{n}}(\rho_{B^d})}$  is defined in Section 2.2. The decay for  $K(f^*, \lambda)$  has recently been discussed in [36].

We then have the following Corollary 2.1.

**Corollary 2.1.** Let Assumptions A and B hold and let  $f_{z,\lambda}$  be the solution of (17) and  $f^* \in C(B^d)$ . If  $g(x) = \frac{\partial f^*(x)}{\partial \bar{n}}$  for  $x \in S^{d-1}$ . Then, for any  $\delta \in (0,1)$ , with confidence  $1 - \delta$ , holds:

$$\begin{aligned} & \|f_{z,\lambda} - f^*\|_{H^{\bar{n}}(\rho_{B^d})} \\ &= O \left( \frac{\sigma}{\lambda\sqrt{m\delta}} + \left( \frac{\sqrt{K(f^*, \lambda)}}{\lambda^{\frac{3}{2}}\sqrt{m}} + \frac{\|f^*\|_{C(B^d)}}{\lambda\sqrt{m}} \right) \log \frac{4}{\delta} + \frac{\sqrt{K(f^*, g, \lambda)}}{\lambda\sqrt{l\delta}} + \sqrt{K(f^*, \lambda)} \right). \end{aligned} \tag{25}$$

By (25), we know if  $\lambda = \lambda(m) \downarrow_{0^+}$  ( $m \rightarrow +\infty$ ) is chosen in such a way that

$\lim_{\lambda \rightarrow 0^+} K(f^*, \lambda) = 0$ ,  $\lambda^{\frac{3}{2}}\sqrt{m} \rightarrow +\infty$  when  $m \rightarrow +\infty$ , then, with confidence  $1 - \delta$ , holds (since  $m \sim l$ ):

$$\lim_{m \rightarrow +\infty} \|f_{z,\lambda} - f^*\|_{H^{\bar{n}}(\rho_{B^d})} = 0. \tag{26}$$

### 2.5. Further Discussions

We now give some explanation on the results and the assumptions.

- 1) There are three reasons encouraging us to choose  $D = B^d$  and  $\partial D = S^{d-1}$

to show the kernel-regularized regression model for solving the Neumann boundary problem (5).

i) When  $D = B^d$  and  $\partial D = S^{d-1}$ , we can easily give explicit representation for outward normal derivatives (see (9)). Therefore, we can extend the method used in this paper to the domains whose outward normal derivatives may be computed easily.

ii) By the statement in Section 2.1, we know a tool for us to construct a reproducing kernel Hilbert space is the orthonormal basis. We notice that function orthogonal basis theory has been developed not only on the unit ball  $B^d$  (see [37]) and the unit sphere  $S^{d-1}$  (see [29]), but also on the some Sobolev space associated with both  $B^d$  and  $S^{d-1}$  (see e.g. [30]). These facts also encourage us to choose  $D = B^d$  and  $\partial D = S^{d-1}$  for problem (5).

iii) In learning theory, the learning rate estimate for the kernel-regularized learning algorithm are sum up to bound the error bounds, which belong to the scope of approximation theory. One can make use of the rich spherical approximation theory skills to bound the learning rates (see [24]).

2) In the present paper, we give two Assumptions A and B. They are reasonable.

Since  $p_k(x, y)$  is a bivariate polynomial on  $B^d \times B^d$  for a given  $k$ ,

$\frac{\partial_x p_k(x, x)}{\partial \vec{n}}$  is still a polynomial whose *sup* can be attained on  $S^{d-1}$ . The Assumption B is reasonable. By the same way, we know  $d_k = \sup_{x, y \in B^d} \partial^\alpha p_k(x, y)$  can

be attained. If we choose  $\lambda_l = e^{-k^\beta}$  ( $\beta > 0$ ), then, we can have

$K^\mu \in C^{(1)}(B^d \times B^d)$ . So, the Assumption A is reasonable as well.

3) The convergence in the present paper are uniform convergence (see (26)), which admits the reliability for the proposed method.

### 3. Lemmas

To give the feature description of the optimal solutions of Algorithms (17)-(19), we need the concept of Gâteaux derivative.

Let  $(\mathcal{H}, \|\cdot\|_{\mathcal{H}})$  be a Hilbert space,  $F(f) : \mathcal{H} \rightarrow R \cup \{\mp\infty\}$  be a real function. We say  $F$  is Gâteaux differentiable at  $f \in \mathcal{H}$  if there is a  $\xi \in \mathcal{H}$  such that for any  $g \in \mathcal{H}$  there holds:

$$\lim_{t \rightarrow 0} \frac{F(f + tg) - F(f)}{t} = \langle g, \xi \rangle_H$$

and write  $\nabla_f F(f) = \xi$  as the Gâteaux derivative of  $F(f)$  at  $f$ .

To prove Theorem 2.1, we need some lemmas.

**Lemma 3.1.** There hold the following equations:

1)  $f_{z,\lambda}$  satisfies equation:

$$\begin{aligned} & \frac{1}{m} \sum_{i=1}^m (f_{z,\lambda}(x_i) - y_{x_i}) K_{x_i}^\mu(\cdot) + \frac{1}{l} \sum_{k=1}^l \left( \frac{\partial f_{z,\lambda}(x_{m+k})}{\partial \vec{n}} - g(x_{m+k}) \right) \frac{\partial K_{x_{m+k}}^\mu(\cdot)}{\partial \vec{n}} \\ & + \lambda f_{z,\lambda}(\cdot) = 0. \end{aligned} \tag{27}$$

2)  $f_{\bar{x},\lambda}$  satisfies equation:

$$\frac{1}{m} \sum_{i=1}^m (f_{\bar{x},\lambda}(x_i) - f^*)(x_i) K_{x_i}^\mu(\cdot) + \frac{1}{l} \sum_{k=1}^l \left( \frac{\partial f_{\bar{x},\lambda}(x_{m+k})}{\partial \bar{n}} - g(x_{m+k}) \right) \frac{\partial K_{x_{m+k}}^\mu(\cdot)}{\partial \bar{n}} \quad (28)$$

$$+ \lambda f_{\bar{x},\lambda}(\cdot) = 0.$$

3)  $f_\lambda$  satisfies equation:

$$\int_{B^d} (f_\lambda(x) - f^*(x)) K_x^\mu(\cdot) d\rho_{B^d} + \int_{S^{d-1}} \left( \frac{\partial f_\lambda(x)}{\partial \bar{n}} - g(x) \right) \frac{\partial K_x^\mu(\cdot)}{\partial \bar{n}} d\rho_{S^{d-1}} \quad (29)$$

$$+ \lambda f_\lambda(\cdot) = 0.$$

*Proof. Proof of 1).* Define  $\Omega_z(f)$  as:

$$\Omega_z(f) = \mathcal{E}_z(f) + \frac{1}{l} \sum_{k=1}^l \left( \frac{\partial f(x_{k+m})}{\partial \bar{n}} - g(x_{k+m}) \right)^2 + \lambda \|f\|_{K^\mu}^2, \quad f \in \mathcal{H}_{K^\mu}^{\bar{n}}.$$

Then,

$$\lim_{t \rightarrow 0^+} \frac{\Omega_z(f+th) - \Omega_z(f)}{t}$$

$$= \lim_{t \rightarrow 0^+} \frac{\mathcal{E}_z(f+th) - \mathcal{E}_z(f)}{t} + 2\lambda \langle h, f \rangle_{K^\mu},$$

$$+ \frac{1}{l} \sum_{k=1}^l \lim_{t \rightarrow 0^+} \frac{\left( \frac{\partial f(x_{k+m})}{\partial \bar{n}} + t \frac{\partial h(x_{k+m})}{\partial \bar{n}} - g(x_{k+m}) \right)^2 - \left( \frac{\partial f(x_{k+m})}{\partial \bar{n}} - g(x_{k+m}) \right)^2}{t},$$

where

$$\lim_{t \rightarrow 0^+} \frac{\mathcal{E}_z(f+th) - \mathcal{E}_z(f)}{t} = \frac{1}{m} \sum_{i=1}^m \lim_{t \rightarrow 0^+} \frac{(f(x_i) + th(x_i) - y_{x_i})^2 - (f(x_i) - y_{x_i})^2}{t}$$

$$= \frac{2}{m} \sum_{i=1}^m (f(x_i) - y_{x_i}) h(x_i)$$

$$= \frac{2}{m} \sum_{i=1}^m (f(x_i) - y_{x_i}) \langle h, K_{x_i}^\mu(\cdot) \rangle_{K^\mu}$$

$$= \left\langle h, \frac{2}{m} \sum_{i=1}^m (f(x_i) - y_{x_i}) K_{x_i}^\mu(\cdot) \right\rangle_{K^\mu}$$

and

$$\frac{1}{l} \sum_{k=1}^l \lim_{t \rightarrow 0^+} \frac{\left( \frac{\partial f(x_{k+m})}{\partial \bar{n}} + t \frac{\partial h(x_{k+m})}{\partial \bar{n}} - g(x_{k+m}) \right)^2 - \left( \frac{\partial f(x_{k+m})}{\partial \bar{n}} - g(x_{k+m}) \right)^2}{t} \quad (30)$$

$$= \left\langle h, \frac{2}{l} \sum_{k=1}^l \left( \frac{\partial f(x_{k+m})}{\partial \bar{n}} - g(x_{k+m}) \right) \frac{\partial K_{x_{k+m}}^\mu(\cdot)}{\partial \bar{n}} \right\rangle_{K^\mu}.$$

It follows

$$\lim_{t \rightarrow 0^+} \frac{\Omega_z(f+th) - \Omega_z(f)}{t}$$

$$= \left\langle h, \frac{2}{m} \sum_{i=1}^m (f(x_i) - y_{x_i}) K_{x_i}^\mu(\cdot) + \frac{2}{l} \sum_{k=1}^l \left( \frac{\partial f(x_{k+m})}{\partial \bar{n}} - g(x_{k+m}) \right) \frac{\partial K_{x_{k+m}}^\mu(\cdot)}{\partial \bar{n}} + 2\lambda f \right\rangle_{K^\mu}.$$

By the definition of Gâteaux derivative, we have:

$$\begin{aligned} \nabla_f \Omega_z(f) &= \frac{2}{m} \sum_{i=1}^m (f(x_i) - y_{x_i}) K_{x_i}^\mu(\cdot) \\ &\quad + \frac{2}{l} \sum_{k=1}^l \left( \frac{\partial f(x_{k+m})}{\partial \bar{n}} - g(x_{k+m}) \right) \frac{\partial K_{x_{k+m}}^\mu(\cdot)}{\partial \bar{n}} + 2\lambda f(\cdot). \end{aligned}$$

By Fermat’s rule (see 1) in Proposition A1) and the definition of  $f_{z,\lambda}$ , we have  $\nabla_f \Omega_z(f)|_{f=f_{z,\lambda}} = 0$ , i.e. (27) holds.

**Lemma 3.2.** *There hold the inequality:*

$$\|f_{z,\lambda} - f_{\bar{x},\lambda}\|_{K^\mu} \leq \frac{2}{\lambda m} \left\| \sum_{i=1}^m \eta_{x_i}(y) K_{x_i}^\mu(\cdot) \right\|_{K^\mu} \tag{31}$$

and the inequality:

$$\begin{aligned} &\|f_{\bar{x},\lambda} - f_{\rho,\lambda}\|_{K^\mu} \\ &\leq \frac{2}{\lambda} \left( \left\| \int_{B^d} (f_\lambda(x) - f^*(x)) K_x^\mu(\cdot) d\rho_{B^d} - \frac{1}{m} \sum_{i=1}^m (f_\lambda(x_i) - f^*(x_i)) K_{x_i}^\mu(\cdot) \right\|_{K^\mu} \right. \\ &\quad \left. + \left\| \int_{S^{d-1}} \left( \frac{\partial f(x)}{\partial \bar{n}} - g(x) \right) K_x^\mu(\cdot) d\rho_{S^{d-1}} - \frac{1}{l} \sum_{k=1}^l \left( \frac{\partial f(x_{k+m})}{\partial \bar{n}} - g(x_{k+m}) \right) K_{x_{k+m}}^\mu(\cdot) \right\|_{K^\mu} \right). \end{aligned} \tag{32}$$

*Proof.* The definition of  $f_{z,\lambda}$  and the inequality (43) give:

$$\begin{aligned} 0 &\geq \Omega_z(f_{z,\lambda}) - \Omega_z(f_{\bar{x},\lambda}) \\ &\geq \frac{2}{m} \sum_{i=1}^m (f_{\bar{x},\lambda}(x_i) - y_{x_i}) (f_{z,\lambda}(x_i) - f_{\bar{x},\lambda}(x_i)) \\ &\quad + \frac{2}{l} \sum_{k=1}^l \left( \frac{\partial f_{\bar{x},\lambda}(x_{k+m})}{\partial \bar{n}} - g(x_{k+m}) \right) \left( \frac{\partial f_{z,\lambda}(x_{k+m})}{\partial \bar{n}} - \frac{\partial f_{\bar{x},\lambda}(x_{k+m})}{\bar{n}} \right) \\ &\quad + \lambda \left( \|f_{z,\lambda}\|_{K^\mu}^2 - \|f_{\bar{x},\lambda}\|_{K^\mu}^2 \right) \\ &\geq \left\langle f_{z,\lambda} - f_{\bar{x},\lambda}, \frac{2}{m} \sum_{i=1}^m (f_{\bar{x},\lambda}(x_i) - y_{x_i}) K_{x_i}^\mu(\cdot) \right\rangle \\ &\quad + \frac{2}{l} \sum_{k=1}^l \left\langle \left( \frac{\partial f_{\bar{x},\lambda}(x_{k+m})}{\partial \bar{n}} - g(x_{k+m}) \right) \frac{\partial K_{x_{k+m}}^\mu(\cdot)}{\partial \bar{n}} \right\rangle_{K^\mu} \\ &\quad + \left\langle f_{z,\lambda} - f_{\bar{x},\lambda}, 2\lambda f_{\bar{x},\lambda} \right\rangle_{K^\mu} + \lambda \|f_{z,\lambda} - f_{\bar{x},\lambda}\|_{K^\mu}^2, \end{aligned}$$

where we have used (44). Since (28) and  $y_{x_i} = f^*(x_i) + \eta_{x_i}$ , we have:

$$\begin{aligned} 0 &\geq \Omega_z(f_{z,\lambda}) - \Omega_z(f_{\bar{x},\lambda}) \\ &\geq \left\langle f_{z,\lambda} - f_{\bar{x},\lambda}, \frac{2}{m} \sum_{i=1}^m \eta_{x_i}(y) K_{x_i}^\mu(\cdot) \right\rangle_{K^\mu} + \lambda \|f_{z,\lambda} - f_{\bar{x},\lambda}\|_{K^\mu}^2. \end{aligned}$$

It follows:

$$\begin{aligned} \lambda \|f_{z,\lambda} - f_{\bar{x},\lambda}\|_{K^\mu}^2 &\leq \left\langle f_{\bar{x},\lambda} - f_{z,\lambda}, \frac{2}{m} \sum_{i=1}^m \eta_{x_i}(y) K_{x_i}^\mu(\cdot) \right\rangle_{K^\mu} \\ &\leq \left\| \frac{2}{m} \sum_{i=1}^m \eta_{x_i}(y) K_{x_i}^\mu(\cdot) \right\|_{K^\mu} \times \|f_{\bar{x},\lambda} - f_{z,\lambda}\|_{K^\mu}. \end{aligned}$$

(31) thus holds.

*Proof of (32).* Define  $\Omega_{\bar{x}}(f)$  as:

$$\Omega_{\bar{x}}(f) = \mathcal{E}_{\bar{x}}(f) + \frac{1}{l} \sum_{k=1}^l \left( \frac{\partial f(x_{m+k})}{\partial \bar{n}} - g(x_{m+k}) \right)^2 + \lambda \|f\|_{K^\mu}^2, \quad f \in \mathcal{H}_{K^\mu}^{\bar{n}}.$$

Then, by the definition of  $f_{\bar{x},\lambda}$  and inequalities (43) and (44), we have:

$$\begin{aligned} 0 &\geq \Omega_{\bar{x}}(f_{\bar{x},\lambda}) - \Omega_{\bar{x}}(f_\lambda) \\ &\geq \left\langle f_{\bar{x},\lambda} - f_\lambda, \frac{2}{m} \sum_{i=1}^m (f_\lambda(x_i) - f^*(x_i)) K_{x_i}^\mu(\cdot) \right\rangle_{K^\mu} \\ &\quad + \left\langle f_{\bar{x},\lambda} - f_\lambda, \frac{2}{l} \sum_{k=1}^l \left( \frac{\partial f_\lambda(x_{m+k})}{\partial \bar{n}} - g(x_{m+k}) \right) \frac{\partial K_{x_{m+k}}^\mu(\cdot)}{\partial \bar{n}} \right\rangle_{K^\mu} \\ &\quad + \left\langle f_{\bar{x},\lambda} - f_\lambda, 2\lambda f_\lambda \right\rangle_{K^\mu} + \lambda \|f_{\bar{x},\lambda} - f_\lambda\|_{K^\mu}^2. \end{aligned}$$

By (29), we have:

$$\begin{aligned} 0 &\geq \left\langle f_{\bar{x},\lambda} - f_\lambda, \int_{B^d} (f_\lambda(x) - f^*(x)) K_x^\mu(\cdot) d\rho_{B^d} \right. \\ &\quad - \frac{1}{m} \sum_{i=1}^m (f_\lambda(x_i) - f^*(x_i)) K_{x_i}^\mu(\cdot) + \int_{S^{d-1}} \left( \frac{\partial f_\lambda(x)}{\partial \bar{n}} - g(x) \right) \frac{\partial K_x^\mu(\cdot)}{\partial \bar{n}} d\rho_{S^{d-1}} \\ &\quad \left. - \frac{1}{l} \sum_{k=1}^l \left( \frac{\partial f_\lambda(x_{m+k})}{\partial \bar{n}} - g(x_{m+k}) \right) \frac{\partial K_{x_{m+k}}^\mu(\cdot)}{\partial \bar{n}} \right\rangle_{K^\mu} + \lambda \|f_{\bar{x},\lambda} - f_\lambda\|_{K^\mu}^2. \end{aligned}$$

It follows:

$$\begin{aligned} &\lambda \|f_{\bar{x},\lambda} - f_\lambda\|_{K^\mu}^2 \\ &\leq \left\langle f_{\bar{x},\lambda} - f_\lambda, \frac{1}{m} \sum_{i=1}^m (f_\lambda(x_i) - f^*(x_i)) K_{x_i}^\mu(\cdot) - \int_{B^d} (f_\lambda(x) - f^*(x)) K_x^\mu(\cdot) d\rho_{B^d} \right. \\ &\quad - \int_{S^{d-1}} \left( \frac{\partial f_\lambda(x)}{\partial \bar{n}} - g(x) \right) \frac{\partial K_x^\mu(\cdot)}{\partial \bar{n}} d\rho_{S^{d-1}} \\ &\quad \left. + \frac{1}{l} \sum_{k=1}^l \left( \frac{\partial f_\lambda(x_{m+k})}{\partial \bar{n}} - g(x_{m+k}) \right) \frac{\partial K_{x_{m+k}}^\mu(\cdot)}{\partial \bar{n}} \right\rangle_{K^\mu} \\ &\leq \left( \left\| \frac{1}{m} \sum_{i=1}^m (f_\lambda(x_i) - f^*(x_i)) K_{x_i}^\mu(\cdot) - \int_{B^d} (f_\lambda(x) - f^*(x)) K_x^\mu(\cdot) d\rho_{B^d} \right\|_{K^\mu} \right. \\ &\quad \times \|f_{\bar{x},\lambda} - f_\lambda\|_{K^\mu} + \left\| \int_{S^{d-1}} \left( \frac{\partial f_\lambda(x)}{\partial \bar{n}} - g(x) \right) \frac{\partial K_x^\mu(\cdot)}{\partial \bar{n}} d\rho_{S^{d-1}} \right. \\ &\quad \left. - \frac{1}{l} \sum_{k=1}^l \left( \frac{\partial f_\lambda(x_{m+k})}{\partial \bar{n}} - g(x_{m+k}) \right) \frac{\partial K_{x_{m+k}}^\mu(\cdot)}{\partial \bar{n}} \right\|_{K^\mu} \Big). \end{aligned}$$

Above inequality gives (32).

**Lemma 3.3.** *There hold following inequalities.*

1) For any  $\delta \in (0,1)$ , with confidence  $1 - \delta$ , holds:

$$\|f_{z,\lambda} - f_{\bar{x},\lambda}\|_{K^\mu} \leq \frac{2k\sigma}{\lambda\sqrt{m\delta}}. \tag{33}$$

2) For any  $\delta \in (0,1)$ , with confidence  $1-\delta$ , holds:

$$\|f_{\bar{x},\lambda} - f_{\lambda}\|_{K^{\mu}} = \left( \frac{2k^2 \sqrt{K(f^*, g, \lambda)}}{\lambda \sqrt{m\lambda}} + \frac{2k \|f^*\|_{C(B^d)}}{\lambda \sqrt{m}} \right) \log \frac{2}{\delta} + \frac{k \sqrt{K(f^*, g, \lambda)}}{\sqrt{l\delta}}. \quad (34)$$

*Proof. Proof of (33).* The definition of  $\|\cdot\|_{K^{\mu}(B^d)}$  and (14) give:

$$\begin{aligned} \left\| \frac{1}{m} \sum_{i=1}^m \eta_{x_i}(y) K_{x_i}^{\mu}(\cdot) \right\|_{K^{\mu}}^2 &= \left\langle \frac{1}{m} \sum_{i=1}^m \eta_{x_i}(y) K_{x_i}^{\mu}(\cdot), \frac{1}{m} \sum_{j=1}^m \eta_{x_j}(y) K_{x_j}^{\mu}(\cdot) \right\rangle_{K^{\mu}} \\ &= \frac{1}{m^2} \sum_{i,j=1}^m \eta_{x_i}(y) \eta_{x_j}(y) K^{\mu}(x_i, x_j). \end{aligned}$$

By Markov inequality, we have:

$$P\left(\|f_{z,\lambda} - f_{\bar{x},\lambda}\|_{K^{\mu}} > \varepsilon\right) \leq \frac{E\left(\|f_{z,\lambda} - f_{\bar{x},\lambda}\|_{K^{\mu}}^2\right)}{\varepsilon^2}. \quad (35)$$

Then, by (31), we have:

$$\begin{aligned} \|f_{z,\lambda} - f_{\bar{x},\lambda}\|_{K^{\mu}}^2 &\leq \frac{4}{\lambda^2 m^2} \left\| \sum_{i=1}^m \eta_{x_i}(y) K_{x_i}^{\mu}(\cdot) \right\|_{K^{\mu}}^2 \\ &\leq \frac{4}{\lambda^2 m^2} \left\langle \sum_{i=1}^m \eta_{x_i}(y) K_{x_i}^{\mu}(\cdot), \sum_{j=1}^m \eta_{x_j}(y) K_{x_j}^{\mu}(\cdot) \right\rangle_{K^{\mu}} \\ &= \frac{4}{\lambda^2 m^2} \sum_{i,j=1}^m \eta_{x_i}(y) \eta_{x_j}(y) K^{\mu}(x_i, x_j). \end{aligned}$$

Since  $E_{\rho(\cdot|x)}(\eta_x) = 0$ , we have:

$$E\left(\|f_{z,\lambda} - f_{\bar{x},\lambda}\|_{K^{\mu}}^2\right) \leq \frac{4k^2}{\lambda^2 m} \int_{B^d} \eta_x^2 d\rho_{B^d} = \frac{4k^2 \sigma^2}{\lambda^2 m}.$$

By (35), we have:

$$P\left(\|f_{z,\lambda} - f_{\bar{x},\lambda}\|_{K^{\mu}} \leq \varepsilon\right) \geq 1 - \frac{4k^2 \sigma^2}{\lambda^2 m \varepsilon^2}.$$

Taking  $\delta = \frac{4k^2 \sigma^2}{\lambda^2 m \varepsilon^2}$ . We have  $\varepsilon = \frac{2k\sigma}{\lambda \sqrt{m\delta}}$  and (33) thus holds.

We now show (34). Take

$$A = \left\| \int_{B^d} (f_{\lambda}(x) - f^*(x)) K_x^{\mu}(\cdot) d\rho_{B^d} - \frac{1}{m} \sum_{i=1}^m (f_{\lambda}(x_i) - f^*(x_i)) K_{x_i}^{\mu}(\cdot) \right\|_{K^{\mu}}$$

and

$$\begin{aligned} B &= \left\| \int_{S^{d-1}} \left( \frac{\partial f_{\lambda}(x)}{\partial \bar{n}} - g(x) \right) \frac{\partial K_x^{\mu}(\cdot)}{\partial \bar{n}} d\rho_{S^{d-1}} \right. \\ &\quad \left. - \frac{1}{l} \sum_{k=1}^l \left( \frac{\partial f_{\lambda}(x_{m+k})}{\partial \bar{n}} - g(x_{m+k}) \right) \frac{\partial K_{x_{m+k}}^{\mu}(\cdot)}{\partial \bar{n}} \right\|_{K^{\mu}}. \end{aligned}$$

Then,

$$\|f_{\bar{x},\lambda} - f_{\rho,\lambda}\|_{K^{\mu}} \leq \frac{2}{\lambda} (A + B), \quad (36)$$



where

$$A \leq \left\| \int_{B^d} f_\lambda(x) K_x^\mu(\cdot) d\rho_{B^d} - \frac{1}{m} \sum_{i=1}^m f_\lambda(x_i) K_{x_i}^\mu(\cdot) \right\|_{K^\mu} + \left\| \int_{B^d} f^*(x) K_x^\mu(\cdot) d\rho_{B^d} - \frac{1}{m} \sum_{i=1}^m f^*(x_i) K_{x_i}^\mu(\cdot) \right\|_{K^\mu}.$$

Since

$$\begin{aligned} \|f_\lambda(x) K_x^\mu(\cdot)\|_{K^\mu} &= |f_\lambda(x)| \sqrt{K_x^\mu(x)} \leq k |f_\lambda(x)| \\ &\leq k^2 \|f\|_{K^\mu} \leq \frac{k^2 \sqrt{K(f^*, g, \lambda)}}{\sqrt{\lambda}} \end{aligned}$$

and  $\|f^*(x) K_x^\mu(\cdot)\|_{K^\mu} \leq k |f^*(x)| \leq k \|f^*\|_{C(B^d)}$ , we have by (47) that, with confidence  $1 - \delta$ , holds:

$$A \leq \left( \frac{2k^2 \sqrt{K(f^*, g, \lambda)}}{\lambda \sqrt{m\lambda}} + \frac{2k \|f^*\|_{C(B^d)}}{\lambda \sqrt{m}} \right) \log \frac{2}{\delta}. \tag{37}$$

On the other hand, take  $\xi(x) = \left( \frac{\partial f_\lambda(x)}{\partial \bar{n}} - g(x) \right)$ . Then,

$$B = \left\| \int_{S^{d-1}} \xi(x) \frac{\partial K_x^\mu(\cdot)}{\partial \bar{n}} d\rho_{S^{d-1}} - \frac{1}{l} \sum_{k=1}^l \xi(x_{m+k}) \frac{\partial K_{x_{m+k}}^\mu(\cdot)}{\partial \bar{n}} \right\|_{K^\mu}.$$

By the definition of  $\|\cdot\|_{K^\mu}$ , we have:

$$\begin{aligned} B^2 &= \left\langle \int_{S^{d-1}} \xi(x) \frac{\partial K_x^\mu(\cdot)}{\partial \bar{n}} d\rho_{S^{d-1}} - \frac{1}{l} \sum_{k=1}^l \xi(x_{m+k}) \frac{\partial K_{x_{m+k}}^\mu(\cdot)}{\partial \bar{n}}, \right. \\ &\quad \left. \int_{S^{d-1}} \xi(u) \frac{\partial K_u^\mu(\cdot)}{\partial \bar{n}} d\rho_{S^{d-1}} - \frac{1}{l} \sum_{i=1}^l \xi(x_{m+i}) \frac{\partial K_{x_{m+i}}^\mu(\cdot)}{\partial \bar{n}} \right\rangle_{K^\mu} \\ &= \int_{S^{d-1}} \int_{S^{d-1}} \xi(x) \xi(u) \left\langle \frac{\partial K_x^\mu(\cdot)}{\partial \bar{n}}, \frac{\partial K_u^\mu(\cdot)}{\partial \bar{n}} \right\rangle_{K^\mu} d\rho_{S^{d-1}}(x) d\rho_{S^{d-1}}(u) \\ &\quad - 2 \int_{S^{d-1}} \xi(x) \left( \frac{1}{l} \sum_{i=1}^l \xi(x_{m+i}) \left\langle \frac{\partial K_x^\mu(\cdot)}{\partial \bar{n}}, \frac{\partial K_{x_{m+i}}^\mu(\cdot)}{\partial \bar{n}} \right\rangle_{K^\mu} \right) d\rho_{S^{d-1}}(x) \\ &\quad + \frac{1}{l^2} \sum_{k=1}^l \xi^2(x_{m+k}) \left\langle \frac{\partial K_{x_{m+k}}^\mu(\cdot)}{\partial \bar{n}}, \frac{\partial K_{x_{m+k}}^\mu(\cdot)}{\partial \bar{n}} \right\rangle_{K^\mu} \\ &\quad + \frac{1}{l^2} \sum_{k,i=1, k \neq i}^l \xi(x_{m+i}) \xi(x_{m+j}) \left\langle \frac{\partial K_{x_{m+k}}^\mu(\cdot)}{\partial \bar{n}}, \frac{\partial K_{x_{m+i}}^\mu(\cdot)}{\partial \bar{n}} \right\rangle_{K^\mu}. \end{aligned}$$

Since (14), we have:

$$\begin{aligned} \left\langle \frac{\partial K_x^\mu(\cdot)}{\partial \bar{n}}, \frac{\partial K_u^\mu(\cdot)}{\partial \bar{n}} \right\rangle_{K^\mu} &= \frac{\partial_u}{\partial \bar{n}} \frac{\partial_x}{\partial \bar{n}} K_x^\mu(u), \\ \left\langle \frac{\partial K_x^\mu(\cdot)}{\partial \bar{n}}, \frac{\partial K_{x_{m+i}}^\mu(\cdot)}{\partial \bar{n}} \right\rangle_{K^\mu} &= \frac{\partial_u}{\partial \bar{n}} \frac{\partial_x}{\partial \bar{n}} K_x^\mu(u) \Big|_{u=x_{m+i}} \end{aligned}$$

and

$$\left\langle \frac{\partial K_{x_{m+k}}^\mu(\cdot)}{\partial \bar{n}}, \frac{\partial K_{x_{m+k}}^\mu(\cdot)}{\partial \bar{n}} \right\rangle_{K^\mu} = \frac{\partial_u}{\partial \bar{n}} \frac{\partial_x}{\partial \bar{n}} K_u^\mu(u) \Big|_{u=x_{m+k}},$$

$$\left\langle \frac{\partial K_{x_{m+k}}^\mu(\cdot)}{\partial \bar{n}}, \frac{\partial K_{x_{m+i}}^\mu(\cdot)}{\partial \bar{n}} \right\rangle_{K^\mu} = \frac{\partial_u}{\partial \bar{n}} \frac{\partial_x}{\partial \bar{n}} K_x^\mu(u) \Big|_{u=x_{m+i}, x=x_{m+k}}.$$

It follows that:

$$\begin{aligned} B^2 &= \int_{S^{d-1}} \int_{S^{d-1}} \xi(x) \xi(u) \frac{\partial_u}{\partial \bar{n}} \frac{\partial_x}{\partial \bar{n}} K_x^\mu(u) d\rho_{S^{d-1}}(x) d\rho_{S^{d-1}}(u) \\ &\quad - 2 \int_{S^{d-1}} \xi(x) \left( \frac{1}{l} \sum_{i=1}^l \xi(x_{m+i}) \frac{\partial_u}{\partial \bar{n}} \frac{\partial_x}{\partial \bar{n}} K_x^\mu(u) \Big|_{u=x_{m+i}} \right) d\rho_{S^{d-1}}(x) \\ &\quad + \frac{1}{l^2} \sum_{k=1}^l \xi^2(x_{m+k}) \frac{\partial_u}{\partial \bar{n}} \frac{\partial_x}{\partial \bar{n}} K_u^\mu(u) \Big|_{u=x_{m+k}} \\ &\quad + \frac{1}{l^2} \sum_{k,i=1, k \neq i}^l \xi(x_{m+i}) \xi(x_{m+j}) \frac{\partial_u}{\partial \bar{n}} \frac{\partial_x}{\partial \bar{n}} K_x^\mu(u) \Big|_{u=x_{m+i}, x=x_{m+k}}. \end{aligned}$$

Since  $(x_{m+1}, x_{m+2}, \dots, x_{m+l})$  are i.i.d. according to  $\rho_{S^{d-1}}$ , we have:

$$\begin{aligned} E(B^2) &= \frac{1}{l} \left( \int_{S^{d-1}} \xi^2(x) \frac{\partial_x}{\partial \bar{n}} \frac{\partial_x}{\partial \bar{n}} K_x^\mu(x) d\rho_{S^{d-1}}(x) \right. \\ &\quad \left. - \int_{S^{d-1}} \int_{S^{d-1}} \xi(x) \xi(u) \frac{\partial_u}{\partial \bar{n}} \frac{\partial_x}{\partial \bar{n}} K_x^\mu(u) d\rho_{S^{d-1}}(x) d\rho_{S^{d-1}}(u) \right) \\ &\leq \frac{1}{l} \left( \int_{S^{d-1}} \xi^2(x) \frac{\partial_x}{\partial \bar{n}} \frac{\partial_x}{\partial \bar{n}} K_x^\mu(x) d\rho_{S^{d-1}}(x) \right), \end{aligned}$$

where we have used the fact that:

$$\int_{S^{d-1}} \int_{S^{d-1}} \xi(x) \xi(u) \frac{\partial_u}{\partial \bar{n}} \frac{\partial_x}{\partial \bar{n}} K_x^\mu(u) d\rho_{S^{d-1}}(x) d\rho_{S^{d-1}}(u) \geq 0$$

since  $\frac{\partial_u}{\partial \bar{n}} \frac{\partial_x}{\partial \bar{n}} K_x^\mu(u)$  is a positive definition function about  $u$  and  $x$ . According to Assumption B, we have:

$$\left| \frac{\partial_u}{\partial \bar{n}} \frac{\partial_x}{\partial \bar{n}} K_x^\mu(u) \right| \leq k^2,$$

It follows that:

$$E(B^2) \leq \frac{k^2}{l} \left( \int_{S^{d-1}} \xi^2(x) d\rho_{S^{d-1}}(x) \right).$$

By Markov inequality, we have:

$$\begin{aligned} P(B > \varepsilon) &\leq \frac{E(B^2)}{\varepsilon^2} \\ &\leq \frac{k^2}{l\varepsilon^2} \left( \int_{S^{d-1}} \xi^2(x) d\rho_{S^{d-1}}(x) \right) \\ &\leq \frac{k^2}{l\varepsilon^2} K(f^*, g, \lambda). \end{aligned}$$

Take  $\delta = \frac{k^2}{l\varepsilon^2} K(f^*, g, \lambda)$ . Then,  $\varepsilon = \frac{k\sqrt{K(f^*, g, \lambda)}}{\sqrt{l\delta}}$ . It follows that with confidence  $1 - \delta$ :

$$B \leq \frac{k\sqrt{K(f^*, g, \lambda)}}{\sqrt{l\delta}}. \tag{38}$$

Collecting (38), (37) and (36), we arrive at (34).

### 4. Proofs

*Proof of Proposition 2.1. Proof of 1).* For any

$f(x) = \sum_{k=0}^{\infty} \sum_{l=1}^{a_k^d} a_{k,l}(f) Q_{k,l}(x) \in \mathcal{H}_{K^\mu}^{\bar{n}}$ . We rewrite  $K^\mu(x, y)$  as:

$$K^\mu(x, y) = K_y^\mu(x) = \sum_{k=0}^{\infty} \sum_{l=1}^{a_k^d} (\lambda_{k,l} Q_{k,l}(y)) Q_{k,l}(x).$$

Then,

$$\langle f, K_y^\mu \rangle_{K^\mu} = \sum_{k=0}^{\infty} \sum_{l=1}^{a_k^d} \frac{(\lambda_{k,l} Q_{k,l}(y)) a_{k,l}(f)}{\lambda_{k,l}} = f(y),$$

which yields (13).

*Proof of 2).* Since (9) and (46), we have:

$$\begin{aligned} \frac{\partial f(x)}{\partial \bar{n}} &= \left\langle f, \sum_{i=1}^d x^i \frac{\partial K(x, \cdot)}{\partial x^i}(x) \right\rangle_{K^\mu} \\ &= \left\langle f, \frac{\partial}{\partial \bar{n}} K(x, \cdot) \right\rangle_{K^\mu}, \quad x = (x^1, \dots, x^d) \in S^{d-1}. \end{aligned}$$

(14) thus holds.

*Proof of 3).* By (14), we have:

$$\left| \frac{\partial f(x)}{\partial \bar{n}} \right| = \left| \left\langle f, \frac{\partial K_x^\mu(\cdot)}{\partial \bar{n}} \right\rangle_{K^\mu} \right| \leq \|f\|_{K^\mu} \left\| \frac{\partial K_x^\mu(\cdot)}{\partial \bar{n}} \right\|_{K^\mu} \leq k \|f\|_{K^\mu}.$$

By the same method, we have by (13) that:

$$\begin{aligned} |f(x)| &= \left| \left\langle f, K_x^\mu(\cdot) \right\rangle_{K^\mu} \right| \\ &\leq \|f\|_{K^\mu} \|K_x^\mu(\cdot)\|_{K^\mu} \\ &\leq k \|f\|_{K^\mu}. \end{aligned}$$

(15) thus holds.

*Proof of Proposition 2.2.* By (27), we have:

$$\begin{aligned} f_{z,\lambda}(\cdot) &= \frac{1}{\lambda m} \sum_{i=1}^m (y_{x_i} - f_{z,\lambda}(x_i)) K_{x_i}^\mu(\cdot) \\ &\quad + \frac{1}{\lambda l} \sum_{k=1}^l \left( g(x_{m+k}) - \frac{\partial f_{z,\lambda}(x_{m+k})}{\partial \bar{n}} \right) \frac{\partial K_{x_{m+k}}^\mu(\cdot)}{\partial \bar{n}}. \end{aligned} \tag{39}$$

Taking  $a_i = \frac{1}{\lambda m} (y_{x_i} - f_{z,\lambda}(x_i))$  and  $a_{m+k} = \frac{1}{\lambda l} \left( g(x_{m+k}) - \frac{\partial f_{z,\lambda}(x_{m+k})}{\partial \bar{n}} \right)$  into (40), we have (20).

By (29), we have:

$$f_{\bar{x},\lambda}(\cdot) = \frac{1}{\lambda m} \sum_{i=1}^m (f^*(x_i) - f_{\bar{x},\lambda}(x_i)) K_{x_i}^\mu(\cdot) + \frac{1}{\lambda l} \sum_{k=1}^l \left( g(x_{m+k}) - \frac{\partial f_{\bar{x},\lambda}(x_{m+k})}{\bar{n}} \right) \frac{\partial K_{x_{m+k}}^\mu(\cdot)}{\partial \bar{n}}. \tag{40}$$

Taking  $b_i = \frac{1}{\lambda m} (f^*(x_i) - f_{\bar{x},\lambda}(x_i))$  and  $b_{m+k} = \frac{1}{\lambda l} \left( g(x_{m+k}) - \frac{\partial f_{\bar{x},\lambda}(x_{m+k})}{\bar{n}} \right)$  into (40), we have (21).

By (29), we have:

$$f_\lambda(\cdot) = \frac{1}{\lambda} \int_{B^d} (f^*(x) - f_\lambda(x)) K_x^\mu(\cdot) d\rho_{B^d} + \frac{1}{\lambda} \int_{S^{d-1}} \left( g(x) - \frac{\partial f_\lambda(x)}{\bar{n}} \right) \frac{\partial K_x^\mu(\cdot)}{\partial \bar{n}} d\rho_{S^{d-1}}. \tag{41}$$

Taking  $G_{\lambda,f^*}(x) = \frac{1}{\lambda} (f^*(x) - f_\lambda(x))$  and  $p_{\lambda,g}(x) = \frac{1}{\lambda} \left( g(x) - \frac{\partial f_\lambda(x)}{\partial \bar{n}} \right)$  into (41), we have (22).

*Proof of Theorem 2.1.* Collecting (33), (34) and (23), we have:

$$\begin{aligned} & \|f_{z,\lambda} - f^*\|_{L^2(\rho_{B^d})} + \left\| \frac{\partial f_{z,\lambda}}{\partial \bar{n}} - g \right\|_{L^2(\rho_{S^{d-1}})} \\ & \leq \frac{2k^2\sigma}{\lambda\sqrt{m\delta}} + k \left( \frac{\sqrt{K(f^*,g,\lambda)}}{\lambda^{\frac{3}{2}}\sqrt{m}} + \frac{\|f^*\|_{C(B^d)}}{\lambda\sqrt{m}} \right) \log \frac{4}{\delta} \\ & \quad + \frac{k\sqrt{K(f^*,g,\lambda)}}{\sqrt{l\delta}} + \sqrt{2}\sqrt{K(f^*,g,\lambda)}. \end{aligned} \tag{42}$$

(42) yields (24).

### Founding

Supported partially by the NSF (Project No. 61877039), the NSFC/RGC Joint Research Scheme (Project No. 12061160462 and N\_CityU102/20) of China.

### Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

### References

[1] Atkinson, K., Hansen, O. and Chien, D. (2011) A Spectral Method for Elliptic Equations: The Neumann Problem. *Advances in Computational Mathematics*, **34**, 295-317.

- <https://doi.org/10.1007/s10444-010-9154-3>
- [2] Atkinson, K., Chien, D. and Hansen, O. (2010) A Spectral Method for Elliptic Equation: The Dirichlet Problem. *Advances in Computational Mathematics*, **33**, 169-189. <https://doi.org/10.1007/s10444-009-9125-8>
- [3] Atkinson, K. and Hansen, O. (2010) A Spectral Method for the Eigenvalue Problem for Elliptic Equations. *Electronic Transactions on Numerical Analysis*, **37**, 386-412.
- [4] Li, X. (2009) Approximation of Potential Integral by Radial Bases for Solutions of Helmholtz Equation. *Advances in Computational Mathematics*, **30**, 201-230. <https://doi.org/10.1007/s10444-008-9065-8>
- [5] Li, X. (2008) Rate of Convergence of the Method of Fundamental Solutions and Hyperinterpolation for Modified Helmholtz Equations on the Unit Ball. *Advances in Computational Mathematics*, **29**, 393-413. <https://doi.org/10.1007/s10444-007-9056-1>
- [6] Li, X. (2008) Convergence of the Method of Fundamental Solutions for Poisson's Equation on the Unit Sphere. *Advances in Computational Mathematics*, **28**, 269-282. <https://doi.org/10.1007/s10444-006-9022-3>
- [7] Cialenco, I., Fasshauer, G.E. and Ye, Q. (2012) Approximation of Stochastic Partial Differential Equations by a Kernel-Based Collocation Method. *International Journal of Computer Mathematics*, **89**, 2543-2561. <https://doi.org/10.1080/00207160.2012.688111>
- [8] Ding, L.L., Liu, Z.Y. and Xu, Q.Y. (2021) Multilevel RBF Collocation Method for the Fourth-Order Thin Plate Problem. *International Journal of Wavelets, Multiresolution and Information Processing*, **19**, Article ID: 2050079. <https://doi.org/10.1142/S0219691320500794>
- [9] Fasshauer, G.E. and Ye, Q. (2013) Kernel-Based Collocation Methods versus Galerkin Finite Element Methods for Approximating Elliptic Stochastic Partial Differential Equation. In: Griebel, M. and Schweitzer, M., Eds., *Meshfree Methods for Partial Differential Equations VI*, Springer, Berlin, 155-170. [https://doi.org/10.1007/978-3-642-32979-1\\_10](https://doi.org/10.1007/978-3-642-32979-1_10)
- [10] Fasshauer, G.E. and Ye, Q. (2012) A Kernel-Based Collocation Method for Elliptic Partial Differential Equations with Random Coefficients. In: Dick, J., Kuo, F., Peters, G. and Sloan, I., Eds., *Monte Carlo and Quasi-Monte Carlo Methods 2012*, Springer, Berlin, 331-347. [https://doi.org/10.1007/978-3-642-41095-6\\_14](https://doi.org/10.1007/978-3-642-41095-6_14)
- [11] Ye, Q. (2014) Approximation of Nonlinear Stochastic Partial Differential Equations by a Kernel-Based Collocation Method. *International Journal of Applied Nonlinear Science*, **1**, 156-172. <https://doi.org/10.1504/IJANS.2014.061018>
- [12] Zhou, D.X. (2008) Derivative Reproducing Properties for Kernel Methods in Learning Theory. *Journal of Computational and Applied Mathematics*, **220**, 456-463. <https://doi.org/10.1016/j.cam.2007.08.023>
- [13] Bao, K.J., Qian, X., Liu, Z.Y. and Song, S.B. (2022) An Operator Learning Approach via Function—Valued Reproducing Kernel Hilbert Space for Differential Equations. arXiv: 2202.09488.
- [14] Mo, Y. and Qian, T. (2014) Support Vector Machine Adapted Tikhonov Regularization Method to Solve Dirichlet Problem. *Applied Mathematics and Computation*, **245**, 509-519. <https://doi.org/10.1016/j.amc.2014.07.089>
- [15] Sheng, B.H., Zhou, D.P. and Wang, S.H. (2022) The Kernel Regularized Learning Algorithm for Solving Laplace Equation with Dirichlet Boundary. *International Journal of Wavelets, Multiresolution and Information Processing*, **20**, Article ID: 2250031. <https://doi.org/10.1142/S021969132250031X>

- [16] Stepaniants, G. (2023) Learning Partial Differential Equations in Reproducing Kernel Hilbert Spaces. *Journal of Machine Learning Research*, **24**, 1-72
- [17] Harosko, D.D. and Triebel, H. (2008) Distributions, Sobolev Spaces, Elliptic Equations. European Mathematical Society, Helsinki. <https://doi.org/10.4171/042>
- [18] Belkin, M. and Niyogi, P. (2004) Semi-Supervised Learning on Riemannian Manifolds. *Machine Learning*, **56**, 209-239. <https://doi.org/10.1023/B:MACH.0000033120.25363.1e>
- [19] Belkin, M., Niyogi, P. and Sindhvani, V. (2006) Manifold Regularization: A Geometric Framework for Learning from Labeled and Unlabeled Examples. *Journal of Machine Learning Research*, **7**, 2399-2434.
- [20] Niyogi, P. (2013) Manifold Regularization and Semi-Supervised Learning: Some Theoretical Analysis, *Journal of Machine Learning Research*, **14**, 1229-1250.
- [21] Sheng, B.H. and Zhu, H.C. (2018) The Convergence Rate of Semi-Supervised Regression with Quadratic Loss. *Applied Mathematics and Computation*, **321**, 11-24. <https://doi.org/10.1016/j.amc.2017.10.033>
- [22] Smale, S. and Zhou, D.X. (2004) Shannon Sampling and Function Reconstruction from Point Values. *Bulletin of the American Mathematical Society*, **41**, 279-305. <https://doi.org/10.1090/S0273-0979-04-01025-0>
- [23] Cucker, F. and Smale, S. (2002) On the Mathematical Foundations of Learning Theory. *Bulletin of the American Mathematical Society*, **39**, 1-49. <https://doi.org/10.1090/S0273-0979-01-00923-5>
- [24] Cucker, F. and Zhou, D.X. (2007) Learning Theory: An Approximation Theory Viewpoint. Cambridge University Press, Cambridge. <https://doi.org/10.1017/CBO9780511618796>
- [25] Chen, H., Zhou, Y.C., Tang, Y.Y., Li, L.Q. and Pan, Z.B. (2013) Convergence Rate of the Semi-Supervised Greedy Algorithm. *Neural Networks*, **44**, 44-50. <https://doi.org/10.1016/j.neunet.2013.03.001>
- [26] Sheng, B.H. and Xiang, D.H. (2017) The Performance of Semi-Supervised Laplacian Regularized Regression with Least Square Loss. *International Journal of Wavelets, Multiresolution and Information Processing*, **15**, Article ID: 1750016. <https://doi.org/10.1142/S0219691317500163>
- [27] Sheng, B.H., Xiang, D.H. and Ye, P.X. (2015) Convergence Rate of Semi-Supervised Gradient Learning. *International Journal of Wavelets, Multiresolution and Information Processing*, **13**, Article ID: 1550021. <https://doi.org/10.1142/S0219691315500216>
- [28] Sheng, B.H. and Zhang, H.Z. (2020) Performance Analysis of the LapRSSLG Algorithm in Learning Theory. *Analysis and Applications*, **18**, 79-108. <https://doi.org/10.1142/S0219530519410033>
- [29] Wang, K.Y. and Li, L.Q. (2000) Harmonic Analysis and Approximation on the Unit Sphere. Science Press, Beijing.
- [30] Delgado, A.M., Fernández, L., Lubinsky, D., Pérez, T.E. and Piñar, M.A. (2016) Sobolev Orthogonal Polynomials on the Unit Ball via Outward Normal Derivatives. *Journal of Mathematical Analysis and Applications*, **440**, 716-740. <https://doi.org/10.1016/j.jmaa.2016.03.041>
- [31] Jordao, T. and Menegatto, V.A. (2012) Reproducing Properties of Differentiable Mercer-Like Kernels on the Sphere. *Numerical Functional Analysis and Optimization*, **33**, 1221-1243. <https://doi.org/10.1080/01630563.2012.660590>
- [32] Castro, M.H., Menegatto, V.A. and Oliveira, C.P. (2013) Laplace-Beltrami Differen-

- tiability of Positive Definite Kernels on the Sphere. *Acta Mathematica Sinica, English Series*, **29**, 93-104. <https://doi.org/10.1007/s10114-012-1067-2>
- [33] Ferreira, J.C. and Menegatto, V.A. (2013) Positive Definiteness, Reproducing Kernel Hilbert Spaces and Beyond. *Annals of Functional Analysis*, **4**, 64-88. <https://doi.org/10.15352/afa/1399899838>
- [34] Sun, H.W. and Wu, Q. (2011) Least Square Regression with Independent Kernels and Coefficient Regularization. *Applied and Computational Harmonic Analysis*, **30**, 96-109. <https://doi.org/10.1016/j.acha.2010.04.001>
- [35] Zhang, J., Wang, J.L. and Sheng, B.H. (2011) Learning from Regularized Regression Algorithms with P-Order Markov Chain Sampling. *Applied Mathematics: A Journal of Chinese Universities*, **226**, 295-306. <https://doi.org/10.1007/s11766-011-2701-y>
- [36] Sheng, B.H. and Wang, J.L. (2024) Moduli of Smoothness,  $K$ -Functionals and Jackson-Type Inequalities Associated with Kernel Function Approximation in Learning Theory. *Analysis and Applications*. <https://doi.org/10.1142/S021953052450009X>
- [37] Dai, F. and Xu, Y. (2013) Approximation Theory and Harmonic Analysis on Spheres and Balls. Springer-Verlag, New York. <https://doi.org/10.1007/978-1-4614-6660-4>
- [38] Bauschke, H.H. and Combettes, P.L. (2010) Convex Analysis and Monotone Operator Theory in Hilbert Spaces. Springer-Verlag, New York. <https://doi.org/10.1007/978-1-4419-9467-7>
- [39] Aronszajn, N. (1950) Theory of Reproducing Kernels. *Transactions of the American Mathematical Society*, **68**, 337-404. <https://doi.org/10.1090/S0002-9947-1950-0051437-7>
- [40] Ferreira, J.C. and Menegatto, V.A. (2012) Reproducing Properties of Differentiable Mercer-Like Kernels. *Mathematische Nachrichten*, **285**, 959-973. <https://doi.org/10.1002/mana.201100072>
- [41] Smale, S. and Zhou, D.X. (2007) Learning Theory Estimates via Integral Operators and Their Applications. *Constructive Approximation*, **26**, 153-172. <https://doi.org/10.1007/s00365-006-0659-y>

## Appendices

### Appendix 1. Gâteaux Derivative and the Convex Function

Following Proposition A1 can be found from the Proposition 17.4, Proposition 17.10 and Proposition 17.12 of [38].

**Proposition A1.** Let  $(\mathcal{H}, \|\cdot\|_{\mathcal{H}})$  be a Hilbert space and

$F(f): \mathcal{H} \rightarrow \mathbb{R} \cup \{\mp\infty\}$  be a function defined on  $\mathcal{H}$ . Then,

1) If  $F(f)$  is a convex function, then,  $F(f)$  attains minimal value at  $f_0$  if and only if  $\nabla_f F(f_0) = 0$ .

2) If  $F(f): \mathcal{H} \rightarrow \mathbb{R} \cup \{\mp\infty\}$  is a Gâteaux differentiable function, then  $F(f)$  is a convex on  $\mathcal{H}$  if and only if for any  $f, g \in \mathcal{H}$ , there holds:

$$F(g) - F(f) \geq \langle g - f, \nabla_f F(f) \rangle_H.$$

In particular, we have:

$$x^2 - y^2 \geq 2y(x - y), \quad \forall x, y \in \mathbb{R}. \tag{43}$$

3) For function  $F(f) = \|f\|_{\mathcal{H}}^2$ , there holds  $\nabla_f F(f) = 2f$  and there holds equality:

$$\|f\|_H^2 - \|g\|_H^2 = \langle f - g, 2g \rangle_H + \|f - g\|_H^2, \tag{44}$$

i.e.

$$\|f\|_H^2 - \|g\|_H^2 = \langle f - g, \nabla_g F(g) \rangle_H + \|f - g\|_H^2.$$

### Appendix 2. Derivatives Reproducing Property

Let  $\Omega \subset \mathbb{R}^d$  be a compact subset which is the closure of its nonempty interior  $\Omega^0$ . Let  $K(x, y)$  be a Mercer kernel on  $\Omega \times \Omega$  having the expansion (see e.g. [39]):

$$K(x, y) = \sum_{k=0}^{+\infty} \lambda_k \varphi_k(x) \varphi_k(y), \quad x, y \in \Omega, \tag{45}$$

where the convergence is absolute (for each  $x, y \in \Omega$ ) and uniform on  $\Omega \times \Omega$ . Then, we have a proposition.

**Proposition A2.** Let  $K(x, y)$  be a Mercer kernel of form (45) and  $K \in C^{(1)}(\Omega \times \Omega)$ . If  $\mathcal{H}_K$  is a reproducing kernel Hilbert space such that:

$$f(x) = \langle f, K(\cdot, x) \rangle_{\mathcal{H}_K}, \quad f \in H_K, x \in \Omega.$$

Then,

$$\partial_x^\alpha f(x) = \langle f, \partial_x^\alpha K(\cdot, x) \rangle_{\mathcal{H}_K(\rho_\Omega)}, \quad |\alpha| \leq 1, f \in \mathcal{H}_K, x \in \Omega, \tag{46}$$

where  $|\alpha| = \sum_{i=1}^d \alpha_i \leq 1$ .

*Proof.* It can be found from Theorem 1 in [12], or see (v) in Theorem 4.7 in [40].



### Appendix 3. A Probability Inequality

**Proposition A4.** [41] *Let  $\xi$  be a random variable taking values in a real separable Hilbert space  $H$  on a probability space  $(\Omega, \mathcal{F}, P)$ . Assume that there are a positive constant  $L$  such that  $\|\xi\|_H \leq L$ . Then, for all  $n \geq 1$  and  $0 < \eta < 1$ , it holds, with confidence  $1 - \eta$ , that:*

$$\left\| \frac{1}{n} \sum_{i=1}^n \xi(\Omega_i) - E(\xi) \right\|_H \leq \frac{4L}{\sqrt{n}} \log \frac{2}{\eta}. \quad (47)$$