



## SOME NOTES ON ERROR ANALYSIS FOR KERNEL BASED REGULARIZED INTERPOLATION

QING ZOU\*

*Applied Mathematics and Computational Sciences, The University of Iowa, IA 52242, USA*

*\*Corresponding author: [zou-qing@uiowa.edu](mailto:zou-qing@uiowa.edu)*

**ABSTRACT.** Kernel based regularized interpolation is one of the most important methods for approximating functions. The theory behind the kernel based regularized interpolation is the well-known Representer Theorem, which shows the form of approximation function in the reproducing kernel Hilbert spaces. Because of the advantages of the kernel based regularized interpolation, it is widely used in many mathematical and engineering applications, for example, dimension reduction and dimension estimation. However, the performance of the approximation is not fully understood from the theoretical perspective. In other word, the error analysis for the kernel based regularized interpolation is lacking. In this paper, some error bounds in terms of the reproducing kernel Hilbert space norm and Sobolev space norm are given to understand the behavior of the approximation function.

### 1. INTRODUCTION

Approximating functions in high dimensional spaces is one of the central problems in both mathematics and engineering. Many real-world problems can be viewed as a function approximation problem. For example, the classification problem in engineering can be viewed as approximating a function whose function values give the classes that the inputs belong to [7] and in image processing, the patch-based image denoising problem can be seen as approximating a function from noisy patches to clean pixels.

---

Received May 12<sup>th</sup>, 2020; accepted June 4<sup>th</sup>, 2020; published June 17<sup>th</sup>, 2020.

2010 *Mathematics Subject Classification.* 65D05, 41A05, 41A10, 65J05, 65G99.

*Key words and phrases.* error bound; kernel based regularized interpolation; reproducing kernel Hilbert space; Sobolev space.

©2020 Authors retain the copyrights of their papers, and all open access articles are distributed under the terms of the Creative Commons Attribution License.

Mathematically speaking, the function approximation problem is to approximate an unknown continuous function  $f : X \rightarrow \mathbb{R}$  from the knowledge of some observations  $\{(x_i, y_i)\}_{i=1}^n \subset X \times \mathbb{R}$ , where  $X \subset \mathbb{R}^d$ ,  $d \geq 1$  is the input space and  $n \in \mathbb{N}$  is the number of observations.

One of the classical ways for approximating functions to explain the real-world phenomenon is called Kriging (a.k.a Gaussian process regression) [4]. However, to use Kriging, we need a large amount of observations, which is expensive to achieve. When only a few samples are given, Kriging will yield non-stationary behavior [10].

Another alternative for approximating multivariate functions is the inverse distance weighting (IDW) method, which was originally proposed in [9]. The key assumption of IDW is that the points that are close to each other are more alike than those that are far away from each other. Therefore, IDW produces prediction at a point relying on observed points close to that point. In other words, the measured values close to the prediction location have more effects on the predicted value than those which are far away. Based on which, the approximation function can be represented as

$$\hat{f}(x) = \frac{\sum_{i=1}^n \|x - x_i\|^{-\alpha} y_i}{\sum_{i=1}^n \|x - x_i\|^{-\alpha}},$$

where  $\alpha$  is a parameter and usually it takes the value 2. IDW can produce good accuracy for the points that are near the observations. However, for the points that are away from the observations, IDW cannot work well. This is one drawback of IDW. Another drawback of IDW is that the gradient vanishes at the observations.

Considering the drawbacks of these methods, the kernel based regularized interpolation (see (2.2) for details) was proposed. The kernel based regularized interpolation in 1D is similar to the spline interpolation. It has several advantages. First of all, the kernel based regularized interpolation works for any dimensions with many different situations [1, 2]. It also enjoys solid theoretical understandings — the representer theorem — which we will show it below. Furthermore, comparing to other methods, one can have the same approximation accuracy using kernel based regularized interpolation by solving a better conditioned linear system. The kernel based regularized interpolation also works for the case that the observations have some noise (see Section 2 for analysis). This is one of the biggest advantages that other interpolation methods do not have. Last but not least, we have much flexibility in choosing kernels when using kernel based regularized interpolation. Kernel based regularized interpolation is also widely used in other engineering problems, for example, clustering, dimension reduction [1] and dimension estimation [11].

In view of the advantages of kernel based regularized interpolation, there is pretty much work on using kernel based regularized interpolation to deal with different situations, as we mentioned above. However, the work on the error analysis for the kernel based regularized interpolation is very limited. It is actually worthwhile to reflect on the nature of error bounds that are needed to characterize the behavior of the

learned functions from observations. In this paper, we would like to provide some error analysis for the kernel based regularized interpolation from the function space point of view. In other word, we focus specifically on deriving Hilbert space type and Sobolev space type error estimates for the kernel based regularized interpolation.

## 2. PRELIMINARY

Let us recall here some basic facts from kernel methods for our analysis. We consider a positive definite kernel  $K : X \times X \rightarrow \mathbb{R}$ , i.e., for  $n \in \mathbb{N}$ ,  $x_1, \dots, x_n \in X$  and  $a_1, \dots, a_n \in \mathbb{R}$ ,

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

Associated with the kernel  $K$ , there exists a unique native Hilbert space  $\mathcal{H}_K$  of functions from  $X$  to  $\mathbb{R}$ . The elements in  $\mathcal{H}_K$  are of the form

$$f(\cdot) = \sum_{i \in I} \alpha_i K(\cdot, x_i),$$

where  $I$  is a countable set. The norm defined for  $\mathcal{H}_K$  is given by

$$\langle f, g \rangle_{\mathcal{H}_K} =: \sum_{i \in I} \sum_{j \in J} \alpha_i \beta_j K(x_i, x_j),$$

where  $g = \sum_{j \in J} \beta_j K(\cdot, x_j) \in \mathcal{H}_K$ . The Hilbert space  $\mathcal{H}_K$  is called reproducing kernel Hilbert space (RKHS) because the kernel  $K$  acts as a reproducing kernel on  $\mathcal{H}_K$ , i.e.,

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

We require the kernel to be positive definite because when we approximate the functions, linear systems will be involved. The positive-definite property guarantees the linear systems are well conditioned and the approximation problem can be solved in  $\mathcal{H}_K$  using the given observations  $\{(x_i, y_i)\}_{i=1}^n$ . Indeed, one can solve the functions approximation problem by considering the regularization problem with the regularization parameter  $\lambda > 0$ ,

$$\min_{f \in \mathcal{H}_K} \sum_{i=1}^n \|f(x_i) - y_i\|^2 + \lambda \|f\|_{\mathcal{H}_K}^2. \quad (2.1)$$

The solution of this regularization problem is characterized by the Representer Theorem [8], which states that the solution of (2.1) is given as

$$f = \sum_{i=1}^n \alpha_i K(x, x_i). \quad (2.2)$$

This expression is called *kernel based regularized interpolation*. In the rest of this paper, we use the notation  $I_X f$  to stand for the kernel based regularized interpolation, i.e.,

$$I_X f = \sum_{i=1}^n \alpha_i K(x, x_i).$$

To obtain error analysis, we need to have the true function. We use the notation  $f$  to represent the true function that is only known at the sampling locations, i.e., we only have the information  $y_i = f(x_i)$  about the true function  $f$ . The goal of this work is to obtain the bounds for the error between the true function  $f$  and the kernel based regularized interpolation  $I_X f$ .

It is worth mentioning that the kernel based regularized interpolation is well defined for positive definite kernels, but it is not a pure interpolation. This means that we may not have  $I_X f(x_i) = y_i = f(x_i)$ . But in the case of strictly positive definite kernels, the regularization problem (2.1) will become an interpolation problem which exactly interpolates  $f$  at the observations and hence the resulted kernel interpolation is a pure interpolation. However, there is no need to consider strictly positive definite kernels. Using positive definite kernel to obtain kernel based regularized interpolation is more general. This is because we have a tunable parameter  $\lambda$  in the regularization problem and it provides a trade-off between pointwise accuracy and stability. Another important reason is that in real-world applications, the observations may be corrupted by noise and it makes no sense to have pure interpolations anymore.

### 3. HILBERT SPACE TYPE ERROR ANALYSIS

In this section, we would like to provide a special case about the error analysis for the kernel based regularized interpolation. The term “special case” is used here because we will assume in this section that the true function  $f$  is living in the RKHS. While in some cases, we cannot guarantee that the true function is within the RKHS. We will show the error analysis for the general case in the next section.

Before proceeding to the error analysis, we first recall an important concept in numerical analysis: best approximation [5]. Given  $f \in \mathcal{H}_K$ , define

$$\rho := \inf_{p \in S} \|f - p\|_{\mathcal{H}_K},$$

where  $S \subset \mathcal{H}_K$ . Then the number  $\rho$  is called the minimax error for the approximation of the function  $f$  by functions in  $S$ . In fact [5], there is a unique  $\hat{p} \in S$  such that

$$\|f - \hat{p}\|_{\mathcal{H}_K} = \rho = \inf_{p \in S} \|f - p\|_{\mathcal{H}_K}.$$

The function  $\hat{p} \in S$  is called the best approximation of  $f$  w.r.t. the  $\mathcal{H}_K$ -norm.

Now, let us look at the inner product of the RKHS. By Cauchy-Schwarz inequality [3], we have that

$$|\langle h, K(\cdot, x) \rangle_{\mathcal{H}_K}| \leq \|h\|_{\mathcal{H}_K} \cdot \|K(\cdot, x)\|_{\mathcal{H}_K}, \quad \forall h \in \mathcal{H}_K. \quad (3.1)$$

Based on which, we can claim from (3.1) that there exists a constant  $0 < \beta \leq 1$  such that

$$\sup_{x \in X} \frac{|\langle h, K(\cdot, x) \rangle_{\mathcal{H}_K}|}{\|K(\cdot, x)\|_{\mathcal{H}_K}} \geq \beta \cdot \|h\|_{\mathcal{H}_K}. \quad (3.2)$$

We then have the first error estimation.

**Theorem 3.1.** *Let  $S = \text{span}\{K(x, x_i)\}_{i=1}^n \subset \mathcal{H}_K$ . Assume that the true function  $f \in \mathcal{H}_K$ .  $I_X f$  is the kernel based regularized interpolation. Then we have the following error bound*

$$\|f - I_X f\|_{\mathcal{H}_K} \leq \left(1 + \frac{C}{\beta}\right) \cdot \rho,$$

where  $C > 1$ ,  $0 < \beta \leq 1$  are two constants and  $\rho$  is the minimax error for the approximation of the function  $f$  by functions in  $S$  in terms of the  $\mathcal{H}_K$ -norm.

*Proof.* First of all, let us consider the trivial case that the true function  $f \in S \subset \mathcal{H}_K$ . In this case, the kernel based regularized interpolation is exact, meaning that we have  $I_X f = f$ . Then both  $\|f - I_X f\|_{\mathcal{H}_K}$  and  $\rho$  are zero. So the conclusion becomes trivial in this case.

Next, we assume that  $f \in \mathcal{H}_K$  but  $f \notin S$ . Then for any  $g \in S$ , we have

$$\|f - I_X f\|_{\mathcal{H}_K} \leq \|f - g\|_{\mathcal{H}_K} + \|I_X f - g\|_{\mathcal{H}_K}. \tag{3.3}$$

Replacing  $h$  by  $I_X f - g$  in (3.2), we have

$$\beta \|I_X f - g\|_{\mathcal{H}_K} \leq \sup_{x \in X} \frac{|\langle I_X f - g, K(\cdot, x) \rangle_{\mathcal{H}_K}|}{\|K(\cdot, x)\|_{\mathcal{H}_K}}.$$

Because we have  $f \neq g$ , then for the right-hand side (RHS) of the above inequality, there exists a constant  $C > 1$  such that

$$\begin{aligned} RHS &= \sup_{x \in X} \frac{|\langle f - g + I_X f - f, K(\cdot, x) \rangle_{\mathcal{H}_K}|}{\|K(\cdot, x)\|_{\mathcal{H}_K}} \\ &\leq \sup_{x \in X} \left[ \frac{|\langle f - g, K(\cdot, x) \rangle_{\mathcal{H}_K}|}{\|K(\cdot, x)\|_{\mathcal{H}_K}} + \frac{|\langle I_X f - f, K(\cdot, x) \rangle_{\mathcal{H}_K}|}{\|K(\cdot, x)\|_{\mathcal{H}_K}} \right] \\ &= \sup_{x \in X} \frac{|\langle f - g, K(\cdot, x) \rangle_{\mathcal{H}_K}|}{\|K(\cdot, x)\|_{\mathcal{H}_K}} + \sup_{x \in X} \frac{|\langle I_X f - f, K(\cdot, x) \rangle_{\mathcal{H}_K}|}{\|K(\cdot, x)\|_{\mathcal{H}_K}} \\ &\leq C \cdot \sup_{x \in X} \frac{|\langle f - g, K(\cdot, x) \rangle_{\mathcal{H}_K}|}{\|K(\cdot, x)\|_{\mathcal{H}_K}} \\ &\leq C \cdot \|f - g\|_{\mathcal{H}_K}. \end{aligned}$$

The last inequality is due to the Cauchy-Schwarz inequality.

Therefore, we have

$$\|I_X f - g\|_{\mathcal{H}_K} \leq \frac{C}{\beta} \|f - g\|_{\mathcal{H}_K}.$$

Then by (3.3), we can obtain that

$$\|f - I_X f\|_{\mathcal{H}_K} \leq \left(1 + \frac{C}{\beta}\right) \|f - g\|_{\mathcal{H}_K}, \quad \forall g \in S,$$

which implies that

$$\|f - I_X f\|_{\mathcal{H}_K} \leq \left(1 + \frac{C}{\beta}\right) \inf_{g \in S} \|f - g\|_{\mathcal{H}_K} = \left(1 + \frac{C}{\beta}\right) \cdot \rho.$$

This completes the proof. □

In fact, this error bound can be further tightened. The improvement is due to the observation that if the true function  $f$  is within  $\mathcal{H}_K$ , the kernel based regularized interpolation is actually a projection. Let  $S = \text{span}\{K(x, x_i)\}_{i=1}^n \subset \mathcal{H}_K$  as defined in Theorem 3.1. Then if  $f \in \mathcal{H}_K$ , we can see that the kernel based regularized interpolation is a projection  $\varphi : \mathcal{H}_K \rightarrow S$ . The most important property of a projection is the idempotence. If the function  $f$  is already in  $S$ , the projection does nothing but to keep the function, i.e.,  $\varphi(g) = g, \forall g \in S$ . This implies the idempotence of the projector:

$$\varphi^2(f) = \varphi(\varphi(f)) = \varphi(f), \quad \forall f \in \mathcal{H}_K.$$

If we define  $\|\varphi\|$  to be the operator norm, which is given by

$$\|\varphi\| =: \max_{0 \neq g \in \mathcal{H}_K} \frac{\|\varphi(g)\|_{\mathcal{H}_K}}{\|g\|_{\mathcal{H}_K}}.$$

Then we have the identity

$$\|\varphi\| = \|I - \varphi\|,$$

where  $I$  is the identity operator. This conclusion is due to the following result given in [6].

**Lemma 3.1.** *Let  $\mathcal{H}$  be a Hilbert space and  $U \subset \mathcal{H}$ . Let  $P : \mathcal{H} \rightarrow U$  be an (oblique) projection ( $P^2 = P$ ) and  $0 \neq P \neq I$ . Then*

$$\|P\| = \|I - P\|.$$

Using the observation that the kernel based regularized interpolation is a projection if  $f \in \mathcal{H}_K$ , we then have the following tightened error bound.

**Theorem 3.2.** *Suppose  $S = \text{span}\{K(x, x_i)\}_{i=1}^n \subset \mathcal{H}_K$ . The kernel based regularized interpolation is given by  $I_X f = \varphi(f)$ , where  $\varphi$  is a projector from  $\mathcal{H}_K$  to  $S$ . Then we have the following error bounds*

$$\|f - I_X f\|_{\mathcal{H}_K} \leq \frac{\rho}{\beta},$$

where  $0 < \beta \leq 1$  is a constant and  $\rho$  is the minimax error for the approximation of the function  $f$  by functions in  $S$  in terms of the  $\mathcal{H}_K$ -norm.

*Proof.* Since  $\varphi$  is a projection, we have  $\|\varphi\| = \|I - \varphi\|$ . Based on which, we can derive that for all  $g \in S$ ,

$$\begin{aligned} \|f - I_X f\|_{\mathcal{H}_K} &= \|f - g - I_X f + g\|_{\mathcal{H}_K} \\ &= \|f - g - \varphi(f - g)\|_{\mathcal{H}_K} \\ &= \|(I - \varphi)(f - g)\|_{\mathcal{H}_K} \\ &\leq \|I - \varphi\| \cdot \|f - g\|_{\mathcal{H}_K} \\ &= \|\varphi\| \cdot \|f - g\|_{\mathcal{H}_K}. \end{aligned}$$

To show the error bound, we need to show that  $\|\varphi\| \leq \frac{1}{\beta}$ . First, it is straightforward that if  $f \in \mathcal{H}_K$ , we have

$$\|I_X f\|_{\mathcal{H}_K} \leq \|f\|_{\mathcal{H}_K}.$$

Since  $0 < \beta \leq 1$ , we can get  $\frac{1}{\beta} \geq 1$ . So for  $f \in \mathcal{H}_K$ ,

$$\|I_X f\|_{\mathcal{H}_K} \leq \frac{1}{\beta} \|f\|_{\mathcal{H}_K}.$$

Therefore, we can obtain that

$$\begin{aligned} \|\varphi\| &= \max_{0 \neq f \in \mathcal{H}_K} \frac{\|\varphi(f)\|_{\mathcal{H}_K}}{\|f\|_{\mathcal{H}_K}} \\ &= \max_{0 \neq f \in \mathcal{H}_K} \frac{\|I_X f\|_{\mathcal{H}_K}}{\|f\|_{\mathcal{H}_K}} \leq \frac{\frac{1}{\beta} \|f\|_{\mathcal{H}_K}}{\|f\|_{\mathcal{H}_K}} \\ &= \frac{1}{\beta}. \end{aligned}$$

This gives us the desired bound. □

#### 4. SOBOLEV SPACE TYPE ERROR ANALYSIS

In the previous section, we have introduced a special case for the error analysis about kernel based regularized interpolation, where we assumed that the true function  $f$  lives in the reproducing kernel Hilbert space  $\mathcal{H}_K$ . However, in general we do not have this guarantee. The true functions are usually within a larger function space. In this section, we consider the case that the true function  $f$  is in the Sobolev spaces, which are in fact the spaces that consist of all  $f \in \mathcal{L}^p(X)$  with certain properties. The formal definition of the Sobolev space is given as follows.

**Definition 4.1.** *Let  $k$  be a nonnegative integer,  $p \in [1, \infty]$ . The Sobolev space  $W^{k,p}(X)$  is the set of all the functions  $f \in \mathcal{L}^p(X)$  such that for each multi-index  $\alpha$  with  $|\alpha| \leq k$ , the  $\alpha$ th weak derivative  $\partial^\alpha f$  exists and  $\partial^\alpha f \in \mathcal{L}^p(X)$ . The norm of the sobolev space is defined as*

$$\|f\|_{W^{k,p}(X)} = \begin{cases} \left[ \sum_{|\alpha| \leq k} \|\partial^\alpha f\|_{\mathcal{L}^p(X)}^p \right]^{1/p}, & 1 \leq p < \infty \\ \max_{|\alpha| \leq k} \|\partial^\alpha f\|_{\mathcal{L}^\infty(X)}, & p = \infty \end{cases}.$$

When  $p = 2$ , we usually write  $W^{k,2}(X) := H^k(X)$ . For simplicity, we replace  $\|f\|_{W^{k,p}(X)}$  by  $\|f\|_{k,p,X}$  when no confusion and we ignore the  $p$  when  $p = 2$ .

Except for the standard norm defined for the Sobolev space, we also have the definition of the seminorm. The standard seminorm over the space  $W^{k,p}(X)$  is given as

$$|f|_{W^{k,p}(X)} = \begin{cases} \left[ \sum_{|\alpha|=k} \|\partial^\alpha f\|_{\mathcal{L}^p(X)}^p \right]^{1/p}, & 1 \leq p < \infty \\ \max_{|\alpha|=k} \|\partial^\alpha f\|_{\mathcal{L}^\infty(X)}, & p = \infty \end{cases}.$$

Similarly, we write  $|f|_{W^{k,p}(X)}$  as  $|f|_{k,p,X}$  when no confusion and ignore the  $p$  when  $p = 2$ .

We now define the continuous embedding between two Banach spaces. Let  $W, V$  be two Banach spaces with  $V \subset W$ . We say the space  $V$  is continuously embedded in  $W$  and write  $V \hookrightarrow W$ , if there exists a constant  $c > 0$  such that for all  $v \in V$ , we have  $\|v\|_W \leq c\|v\|_V$ .

With these concepts in the theory of Sobolev spaces, we can then have the error bound for the kernel based regularized interpolation. We assume that the true function  $f$  is within  $H^k(X)$ .

**Theorem 4.1.** *Let  $k$  and  $m$  be nonnegative integers with  $k \geq 0, k + 1 \geq m$ . Suppose  $\mathcal{H}_K \supset S = \text{span}\{K(x, x_i)\}$ . Then there exists a constant  $c > 0$  such that*

$$|f - I_X f|_{m,X} \leq c \inf_{g \in S} \|f + g\|_{k+1,X}.$$

When the positive definite kernel is chosen to be the polynomial kernel with degree  $d$  and  $k + 1 > d$ , we have

$$|f - I_X f|_{m,X} \leq c|f|_{k+1,X}.$$

*Proof.* Since  $k > 0$ , we then have  $H^{k+1}(X) \hookrightarrow C(X)$ , where  $C(X)$  is the set of all the continuous functions defined on  $X$ . So  $f \in H^{k+1}(X)$  is continuous and  $I_X f$  is well defined. Besides, there exists a constant  $c_1 > 0$  such that

$$\|I_X f\|_{m,X} \leq \sum_{i=1}^n |\alpha_i| \cdot \|K(x, x_i)\|_{m,X} \leq c_1 \|f\|_{C(X)}.$$

Since  $H^{k+1}(X) \hookrightarrow C(X)$ , we then have that there exists a constant  $c_2 > 0$  such that

$$\|f\|_{C(X)} \leq c_2 \|f\|_{k+1,X}.$$

Therefore, there exists some  $c_3 > 0$  such that

$$\|I_X f\|_{m,X} \leq c_3 \|f\|_{k+1,X}. \tag{4.1}$$

Now, for all  $f \in H^{k+1}$  and  $g \in S$ , we can obtain from (4.1) that there exist some constant  $c > 0$ ,

$$\begin{aligned} |f - I_X f|_{m,X} &\leq \|f - I_X f\|_{m,X} \\ &= \|f - I_X f + g - I_X g\|_{m,X} \\ &= \|(f + g) - I_X(f + g)\|_{m,X} \\ &\leq \|f + g\|_{m,X} + \|I_X(f + g)\|_{m,X} \\ &\leq c\|f + g\|_{k+1,X}. \end{aligned}$$

By which we have

$$|f - I_X f|_{m,X} \leq c \inf_{g \in S} \|f + g\|_{k+1,X}.$$

We now consider the case that the positive definite kernel  $K$  is the polynomial kernel. First, we have the following inequality using the norm equivalence theorem [12]

$$\|f\|_{k+1,X} \leq c_4 \left[ |f|_{k+1,X} + \sum_{|\alpha| \leq k} \left| \int_X \partial^\alpha f(x) dx \right| \right], \tag{4.2}$$

$$\forall f \in H^{k+1}(X).$$

Since the kernel is a polynomial kernel and  $k + 1 > d$ , we have that for all  $g \in S$ ,  $\partial^\alpha g = 0$  for  $|\alpha| = k + 1$ . Replacing  $f$  by  $f + g$  in (4.2) gives us that  $\forall f \in H^{k+1}(X), g \in S$ ,

$$\begin{aligned} & \|f + g\|_{k+1,X} \\ & \leq c_4 \left[ |f + g|_{k+1,X} + \sum_{|\alpha| \leq k} \left| \int_X \partial^\alpha (f + g) dx \right| \right] \\ & = c_4 \left[ |f|_{k+1,X} + \sum_{|\alpha| \leq k} \left| \int_X \partial^\alpha (f + g) dx \right| \right]. \end{aligned}$$

For the term  $\int_X \partial^\alpha (f + g) dx$ , we should note that for any  $f \in H^{k+1}(X)$ , one can always have a  $g \in S$  such that for  $|\alpha| \leq k$ ,

$$\int_X \partial^\alpha (f + g) dx = 0. \tag{4.3}$$

This is because when we set  $|\alpha| = k, k - 1, k - 2, \dots$ , we will have a set of equations from (4.3). We can then solve the set of equations to obtain the coefficients of the polynomial  $g \in S$ . Besides, we note that such set of equations is always solvable because we assumed that  $k + 1 > d$ . Then the  $g$  which was constructed by such way will annihilate the integral  $\int_X \partial^\alpha (f + g) dx$  for  $|\alpha| \leq k$ .

Therefore, we have that

$$\inf_{g \in S} \|f + g\|_{k+1,X} \leq c_4 |f|_{k+1,X},$$

for some constant  $c_4 > 0$ . Then by the conclusion we obtained from the first part, we know that there exists some  $c > 0$  such that

$$|f - I_X f|_{m,X} \leq c |f|_{k+1,X}.$$

This completes the proof of the result. □

From this error bound, we can see that if we consider the polynomial kernel and the true function is in the RKHS  $\mathcal{H}_K$ , we will have

$$|f - I_X f|_{m,X} \leq c \|f\|_{\mathcal{H}_K},$$

for some constant  $c > 0$ . This is because the RKHS  $\mathcal{H}_K$  is continuously embedded in  $H^{k+1}(X)$  and we then have  $|f|_{k+1,X} \leq c \|f\|_{\mathcal{H}_K}$ .

## 5. CONCLUSION

In this paper, some error bounds for the kernel based regularized interpolation are provided. We derived these error bounds in terms of the reproducing kernel Hilbert space norm and Sobolev space norm. We also introduced a error bound for a special type of a commonly used kernel: polynomial kernel. The error bounds then provide theoretical understandings of the approximation performance.

**Conflicts of Interest:** The author(s) declare that there are no conflicts of interest regarding the publication of this paper.

## REFERENCES

- [1] M. Belkin and P. Niyogi, Semi-supervised learning on riemannian manifolds, *Mach. Learn.* 56 (2004), 209–239.
- [2] M. Belkin, P. Niyogi, and V. Sindhwani, Manifold regularization: A geometric framework for learning from labeled and unlabeled examples, *J. Mach. Learn. Res.* 7 (2006), 2399–2434.
- [3] R. Bhatia and C. Davis, A cauchy-schwarz inequality for operators with applications, *Linear Algebra Appl.* 223 (1995), 119–129.
- [4] N. Cressie, The origins of kriging, *Math. Geol.* 22 (1990), 239–252.
- [5] F. R. Deutsch, *Best approximation in inner product spaces*, Springer Science & Business Media, 2012.
- [6] T. Kato, Estimation of iterated matrices, with application to the von neumann condition, *Numer. Math.* 2 (1960), 22–29.
- [7] H.-J. Rong, G.-B. Huang, N. Sundararajan, and P. Saratchandran, Online sequential fuzzy extreme learning machine for function approximation and classification problems, *IEEE Trans. Syst. Man Cybern. Part B (Cybernetics)*, 39 (2009), 1067–1072.
- [8] B. Schölkopf, R. Herbrich, and A. J. Smola, A generalized representer theorem, in *International conference on computational learning theory*, Springer, 2001, 416–426.
- [9] D. Shepard, A two-dimensional interpolation function for irregularly-spaced data, in *Proceedings of the 1968 23rd ACM national conference*, 1968, 517–524.
- [10] M. Thakur, B. Samanta, and D. Chakravarty, A non-stationary geostatistical approach to multigaussian kriging for local reserve estimation, *Stoch. Environ. Res. Risk Assess.* 32 (2018), 2381–2404.
- [11] J. J. Thiagarajan, P.-T. Bremer, and K. N. Ramamurthy, Multiple kernel interpolation for inverting non-linear dimensionality reduction and dimension estimation, in *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, 2014, 6751–6755.
- [12] A. Wannebo, Equivalent norms for the sobolev space  $W_0^{m,p}(\Omega)$ , *Ark. Mat.* 32 (1994), 245–254.