On the Convergence of Irregular Sampling in Reproducing Kernel Hilbert Spaces

Armin Iske
Department of Mathematics, University of Hamburg, Germany
armin.iske@uni-hamburg.de

***Abstract*—We analyse the convergence of sampling algorithms for functions in reproducing kernel Hilbert spaces (RKHS). To this end, we discuss approximation properties of kernel regression under minimalistic assumptions on both the kernel and the input data. We first prove error estimates in the kernel's RKHS norm. This leads us to new results concerning uniform convergence of kernel regression on compact domains. For Lipschitz continuous and Hölder continuous kernels, we prove convergence rates.**

I. Introduction

Learning theory [4], [5] requires the approximation of an unknown function $f : \Omega \longrightarrow \mathbb{R}$ from irregular samples of f on a compact domain $\Omega \subset \mathbb{R}^d$, for $d > 1$. From the viewpoint of statistical learning theory [13], the general purpose of learning is referred to *data regression* (or *data fitting*), where the basic task is to determine a regression function $g : \Omega \longrightarrow \mathbb{R}$ from samples $f_X = \{f(x)\}_{x \in X}$ taken at sampling points $X \subset \Omega$.

Reproducing kernels provide popular concepts for data regression in machine learning [9], in particular for support vector machines [11]. In this case, the target f is assumed to lie in a Hilbert space $\mathcal{H}_{K,\Omega}$ of functions, being generated by a (conditionally) positive definite kernel function K on Ω.

The theory on *reproducing kernel Hilbert spaces* (RKHS) is dating back to the seminal work [2] of Aronszajn (in 1950). Contemporary questions on kernel-based learning are concerning approximation properties of kernel regression [5], [10], [15]. Quite recently, dimensionality reduction in kernel regression has been investigated in [7] from the viewpoint of statistics. Moreover, convergence rates and stability results for a general high-dimensional kernel regression framework were proven in [6], where rather specific assumptions on both the kernel K and the sampling points X were essentially needed.

In this work, we analyse the convergence of kernel regression in RKHS under *minimalistic* assumptions on the kernel K, and so on the RKHS $\mathcal{H}_{K,\Omega}$, and on the sampling points X.

The outline of this paper is as follows. We first explain key features on RKHS (in Section II) and on kernel regression (in Section III). Then, we formulate minimalistic assumptions for K and X (in Section IV), under which we can prove convergence of kernel regression (in Section V) with respect to the kernel's RKHS norm and for uniform convergence.

II. Three Key Features of Kernel Regression

Starting point for our discussion on kernel regression are positive definite functions (for details refer to [3], [8], [14]).

Definition 1. *For $\Omega \subset \mathbb{R}^d$, a continuous and symmetric function $K : \Omega \times \Omega \longrightarrow \mathbb{R}$ is said to be a* positive definite kernel *on Ω, $K \in \mathbf{PD}(\Omega)$, if for any finite set of pairwise distinct points $X = \{x_1, \ldots, x_n\} \subset \Omega$, $n \in \mathbb{N}$, the matrix*

$$A_{K,X} = (K(x_k, x_j))_{1 \le j,k \le n} \in \mathbb{R}^{n \times n}$$

is symmetric and positive definite.

Positive definite kernels on $\mathbb{R}^d$ are often required to be *translation invariant*, i.e., K is assumed to have the form

$$K(x, y) = \Phi(x - y) \qquad \text{for } x, y \in \mathbb{R}^d \tag{1}$$

for an even function $\Phi : \mathbb{R}^d \longrightarrow \mathbb{R}$. Popular examples for translation invariant kernels $K \in \mathbf{PD}(\mathbb{R}^d)$ include the *Gaussian* $K(x, y) = \Phi(x - y) = \exp(-\|x - y\|_2^2)$, and the *inverse multiquadric* $K(x, y) = \Phi(x - y) = (1 + \|x - y\|_2^2)^{-1/2}$, where $\|\cdot\|_2$ denotes as usual the Euclidean norm on $\mathbb{R}^d$.

Next we explain the basic setup of kernel regression in learning theory [15]. To this end, for fixed domain $\Omega \subset \mathbb{R}^d$, let $K : \Omega \times \Omega \longrightarrow \mathbb{R}$ be positive definite on Ω, i.e., $K \in \mathbf{PD}(\Omega)$. In the following discussion, it will be convenient to let the function $K_x : \Omega \longrightarrow \mathbb{R}$, for $x \in \Omega$, be defined as

$$K_x(y) := K(x, y) \qquad \text{for } x, y \in \Omega.$$

Then, according to the seminal work of Aronszajn [2], the *reproducing kernel Hilbert space* (RKHS) $\mathcal{H}_{K,\Omega}$ associated with $K \in \mathbf{PD}(\Omega)$ is the closure

$$\mathcal{H}_{K,\Omega} := \overline{\operatorname{span}\{K_x : x \in \Omega\}}$$

with respect to the inner product $(\cdot,\cdot)_K \equiv (\cdot,\cdot)_{\mathcal{H}_{K,\Omega}}$ satisfying

$$(K_x, K_y)_K = K(x, y) \qquad \text{for all } x, y \in \mathbb{R}^d,$$

whereby we have

$$\begin{aligned} \left\| \sum_{j=1}^n c_j K_{x_j} \right\|_K^2 \quad &:= \quad \left(\sum_{j=1}^n c_j K_{x_j}, \sum_{k=1}^n c_k K_{x_k} \right)_K \\ &= \quad \sum_{j,k=1}^n c_j K(x_j, x_k) c_k = c^\top A_{K,X} c, \end{aligned}$$

for all $X = \{x_1, \ldots, x_n\} \subset \Omega$ and $c = (c_1, \ldots, c_n)^\top \in \mathbb{R}^n$.

Now let us recall three key features of kernel regression.

Feature 1: The *reproducing kernel property*

$$f(x) = (K_x, f)_K \qquad \text{for all } x \in \Omega \tag{2}$$

holds for all $f \in \mathcal{H}_{K,\Omega}$. In particular, for any $K_x \in \mathcal{H}_{K,\Omega}$,

$$K_x(y) = (K_y, K_x)_K = K(x,y) \text{ for all } x, y \in \Omega, \quad (3)$$

holds, whereby we have $K_x(y) = K_y(x)$, for all $x, y \in \Omega$.

The reproducing kernel properties (2) and (3) lead us to

$$\begin{aligned} |f(x) - f(y)|^2 &= |(K_x - K_y, f)_K|^2 \\ &\leq \|K_x - K_y\|_K^2 \cdot \|f\|_K^2 \\ &= (K_x(x) - 2K_x(y) + K_y(y)) \cdot \|f\|_K^2, \end{aligned}$$

which immediately implies the continuity of $f \in \mathcal{H}_{K,\Omega}$, from the continuity of K on $\Omega \times \Omega$. Therefore, the reproducing kernel Hilbert space $\mathcal{H}_{K,\Omega}$ of K is embedded in the continuous functions on Ω, i.e., $\mathcal{H}_{K,\Omega} \subset \mathscr{C}(\Omega)$.

Feature 2: For $X = \{x_1, \ldots, x_n\} \subset \Omega$ we let

$$\mathcal{S}_{K,X} := \operatorname{span}\{K_x : x \in X\} \subset \mathcal{H}_{K,\Omega}$$

denote the n-dimensional subspace of $\mathcal{H}_{K,\Omega}$ spanned by X. Then, the *orthogonal projection* of $f \in \mathcal{H}_{K,\Omega}$ onto $\mathcal{S}_{K,X}$ is the unique interpolant $s_{f,X} \in \mathcal{S}_{K,X}$ to f on X. In other words, the interpolant $s_{f,X}$ to f on X is the unique best approximation to f with respect to the RKHS norm $\|\cdot\|_K = (\cdot,\cdot)_K^{1/2}$, i.e.,

$$\|s_{f,X} - f\|_K \leq \|s - f\|_K \qquad \text{for all } s \in \mathcal{S}_{K,X}.$$

In conclusion, the interpolant $s_{f,X}$ is the best regression fit to $f \in \mathcal{H}_{K,\Omega}$ from data $f_X = (f(x_1), \ldots, f(x_n))^\top \in \mathbb{R}^n$. Moreover, $s_{f,X} \in \mathcal{S}_{K,X}$ has the form

$$s_{f,X} = \sum_{j=1}^{n} c_j K_{x_j}$$

where the coefficient vector $c = (c_1, \ldots, c_n)^\top \in \mathbb{R}^n$ is the unique solution of the linear system $A_{K,X} c = f_X$, due to the interpolation conditions $s_{f,X}(x_k) = f(x_k)$, for all $1 \leq k \leq n$.

Feature 3: The orthogonality $s_{f,X} - f \perp \mathcal{S}_{K,X}$ implies

$$\|s_{f,X}\|_K \leq \|f\|_K \quad \text{and} \quad \|s_{f,X} - f\|_K \leq \|f\|_K, \quad (4)$$

due to the *Pythagoras theorem*

$$\|f\|_K^2 = \|s_{f,X} - f\|_K^2 + \|s_{f,X}\|_K^2.$$

In other words, the kernel regression $s_{f,X} \in \mathcal{S}_{K,X}$ minimizes the RKHS norm $\|\cdot\|_K$ among all interpolants to the samples f_X from $\mathcal{H}_{K,\Omega}$. Therefore, kernel regression can be viewed as a spline approximation method.

III. Problem Formulation and Further Notations

Let $X = (x_k)_{k \in \mathbb{N}}$ be a sequence of pairwise distinct points in Ω. We use the notation $X_n = \{x_1, \ldots, x_n\} \subset \Omega$ for the (ordered) point set containing the first n points in X.

Recall that each point set $X_n \subset \Omega$ spans a finite dimensional regression space $\mathcal{S}_{K,X_n}$. Moreover, recall that for any target $f \in \mathcal{H}_{K,\Omega}$ there is one unique minimizer $s_{f,X_n} \in \mathcal{S}_{K,X_n}$ of the kernel regression error

$$\eta_n \equiv \eta_n(f, \mathcal{S}_{K,X_n}) := \|s_{f,X_n} - f\|_K \qquad \text{for } n \in \mathbb{N}. \quad (5)$$

For notational brevity, we let $s_n := s_{f,X_n}$, for $n \in \mathbb{N}$.

Problem Formulation: We analyze the convergence of kernel regression under minimalistic assumptions. To be more precise, we prove convergence results of the form

$$\|s_n - f\| \longrightarrow 0 \quad \text{for } n \to \infty \quad (6)$$

under mild as possible conditions on the kernel $K \in \mathbf{PD}(\Omega)$, the target $f \in \mathcal{H}_{K,\Omega}$ and the sample points $X = (x_k)_{k \in \mathbb{N}}$. Our convergence analysis is first done with respect to the RKHS norm $\|\cdot\|_K$, before we turn to uniform convergence. For the case of uniform convergence, we prove convergence rates under slightly more restrictive conditions on $K \in \mathbf{PD}(\Omega)$.

In our analysis, the sequence $(h_n)_{n \in \mathbb{N}}$ of *fill distances*

$$h_n \equiv h(X_n, \Omega) := \sup_{y \in \Omega} \min_{x \in X_n} \|y - x\|_2 \qquad \text{for } n \in \mathbb{N} \quad (7)$$

of X_n in Ω will play an important role. Note that the (non-negative) fill distances $(h_n)_{n \in \mathbb{N}}$ of the sequence $X = (x_k)_{k \in \mathbb{N}}$ are monotonically decreasing. We remark already at this point that we can only obtain convergence in (6), if $(h_n)_{n \in \mathbb{N}}$ is a zero sequence, i.e., if $h_n \searrow 0$ for $n \to \infty$.

IV. Minimalistic Assumptions

A. Minimalistic Assumptions on the Kernel

We remark that the required continuity of $K \in \mathbf{PD}(\Omega)$, as stated at the outset of this work, is necessary for the well-posedness of kernel regression on (truly multi-dimensional) domains Ω. This is due to the classical theorem of Mairhuber-Curtis from approximation theory, according to which there are no non-trivial Haar systems on domains $\Omega \subset \mathbb{R}^d$, for $d > 1$, containing bifurcations (cf. [8, Theorem 5.25]).

To prove convergence of kernel regression with respect to $\|\cdot\|_K$, we won't require any further (stricter) assumptions on $K \in \mathbf{PD}(\Omega)$ other than its continuity on $\Omega \times \Omega$. Moreover, we won't require any conditions on Ω. To prove decay rates for uniform convergence, we will merely require local Hölder continuity for $K \in \mathbf{PD}(\Omega)$, cf. Definition 3.

B. Minimalistic Assumptions on the Target Functions

We recall the inclusion $\mathcal{H}_{K,\Omega} \subset \mathscr{C}(\Omega)$ from our discussion on Feature 1 in Section II. In other words, any (admissible) target $f \in \mathcal{H}_{K,\Omega}$ must necessarily be a continuous function. To prove convergence of kernel regression with respect to $\|\cdot\|_K$, we won't require any stricter assumptions on $f \in \mathcal{H}_{K,\Omega}$.

Nevertheless, this gives rise to the question whether or not there is a kernel $K \in \mathbf{PD}(\Omega)$ satisfying $f \in \mathcal{H}_{K,\Omega}$, on given $f \in \mathscr{C}(\Omega)$. The kernel $K(x,y) := f(x) \cdot f(y)$ is only one (trivial) example to give a positive answer for this question.

Another relevant question is the inclusion $\mathscr{C}(\Omega) \subset \mathcal{H}_{K,\Omega}$, i.e., is there a kernel $K \in \mathbf{PD}(\Omega)$, whose RKHS $\mathcal{H}_{K,\Omega}$ contains *all* continuous functions on Ω? If so, this would yield the equality $\mathscr{C}(\Omega) = \mathcal{H}_{K,\Omega}$. Just very recently, Steinwart [12] gave a negative answer on this important question.

C. Minimalistic Assumptions on the Sampling Points

We require that the monotonically decreasing sequence $(h_n)_{n\in\mathbb{N}}$ of fill distances in (7) is a zero sequence, which is a *necessary* condition for the convergence of kernel regression.

In fact, if $(h_n)_{n\in\mathbb{N}}$ is not a zero sequence, then there must be one $h_0 > 0$ satisfying $h_n \geq h_0$ for all $n \in \mathbb{N}$. But this implies that there is one open ball $B(y, h_0) \subset \mathbb{R}^d$ centered at $y \in \Omega$ with radius $h_0 > 0$ which does not contain any point from the sequence $X = (X_k)_{k\in\mathbb{N}}$. Now let $f \in \mathcal{H}_{K,\Omega}$ be compactly supported with $\mathrm{supp}(f) \subset B(y, h_0)$ and $f \not\equiv 0$. In this case, we have $f_{X_n} = 0$, which implies $s_n = s_{f,X_n} \equiv 0$, and so $\|s_n - f\|_K = \|f\|_K > 0$, for all $n \in \mathbb{N}$, i.e., the sequence of kernel regressions $(s_n)_{n\in\mathbb{N}}$ cannot convergence to f.

V. Convergence of Kernel Regression

Now let us analyze the asymptotic behaviour of the kernel regression errors $(\eta_n)_{n\in\mathbb{N}}$ in (5) for the RKHS norm $\|\cdot\|_K$ and for the maximum norm $\|\cdot\|_\infty$, respectively. To this end, we rely on our previous work [8, Section 8.4.2]. For more recent results concerning the convergence of generalized kernel-based interpolation schemes, we refer to [1].

A. Convergence with respect to the RKHS Norm

The following result (cf. [8, Theorem 8.37]) relies on *minimalistic* assumptions on the sampling points $(x_n)_{n\in\mathbb{N}}$ and on the kernel $K \in \mathbf{PD}(\Omega)$, as they were stated in Section IV.

Theorem 2. *Let $X = (x_n)_{n\in\mathbb{N}}$ be a sequence of pairwise distinct points, whose associated fill distances $(h_n)_{n\in\mathbb{N}}$ in* (7) *are a zero sequence. Then, for any $f \in \mathcal{H}_{K,\Omega}$ we have*

$$\eta_K(f, \mathcal{S}_{K,X_n}) = \|s_n - f\|_K \longrightarrow 0 \quad \textit{for } n \to \infty.$$

Proof. Let $y \in \Omega$. By our assumption on X, there is a subsequence $(x_{n_k})_{k\in\mathbb{N}}$ of sampling points $x_{n_k} \in \Omega$ satisfying $\|y - x_{n_k}\|_2 \leq h_{n_k} \longrightarrow 0$ for $k \to \infty$. This implies

$$\begin{aligned}\eta_K^2(K_y, \mathcal{S}_{K,X_{n_k}}) &\leq \|K_{x_{n_k}} - K_y\|_K^2 \\ &= K(x_{n_k}, x_{n_k}) - 2K(y, x_{n_k}) + K(y,y) \\ &\longrightarrow 0\end{aligned}$$

for $k \to \infty$, due to the continuity of $K \in \mathbf{PD}(\Omega)$ on $\Omega \times \Omega$.

Now, for a finite sequence $Y = (y_1, \ldots, y_m) \in \Omega^m$ of pairwise distinct points in Ω, we regard the function

$$f_{c,Y} := \sum_{j=1}^m c_j K_{y_j} \in \mathcal{S}_{K,Y} \subset \mathcal{H}_{K,\Omega}$$

whose coefficient vector is $c = (c_1, \ldots, c_m)^\top \in \mathbb{R}^m$.

For any y_j, $1 \leq j \leq m$, there is a subsequence $(x_n^{(j)})_{n\in\mathbb{N}}$ in X satisfying $\|y_j - x_n^{(j)}\|_2 \leq h_n$. Then, the sequence $(s_{c,n})_{n\in\mathbb{N}}$ of kernel regressions

$$s_{c,n} := \sum_{j=1}^m c_j K_{x_n^{(j)}} \qquad \text{for } n \in \mathbb{N}$$

converges to $f_{c,Y}$, i.e., $s_{c,n} \longrightarrow f_{c,Y}$, for $n \to \infty$, by

$$\begin{aligned}\|s_{c,n} - f_{c,Y}\|_K &= \left\| \sum_{j=1}^m c_j \left(K_{x_n^{(j)}} - K_{y_j} \right) \right\|_K \\ &\leq \sum_{j=1}^m |c_j| \cdot \|K_{x_n^{(j)}} - K_{y_j}\|_K \longrightarrow 0.\end{aligned}$$

Thereby, kernel regression converges on the dense subset

$$\mathcal{S}_{K,\Omega} := \{f_{c,Y} \in \mathcal{S}_{K,Y} : |Y| < \infty\} \subset \mathcal{H}_{K,\Omega},$$

and so, as stated, also on $\mathcal{H}_{K,\Omega}$ by continuous extension. □

We remark that the convergence of Theorem 2 may be arbitrarily slow. Indeed, for any monotonically decreasing zero sequence $(\eta_n)_{n\in\mathbb{N}}$ of non-negative numbers, i.e., $\eta_n \searrow 0$, there is a point sequence $X = (x_k)_{k\in\mathbb{N}}$ in Ω satisfying $h_n \searrow 0$, and $f \in \mathcal{H}_{K,\Omega}$ satisfying $\eta_K(f, \mathcal{S}_{K,X_n}) \geq \eta_n$ for large enough $n \in \mathbb{N}$. Since this is immaterial here, we omit further details.

B. Uniform Convergence

Now we analyze the convergence of kernel regression with respect to the maximum norm $\|\cdot\|_\infty$. Recall the inclusion $\mathcal{H}_{K,\Omega} \subset \mathscr{C}(\Omega)$, whereby $\|\cdot\|_\infty$ is well-defined on $\mathcal{H}_{K,\Omega}$. We further remark that $\|\cdot\|_\infty$ is *weaker* than the RKHS norm $\|\cdot\|_K$, provided that $K \in \mathbf{PD}(\Omega)$ is bounded on $\Omega \times \Omega$. This is due to the reproducing kernel property in (2), whereby

$$\begin{aligned}|(s_n - f)(x)|^2 &= |(s_n - f, K_x)_K|^2 \\ &\leq \|s_n - f\|_K^2 \cdot \|K_x\|_K^2 \\ &= \|s_n - f\|_K^2 \cdot K(x,x)\end{aligned}$$

for all $x \in \Omega$, which in turn implies

$$\|s_n - f\|_\infty \leq \|s_n - f\|_K \cdot \|\sqrt{K}\|_\infty \qquad \text{for all } f \in \mathcal{H}_{K,\Omega}.$$

To prove uniform convergence of kernel regression, we rely on *local* α-Hölder continuity for $K \in \mathbf{PD}(\Omega)$, where $\alpha > 0$.

Definition 3. *For $\Omega \subset \mathbb{R}^d$, let $K \in \mathbf{PD}(\Omega)$. Then, K is said to be* locally α-Hölder continuous *on Ω, for $\alpha > 0$, if every function K_x, for $x \in \Omega$, is locally α-Hölder continuous on Ω, i.e., for any $x \in \Omega$ we have*

$$|K_x(y_1) - K_x(y_2)| \leq C\|y_1 - y_2\|_2^\alpha$$

for all $y_1, y_2 \in \Omega$ satisfying $\|y_1 - y_2\|_2 < r$, for small enough $r > 0$, and for some $C > 0$. For $\alpha = 1$, we say that K is locally Lipschitz continuous *on Ω.*

Before we continue our error analysis on kernel regression, let us first remark two relevant properties of locally α-Hölder continuous kernels $K \in \mathbf{PD}(\Omega)$.

Remark 4. *A translation invariant kernel $K \in \mathbf{PD}(\Omega)$ of the form (1), i.e., $K(x,y) = \Phi(x - y)$ for $x, y \in \Omega$, is locally α-Hölder continuous on Ω, iff the function Φ is locally α-Hölder continuous by satisfying the growth condition*

$$|\Phi(z_1) - \Phi(z_2)| \leq C\|z_1 - z_2\|_2^\alpha$$

for all $z_1, z_2 \in \Omega$ with $\|z_1 - z_2\|_2 < r$, for small enough $r > 0$ and for some $C > 0$.

Remark 5. *A positive definite kernel $K \in \mathbf{PD}(\Omega)$ on open $\Omega \subset \mathbb{R}^d$ can only be locally α-Hölder continuous for $\alpha \leq 1$. Indeed, for $\alpha > 1$, and for any (fixed) $x \in \Omega$, the local estimate*

$$\frac{|K_x(y_1) - K_x(y_2)|}{\|y_1 - y_2\|_2} \leq C \|y_1 - y_2\|_2^{\alpha - 1}$$

holds for all $y_1, y_2 \in \Omega$, $y_1 \neq y_2$, with $\|y_1 - y_2\|_2 < r$ for small enough $r > 0$ and for some $C > 0$. But this means that all directional derivatives of K_x must vanish at all points in Ω, due to the mean value theorem. In this case, K is constant on $\Omega \times \Omega$, so that K cannot be positive definite, i.e., $K \notin \mathbf{PD}(\Omega)$.

From now on, we assume that $K \in \mathbf{PD}(\Omega)$ is locally α-Hölder continuous for $\alpha \in (0, 1]$. Note that this condition on K is slightly more restrictive than the *minimalistic* assumption of continuity for K on $\Omega \times \Omega$ in Section IV-A.

Now we show that all functions in the RKHS $\mathcal{H}_{K,\Omega}$ are *locally* $\alpha/2$-Hölder continuous, if $K \in \mathbf{PD}(\Omega)$ is locally α-Hölder continuous on Ω.

Lemma 6. *For $\Omega \subset \mathbb{R}^d$, let $K \in \mathbf{PD}(\Omega)$ be locally α-Hölder continuous on Ω, for some $\alpha \in (0, 1]$. Then, all functions in $\mathcal{H}_{K,\Omega}$ are locally $\alpha/2$-Hölder continuous on Ω.*

Proof. Let $f \in \mathcal{H}_{K,\Omega}$ and $x \in \Omega$ be fixed. Then, we have

$$\begin{aligned}
|f(x) - f(y)|^2 &= |(K_x - K_y, f)_K|^2 \leq \|K_x - K_y\|_K^2 \, \|f\|_K^2 \\
&= (K_x(x) - K_x(y) + K_y(y) - K_y(x)) \, \|f\|_K^2 \\
&\leq 2\, C \|x - y\|_2^\alpha \, \|f\|_K^2
\end{aligned}$$

for some $C > 0$, and where $y \in \Omega$, $x \neq y$, is required to satisfy $\|x - y\|_2 < r$, for $r > 0$ small enough. ☐

From Lemma 6, we can directly conclude the following error estimate for kernel regression from finite sampling points.

Proposition 7. *For $\alpha \in (0, 1]$, let $K \in \mathbf{PD}(\Omega)$ be locally α-Hölder continuous on $\Omega \subset \mathbb{R}^d$. Moreover, let $X \subset \Omega$ be a finite subset of Ω. Then, we have for any $f \in \mathcal{H}_{K,\Omega}$ the error estimate*

$$\|s_{f,X} - f\|_\infty \leq \sqrt{2 C h_{X,\Omega}^\alpha} \cdot \|f\|_K,$$

where $s_{f,X}$ denotes the interpolant to f on X.

Proof. Suppose $y \in \Omega$. Then, there is one $x \in X$ satisfying $\|y - x\|_2 \leq h_{X,\Omega}$. By using $(s_{f,X} - f)(x) = 0$ we can conclude

$$\begin{aligned}
|(s_{f,X} - f)(y)|^2 &= |(s_{f,X} - f)(y) - (s_{f,X} - f)(x)|^2 \\
&\leq 2 C h_{X,\Omega}^\alpha \cdot \|s_{f,X} - f\|_K^2 \\
&\leq 2 C h_{X,\Omega}^\alpha \cdot \|f\|_K^2
\end{aligned}$$

from Lemma 6, where we further used

$$\|s_{f,X} - f\|_K \leq \|f\|_K,$$

from the stability estimates in (4). ☐

Finally, our next result follows directly from Proposition 7.

Corollary 8. *For $\alpha \in (0, 1]$, let $K \in \mathbf{PD}(\Omega)$ be locally α-Hölder continuous on $\Omega \subset \mathbb{R}^d$. Moreover, let $X = (x_k)_{k \in \mathbb{N}}$ be a sequence of pairwise distinct points in Ω, whose corresponding sequence $(h_n)_{n \in \mathbb{N}}$ of fill distances $h_n = h(X_n, \Omega)$, as in (7), is a zero sequence, i.e., $h_n \searrow 0$, for $n \to \infty$. Then, the uniform convergence*

$$\|s_n - f\|_\infty = \mathcal{O}\left(h_n^{\alpha/2}\right) \quad \text{for } n \to \infty$$

holds for all $f \in \mathcal{H}_{K,\Omega}$ at convergence rate $\alpha/2$.

VI. Conclusion and Future Work

We have proven convergence of kernel regression from irregular samples in reproducing kernel Hilbert spaces (RKHS), under minimalistic assumptions (cf. Section IV) on the kernel K, its RKHS $\mathcal{H}_{K,\Omega}$, and the sampling points $X \subset \Omega$.

Now it may be inspiring to work on even weaker conditions for K and X, under which kernel regression is convergent.

Yet, it remains to analyse conditions for $K \in \mathbf{PD}(\Omega)$, under which given functions $f \in \mathscr{C}(\Omega)$ lie in $\mathcal{H}_{K,\Omega}$, so that kernel regression converges to f (due to Theorem 2). And if so, i.e., if $f \in \mathcal{H}_{K,\Omega}$, can we then conclude properties of K from properties of f? E.g. if $f \in \mathcal{H}_{K,\Omega}$ is (locally) α-Hölder continuous, for $\alpha \in (0, 1]$, can we then conclude that K is also (locally) α-Hölder continuous, so that f can be approximated at convergence rate $\alpha/2$ (due to Corollary 8)?

References

[1] K. Albrecht and A. Iske: On the convergence of generalized kernel-based interpolation by greedy data selection algorithms. *BIT Numer. Math.* **65**, 5 (2025). https://doi.org/10.1007/s10543-024-01048-3.

[2] N. Aronszajn: Theory of reproducing kernels. *Transactions of the American Mathematical Society* **68** (3), 1950, 337–404.

[3] M.D. Buhmann: *Radial Basis Functions.* Cambridge University Press, Cambridge, UK, 2003.

[4] F. Cucker and S. Smale: On the mathematical foundations of learning. *Bull. Amer. Math. Soc.* **39** 2001, 1–49,.

[5] F. Cucker and D.-X. Zhou: *Learning Theory: An Approximation Theory Viewpoint.* Cambridge University Press, 2007.

[6] S. Eckstein, A. Iske, and M. Trabs: Dimensionality reduction and Wasserstein stability for kernel regression. *Journal of Machine Learning Research* **24** (334), 2023, 1–35.

[7] K. Fukumizu, F. Bach, and M. Jordan: Kernel dimension reduction in regression. *Annals of Statistics* **37**, 2009, 1871–1905.

[8] A. Iske: *Approximation Theory and Algorithms for Data Analysis.* Texts in Applied Mathematics, vol. 68, Springer, Cham, 2018.

[9] B. Schölkopf and A.J. Smola: *Learning with Kernels.* MIT Press, Cambridge, 2002.

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

[11] I. Steinwart and A. Christmann: *Support Vector Machines.* Springer, New York, 2008.

[12] I. Steinwart: Reproducing kernel Hilbert spaces cannot contain all continuous functions on a compact metric space. *Archiv der Mathematik* **122**, 2024, 553–557.

[13] V. Vapnik: *Statistical Learning Theory*. Wiley, New York, 1998.

[14] H. Wendland: *Scattered Data Approximation.* Cambridge University Press, Cambridge, UK, 2005.

[15] D.-X. Zhou: Capacity of reproducing kernel spaces in learning theory. *IEEE Transactions on Information Theory* **49** (7), July 2003, 1743–1752.