Max-Planck-Institut
für Mathematik
in den Naturwissenschaften
Leipzig

Sampling Inequalities for Infinitely Smooth
Functions, with Applications to Interpolation
and Machine Learning

(revised version: April 2008)

by

Christian Rieger, and Barbara Zwicknagl

Sampling Inequalities for Infinitely Smooth Functions, with Applications to Interpolation and Machine Learning

Christian Rieger* and Barbara Zwicknagl†

April 4, 2008

Abstract

Sampling inequalities give a precise formulation of the fact that a differentiable function cannot attain large values, if its derivatives are bounded and if it is small on a sufficiently dense discrete set. Sampling inequalities can be applied to the difference of a function and its reconstruction in order to obtain (sometimes optimal) convergence orders for very general possibly regularized recovery processes. So far, there are only sampling inequalities for finitely smooth functions, which lead to algebraic convergence orders. In this paper, the case of infinitely smooth functions is investigated, in order to derive error estimates with exponential convergence orders.

Keywords: Gaussians, inverse multiquadrics, smoothing, approximation, error bounds, radial basis functions, convergence orders

Classification: 41A05, 41A25, 41A63, 65D10, 68T05

1 Introduction

Sampling inequalities provide a quantitative formulation of the observation that a differentiable function cannot attain large values anywhere if its derivatives are bounded and if it is small on a sufficiently dense discrete set. Inequalities of this kind can be used to derive a priori error estimates for various regularized approximation problems as they occur for instance in many machine learning algorithms or PDE solvers (see, e.g., [11, 10] and the references therein).

Recently several such *sampling inequalities* for functions u from certain Sobolev spaces $W_p^k(\Omega)$ with $1 \leq p < \infty$ on a bounded domain $\Omega \subset \mathbb{R}^d$ were obtained. The Sobolev space $W_p^k(\Omega)$ consists of all functions u with distributional derivatives $D^\alpha u \in L_p(\Omega)$ for all $\alpha \in \mathbb{N}_0^d$ with $|\alpha| \leq k$. Associated with these spaces are the (semi-)norms

$$|u|_{W_p^k(\Omega)} = \left(\sum_{|\alpha|=k} \|D^\alpha u\|_{L_p(\Omega)}^p \right)^{1/p} \quad \text{and} \quad \|u\|_{W_p^k(\Omega)} = \left(\sum_{|\alpha|\leq k} \|D^\alpha u\|_{L_p(\Omega)}^p \right)^{1/p}.$$

---

*Institut für Numerische und Angewandte Mathematik, Universität Göttingen, Lotzestr. 16-18, 37083 Göttingen, Germany, email: crieger@math.uni-goettingen.de

†Max Planck Institute for Mathematics in the Sciences, Inselstr. 22, 04103 Leipzig, Germany, email: Barbara.Zwicknagl@mis.mpg.de

For a discrete subset $X = \{x_1, \ldots, x_N\}$ of Ω, we denote the associated fill distance with

$$h := h_{X,\Omega} := \sup_{x \in \Omega} \min_{x_j \in X} \|x - x_j\|_2.$$

If we assume $1 \le q \le \infty$, $\alpha \in \mathbb{N}_0^d$, $k \in \mathbb{N}$, and $1 \le p < \infty$ with $k > |\alpha| + d/p$ if $p > 1$, or with $k \ge |\alpha| + d$ if $p = 1$, typical sampling inequalities take the form (see [17, Theorem 2.6])

$$\|D^\alpha u\|_{L_q(\Omega)} \le C \left(h^{k-|\alpha|-d(\frac{1}{p}-\frac{1}{q})_+} |u|_{W_p^k(\Omega)} + h^{-|\alpha|} \|u|_X\|_{\ell_\infty(X)} \right),$$

or for $\alpha = 0$ (see [6, Theorem 3.5])

$$\|u\|_{L_p(\Omega)} \le C \left(h^k |u|_{W_p^k(\Omega)} + h^{d/p} \|u|_X\|_{\ell_p(X)} \right),$$

for all $u \in W_p^k(\Omega)$ and all discrete sets $X \subset \Omega$ with sufficiently small fill distance $h := h_{X,\Omega}$, where the constants C do not depend on u, X or h. In [17, Section 3], these bounds were used to derive optimal algebraic convergence orders for kernel based smoothed interpolation methods in a finitely smooth setting.

In this paper, we derive sampling inequalities for infinitely smooth functions where the convergence orders turn out to depend exponentially on the fill distance h. We are handling infinitely smooth functions by normed linear function spaces $\mathcal{H}(\Omega)$ on domains $\Omega \subset \mathbb{R}^d$ that can for some $1 \le p < \infty$ be continuously embedded into every classical Sobolev space $W_p^k(\Omega)$. More precisely, for a fixed $p \in [1, \infty)$ and all $k \in \mathbb{N}$ we assume that there are embedding operators $I_k^{(p)}$ and constants $E(k)$ such that

$$I_k^{(p)} : \mathcal{H}(\Omega) \to W_p^k(\Omega) \quad \text{with}$$
$$\left\| I_k^{(p)} \right\|_{\{\mathcal{H}(\Omega) \to W_p^k(\Omega)\}} \le E(k) \quad \text{for all } k \in \mathbb{N}. \tag{1}$$

There are various examples for spaces with this property, e.g., Sobolev spaces of infinite order as they occur in the study of partial differential equations of infinite order [1], or reproducing kernel Hilbert spaces of Gaussians and inverse multiquadrics (see Section 5).

In the case of infinitely smooth functions the shape of the domain Ω crucially influences our sampling inequalities. For bounded domains Ω obeying a uniform interior cone condition we use a polynomial reproduction from [16], which accepts slight oversampling, to bound the Lebesgue-constants. This results in a good behaviour of the term with the discrete norm. A typical result in this case is that for $\alpha \in \mathbb{N}_0^d$ and $1 \le q \le \infty$, there are constants $C, \tilde{C} > 0$ such that for all discrete sets $X \subset \Omega$ with sufficiently small fill distance h the inequality

$$\|D^\alpha u\|_{L_q(\Omega)} \le e^{C \log(h)/\sqrt{h}} \|u\|_{\mathcal{H}(\Omega)} + \tilde{C} h^{-|\alpha|} \|u|_X\|_{\ell_\infty(X)}$$

holds for all $u \in \mathcal{H}(\Omega)$. Here and throughout the paper, e denotes Euler's constant. The best approximation orders for the first term can be obtained on compact cubes since we can then use a polynomial reproduction based on [5, Lemma 1]. Unfortunately, this approach is limited to cubes and cannot cope with derivatives on the left-hand sides of our sampling inequalities. Nevertheless, we obtain as a typical result, which applies, e.g., to functions

2

from the native spaces of Gaussian kernels, that there are constants $C, \tilde{C} > 0$, such that the inequality

$$\|u\|_{L_q(\Omega)} \le e^{C \log(h)/h} \|u\|_{\mathcal{H}(\Omega)} + e^{\tilde{C}/h} \|u|_X\|_{\ell_\infty(X)}$$

holds for all $u \in \mathcal{H}(\Omega)$, and for all sets $X \subset \Omega$ with sufficiently small fill distance h.

Our main applications deal with reconstruction problems in Hilbert spaces. Therefore, in the second part we will focus on the native Hilbert spaces of Gaussian and inverse multiquadric kernels. In this case, we suppose u to be an error function $u = f - Rf$, where f denotes the function to be reconstructed and Rf is the reconstruction. To obtain optimal order error bounds, one needs two properties of the reconstruction, namely some kind of stability and consistency, that is,

$$\|Rf\|_{\mathcal{H}(\Omega)} \le C\|f\|_{\mathcal{H}(\Omega)} \quad \text{and} \quad \|(Rf - f)|_X\|_{\ell_p(X)} \le g(f, h),$$

where g determines the expected approximation order. These conditions are satisfied in most of the commonly used kernel-based reconstruction methods. The theory presented here in particular reproduces the well-known error estimates for the standard interpolation problem in the native Hilbert spaces of the inverse multiquadrics and Gaussian kernels on cubes [16, Theorem 11.22], and generalizes them to a wider class of domains and to derivatives. One advantage of our estimates in the case of Gaussian kernels on general Lipschitz domains is, that the constant C, crucially influencing the speed of convergence of $\exp\left(C \log(h)/\sqrt{h}\right)$ for $h \to 0$, does not depend on the space dimension d in contrast to the exponential dependence in the well-known estimates $\exp\left(C_d \log(h)/h\right)$ on cubes [16].

2   Estimates on Domains Star-Shaped with Respect to a Ball

Following [7], we first derive algebraic sampling inequalities on bounded domains $\mathcal{D} \subset \mathbb{R}^d$ that are star shaped with respect to a ball. We shall obtain exponential approximation orders by appropriately relating the smoothness k to the fill distance h. Therefore, we repeat some arguments from [7, Section 2] taking special care of the k-dependence of the various constants involved. We assume a domain $\mathcal{D}$ that is star shaped with respect to a ball $B(x_c, r) := \left\{x \in \mathbb{R}^d : |x - x_c| < r\right\}$ and that is contained in a ball $B(x_c, R)$. Then by [7, Proposition 2.1], $\mathcal{D}$ satisfies an interior cone condition with radius r and angle

$$\theta = 2 \arcsin\left(\frac{r}{2R}\right). \tag{2}$$

We shall make use of this property throughout this section. We denote the associated chunkiness parameter with $\gamma = \frac{\delta_{\mathcal{D}}}{\rho_{\max}}$, where $\rho_{\max}$ is the radius of the largest ball relative to which $\mathcal{D}$ is star shaped, and $\delta_{\mathcal{D}}$ denotes the diameter of $\mathcal{D}$.

Let $\left\{a_j^{(\alpha)} : j = 1, \ldots, N\right\}$ be a polynomial reproduction of degree k with respect to a discrete set $X = \{x_1, \ldots, x_N\} \subset \mathcal{D}$, i.e.,

$$D^\alpha p(x) = \sum_{j=1}^{N} a_j^{(\alpha)}(x) p(x_j)$$

3

for all $\alpha \in \mathbb{N}_0^d$ with $|\alpha| \leq k$, all $x \in \mathcal{D}$ and all polynomials $p \in \mathbb{P}_k^d(\mathcal{D})$, where $\mathbb{P}_k^d$ denotes the space of all d-variate polynomials of degree not exceeding k. Then we have

$$|D^\alpha u(x)| \leq |D^\alpha u(x) - D^\alpha p(x)| + |D^\alpha p(x)|$$

$$\leq \|D^\alpha u - D^\alpha p\|_{L_\infty(\mathcal{D})} + \sum_{j=1}^N \left| a_j^{(\alpha)}(x) \right| |p(x_j)|$$

$$\leq \|D^\alpha u - D^\alpha p\|_{L_\infty(\mathcal{D})} + \sum_{j=1}^N \left| a_j^{(\alpha)}(x) \right| \|p|_X\|_{\ell_\infty(X)}$$

$$\leq \|D^\alpha u - D^\alpha p\|_{L_\infty(\mathcal{D})} + \sum_{j=1}^N \left| a_j^{(\alpha)}(x) \right| \left(\|u - p\|_{L_\infty(\mathcal{D})} + \|u|_X\|_{\ell_\infty(X)} \right) \quad (3)$$

for all $u \in W_p^k(\mathcal{D})$, all $x \in \mathcal{D}$, and all polynomials $p \in \mathbb{P}_k^d(\mathcal{D})$.

As a polynomial approximation we use averaged Taylor polynomials $Q^k u \in \mathbb{P}_{k-1}\left(\mathbb{R}^d\right)$ (see [2, Chapter 4] for a detailed overview). They are defined as

$$Q^k u(x) := \sum_{|\alpha| < k} \frac{1}{\alpha!} \int_{B_\rho} D^\alpha u(y)(x - y)^\alpha \phi(y) dy \, .$$

Here, B_ρ is a ball relative to which $\mathcal{D}$ is star shaped and having radius $\rho \geq \frac{1}{2}\rho_{\max}$, the largest radius of a ball relative to which $\mathcal{D}$ is star shaped. Further, $\phi \geq 0$ is a C^∞ "bump" function supported on B_ρ satisfying both $\int_B \phi(y) dy = 1$ and $\max \phi \leq C_d \, \mathrm{diam}(B)^{-d}$. For the remainder $R^k := u - Q^k u$ we use a bound from [2, Proposition 4.3.2], where the explicit constants can be found in [7, Proposition 2.5]. As pointed out earlier, we are particularly interested in the k-dependence of the constants. Further, we have to be careful about how the constants depend on the domain, because in the next section, we will prove estimates on a bounded Lipschitz domain Ω by decomposing Ω into domains $\mathcal{D}_t$ star shaped with respect to a ball.

Lemma 2.1 *Suppose that $\mathcal{D} \subset \mathbb{R}^d$ is bounded, that is, $\mathcal{D} \subset B(x_c, R)$, and star shaped with respect to a ball $B(x_c, r)$, which implies that $\mathcal{D}$ satisfies an interior cone condition with radius r and angle θ as given in equation (2). Suppose $\alpha \in \mathbb{N}_0^d$ and $1 \leq p < \infty$. Then for all $k \in \mathbb{N}$ with $k > |\alpha| + \frac{d}{p} + 1$ if $p > 1$, or with $k \geq |\alpha| + d$ if $p = 1$,*

$$\left\| D^\alpha u - D^\alpha Q^k u \right\|_{L_\infty(\mathcal{D})} \leq \frac{C_{d,p,\theta}^k}{(k - |\alpha|)!} \delta_{\mathcal{D}}^{k - |\alpha| - d/p} |u|_{W_p^k(\mathcal{D})}$$

for all $u \in W_p^k(\mathcal{D})$, where the constant $C_{d,p,\theta}$ depends only on the space dimension d, the angle θ, and on p, but not on k, u, α, $\delta_{\mathcal{D}}$ or r.

Proof: For $|\beta| \leq k - 1$, we use the identity $D^\beta Q^k u = Q^{k-|\beta|} D^\beta u$ from [2, Proposition 4.1.17]. Then [7, Proposition 2.5] gives with the convention $0^0 = 1$ in the case $p = 1$,

$$\left\| D^\alpha u - D^\alpha Q^k u \right\|_{L_\infty(\mathcal{D})} = \left\| D^\alpha u - Q^{k-|\alpha|} D^\alpha u \right\|_{L_\infty(\mathcal{D})}$$

$$\leq c_{d,p}(1 + \gamma)^d \frac{d^{k-|\alpha|-1}}{(k - |\alpha| - 1)!}(k - |\alpha| - \frac{d}{p})^{\frac{1}{p}-1} \delta_{\mathcal{D}}^{k-|\alpha|-\frac{d}{p}} |D^\alpha u|_{W_p^{k-|\alpha|}(\mathcal{D})} \, ,$$

4

where the constant $c_{d,p}$ depends only on d and p. Using $\frac{1}{(k-|\alpha|-1)!} \leq \frac{e^k}{(k-|\alpha|)!}$ with Euler's constant e, we finally obtain

$$\left\| D^\alpha u - D^\alpha Q^k u \right\|_{L_\infty(\mathcal{D})} \leq c_{d,p,\theta} \frac{e^k d^k}{(k-|\alpha|)!} \delta_{\mathcal{D}}^{k-|\alpha|-d/p} |u|_{W_p^k(\mathcal{D})} .$$

Here we used the bound $1 \leq \gamma \leq \csc\left(\frac{\theta}{2}\right)$ from [7, Section 2.1.4], which implies that we can find a constant $c_{d,p,\theta}$ that depends on the domain $\mathcal{D}$ only via the angle θ. Setting $C_{d,p,\theta} := ed\,(c_{d,p,\theta}+1)$ finishes the proof. $\qquad\square$

Note that here and in the following we suppose $k > |\alpha| + \frac{d}{p} + 1$ for $1 < p < \infty$ only for technical reasons to simplify the k-dependence of the constants. The condition $k > |\alpha| + \frac{d}{p}$ is sufficient to derive a constant C depending on k, p, d, θ and α such that

$$\left\| D^\alpha u - D^\alpha Q^k u \right\|_{L_\infty(\mathcal{D})} \leq C \delta_{\mathcal{D}}^{k-|\alpha|-d/p} |u|_{W_p^k(\mathcal{D})} .$$

We shall use a local polynomial reproduction from [15] (see also [16, Theorem 3.8]).

Theorem 2.2 *Suppose $\mathcal{D} \subset \mathbb{R}^d$ is bounded and satisfies an interior cone condition with angle $\theta \in (0, \pi/2)$ and radius r. Let $\ell \in \mathbb{N}_0$ and $\alpha \in \mathbb{N}_0^d$ with $|\alpha| \leq \ell$. Suppose the set $X = \{x_1, \ldots, x_N\} \subset \mathcal{D}$ and $h > 0$ with*

$$h \leq \frac{r \sin \theta}{4 \left(1 + \sin \theta\right) \ell^2} \tag{4}$$

satisfy the condition that every ball $B\,(x, h) \subset \mathcal{D}$ contains at least one point in X. Then, for every $x \in \mathcal{D}$ there exist numbers $\tilde{a}_1^{(\alpha)}(x), \ldots, \tilde{a}_N^{(\alpha)}(x)$ with

1. $\sum_{j=1}^N p(x_j) \tilde{a}_j^{(\alpha)}(x) = D^\alpha p(x)$ *for all* $p \in \mathbb{P}_\ell^d(\mathcal{D})$,

2. $\sum_{j=1}^N |\tilde{a}_j^{(\alpha)}(x)| \leq 2 \left(\frac{2\ell^2}{r \sin \theta}\right)^{|\alpha|} \leq 2 \left(\frac{1}{2(1+\sin\theta)}\right)^{|\alpha|} h^{-|\alpha|} \leq 2h^{-|\alpha|}$.

Remark 2.3 *In the proof of the result in [15], h is only required to satisfy the condition that every closed ball $\overline{B\,(x,h)} \subset \mathcal{D}$ contains at least one point in X. This condition is certainly satisfied if h is the fill distance $h_{X,\mathcal{D}}$. However, when we later cover a domain Ω by local regions $\mathcal{D}$, we may use the global fill distance $h_{X,\Omega}$ instead of the local fill distances $h_{X\cap\mathcal{D},\mathcal{D}}$.*

Inserting the bounds of Lemma 2.1 and Theorem 2.2 into the estimate (3) leads to the following result.

Theorem 2.4 *Suppose $\mathcal{D}$ is bounded, and star shaped with respect to a ball, which implies that $\mathcal{D}$ satisfies an interior cone condition with angle θ and radius r. Then, for all sets $X \subset \Omega$ and all $h > 0$ satisfying the conditions of Theorem 2.2 with $k \in \mathbb{N}$, and for all $u \in W_p^k(\mathcal{D})$,*

$$\|D^\alpha u\|_{L_\infty(\mathcal{D})} \leq \frac{C_S^k}{(k-|\alpha|)!} \delta_{\mathcal{D}}^{k-d/p} \left(\delta_{\mathcal{D}}^{-|\alpha|} + h^{-|\alpha|}\right) |u|_{W_p^k(\mathcal{D})} + 2h^{-|\alpha|} \|u|_X\|_{\ell_\infty(X)} \tag{5}$$

provided that $\alpha \in \mathbb{N}_0^d$ satisfies $k > |\alpha| + \frac{d}{p} + 1$ if $1 < p < \infty$, or $k > |\alpha| + d$ if $p = 1$. The constant C_S depends only on d, p and θ, but not on k, h, u, α, $\delta_{\mathcal{D}}$ or r.

Corollary 2.5 *Under the assumptions from Theorem 2.4 we get for $1 \leq q \leq \infty$*

$$\|D^\alpha u\|_{L_q(\mathcal{D})} \leq vol(\mathcal{D})^{1/q} \|D^\alpha u\|_{L_\infty(\mathcal{D})} \leq \delta_{\mathcal{D}}^{d/q} \|D^\alpha u\|_{L_\infty(\mathcal{D})}$$

$$\leq \frac{C_S^k}{(k-|\alpha|)!} \delta_{\mathcal{D}}^{k-d\left(\frac{1}{p}-\frac{1}{q}\right)} \left(\delta_{\mathcal{D}}^{-|\alpha|} + h^{-|\alpha|}\right) |u|_{W_p^k(\mathcal{D})} + 2\delta_{\mathcal{D}}^{d/q} h^{-|\alpha|} \|u|_X\|_{\ell_\infty(X)} .$$

For infinitely smooth functions $f \in \mathcal{H}(\mathcal{D})$, i.e., $f \in W_p^k(\mathcal{D})$ for all $k \in \mathbb{N}$, we obtain sampling inequalities as in Theorem 2.4 and Corollary 2.5 for almost all $k \in \mathbb{N}$. For such infinitely smooth functions, we shall now derive sampling inequalities with exponential approximation orders by appropriately relating the smoothness k to the fill distance h.

Theorem 2.6 *Suppose that $\mathcal{D} \subset \mathbb{R}^d$ is bounded, and star-shaped with respect to a ball, which implies that $\mathcal{D}$ satisfies an interior cone condition with radius r and angle θ. Suppose further that $\mathcal{H}(\mathcal{D})$ consists of smooth functions $f : \mathcal{D} \to \mathbb{R}$, i.e., for some $1 \leq p < \infty$ and all $k \in \mathbb{N}$ there exist numbers $E(k)$ that may depend on k, $\mathcal{D}$, p and d such that for all $u \in \mathcal{H}(\mathcal{D})$,*

$$\|u\|_{W_p^k(\mathcal{D})} \leq E(k) \|u\|_{\mathcal{H}(\mathcal{D})} .$$

If there are constants $\epsilon, C_E > 0$ such that $E(k) \leq C_E^k k^{(1-\epsilon)k}$ for all $k \in \mathbb{N}$, then, for all $1 \leq q \leq \infty$, there are constants C and $h_0 > 0$ such that for all data sets $X \subset \mathcal{D}$ with fill distance $h \leq h_0$ the inequality

$$\|D^\alpha u\|_{L_q(\mathcal{D})} \leq e^{C\log(h)/\sqrt{h}} \|u\|_{\mathcal{H}(\mathcal{D})} + 2\delta_{\mathcal{D}}^{d/q} h^{-|\alpha|} \|u|_X\|_{\ell_\infty(X)}$$

holds for all $u \in \mathcal{H}(\mathcal{D})$. The constant $C = \epsilon\sqrt{c_0}/4$, with $c_0 = \min\left\{1, \frac{r\sin\theta}{4(1+\sin\theta)}\right\}$, depends only on ϵ, r and θ, but in particular not on d, α, X, h or u, while h_0 may depend on d, p, q, r, θ, α, $\delta_{\mathcal{D}}$, ϵ and C_E, but not on h, X or u.

Proof: We use Stirling's formula to estimate

$$\frac{1}{(k-|\alpha|)!} \leq \frac{k^{|\alpha|}}{k!} \leq \frac{k^{|\alpha|}e^k}{k^k} . \tag{6}$$

If $\|u\|_{W_p^k(\Omega)} \leq C_E^k k^{(1-\epsilon)k} \|u\|_{\mathcal{H}(\Omega)}$ holds for all $k \in \mathbb{N}$, Corollary 2.5 gives with $c_0 := \min\left\{1, \frac{r\sin\theta}{4(1+\sin\theta)}\right\}$ for $h \leq \frac{c_0}{k^2}$ with $k > |\alpha| + 1 + d/p$,

$$\|D^\alpha u\|_{L_q(\mathcal{D})} \leq \mathfrak{c}^k \left(\frac{k}{h}\right)^{|\alpha|} k^{-\epsilon k} \|u\|_{\mathcal{H}(\mathcal{D})} + 2\delta_{\mathcal{D}}^{d/q} h^{-|\alpha|} \|u|_X\|_{\ell_\infty(X)} , \tag{7}$$

where $\mathfrak{c} := C_S C_E e \delta_{\mathcal{D}} \left(\delta_{\mathcal{D}}^{-d\left(\frac{1}{p}-\frac{1}{q}\right)} + 1\right)\left(\delta_{\mathcal{D}}^{-|\alpha|} + 1\right)$ depends only on d, r, θ, $\delta_{\mathcal{D}}$, p, q, C_E and α, but not on k, h or u. If the fill distance h is sufficiently small, i.e.,

$$h \leq c_0 (|\alpha| + 1 + d/p)^{-2} ,$$

6

we can choose $k \in \mathbb{N}$ with $k > |\alpha| + 1 + d/p$ such that

$$\frac{c_0}{2k^2} \leq h \leq \frac{c_0}{k^2} \; .$$

Using $h^{-3|\alpha|/2} \leq e^{3|\alpha|/\sqrt{h}}$, the first term of the right-hand side of (7) can be bounded by

$$\mathfrak{c}^k \left(\frac{k}{h}\right)^{|\alpha|} k^{-\epsilon k} \|u\|_{\mathcal{H}(\mathcal{D})} \leq B^{1/\sqrt{h}} h^{\epsilon\sqrt{2c_0}/\left(4\sqrt{h}\right)} \|u\|_{\mathcal{H}(\mathcal{D})}$$

with $B := (\mathfrak{c}+1)^{\sqrt{c_0}} e^{3|\alpha|} \left(\frac{2}{c_0}\right)^{\epsilon\sqrt{c_0}/2}$. If we set

$$h_0 := \min\left\{ B^{4/\left(\epsilon\sqrt{c_0}\left(1-\sqrt{2}\right)\right)}, \; c_0 \left(|\alpha| + 1 + d/p\right)^{-2} \right\}$$

we finally obtain for all $h \leq h_0$,

$$\|D^\alpha u\|_{L_q(\mathcal{D})} \leq e^{C \log h/\sqrt{h}} \|u\|_{\mathcal{H}(\mathcal{D})} + 2\delta_{\mathcal{D}}^{d/q} h^{-|\alpha|} \|u|_X\|_{\ell_\infty(X)} \; ,$$

where the constant $C := \epsilon\sqrt{c_0}/4$ does not depend on d, X, h, u or α. $\qquad\square$

3 Estimates on Lipschitz Domains Obeying a Cone Condition

We now consider a domain $\Omega \subset \mathbb{R}^d$ that is bounded, has a Lipschitz boundary and satisfies a uniform interior cone condition with maximum radius $R_{\max}$ and angle $\phi \in (0, \pi/2)$. To decompose Ω into domains $\{\mathcal{D}_t\}$ that are star shaped with respect to a ball we use a construction due to Duchon [3], described in detail in [7, Section 2.2]. We define

$$\theta := \theta(\phi) = 2\arcsin\left(\frac{\sin\phi}{4(1+\sin\phi)}\right) \quad \text{and} \quad Q(\phi) := \frac{\sin\theta \sin\phi}{8(1+\sin\phi)(1+\sin\theta)} \; .$$

Note, that $\theta = \theta(\phi)$ implies that $Q(\phi)$ depends only on ϕ. If $k \in \mathbb{N}$ and $h > 0$ satisfy

$$h \leq Q(\phi) R_{\max} k^{-2}$$

we can further define

$$R := \frac{k^2 h}{Q(\phi)} \quad \text{and} \quad r := \frac{\sin\phi}{2(1+\sin\phi)} R = 4\frac{1+\sin\theta}{\sin\theta} k^2 h \; .$$

Finally, we set

$$T_r := \left\{ t \in \frac{2r}{\sqrt{d}}\mathbb{Z}^d \; : \; B(t, r) \subset \Omega \right\} \; ,$$

and for $t \in T_r$,

$$\mathcal{D}_t := \{ x \in \Omega \; : \; co(\{x\} \cap B(t, r)) \text{ is contained in } \Omega \cap B(t, R) \} \; ,$$

where $co(A)$ denotes the convex hull of a set A. Then $\{\mathcal{D}_t\}_{t \in T_r}$ is a covering of Ω with several useful properties, which are summarized in the following theorem.

Theorem 3.1 *[7] With the notation introduced above, $\{\mathcal{D}_t\}_{t \in T_r}$ is a covering of Ω, i.e., $\Omega \subset \cup_{t \in T_r} \mathcal{D}_t$, with the following properties:*

- *Each set $\mathcal{D}_t$ is star shaped with respect to a ball $B(t, r)$ with $B(t, r) \subset \mathcal{D}_t \subset B(t, R) \cap \Omega$.*

- *Each set $\mathcal{D}_t$ satisfies an interior cone condition with radius r and angle θ.*

- *If we set $A_\phi := 4 \frac{1 + \sin \theta(\phi)}{\sin \theta(\phi)}$, and $\tilde{A}_\phi := \frac{2}{Q(\phi)} > 1$, we have*

$$A_\phi \cdot hk^2 = r \leq \delta_{\mathcal{D}_t} \leq 2R = \tilde{A}_\phi \cdot hk^2 \ .$$

- *Let χ_S denote the characteristic function of a set S. Then there is a constant $M_1 > 0$ that depends only on ϕ and d such that*

$$\sum_{t \in T_r} \chi_{\mathcal{D}_t} \leq M_1 \ .$$

- *There is a constant $M_2 > 0$ that depends only on d, ϕ and $vol(\Omega)$ such that the cardinality of T_r is bounded by*

$$\#T_r \leq \frac{vol(\Omega)}{vol(B(t, r))} \leq M_2 (hk^2)^{-d} \ .$$

We use this covering to derive sampling inequalities on the global domain Ω. From now on, we need a constant $C_{\min}$ depending only on $R_{\max}$ and ϕ, but in particular not on d,

$$C_{\min} := \min\{1 \ , \ Q_\phi R_{\max}\} \ . \tag{8}$$

Theorem 3.2 *Suppose $\Omega \subset \mathbb{R}^d$ is bounded with Lipschitz boundary and satisfies an interior cone condition with maximum radius $R_{\max}$ and angle ϕ. Let $\alpha \in \mathbb{N}_0^d$, $1 \leq q \leq \infty$, and $1 \leq p < \infty$. Then there are constants C_Q and $\mathfrak{C}_Q > 0$ such that for all $k \in \mathbb{N}$ with $k > |\alpha| + d/p + 1$ if $p > 1$, or $k \geq |\alpha| + d$ if $p = 1$, for all discrete sets $X \subset \Omega$ with fill distance $h \leq C_{\min}/k^2$, and for all $u \in W_p^k(\Omega)$,*

$$\|D^\alpha u\|_{L_q(\Omega)} \leq \frac{C_Q^k h^{-|\alpha|}}{(k - |\alpha|)!} (hk^2)^{k - d\left(\frac{1}{p} - \frac{1}{q}\right)_+} |u|_{W_p^k(\Omega)} + \mathfrak{C}_Q h^{-|\alpha|} (hk^2)^{\frac{d}{q}} \|u|_X\|_{\ell_q(X)} \ , \tag{9}$$

where C_Q and $\mathfrak{C}_Q$ depend only on d, ϕ, $vol(\Omega)$, p, q and α, but not on k, h, or u.

Proof: As pointed out in [7, Section 2.2.1], by construction every ball $\overline{B(x, h)} \subset \Omega$ contains at least one point in X. Thus, since $h = h_{X,\Omega}$ satisfies (4), the results of Theorem 2.4 and Corollary 2.5 hold with this h. Hence, Corollary 2.5 and the estimate $\delta_{\mathcal{D}_t}^{|\alpha|} \leq A_\phi^{-|\alpha|} h^{-|\alpha|}$

give with the notation introduced above for $u \in W_p^k(\Omega)$

$$\|D^\alpha u\|_{L_q(\Omega)} = \left(\int_\Omega |D^\alpha u(x)|^q \, dx\right)^{1/q} \leq \left(\sum_{t \in T_r} \|D^\alpha u\|^q_{L_q(\mathcal{D}_t)}\right)^{1/q}$$

$$\leq \quad \frac{C_S^k}{(k-|\alpha|)!} \left(\tilde{A}_\phi h k^2\right)^{k-d\left(\frac{1}{p}-\frac{1}{q}\right)} h^{-|\alpha|} \left(A_\phi^{-|\alpha|}+1\right) \left(\sum_{t \in T_r} |u|^q_{W_p^k(\mathcal{D}_t)}\right)^{1/q} +$$

$$+ \quad 2\left(\tilde{A}_\phi h k^2\right)^{d/q} h^{-|\alpha|} \left(\sum_{t \in T_r} \|u|_{X \cap \mathcal{D}_t}\|^q_{\ell_\infty(X \cap \mathcal{D}_t)}\right)^{1/q}$$

$$\leq \quad \frac{\left(C_S \tilde{A}_\phi\right)^k \left(A_\phi^{-|\alpha|}+1\right) M_2^{\left(\frac{1}{q}-\frac{1}{p}\right)_+}}{(k-|\alpha|)!} h^{-|\alpha|} \left(h k^2\right)^{k-d\left(\frac{1}{p}-\frac{1}{q}\right)-d\left(\frac{1}{q}-\frac{1}{p}\right)_+} \left(\sum_{t \in T_r} |u|^p_{W_p^k(\mathcal{D}_t)}\right)^{\frac{1}{p}}$$

$$+ \quad 2\tilde{A}_\phi^{d/q} h^{-|\alpha|} \left(h k^2\right)^{d/q} \left(\sum_{t \in T_r} \|u|_{X \cap \mathcal{D}_t}\|^q_{\ell_\infty(X \cap \mathcal{D}_t)}\right)^{1/q}$$

$$\leq \quad \frac{C_Q^k}{(k-|\alpha|)!} h^{-|\alpha|} \left(h k^2\right)^{k-d\left(\frac{1}{p}-\frac{1}{q}\right)_+} |u|_{W_p^k(\Omega)} + \mathfrak{C}_Q h^{-|\alpha|} (h k^2)^{d/q} \|u|_X\|_{\ell_q(X)} \ ,$$

where $\mathfrak{C}_Q = 2\tilde{A}_\phi^{d/q} M_1^{1/q}$ and $C_Q = C_S \tilde{A}_\phi \left(A_\phi^{-|\alpha|}+1\right)(M_2^{\left(\frac{1}{q}-\frac{1}{p}\right)_+}+1)(M_1^{1/p}+1)$ do not depend on k, h, u or $\mathcal{D}_t$. In the last step, we used the identity $(a-b)+(b-a)_+ = (a-b)_+$ for $a, b \in \mathbb{R}$. □

Remark 3.3 *With the remark following Lemma 2.1, the estimate of Theorem 3.2 generalizes results of [6] to derivatives.*

Corollary 3.4 *Under the assumptions from Theorem 3.2, simply changing the last step of the proof,*

$$\|D^\alpha u\|_{L_q(\Omega)} \leq \frac{C_1^k}{(k-|\alpha|)!} h^{-|\alpha|} (h k^2)^{k-d\left(\frac{1}{p}-\frac{1}{q}\right)_+} |u|_{W_p^k(\Omega)} + C_2 h^{-|\alpha|} \|u|_X\|_{\ell_\infty(X)} \ , \quad (10)$$

where $C_1 = C_Q$ and $C_2 = 2\tilde{A}_\phi^{d/q} M_2^{1/q}$ do not depend on k, h, or u.

We shall now again relate h and k to derive exponential estimates. The actual orders depend on the asymptotic behaviour of the embedding constants $E(k)$ from (1) for $k \to \infty$.

Theorem 3.5 *Suppose that $\Omega \subset \mathbb{R}^d$ is bounded, has a Lipschitz boundary, and satisfies an interior cone condition with a maximum radius $R_{\max}$ and angle $\phi \in (0, \pi/2)$. Suppose further, that $\mathcal{H}(\Omega)$ consists of smooth functions $f : \Omega \to \mathbb{R}$, i.e., for a $1 \leq p < \infty$ and all $k \in \mathbb{N}$ there exist numbers $E(k)$ that may depend on k, Ω, p and d such that for all $u \in \mathcal{H}(\Omega)$,*

$$\|u\|_{W_p^k(\Omega)} \leq E(k) \|u\|_{\mathcal{H}(\Omega)} \ .$$

9

If there are constants $0 < \epsilon \leq 1$ and $C_E > 0$ such that $E(k) \leq C_E^k k^{(1-\epsilon)k}$ for all $k \in \mathbb{N}$, then, for all $1 \leq q \leq \infty$, there are constants C and $h_0 > 0$ such that for all data sets $X \subset \Omega$ with fill distance $h \leq h_0$ the inequality

$$\|D^\alpha u\|_{L_q(\Omega)} \leq e^{C\log(h)/\sqrt{h}}\|u\|_{\mathcal{H}(\Omega)} + C_2 h^{-|\alpha|}\|u|_X\|_{\ell_\infty(X)}$$

holds for all $u \in \mathcal{H}(\Omega)$. The constant $C = \epsilon\sqrt{C_{\min}}/4$ depends only on ϵ, ϕ und $R_{\max}$, but in particular not on d, α, h or u, while h_0 may depend on d, p, q, $R_{\max}$, ϕ, $vol(\Omega)$, α, ϵ and C_E, but not on h or u. The constant C_2 comes from Corollary 3.4.

If there are constants $C_E > 0$ and $s \geq 1$ such that $E(k) \leq C_E^k k^{sk}$ for all $k \in \mathbb{N}$, then, for all $1 \leq q \leq \infty$, there are constants C and $h_0 > 0$ depending only on d, p, q, $R_{\max}$, ϕ, $vol(\Omega)$, α, s and C_E such that for all data sets $X \subset \Omega$ with fill distance $h \leq h_0$ the inequality

$$\|D^\alpha u\|_{L_q(\Omega)} \leq e^{-\frac{C}{h^{1/(1+s)}}}\|u\|_{\mathcal{H}(\Omega)} + C_2 h^{-|\alpha|}\|u|_X\|_{\ell_\infty(X)}$$

holds for all $u \in \mathcal{H}(\Omega)$ with the constant C_2 from Corollary 3.4.

Proof: If $\|u\|_{W_p^k(\Omega)} \leq C_E^k k^{(1-\epsilon)k}\|u\|_{\mathcal{H}(\Omega)}$ for all $k \in \mathbb{N}$, we use estimate (6) for $k \in \mathbb{N}$ provided that $k > |\alpha| + 1 + d/p$ and $h \leq C_{\min}/k^2$ with the constant $C_{\min}$ from (8), to bound the right-hand side of (10) by

$$\left(\mathfrak{c}hk^{2-\epsilon}\right)^k \left(\frac{k}{h}\right)^{|\alpha|} (hk^2)^{-d\left(\frac{1}{p}-\frac{1}{q}\right)_+}\|u\|_{\mathcal{H}(\Omega)} + C_2 h^{-|\alpha|}\|u|_X\|_{\ell_\infty(X)}\;, \tag{11}$$

where the constant $\mathfrak{c} = C_1 C_E e$ depends only on d, R, ϕ, p, q and α, but not on k, h or u. We choose $k \in \mathbb{N}$, $k > |\alpha| + 1 + d/p$, such that $\frac{C_{\min}}{2k^2} \leq h \leq \frac{C_{\min}}{k^2}$. That is always possible if $h \leq C_{\min}(|\alpha| + 1 + d/p)^{-2}$. Then, using $h \leq 1$ and $h^{-3|\alpha|/2} \leq e^{3|\alpha|/\sqrt{h}}$, we can bound the first term of (11) by

$$(1+\mathfrak{c})^{\frac{\sqrt{C_{\min}}}{\sqrt{h}}} C_{\min}^{(2-\epsilon)k/2} h^{\epsilon k/2} h^{-3|\alpha|/2} \left(\frac{2}{C_{\min}}\right)^{d\left(\frac{1}{p}-\frac{1}{q}\right)_+} \leq B^{1/\sqrt{h}} h^{\epsilon\sqrt{C_{\min}}/\left(2\sqrt{2h}\right)}$$

where the constant $B := (\mathfrak{c}+1)^{\sqrt{C_{\min}}} e^{3|\alpha|} (2/C_{\min})^{d\left(\frac{1}{p}-\frac{1}{q}\right)_+}$ depends only on d, p, q, $R_{\max}$, ϕ, α, and C_E, but not on k, h or u. If we now set

$$h_0 := \min\left\{1, \frac{C_{\min}}{(|\alpha| + 1 + d/p)^2}, B^{-4/\left[\epsilon\sqrt{C_{\min}}(\sqrt{2}-1)\right]}\right\},$$

we find for all $h \leq h_0$,

$$\|D^\alpha u\|_{L_q(\Omega)} \leq e^{\epsilon\sqrt{C_{\min}}\log h/\left(4\sqrt{h}\right)}\|u\|_{\mathcal{H}(\Omega)} + C_2 h^{-|\alpha|}\|u|_X\|_{\ell_\infty(X)}\;.$$

Setting $C := \epsilon\sqrt{C_{\min}}/4$ finishes the proof of the first part.

If $E(k) \leq C_E^k k^{sk}$ we can proceed along the lines of the first part and bound the first term of the right-hand side (10) for $k \in \mathbb{N}$ with $h \leq C_{\min}/k^2$ and $k > |\alpha| + 1 + d/p$, by

$$(\tilde{\mathfrak{c}} h k^{1+s})^k (h^{-1}k)^{|\alpha|} (hk^2)^{-d\left(\frac{1}{p} - \frac{1}{q}\right)_+} \|u\|_{\mathcal{H}(\Omega)} , \tag{12}$$

where the constant $\tilde{\mathfrak{c}} := C_1 e C_E$ depends only on d, R, ϕ, p, q and α, but not on k, h or u. We set $T = \min\left\{C_{\min}, \frac{1}{e\tilde{\mathfrak{c}}}\right\}$, and choose $k \in \mathbb{N}$, $k > |\alpha| + 1 + d/p$ such that $\frac{T}{2k^{1+s}} \leq h \leq \frac{T}{k^{1+s}}$ holds. That is always possible if $h \leq T(|\alpha| + 1 + d/p)^{-2}$. We point out that the condition $h \leq \frac{C_{\min}}{k^2}$ is satisfied since $s \geq 1$. Then $\frac{T}{2}k^{1-s} \leq hk^2 \leq T$, and thus with $G := \left(\frac{2}{T}\right)^{|\alpha| + d\left(\frac{1}{p} - \frac{1}{q}\right)_+}$, and $F := |\alpha|(2 + s) + (s - 1)d\left(\frac{1}{p} - \frac{1}{q}\right)_+$ we have for sufficiently small $h \leq h_0$, that is, for sufficiently large $k \in \mathbb{N}$,

$$(\tilde{\mathfrak{c}} h k^{1+s})^k \left(\frac{k}{h}\right)^{|\alpha|} (hk^2)^{-d\left(\frac{1}{p} - \frac{1}{q}\right)_+} \leq e^{-k} G k^F \leq e^{-k/2} \leq e^{-(T/h)^{\frac{1}{1+s}}} .$$

$\square$

Remark 3.6 *Suppose $\mathcal{H}(\Omega)$ is a space of smooth functions with $E(k) \leq C_E^k k^{sk}$ for all $k \in \mathbb{N}$ with some $s \geq 1$. Under the assumptions of Theorem 3.5 we can achieve convergence rates with a constant independent of the space dimension d if we accept slightly worse approximation orders than the ones given in Theorem 3.5, e.g., for all $\delta > 0$ there is a constant C_δ that depends only on δ, r, θ and s, but not on d, such that*

$$\|D^\alpha u\|_{L_q(\Omega)} \leq \exp\left\{\frac{C_\delta \log h}{h^{1/(1+s+\delta)}}\right\} \|u\|_{\mathcal{H}(\Omega)} + C_2 h^{-|\alpha|} \|u|_X\|_{\ell_\infty(X)}$$

holds for all $u \in \mathcal{H}(\Omega)$, all $\alpha \in \mathbb{N}_0^d$ and all sets X with fill distance $h \leq h_0$.

Proof: We follow the proof of Theorem 3.5, but choose $k \in \mathbb{N}$ for sufficiently small h such that $\frac{C_{\min}}{2k^{1+s+\delta}} \leq h \leq \frac{C_{\min}}{k^{1+s+\delta}}$. We set

$$2C_\delta := \frac{\delta \left(\frac{C_{\min}}{2}\right)^{1/(1+s+\delta)}}{(1 + s + \delta)} ,$$

which depends only on δ, s, $R_{\max}$, and ϕ but not on the space dimension d. Then, analogously to the proof of the first part of Theorem 3.5, we find a constant $V > 0$ that depends only on d, R, ϕ, p, q, s, δ and α, but not on k, h or u such that (12) can be bounded by

$$\exp\left\{\frac{V}{h^{1/(1+s+\delta)}} + \frac{2C_\delta \log h}{h^{1/(1+s+\delta)}}\right\} \|u\|_{\mathcal{H}(\Omega)} \leq \exp\left\{\frac{C_\delta \log h}{h^{1/(1+s+\delta)}}\right\} \|u\|_{\mathcal{H}(\Omega)}$$

for sufficiently small h. $\square$

Since $\|u|_X\|_{\ell_\infty(X)} \leq \|u|_X\|_{\ell_Q(X)}$ for all $1 \leq Q \leq \infty$, we can replace the $\ell_\infty(X)$-norm by any $\ell_Q(X)$-norm in all sampling inequalities derived so far. However, using Theorem 3.2 in terms of the $\ell_q(X)$-norm, we can gain a positive power of h in the discrete term from the term $(hk^2)^{d/q}$ by choosing k smaller than required by the condition $h \leq C_{\min}/k^2$. The general guideline is that the worse the exponential rates of the first term, the better is the behaviour of the discrete term for $h \to 0$. Examplarily, we state the following result.

Corollary 3.7 *Under the assumptions of Theorem 3.5, if $E(k) \leq C_E^k k^{sk}$ for all $k \in \mathbb{N}$ with some $s \geq 1$, then for all $1 \leq q \leq \infty$, there are constants C, $\tilde{C}$ and $h_0 > 0$ such that for all sets $X \subset \Omega$ with fill distance $h \leq h_0$ and for all $u \in \mathcal{H}(\Omega)$*

$$\|D^\alpha u\|_{L_q(\Omega)} \leq e^{-\frac{C}{h^{1/(1+s)}}} \|u\|_{\mathcal{H}(\Omega)} + \tilde{C} h^{-|\alpha|} h^{\frac{d}{q}\left(1-\frac{2}{1+s}\right)} \|u|_X\|_{\ell_q(X)} ,$$

where C, $\tilde{C}$ and h_0 may depend on d, q, R, ϕ, α, s and C_E, but not on h or u.

Proof: With the notations of the proof of Theorem 3.5, the first term of ther right-hand side of (9) can be treated as before by simply replacing C_1 by C_Q. For the second term we estimate

$$\mathfrak{C}_Q h^{-|\alpha|} h^{d/q} k^{2d/q} \leq \mathfrak{C}_Q h^{-|\alpha|} h^{d/q} \left(\frac{T}{h}\right)^{2d/[q(1+s)]} = \tilde{C} h^{-|\alpha|} h^{\frac{d}{q}\left(1-\frac{2}{1+s}\right)}$$

with $\tilde{C} := \mathfrak{C}_Q T^{2d/[q(1+s)]}$. $\qquad\qquad\square$

4   Estimates on Compact Cubes

In this section, we shall prove estimates on compact cubes. Since the orientation can be adjusted by rotation, we restrict ourselves to axial parallel cubes and use the notation

$$W(x_0, R) := \left\{ x \in \mathbb{R}^d \ : \ \|x - x_0\|_{\ell_\infty} \leq R \right\}$$

for a compact cube with center $x_0 \in \mathbb{R}^d$ and side length $2R > 0$. To derive estimates for function values on compact cubes, we follow the paths of the previous sections. Here, we can use a polynomial reproduction from [16, Theorem 11.21], which is based on [5, Lemma 1]. The result derived in [16, Theorem 11.21] is stated with h taken to be the fill distance of X in Ω. In the proof, however, h is again only required to satisfy the condition that every closed ball $\overline{B(x,h)} \subset \Omega$ contains at least one point in X. Following the proof, we can choose the constant $C_0 = \frac{R}{6\gamma_d \ell}$. That said, we have the following result.

Theorem 4.1 *Let Ω be a compact cube in $\mathbb{R}^d$ with side length $2R > 0$ and center $x_0 \in \mathbb{R}^d$, and let $\ell \in \mathbb{N}$. Suppose $X = \{x_1, \ldots, x_N\}$ and $0 < h \leq \frac{R}{6\gamma_d \ell}$ satisfy the condition that every closed ball $\overline{B(x,h)} \subset \Omega$ contains at least one point in X. Then there are functions $a_j : \Omega \to \mathbb{R}$ satisfying*

- *$\sum_{j=1}^N a_j(x) p(x_j) = p(x)$ for all $x \in \Omega$ and all $p \in \mathbb{P}_\ell(\mathbb{R}^d)$,*

- *$\sum_{j=1}^N |a_j(x)| \leq e^{2d\gamma_d(\ell+1)}$ for all $x \in \Omega$.*

The numbers γ_d are defined recursively by $\gamma_1 = 2$ and $\gamma_d = 2d(1 + \gamma_{d-1})$.

For any cube $\mathcal{D}$ with diameter $\delta_\mathcal{D}$, Lemma 2.1 gives in the special case $\alpha = 0$ the bound

$$\left\| u - Q^k u \right\|_{L_\infty(\mathcal{D})} \leq \frac{C_{d,p}^k}{k!} \delta_\mathcal{D}^{k-d/p} |u|_{W_p^k(\mathcal{D})} ,$$

12

where $C_{d,p}$ now depends only on d and p, since cubes satisfy cone conditions with a fixed angle. If we insert this estimate and the bound from Theorem 4.1 into the estimate (3) we find

$$\|u\|_{L_\infty(\mathcal{D})} \le \frac{C^k \delta_{\mathcal{D}}^{k-d/p}}{k!} |u|_{W_p^k(\mathcal{D})} + c^k \|u|_X\|_{\ell_\infty(X\cap\mathcal{D})} \ ,$$

and similarly to Corollary 2.5

$$\|u\|_{L_q(\mathcal{D})} \le \frac{C^k \delta_{\mathcal{D}}^{k-d\left(\frac{1}{p}-\frac{1}{q}\right)}}{k!} |u|_{W_p^k(\mathcal{D})} + c^k \delta_{\mathcal{D}}^{d/q} \|u|_X\|_{\ell_\infty(X\cap\mathcal{D})} \ .$$

The constants $C = C_{d,p} e^{4d\gamma_d}$ and $c = e^{4d\gamma_d}$ depend now only on d and p. To derive estimates for arbitrary embedding constants $E(k)$, we use a covering of the big cube Ω with axially parallel small cubes $\mathcal{D}_t$. A similar approach can be found in [6]. We start with a compact cube Ω in $\mathbb{R}^1$ with center x_0 and side length $2R$, i.e., $\Omega = [x_0 - R, x_0 + R]$. For $h \le \frac{R}{6\gamma_d k}$, $k \in \mathbb{N}$, we set $r := 6\gamma_d hk \le R$, and define the set of centers

$$T_r^{(1)} := \{t \in 2r\mathbb{Z} \ : \ W(t,r) \subset \Omega\} \ \cup \ \{x_0 - R + r, \ x_0 + R - r\} \ .$$

For $t \in T_r^{(1)}$ we set $\mathcal{D}_t^{(1)} = W(t,r) \subset \mathbb{R}$. Then $\{\mathcal{D}_t^{(1)}\}_{t \in T_r}$ is a covering of Ω, i.e., $\bigcup_{t \in T_r} \mathcal{D}_t^{(1)} = \Omega$, where the diameters are given by $\delta_{\mathcal{D}_t^{(1)}} = 2r$, no point in Ω lies in more than two small cubes, i.e., $\sum_{t \in T_r} \chi_{\mathcal{D}_t^{(1)}} \le 2$, and the number of cubes is bounded by $\#T_r^{(1)} \le 3\left(\frac{R}{r}\right) \le \frac{R}{2\gamma_d}(hk)^{-1}$. Inductively, we find by a tensorization argument the following result for compact cubes in $\mathbb{R}^d$.

Theorem 4.2 *Suppose $\Omega = W(x_0, R) \subset \mathbb{R}^d$ is a compact cube with side length $2R$. For $h \le \frac{R}{6\gamma_d k}$ we set $r := 6\gamma_d hk$. Then there is a covering $\{\mathcal{D}_t^{(d)}\}_{t \in T_r^{(d)}}$ with a index set $T_r^{(d)} \subset \mathbb{R}^d$ of Ω with the following properties.*

- $\#T_r^{(d)} \le \left(\frac{R}{2\gamma_d}\right)^d (hk)^{-d}$

- *For all $t \in T_r^{(d)}$, the set $\mathcal{D}_t^{(d)}$ is a compact cube with center $t \in T_r^{(d)}$ and side length $2r$, i.e., $\delta_{\mathcal{D}_t^{(d)}} = 2\sqrt{d}r = 12\gamma_d\sqrt{d}hk$.*

- $\sum_{t \in T_r^{(d)}} \chi_{\mathcal{D}_t^{(d)}} \le 2^d.$

As in the previous section we find the following estimate on the compact cube Ω covered by the cubes $\{D_t\}$ from Theorem 4.2. We shall skip the super-index (d) from now on.

Theorem 4.3 *Suppose $\Omega \subset \mathbb{R}^d$ is a compact cube with side length R. Then, for all $1 \le q \le \infty$, and all $1 \le p < \infty$, there are constants C_W, $\mathfrak{C}_W > 0$ depending only on d, p, q and R such that*

$$\|u\|_{L_q(\Omega)} \le \frac{C_W^k}{k!}(hk)^{k-d\left(\frac{1}{p}-\frac{1}{q}\right)} + |u|_{W_p^k(\Omega)} + \mathfrak{C}_W^k (hk)^{d/q} \|u|_X\|_{\ell_q(X)} \tag{13}$$

for all $u \in W_p^k(\Omega)$, provided that $k > \frac{d}{p} + 1$ if $p > 1$, or $k \geq d$ if $p = 1$, and all data sets $X \subset \Omega$ with fill distance $h := h_{X,\Omega} \leq \frac{R}{6\gamma_d k}$. The constants can be chosen as

$$C_W = 2^{d/p} C_{d,p} e^{8d\gamma_d} \left(12\gamma_d \sqrt{d}(R+1)\right)^{1+d\left(\frac{1}{q}-\frac{1}{p}\right)_+} \text{ and } \mathfrak{C}_W = e^{4d\gamma_d} \left(24\gamma_d\sqrt{d}\right)^{d/q} .$$

As in the case of domains obeying a cone condition we get the following corollary.

Corollary 4.4 *Under the assumptions from Theorem 4.3 we obtain*

$$\|u\|_{L_q(\Omega)} \leq \frac{C_3^k}{k!}(hk)^{k-d\left(\frac{1}{p}-\frac{1}{q}\right)_+} |u|_{W_p^k(\Omega)} + C_4^k \|u|_X\|_{\ell_\infty(X)} \tag{14}$$

with $C_3 = C_W$ and $C_4 = \mathfrak{C}_W \left(\frac{R}{2\gamma_d} + 1\right)^{d/q}$.

Now we can derive exponential approximation orders. Again, the actual orders depend on the asymptotic behaviour of the embedding constants $E(k)$ from (1) for $k \to \infty$.

Theorem 4.5 *Suppose $\Omega \subset \mathbb{R}^d$ is a compact cube with side length R. Suppose further, $\mathcal{H}(\Omega)$ consists of smooth functions, i.e., for a $1 \leq p < \infty$ and all $k \in \mathbb{N}$ there exist numbers $E(k)$ that may depend on k, R, p and d such that for all $u \in \mathcal{H}(\Omega)$,*

$$\|u\|_{W_p^k(\Omega)} \leq E(k)\|u\|_{\mathcal{H}(\Omega)} .$$

If there are constants $C_E > 0$, and $0 < \epsilon \leq 1$ such that $E(k) \leq C_E^k k^{(1-\epsilon)k}$ for all $k \in \mathbb{N}$, then for all $1 \leq q \leq \infty$, there exist constants C, $\tilde{C}$, and $h_0 > 0$ such that for all data sets X with fill distance $h \leq h_0$ the estimate

$$\|u\|_{L_q(\Omega)} \leq e^{C\log(h)/h}\|u\|_{\mathcal{H}(\Omega)} + e^{\tilde{C}/h}\|u|_X\|_{\ell_\infty(X)}$$

holds for all $u \in \mathcal{H}(\Omega)$. The constants C, $\tilde{C}$ and h_0 depend only on d, R, p and q, but not on u or h.
If there are constants $C_E > 0$ and $s \geq 1$ such that $E(k) \leq C_E^k k^{sk}$ for all $k \in \mathbb{N}$, then, for all $1 \leq q \leq \infty$, there exist constants $\tilde{C}$, C, $h_0 > 0$ such that for all data sets X with fill distance $h \leq h_0$

$$\|u\|_{L_q(\Omega)} \leq e^{-C/\sqrt[s]{h}}\|u\|_{\mathcal{H}(\Omega)} + e^{\tilde{C}/\sqrt[s]{h}}\|u|_X\|_{\ell_\infty(X)}$$

holds for all $u \in \mathcal{H}$. The constants C, $\tilde{C}$ and h_0 depend on d, R, p and q, but not on u or h.

Proof: If $\|u\|_{W_p^k(\Omega)} \leq C_E^k k^{(1-\epsilon)k}\|u\|_{\mathcal{H}(\Omega)}$ for all $u \in \mathcal{H}(\Omega)$, estimate (14) shows

$$\|u\|_{L_q(\Omega)} \leq \frac{(C_3 C_E)^k}{k!}(hk)^{k-d\left(\frac{1}{p}-\frac{1}{q}\right)_+} k^{(1-\epsilon)k}\|u\|_{\mathcal{H}(\Omega)} + C_4^k \|u|_X\|_{\ell_\infty(X)} ,$$

provided that $h \leq R / (6\gamma_d k)$. Using Stirling's formula $k! \geq \sqrt{2\pi k} \left(\frac{k}{e}\right)^k \geq \left(\frac{k}{e}\right)^k$, we can bound the first term of the right-hand side by

$$\left(\mathfrak{c}hk^{(1-\epsilon)}\right)^k (hk)^{-d\left(\frac{1}{p}-\frac{1}{q}\right)}_+ \|u\|_{\mathcal{H}(\Omega)}$$

with $\mathfrak{c} := C_3 C_E e$. We set $B := \min\left\{\frac{R}{6\gamma_d}, \frac{1}{\mathfrak{c}}, 1\right\}$, and choose $k \in \mathbb{N}$ such that $\frac{B}{2k} \leq h \leq \frac{B}{k}$. That is always possible if $h \leq B$. Then, for $h \leq (1 + 2/B)^{-4}$

$$
\begin{aligned}
\left(\mathfrak{c}hk^{(1-\epsilon)}\right)^k (hk)^{-d\left(\frac{1}{p}-\frac{1}{q}\right)}_+ \|u\|_{\mathcal{H}(\Omega)} &\leq k^{-\epsilon k} (B/2)^{-d\left(\frac{1}{p}-\frac{1}{q}\right)}_+ \|u\|_{\mathcal{H}(\Omega)} \\
&\leq (1 + 2/B)^{\epsilon B/h} h^{\epsilon B/(2h)} \|u\|_{\mathcal{H}(\Omega)} \leq h^{\epsilon B/(4h)} \|u\|_{\mathcal{H}(\Omega)} .
\end{aligned}
\tag{15}
$$

Choosing $C = \epsilon B/4$ and $\tilde{C} = B \log(C_4 + 1)$ finishes the first part.
If $E(k) \leq C_E^k k^{sk}$ for all k with some $s \geq 1$, the first term of the right-hand side of (14) is bounded by

$$(\tilde{\mathfrak{c}}hk^s)^k (hk)^{-d\left(\frac{1}{p}-\frac{1}{q}\right)}_+ \|u\|_{\mathcal{H}(\Omega)}$$

with $\tilde{\mathfrak{c}} := C_3 C_E e$. We set $B := \min\left\{\frac{R}{6\gamma_d}, \frac{1}{e\tilde{\mathfrak{c}}}, 1\right\}$, and choose $k \in \mathbb{N}$ such that $\frac{B}{2k^s} \leq h \leq \frac{B}{k^s}$. That is possible if $h \leq B =: h_0$, and the condition $h \leq \frac{c_0}{2k}$ is satisfied. For sufficiently small h, that is, for sufficiently large k,

$$
\begin{aligned}
(\tilde{\mathfrak{c}}hk^s)^k (hk)^{-d\left(\frac{1}{p}-\frac{1}{q}\right)}_+ \|u\|_{\mathcal{H}(\Omega)} &\leq \left(\frac{B}{2}\right)^{-d\left(\frac{1}{p}-\frac{1}{q}\right)}_+ k^{(s-1)d\left(\frac{1}{p}-\frac{1}{q}\right)_+} e^{-k} \|u\|_{\mathcal{H}(\Omega)} \\
&\leq e^{-k/2} \|u\|_{\mathcal{H}(\Omega)} \leq e^{-\sqrt[s]{B}/\left(2\sqrt[s]{2h}\right)} \|u\|_{\mathcal{H}(\Omega)} .
\end{aligned}
\tag{16}
$$

Choosing $C = \sqrt[s]{B} / \left(2\sqrt[s]{2}\right)$ and $\tilde{C} = \sqrt[s]{B} \log(C_4 + 1)$ finishes the proof. $\qquad\square$

Again, the ℓ_∞-norm can be replaced by any ℓ_Q-norm for $1 \leq Q \leq \infty$. However, in this case the particular choice of the ℓ_q-norm as in Theorem 4.3 does not lead to a significantly better behaviour of the discrete term due to the exponential factor $\mathfrak{C}_W^k$.

5   Kernels and Native Spaces

In this section, we will provide two famous examples of Hilbert spaces of infinitely smooth functions. In the case of a positive definite radial basis function K which possesses a Fourier transform $\hat{K}$ the native Hilbert space, which is the uniquely determined reproducing kernel Hilbert space of K, is defined via [16, Theorem 10.12]

$$\mathcal{N}_K(\mathbb{R}^d) = \left\{ f \in C(\mathbb{R}^d) \cap L_2(\mathbb{R}^d) \;:\; \|f\|_{\mathcal{N}_K}^2 := \int_{\mathbb{R}^d} \frac{|\hat{f}(\omega)|^2}{\left|\hat{K}(\omega)\right|} d\omega < \infty \right\} .$$

In general, the native Hilbert space $\mathcal{N}_K(\Omega)$ on a bounded domain Ω is defined as

$$\mathcal{N}_K(\Omega) := \left\{ f \in \mathcal{N}_K(\mathbb{R}^d) \ : f|_\Omega = 0 \right\}^{\perp_{\mathcal{N}_K}}$$

$$:= \left\{ f|_\Omega : f \in \mathcal{N}_K(\mathbb{R}^d) \text{ and } (f,g)_{\mathcal{N}_K(\mathbb{R}^d)} = 0 \text{ for all } g \in \mathcal{N}_K(\mathbb{R}^d) \text{ s.t. } g|_\Omega = 0 \right\}.$$

On the other hand, the Sobolev spaces on $\mathbb{R}^d$ can be defined via

$$W_2^k(\mathbb{R}^d) := \left\{ f \in L_2(\mathbb{R}^d) \ : \hat{f}(\cdot)(1 + \| \cdot \|_2^2)^{k/2} \in L_2(\mathbb{R}^d) \right\}.$$

This definition is equivalent to the one given in the introduction, and following [18],

$$\frac{1}{2^{2k}} \| \hat{f}(\cdot)(1 + \| \cdot \|_2^2)^{k/2} \|_{L_2(\mathbb{R}^d)} \leq \| u \|_{W_2^k(\mathbb{R}^d)} \leq \| \hat{f}(\cdot)(1 + \| \cdot \|_2^2)^{k/2} \|_{L_2(\mathbb{R}^d)}.$$

The constant 2^{2k} is absorbed into the C_E^k term of the embedding constants $E(k)$, which does not influence the approximation orders, but only the constants $\tilde{C}$ in Theorem 4.5. Since we do not take much care about factors 4 in these constants, we may as well use this definition of Sobolev spaces. Note that for $\Omega = \mathbb{R}^d$ for infinitely smooth kernels an upper bound for $E(k)$ is given by

$$E(k)^2 \leq \sup_{x \in \mathbb{R}^d} \left| \hat{K}(x) \left(1 + \|x\|_2^2 \right)^k \right| = \operatorname{ess\,sup}_{x \in \mathbb{R}^d} \left| (Id - \Delta)^k K(x) \right|. \tag{17}$$

Therefore, the theory presented here for native spaces of smooth radial basis functions appears to be closely related to classical interpolation in native spaces of smooth kernels (see Section 6) as considered in [19], where the exponential approximation orders turn out to depend on the asymptotic behaviour of the numbers

$$C_K^{(2k)} := \sup_{x,y \in \Omega} \max_{|\beta| + |\nu| = 2k} \left| D_x^\beta D_y^\nu K(x,y) \right|$$

for $k \to \infty$, where D_z^α denotes the α-th derivative with respect to the variable z.

By [16, Theorem 10.46], every $f \in \mathcal{N}_K(\Omega)$ has an extension $\mathcal{E}f \in \mathcal{N}_K(\mathbb{R}^d)$ with

$$\| \mathcal{E}f \|_{\mathcal{N}_K(\mathbb{R}^d)} \leq \| f \|_{\mathcal{N}_K(\Omega)}.$$

Thus we have for $f = \mathcal{E}f|_\Omega \in \mathcal{N}_K(\Omega)$ for all $k \in \mathbb{N}$, with the bound on the embedding constant $E(k)$ on $\mathbb{R}^d$ as given in (17),

$$\| f \|_{W_2^k(\Omega)} \leq \| \mathcal{E}f \|_{W_2^k(\mathbb{R}^d)} \leq E(k) \| \mathcal{E}f \|_{\mathcal{N}_K(\mathbb{R}^d)} \leq E(k) \| f \|_{\mathcal{N}_K(\Omega)}. \tag{18}$$

For the special case of the Gaussian kernel $G(x) = e^{-c^2\|x\|_2^2}$ the native space is given by

$$\mathcal{N}_G(\mathbb{R}^d) = \left\{ f \in C(\mathbb{R}^d) \cap L_2(\mathbb{R}^d) \ : \int_{\mathbb{R}^d} |\hat{f}(\omega)|^2 e^{\frac{\|x\|_2^2}{4c^2}} \, d\omega < \infty \right\}.$$

We show in the appendix, Theorem A.1 that there is some $C_E > 0$ depending only on the scaling parameter c and the space dimension d such that for all $k \geq 0$

$$\mathcal{N}_G(\mathbb{R}^d) \subset W_2^k(\mathbb{R}^d) \text{ with } \|f\|_{W_2^k(\mathbb{R}^d)} \leq C_E^k k^{k/2} \|f\|_{\mathcal{N}_G(\mathbb{R}^d)} \, ,$$

which implies $\mathcal{N}_G(\Omega) \subset W_2^k(\Omega)$ and $\|f\|_{W_2^k(\Omega)} \leq C_E^k k^{k/2} \|f\|_{\mathcal{N}_G(\Omega)}$ for all $f \in \mathcal{N}(\Omega)$ and all $k \in \mathbb{N}$. Similarly to Gaussian kernels we can consider inverse multiquadrics

$$K_M(x) = \left(c^2 + \|x\|_2^2\right)^{-\beta} \, , \text{ for } \beta > d/2 \, .$$

The essentially same argument as for Gaussian kernels (see Theorem A.2) leads to

$$\mathcal{N}_M(\mathbb{R}^d) \subset W_2^\tau(\mathbb{R}^d) \text{ and } \|f\|_{W_2^\tau(\Omega)} \leq C_E^k k^k \|f\|_{\mathcal{N}_M(\Omega)} \tag{19}$$

with a constant $C_E > 0$ that depends only on β and d. Now we are able to apply the sampling inequalities from Theorems 2.6, 3.5 and 4.5 and the associated corollaries and remarks to the native Hilbert spaces of Gaussians and inverse multiquadrics, e.g., we have

Corollary 5.1 *Under the assumptions from Theorem 4.5 and with the constant $E_G(k) = C_E^k k^{k/2}$ (see Appendix, Theorem A.1) we obtain for all $u \in \mathcal{N}_G(\Omega)$*

$$\|u\|_{L_q(\Omega)} \leq e^{C \log(h)/h} \|u\|_{\mathcal{N}_G(\Omega)} + e^{\tilde{C}/h} \|u|_X\|_{\ell_\infty(X)} \, .$$

Analogously with $E_M(k) \leq C_E^k k^k$ (see Appendix, Theorem A.2) we find for all $u \in \mathcal{N}_M(\Omega)$

$$\|u\|_{L_q(\Omega)} \leq e^{-C/h} \|u\|_{\mathcal{N}_M(\Omega)} + e^{\tilde{C}/h} \|u|_X\|_{\ell_\infty(X)} \, .$$

6 Application to Smoothed Interpolation

As an application of the results derived in the previous sections, we consider (possibly regularized) kernel based interpolation in the native space. To start with, we briefly summarize the problem. One is given centers $X = \{x_1, \ldots, x_N\} \subset \Omega \subset \mathbb{R}^d$ and data $(f_1, \ldots, f_N)^T \in \mathbb{R}^N$ generated by an unknown function $f \in \mathcal{N}_K(\Omega)$, that is, $f_j = f(x_j)$, $1 \leq j \leq N$. Here, $\mathcal{N}_K(\Omega)$ denotes the native Hilbert space of a positive definite radial basis function K. One has to solve the system

$$(\mathbf{K} + \lambda \text{Id}) \, b = f|_X \, , \tag{20}$$

with $\mathbf{K} = (K(x_i - x_j))_{i,j=1\ldots N}$ to build an approximant

$$s_{\lambda,X,K}(f)(\cdot) := \sum_{x_j \in X} b_j K(\cdot, x_j) \, .$$

Classical interpolant arises as a special case, namely for $\lambda = 0$. We follow the paths of [17, Section 3], where recovery in a setting of finite smoothness is considered. However, in the case of infinitely smooth functions, regularization is even more important due to the bad condition of the Gram matrix $\mathbf{K}$ [9]. We know from [17, Proposition 3.1] the following stability and consistency properties of the approximant,

$$\begin{aligned}
\|s_{\lambda,X,K}(f)\|_{\mathcal{N}_K(\Omega)} &\leq \|f\|_{\mathcal{N}_K(\Omega)} \\
\|s_{\lambda,X,K}(f)|_X - f|_X\|_{\ell_\infty(X)} &\leq \sqrt{\lambda} \|f\|_{\mathcal{N}_K(\Omega)} \, .
\end{aligned} \tag{21}$$

Theorem 6.1 *If Ω is a compact cube with side length R, then for all $1 \leq q \leq \infty$, there exist constants $h_0, C, \tilde{C} > 0$ such that for all data sets $X \subset \Omega$ with fill distance $h \leq h_0$ and for all $f \in \mathcal{N}_G(\Omega)$ in the native space of a Gaussian kernel G,*

$$\|f - s_{\lambda,X,K}(f)\|_{L_q(\Omega)} \leq \left(2e^{C\log(h)/h} + \sqrt{\lambda}e^{\tilde{C}/h} \right) \|f\|_{\mathcal{N}_G(\Omega)} .$$

For all data sets $X \subset \Omega$ with fill distance $h \leq h_0$, and all $f \in \mathcal{N}_M(\Omega)$ in the native space of an inverse multiquadric kernel we have

$$\|f - s_{\lambda,X,K}(f)\|_{L_q(\Omega)} \leq \left(2e^{-C/h} + \sqrt{\lambda}e^{\tilde{C}/h} \right) \|f\|_{\mathcal{N}_M(\Omega)} .$$

Remark 6.2 *In the case $\lambda = 0$, i.e., the standard interpolation we obtain the well known orders for interpolation with Gaussian and inverse multiquadric kernels on compact cubes [16].*

Proof of Theorem 6.1: For Gaussian kernels Theorem 4.5 and estimates (21) give

$$\begin{aligned}
\|f - s_{\lambda,X,K}(f)\|_{L_q(\Omega)} &\leq& e^{C\log(h)/h}\|f - s_{\lambda,X,K}(f)\|_{\mathcal{N}_G(\Omega)} \\
&+& e^{\tilde{C}/h}\|s_{\lambda,X,K}(f)|_X - f|_X\|_{L_\infty(X)} \\
&\leq& \left(2e^{C\log(h)/h} + \sqrt{\lambda}e^{\tilde{C}/h} \right) \|f\|_{\mathcal{N}_G(\Omega)} ,
\end{aligned}$$

and similar considerations apply to the inverse multiquadrics. $\qquad\square$

From here on we restrict ourselves to the case of Gaussian kernels since all results can be carried over to inverse multiquadrics easily. We write abbreviately $\mathcal{N} := \mathcal{N}_G$. Theorem 6.1 suggests an *a priori* choice of the regularization parameter λ, which we state in the following corollary.

Corollary 6.3 *If we choose $\lambda \leq \exp\left(\frac{2\left(C\log(h)-\tilde{C}\right)}{h} \right)$ the approximation error on a compact cube Ω is bounded by*

$$\|f - s_{\lambda,X,K}(f)\|_{L_q(\Omega)} \leq 3e^{C\log(h)/h}\|f\|_{\mathcal{N}(\Omega)} .$$

We point out that we achieve the same approximation order for regularized kernel-based interpolation in the native space on cubes as is known for interpolation [16], provided the regularization parameter λ is chosen appropriately.

For Lipschitz domains obeying a cone condition, we can proceed similarly employing Theorem 3.5.

Theorem 6.4 *If the bounded domain Ω has a Lipschitz boundary, and satisfies an interior cone condition with a maximum radius R and angle ϕ, then for all $1 \leq q \leq \infty$, there exist constants C, C_2 and $h_0 > 0$ such that for all data sets $X \subset \Omega$ with fill distance $h \leq h_0$*

$$\|D^\alpha(f - s_{\lambda,X,K}(f))\|_{L_q(\Omega)} \leq \left(2e^{C\log(h)/\sqrt{h}} + \sqrt{\lambda}C_2 h^{-|\alpha|} \right) \|f\|_{\mathcal{N}(\Omega)} .$$

Here the constants C and C_2 depend only on ϕ, α, d and q but not on h or f.

The constant C crucially influencing the speed of convergence of $e^{C \log(h)/\sqrt{h}}$ for $h \to 0$ is given as $C = \sqrt{C_{\min}}/4$ with the constant $C_{\min}$ as defined in (8). Thus, it depends only on the parameters of the cone condition and in particular not on the space dimension d or α. This a great advantage compared to the well known error estimates on compact cubes, where the constants grow exponentially with the space dimension. Again, we can choose λ appropriately *a priori* to obtain the same approximation orders as for classical interpolation.

Corollary 6.5 *If we choose $\lambda \leq C_2^{-2} e^{2C \log(h)/\sqrt{h}} h^{2|\alpha|}$, the approximation error for regularized interpolation is for all sets $X \subset \Omega$ with fill distance $h \leq h_0$ bounded by*

$$\|f - s_{\lambda,X,K}(f)\|_{L_q(\Omega)} \leq 3 e^{C \log(h)/\sqrt{h}} \|f\|_{\mathcal{N}(\Omega)} \,.$$

This shows that we can improve the condition number of the system (20) at least to the value of $\lambda = C_2^{-2} e^{2C \log(h)/\sqrt{h}} h^{2|\alpha|}$ instead of e^{-c/q^2} for the Gaussian kernel [16] and still get good approximation orders. We point out that we get rid of the separation distance

$$q := q_X := \frac{1}{2} \min_{1 \leq i,j \leq N} \|x_j - x_i\|_2 \,,$$

which can spoil the condition number in case of badly distributed points.

7 Application to Support Vector Machines

In many kernel-based machine learning algorithms, such as support vector machines, infinitely smooth kernels are quite popular. The most important examples are Gaussians and infinite dor product kernels [4]. Since dot product kernels have been treated in the more general context of power series kernels [19], we restrict ourselves to the native spaces of Gaussian kernels but the results can be easily carried over to other spaces of infinitely smooth function. We consider examplarily the following optimization problem, the so called ν-support vector regression (SVR) [12], in Hilbert space formulation. The function

$$V_\epsilon(x) = |x|_\epsilon + \epsilon \nu$$

is related to Vapnik's ϵ-intensive loss function [13]

$$|x|_\epsilon = \begin{cases} 0 & \text{if } |x| \leq \epsilon \\ |x| - \epsilon & \text{if } |x| > \epsilon \,, \end{cases}$$

but has an additional term with a positive parameter ν. The associated optimization problem ν-SVR takes the form

$$\min_{\substack{f \in \mathcal{N}_K(\Omega) \\ \epsilon \in \mathbb{R}_0^+}} \frac{1}{N} \sum_{j=1}^{N} |f(x_j) - y_j|_\epsilon + \epsilon \nu + \frac{1}{2C} \|f\|_{\mathcal{N}_K(\Omega)}^2 \tag{22}$$

with positive parameters C and ν. We point out that we use a slightly different notation in this section, which comes from the machine learning literature. In particular, the character C denotes a problem parameter rather than a (generic) constant. The solution of the regression problem (22) can be computed by solving a finite dimensional optimization problem (see [12] for details).

7.1 Exact Data

In the case of exact data given data $y \in \mathbb{R}^N$, generated by a function $f \in \mathcal{N}_K(\Omega)$ from the native space of a kernel K, that is,

$$f(x_j) = y_j \quad \text{for } j = 1, \ldots, N, \tag{23}$$

a solution (f^*, ϵ^*) of the ν-SVR (22) exists and satisfies the estimates [8]

$$\|f^*\|_{\mathcal{N}_K(\Omega)} \leq \|f\|_{\mathcal{N}_K(\Omega)} \quad \text{and}$$

$$\|f^*|_X - y\|_{\ell_\infty(X)} \leq \frac{N}{2C}\|f\|_{\mathcal{N}_K(\Omega)}^2 - \epsilon^* \cdot (N\nu - 1). \tag{24}$$

As in the previous section, we restrict ourselves to the case of Gaussian kernels and write again abbreviately $\mathcal{N} := \mathcal{N}_G$. Applying Theorem 4.5 to the error function yields with estimates (24) the following bound.

Theorem 7.1 *Let Ω be a compact cube with side length $0 < 2R < \infty$. Then, for all $1 \leq q \leq \infty$, there are constants $a, b > 0$ such that for all training sets $X \subset \Omega$ with fill distance $h \leq h_0$ a solution (f^*, ϵ^*) of the optimization problem (22) satisfies*

$$\|f - f^*\|_{L_q(\Omega)} \leq 2e^{a\log(h)/h}\|f\|_{\mathcal{N}(\Omega)} + \frac{N}{2C}e^{b/h}\|f\|_{\mathcal{N}(\Omega)}^2 + (1 - N\nu)\epsilon^* e^{b/h}.$$

The constants a and b depend only on d, R and q, but not on f, N or h.

As in the case of smoothed interpolation, the error bound suggests some parameter choices.

Corollary 7.2 *Under the assumptions of Theorem 7.1 we can choose the problem parameters as*

$$C = \frac{N}{2}\exp\left(\frac{b - a\log(h)}{h}\right)\|f\|_{\mathcal{N}(\Omega)} \quad \text{and} \quad \nu \geq \frac{1}{N},$$

to obtain

$$\|f - f^*\|_{L_q(\Omega)} \leq 3e^{a\log(h)/h}\|f\|_{\mathcal{N}(\Omega)}. \tag{25}$$

Following [12], choosing the parameter $\nu \geq 1$ ensures a non-trivial solution f^*. Since the error analysis for trivial solutions f^* is not very interesting, this seems to be a reasonable assumption.

For learning on general domains satisfying an interior cone condition we can proceed similarly applying Theorem 3.5 to the error function.

Theorem 7.3 *Suppose $\Omega \subset \mathbb{R}^d$ is bounded with a Lipschitz boundary, and satisfies an interior cone condition with a maximum radius R and angle ϕ. Let (f^*, ϵ^*) be a solution of (22). Then the approximation error can be bounded by*

$$\|f - f^*\|_{L_q(\Omega)} \leq e^{A\log(h)/\sqrt{h}}\|f\|_{\mathcal{N}(\Omega)} + \frac{C_2 N}{2C}\|f\|_{\mathcal{N}(\Omega)}^2 - C_2(N\nu - 1)\epsilon^*.$$

Here the constant A depends only on ϕ, and R, but not on d, N, h, α, q or f. The constant C_2 comes from Corollary 3.4. If we choose the parameters as

$$C = \frac{C_2 N}{2}e^{-A\log(h)/\sqrt{h}}\|f\|_{\mathcal{N}(\Omega)} \quad \text{and} \quad \nu \geq \frac{1}{N},$$

we get the estimate

$$\|f - f^*\|_{L_q(\Omega)} \leq 3e^{A\log(h)/\sqrt{h}}\|f\|_{\mathcal{N}(\Omega)}. \tag{26}$$

7.2 Inexact Data

We now consider the case of inexact given data, i.e., the data y generated by a (unknown) function $f \in \mathcal{N}(\Omega)$ is given by

$$f(x_j) = y_j + r_j \quad \text{for } j = 1, \ldots, N, \quad \text{and } f \in \mathcal{N}(\Omega), \tag{27}$$

where $r = (r_1, \ldots, r_N)$ describes some additive error with error level δ satisfying

$$\|r\|_{\ell_\infty(X)} \leq \delta \leq \|f\|_{\mathcal{N}(\Omega)}.$$

Our error analysis is purely deterministic, i.e., there are no assumptions concerning the error distribution. Again we have to bound both the native space norm and the discrete norm. Following [8] we have for a solution (f^*, ϵ^*) and any $\epsilon > 0$,

$$\|f^*\|_{\mathcal{N}(\Omega)} \leq \sqrt{\frac{2C}{N}\sum_{j=1}^{N}|r_j|_\epsilon + 2C\nu\epsilon + \|f\|_{\mathcal{N}(\Omega)}^2} \quad \text{and}$$

$$\|f^* - y\|_{\ell_\infty(X)} \leq \sum_{j=1}^{N}|r_j|_\epsilon + \nu N\epsilon - \epsilon^* \cdot (N\nu - 1) + \frac{N}{2C}\|f\|_{\mathcal{N}(\Omega)}^2.$$

We shall restrict ourselves to the case of general Lipschitz domains, but similar results can be obtained on cubes using the appropriate sampling inequalities from Section 4. Theorem 3.5 immediately yields the following error bound.

Theorem 7.4 *Suppose $\Omega \subset \mathbb{R}^d$ is bounded with Lipschitz boundary, and satisfies an interior cone condition with a maximum radius R and angle ϕ. Then, for all $1 \leq q \leq \infty$, there are constants D and $h_0 > 0$ such that for all sets $X \subset \Omega$ with fill distance $h \leq h_0$, for all $\epsilon > 0$ and all $f \in \mathcal{N}(\Omega)$ a solution (f^*, ϵ^*) of (22) satisfies under the assumption (27),*

$$\|f - f^*\|_{L_q(\Omega)} \leq e^{D\log(h)/\sqrt{h}}\left(\|f\|_{\mathcal{N}(\Omega)} + \sqrt{\frac{2C}{N}\sum_{j=1}^{N}|r_j|_\epsilon + 2C\nu\epsilon + \|f\|_{\mathcal{N}(\Omega)}^2}\right) +$$

$$+ \quad C_2\left(\sum_{j=1}^{N}|r_j|_\epsilon + \nu N\epsilon - \epsilon^*(N\nu - 1) + \frac{N}{2C}\|f\|_{\mathcal{N}(\Omega)}^2\right).$$

Here, the constant C_2 comes from Theorem 3.5, D depends only on the parameters of the cone condition, and h_0 depends on ϕ, R, d and q, but not on h, f, δ or N. If we choose

$$C = \frac{N\|f\|_{\mathcal{N}(\Omega)}^2}{2\delta}, \quad \epsilon = \delta, \quad \nu = \frac{1}{N},$$

then for sufficiently small h, any solution (f^*, ϵ^*) satisfies

$$\|f - f^*\|_{L_2(\Omega)} \leq \left(1 + \sqrt{2}\right)e^{D\log(h)/\sqrt{h}}\|f\|_{\mathcal{N}(\Omega)} + 2C_2\delta. \tag{28}$$

Note that bounds like (28) allow excellent bounds on the number of training samples required in the worst possible case to get required prediction quality.

A Embedding Constants

Theorem A.1 *For the Gaussian $G(x) = e^{-c^2\|x\|_2^2}$ we have for all $f \in \mathcal{N}_G(\mathbb{R}^d)$ and all $k \in \mathbb{N}$*

$$\|f\|_{W_2^k(\mathbb{R}^d)} \le \left((2c^2)^{d/2} + 1\right)^{\frac{k}{2}} \left(\frac{8c^2}{e} + 2\right)^{\frac{k}{2}} k^{\frac{k}{2}} \|f\|_{\mathcal{N}_G(\mathbb{R}^d)} .$$

Proof: By the definition of the spaces in Section 5 it suffices to check for the choice $E(k) = \left(\frac{8c^2}{e} + 2\right)^{\frac{k}{2}}$ if the inequality

$$(1 + r)^k \le E(k)^2 e^{\frac{r}{4c^2}}$$

holds for all $r \ge 0$. Here we substituted $r = \|x\|_2^2$. We split our analysis into two parts. For $r \le 1$ we see that $(1 + r)^k \le 2^k$ and $e^{\frac{r}{4c^2}} \ge 1$, hence $E(k)^2 = 2^k$ will work in this case. For $r > 1$ things are more complicated. First we make the observation $(1 + r)^k \le 2^k r^k$. If we change variables $r \mapsto 4c^2 r$ it remains to check

$$2^k(4c^2 r)^k = (8c^2)^k r^k \le E(k)^2 e^r \Leftrightarrow k \ln\left(8c^2\right) + k \ln r \le r + \ln E(k)^2 .$$

We shall look at the funktion

$$f : (1, \infty) \to \mathbb{R}_+, \quad r \mapsto r + \ln E(k)^2 - \ln k \left(8c^2\right) - k \ln r$$

and compute a minimum of this function. Easy calculation yields

$$f'(r) = 1 - \frac{k}{r} \quad \text{and} \quad f''(r) = \frac{k}{r^2} > 0 .$$

Since $f'(r) = 0$ implies $r = k$ we see that the global minimum is attained for $r_{\min} = k$, with $f(r_{\min}) = f(k) = k + \ln E(k)^2 - k(\ln\left(8c^2\right) + \ln k)$. This shows that

$$E(k)^2 = \left(\frac{8c^2}{e}\right)^k k^k$$

will be sufficient.

Hence taking the maximum $E(k)^2 = \max\left\{2^k, \left(\frac{8c^2}{e}\right)^k k^k\right\} \le \left(\frac{8c^2}{e} + 2\right)^k k^k$ will ensure

$$(1 + r)^k \le E(k)^2 e^{\frac{r}{4c^2}} \quad \text{for all } r \in [0, \infty) .$$

$\square$

Theorem A.2 *For the inverse multiquadrics $K_M(x) = \left(c^2 + \|x\|_2^2\right)^{-\beta}$ with $\beta > \frac{d}{2}$, we have for all $f \in \mathcal{N}_M(\mathbb{R}^d)$ and all $k \in \mathbb{N}$*

$$\|f\|_{W_2^k(\mathbb{R}^d)} \le \max\left\{\frac{2^k}{C_1}, C_2^k k^{2k}\right\}^{1/2} \|f\|_{\mathcal{N}_M(\mathbb{R}^d)} ,$$

where the constants $C_1, C_2 > 0$, depending only on β, c and d, are defined in the proof.

Proof: The Fourier transform of the inverse multiquadrics is given by [16, Theorem 6.13]

$$\widehat{K_M}(\omega) = \frac{2^{1-\beta}}{\Gamma(\beta)} \left(\frac{\|\omega\|_2}{c} \right)^{\beta - d/2} \mathcal{K}_{d/2-\beta} \left(c\|\omega\|_2 \right) \,,$$

where $\mathcal{K}_\sigma$ denotes the modified Bessel function of third kind. As in the case of Gaussians it suffices to check for this choice for $E(k)$ if the inequality

$$(1 + r^2)^k \le E(k)^2 \frac{\Gamma(\beta)}{2^{1-\beta}} \left(\left(\frac{r}{c} \right)^{\beta - d/2} \mathcal{K}_{d/2-\beta}(cr) \right)^{-1}$$

holds for all $r \ge 0$. Again we split our analysis into two parts: $r \in [0, 1]$ and $r > 1$. In the case $r \in [0, 1]$ we use that

$$g : [0, \infty) \to \mathbb{R} \quad , \quad r \mapsto \frac{2^{1-\beta}}{\Gamma(\beta)} \left(\frac{r}{c} \right)^{\beta - d/2} \mathcal{K}_{d/2-\beta}(cr)$$

is a positive and continuous mapping since the singularity of the Bessel function at 0 is of order $\beta - \frac{d}{2}$ [14]. Further, the function g is nonincreasing for $r > 0$ [16, Cor. 5.12], so the function $\frac{1}{g}$ is nondecreasing, hence the minimum of $\frac{1}{g}$ is reached for $r = 0$. This minimum is given by

$$C_1 := \frac{c^{2\beta - d} \cdot 2^{d/2} \cdot \Gamma(\beta)}{\Gamma(\beta - d/2)} \,.$$

Since $(1 + r^2)^k \le 2^k$ for $r \in [0, 1]$ we can conclude that $E(k)^2 = \frac{2^k}{C_1}$ is sufficient in the first case.

For $r > 1$ we use the inequality [16, Lemma 5.13]

$$\left(\left(\frac{r}{c} \right)^{\beta - d/2} \mathcal{K}_{\beta - d/2}(cr) \right)^{-1} \ge r^{-\beta + \frac{d-1}{2}} \cdot c^{\beta - \frac{d+1}{2}} \cdot (2\pi)^{-1/2} e^{cr} e^{-\frac{(\beta - d/2)^2}{2cr}} \,.$$

Since $r > 1$ and $\beta > \frac{d}{2}$ easy calculations yield the bounds

$$e^{-\frac{(d/2 - \beta)^2}{2cr}} \ge e^{-\frac{(d/2 - \beta)^2}{2c}} \quad \text{and}$$

$$e^{rc} \cdot r^{-\beta + \frac{(d+1)}{2}} \ge e^{rc} \cdot r^{-\beta} \ge \left(\frac{ec}{2\beta} \right)^\beta e^{rc/2} \,.$$

If we use $(1 + r^2)^k \le 2^k r^{2k}$, we have to determine $E(k)$ such that for all $r > 1$

$$f(r) := \ln \tilde{C} + \ln E(k)^2 + \frac{rc}{2} - 2k \left(\frac{\ln(2)}{2} + \ln r \right) \ge 0 \,,$$

$$\text{where} \quad \tilde{C} = \frac{\Gamma(\beta)}{2^{1-\beta}} \left(\frac{ec}{2\beta} \right)^\beta c^{\beta - \frac{d+1}{2}} (2\pi)^{-1/2} e^{-\frac{(\beta - d/2)^2}{2c}} \,.$$

We find $f'(r) = -\frac{2k}{r} + \frac{c}{2}$ and $f''(r) = \frac{2k}{r^2}$. This gives $r_{min} = 4k/c$ and

$$f(r_{\min}) = \ln \tilde{C} + \ln E(k)^2 + 2k - 2k(\ln(2)/2 + \ln(k) + \ln(4/c)) \,.$$

Thus, for $E(k)^2 \ge e^{\tilde{C}} 2^k e^{-2k} \left(\frac{4}{c} \right)^{2k} k^{2k}$ the function f is positive. This yields that the constant $E(k)^2 = \max \left\{ \frac{2^k}{C_1}, C_2^k k^{2k} \right\}$ with $C_2 = 16 \left(e^{-\tilde{C}} + 1 \right) (ce)^{-2}$ will ensure the claim. $\qquad \square$

References

[1] A. AGADZHANOV, *Functional properties of Sobolev spaces of infinite order*, Soviet. Math. Dokl., 38, No. 1 (1989), pp. 88–92.

[2] S. BRENNER AND L. SCOTT, *The Mathematical Theory of Finite Element Methods*, Springer, New York, 1994.

[3] J. DUCHON, *Sur l'erreur d' interpolation des fonctions de plusieurs variables par les D^m-splines.*, Rev. Française Automat. Informat. Rech. Opèr. Anal. Numer., 12 (1978), pp. 325–334.

[4] F. LU AND H. SUN, *Positive definite dot product kernels in learning theory*, Advances in Computational Mathematics, 22 (2005), pp. 181–198.

[5] W. MADYCH AND S. NELSON, *Bounds on multivariate polynomials and exponential error estimates for multiquadric interpolation*, J. Approx. Theory, 70 (1992), pp. 94–114.

[6] W. R. MADYCH, *An estimate for multivariate interpolation II*, J. Approx. Theory, 142 (2006), pp. 116–128.

[7] F. NARCOWICH, J. WARD, AND H. WENDLAND, *Sobolev bounds on functions with scattered zeros, with applications to radial basis function surface fitting*, Mathematics of Computation, 74 (2005), pp. 743–763.

[8] C. RIEGER AND B. ZWICKNAGL, *Deterministic Error Analysis of Support Vector Machines and Related Regularized Kernel Methods.* Preprint Göttingen/Leipzig, 2006.

[9] R. SCHABACK, *Error estimates and condition numbers for radial basis function interpolation*, Advances in Computational Mathematics, 3 (1995), pp. 251–264.

[10] R. SCHABACK, *Convergence of Unsymmetric Kernel-Based Meshless Collocation Methods*, SIAM Journal of Numerical Analysis, 45/1 (2007), pp. 333–351.

[11] R. SCHABACK AND H. WENDLAND, *Kernel Techniques: From Machine Learning to Meshless Methods*, Acta Numerica, 15 (2006), pp. 543–639.

[12] B. SCHÖLKOPF, R. C. WILLIAMSON, AND P. L. BARTLETT, *New Support Vector Algorithms*, Neural Computation, 12 (2000), pp. 1207–1245.

[13] V. VAPNIK, *The nature of statistical learning theory*, Springer-Verlag, New York, 1995.

[14] G. WATSON, *A Treatise on the Theory of Bessel Functions*, Cambridge University Press, Cambridge, 1966.

[15] H. WENDLAND, *Local polynomial reproduction and moving least squares approximation*, IMA J. Numer. Anal., 21 (2001), pp. 285–300.

[16] H. WENDLAND, *Scattered Data Approximation*, Cambridge Monographs on Applied and Computational Mathematics, Cambridge University Press, Cambridge, 2005.

[17] H. WENDLAND AND C. RIEGER, *Approximate Interpolation*, Numerische Mathematik, 101 (2005), pp. 643–662.

[18] J. WLOKA, *Partielle Differentialgleichungen: Sobolevräume und Randwertaufgaben*, Mathematische Leitfäden, Teubner, Stuttgart, 1982.

[19] B. ZWICKNAGL, *Power Series Kernels*. to appear in: Constructive Approximation.