Max-Planck-Institut
für Mathematik
in den Naturwissenschaften
Leipzig

Power Series Kernels

(revised version: October 2007)

by

Barbara Zwicknagl

Preprint no.: 155 2006

Constructive Approximation manuscript No.
(will be inserted by the editor)

Power Series Kernels

Barbara Zwicknagl[1]

Max Planck Institute for Mathematics in the Sciences, Inselstr. 22, 04103 Leipzig, Germany,

email: Barbara.Zwicknagl@mis.mpg.de

Received: date / Revised version: date

Abstract We introduce a class of analytic positive definite multivariate kernels which includes infinite dot product kernels as sometimes used in machine learning, certain new nonlinearly factorizable kernels and a kernel which is closely related to the Gaussian. Each such kernel reproduces in a certain ‘native‘ Hilbert space of multivariate analytic functions. If functions from this space are interpolated in scattered locations by translates of the kernel, we prove spectral convergence rates of the interpolants and all derivatives. By truncation of the power series of the kernel-based interpolants, we constructively generalize the classical Bernstein theorem concerning polynomial approximation of analytic functions to the multivariate case. An application to machine learning algorithms is presented.

Keywords: multivariate polynomial approximation, Bernstein theorem, dot product kernels, reproducing kernel Hilbert spaces, error bounds, convergence orders

Classification: 41A05, 41A10, 41A25, 41A58, 41A63, 68T05

1 Introduction

It is well known [1] that non-stationary multivariate interpolation of certain classes of analytic functions by certain positive definite analytic radial basis functions like the Gaussians and inverse Multiquadrics leads to exponential approximation orders. On the other hand, multivariate polynomial interpolation by the technique of de Boor and Ron [2] arises as the ‘flat limit‘ of interpolation by Gaussians [3]. This surprising result was established via an intermediate kernel of the power series form

$$K(x,y) = \sum_{\alpha \in \mathbb{N}_0^d} \frac{x^\alpha y^\alpha}{\alpha!} \ .$$

Starting from this observation we consider general multivariate positive definite kernels of the form

$$K(x,y) = \sum_{\alpha \in \mathbb{N}_0^d} w_\alpha \frac{x^\alpha y^\alpha}{\alpha!^2} \tag{1}$$

for general weights w_α and call them *Power Series Kernels*. It turns out that this class combines several other interesting cases of d-variate analytic kernels, e.g. of the forms

$$K(x,y) = \sum_{j=0}^{\infty} a_j < x, y >^j \quad \text{or}$$
$$K(x,y) = \prod_{k=1}^{d} f(x_k \cdot y_k), \quad f(z) = \sum_{j=0}^{\infty} f_j z^j \ .$$

Kernels of the first type are called ‘infinite dot product‘ kernels or ‘infinite polynomial‘ kernels while the second type seems to be new and can be called ‘nonlinearly factorizable‘. It turns out that the Gaussian kernel G, which plays a key role in the investigation of ‘flat limits‘ [4], is closely related to the class of power series kernels since

$$G(x,y) = e^{-c\|x-y\|_2^2} = \left(e^{-c\|x\|_2^2}\right) \sum_{\alpha\in\mathbb{N}_0^d} \frac{(2c)^{|\alpha|}}{\alpha!} x^\alpha y^\alpha \left(e^{-c\|y\|_2^2}\right)$$

with the power series kernel $K(x,y) = \sum_{\alpha\in\mathbb{N}_0^d} \frac{(2c)^{|\alpha|}}{\alpha!} x^\alpha y^\alpha$.

Based on general theoretic background by Madych and Nelson [1], de Boor and Ron [2], Schaback [3] and Wendland [5] among others on multivariate interpolation of certain classes of analytic functions, we shall investigate this general class of power series kernels. Under weak additional conditions on the weights, the next section of this paper proves existence and positive definiteness of power series kernels and determines the ‘native‘ Hilbert spaces of multivariate functions in which they are reproducing. The following sections derive spectral error bounds for non-stationary multivariate scattered data interpolation using power series kernels, if the data comes from functions of their respective native spaces. The bounds are in terms of the *fill distance*

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

for discrete data sets X in a fixed bounded domain $\Omega \subset \mathbb{R}^d$, and have one of the forms

$$e^{-a/h},\ e^{a\log(bh)/h},\ e^{-a/\sqrt{h}},\ e^{a\log(bh)/\sqrt{h}}$$

with positive constants a, b, depending on the behaviour of the weights in (1) and the presence of derivatives. This generalizes results of [1] to more general analytic kernels and to derivatives, and it will be very useful for proving results on the behaviour of machine learning with dot product kernels [6], see section 7.

Truncation of the power series of kernel-based interpolants leads to good polynomial approximants. If the truncation is properly related to the fill distance h, one can approximate analytic functions from native spaces of power series kernels by multivariate polynomials of total degree not exceeding k in the L_∞-norm at a geometric rate q^k for some $q < 1$. This generalizes a classical theorem of Bernstein to the multivariate situation and allows a constructive realization of good multivariate polynomial approximations.

Section 7 shows how the results of the previous chapters can be used to derive an error analysis for quite general non-interpolatory kernel-based approximation algorithms, such as several popular support vector machines [7].

2 Power series kernels

For all $\alpha \in \mathbb{N}_0^d$ let w_α be a positive real number such that $\sum_{\alpha\in\mathbb{N}_0^d} \frac{w_\alpha}{\alpha!^2} < \infty$ holds. Then we call the kernel

$$K(x,y) = \sum_{\alpha\in\mathbb{N}_0^d} w_\alpha \frac{x^\alpha}{\alpha!}\frac{y^\alpha}{\alpha!} \tag{2}$$

a *power series kernel* (PSK) on $(-1,1)^d$. We restrict ourselves to the unit cube $(-1,1)^d$ since the diameter can be adjusted by rescaling $K(x,y) \to K(\theta x, \theta y)$, and similar considerations apply to translation.

Remark 1 Power series kernels are positive definite on $(-1,1)^d$.

Proof: For arbitrary $N \in \mathbb{N}$, $x_1, \ldots, x_N \in (-1,1)^d$ and $\beta \in \mathbb{R}^N$ we have

$$\sum_{j=1}^{N}\sum_{k=1}^{N}\beta_j\beta_k K(x_j,x_k) = \sum_{\alpha\in\mathbb{N}_0^d}\frac{w_\alpha}{\alpha!^2}\left(\sum_{j=1}^{N}\beta_j x_j^\alpha\right)^2 \geq 0,$$

and $\sum_{j=1}^{N}\sum_{k=1}^{N}\beta_j\beta_k K(x_j,x_k) = 0$ implies $\sum_{j=1}^{N}\beta_j p(x_j) = 0$ for all polynomials p on $(-1,1)^d$. Since polynomials separate points, this implies $\beta_j = 0$ for all $j = 1, \ldots, N$. □

Note that the condition $w_\alpha > 0$ for all $\alpha \in \mathbb{N}_0^d$ is sufficient but not necessary to ensure that a kernel K of the power series form (2) is positive definite. The complete theory presented here applies also to more general positive definite kernels of this form.

There are two main classes of kernels of the form (2). First there are *infinite dot product kernels* which perform well on certain support vector algorithms [8]. Infinite dot product kernels on $(-1,1)^d$ are kernels of the form

$$D(x,y) = \sum_{j=0}^{\infty} a_j \langle x,y\rangle^j = \sum_{\alpha\in\mathbb{N}_0^d} a_{|\alpha|}\frac{|\alpha|!}{\alpha!}x^\alpha y^\alpha\,, \tag{3}$$

where $a_j \geq 0$ for $j = 0,1,\ldots$ with $\sum_{j=0}^{\infty} a_j 2^j < \infty$. If there are infinitely many even integers $j_e \in 2\mathbb{N}$ and infinitely many odd integers $j_o \in 2\mathbb{N}+1$ such that $a_{j_e} > 0$ and $a_{j_o} > 0$, then the dot product kernel (3) is positive definite [8].

Dot product kernels can be written in the form

$$D(x,y) = \sum_{\alpha\in\mathbb{N}_0^d}\tilde{w}_\alpha\frac{x^\alpha}{\alpha!}\frac{y^\alpha}{\alpha!}$$

where the weights $\tilde{w}_\alpha \geq 0$ are specified by (3). All considerations presented here can be applied to positive definite dot product kernels as well if we simply replace the index set $\mathbb{N}_0^d$ by $S := \left\{\alpha\in\mathbb{N}_0^d | \tilde{w}_\alpha \neq 0\right\}$.

In our examples we mainly deal with a second class of power series kernels, namely *nonlinearly factorizable* kernels.

Definition 1 *We call a kernel K of the power series form (2)* nonlinearly factorizable *if there is an analytic function $f : \mathbb{R} \to \mathbb{R}$, $f(z) = \sum_{n=0}^{\infty} f_n z^n$ such that*

$$K(x,y) = \prod_{j=1}^{d} f\left(x_j y_j\right) = \sum_{\alpha \in \mathbb{N}_0^d} \left(\prod_{j=1}^{d} f_{\alpha_j} \right) x^{\alpha} y^{\alpha} .$$

Special cases of factorizable power series kernels arise for all f with positive expansion coefficients f_n, e.g.

$$K(x,y) = \exp < x,y >, \ f(z) = \exp(z) ,$$
$$K(x,y) = \prod_{j=1}^{d} \frac{1}{1 - c x_j y_j}, \ f(z) = \frac{1}{1 - cz} \ \text{ for } \ 0 < c < 1 ,$$
$$K(x,y) = \prod_{j=1}^{d} I_0 \left(2 \left(x_j y_j\right)^{\frac{1}{2}} \right), \ f(z) = I_0 \left(2 z^{\frac{1}{2}} \right)$$

with the modified Bessel function I_0 [9].

For a power series kernel K we define the associated function space [3]

$$\mathcal{N}_K := \left\{ f : (-1,1)^d \to \mathbb{R} \;\middle|\; f(\cdot) = \sum_{\alpha \in \mathbb{N}_0^d} a_\alpha (\cdot)^\alpha \text{ with } \sum_{\alpha \in \mathbb{N}_0^d} \frac{\alpha!^2}{w_\alpha} a_\alpha^2 < \infty \right\}$$

equipped with the inner product

$$(f,g)_{\mathcal{N}_K} = \sum_{\beta \in \mathbb{N}_0^d} \frac{1}{w_\beta} (D^\beta f(0))(D^\beta g(0)) .$$

Then $\mathcal{N}_K$ is well-defined and isometrically isomorphic to a weighted ℓ_2-space, i.e., it is a Hilbert space. Further, by Taylor expansion all elements of $\mathcal{N}_K$ are reproduced on $(-1,1)^d$ by the kernel K via

$$f(x) = (f, K(x,\cdot))_{\mathcal{N}_K} .$$

Therefore, $\mathcal{N}_K$ is the uniquely determined reproducing kernel Hilbert space associated to K, namely the *native space* of K [5].

3 Interpolation in the Native Space

In this chapter we shall analyze the following interpolation problem. For given centers $X = \{x_1, \dots, x_N\} \subset \Omega$ and data $(f_1, \dots, f_N)^T \in \mathbb{R}^N$ generated by a (unknown) function $f \in \mathcal{N}_K$ we consider the uniquely determined interpolant of the form $s_{f,X}(x) = \sum_{j=1}^N \alpha_j K(x, x_j)$ satisfying $s_{f,X}(x_k) = f_k$ for $k = 1, \dots, N$. The error estimates between the unknown function f and its interpolant $s_{f,X}$ are usually expressed in terms of the *fill distance*

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

and convergence is studied in the limit $h \to 0$. We shall show below that convergence occurs for the function and all derivatives.

For any $\alpha \in \mathbb{N}_0^d$ and $x \in (-1,1)^d$ the α-th power function is defined by

$$\left[P_{K,X}^{(\alpha)}(x)\right]^2 := D_1^\alpha D_2^\alpha K(x,x) - 2\sum_{j=1}^N D^\alpha u_j^*(x) D_1^\alpha K(x, x_j)$$

$$+ \sum_{i,j=1}^N D^\alpha u_i^*(x) D^\alpha u_j^*(x) K(x_i, x_j) ,$$

where $u_j^*(\cdot) := \sum_{i=1}^N \alpha_i^{(j)} K(\cdot, x_i)$ satisfy $u_j^*(x_k) = \delta_{jk}$ for $j, k = 1, \dots, N$, and D_ℓ^α denotes the α-th derivative with respect to the ℓ-th argument, $\ell = 1, 2$. The error between f and its interpolant $s_{f,X}$ can be bounded by [5, Theorem 11.4]

$$|D^\alpha f(x) - D^\alpha s_{f,X}(x)| \le P_{K,X}^{(\alpha)}(x) \, \|f\|_{\mathcal{N}_K} \; .$$

Consequently, error bounds for interpolation can be derived from upper bounds of the power function. If we fix X, x and α we can define a quadratic form $\mathcal{Q}_\alpha : \mathbb{R}^N \to \mathbb{R}$ by

$$\mathcal{Q}_\alpha(u) = D_1^\alpha D_2^\alpha K(x,x) - 2\sum_{j=1}^N u_j D_1^\alpha K(x,x_j) + \sum_{i,j=1}^N u_i u_j K(x_i,x_j)\ .$$

The vector $D^\alpha u^*(x) = D^\alpha \left(u_1^*(x),\dots,u_N^*(x)\right)^T$ minimizes this quadratic form [5, Theorem 11.5], i.e. $\left[P_{K,X}^{(\alpha)}(x)\right]^2 := \mathcal{Q}_\alpha\left(D^\alpha u^*(x)\right) \leq \mathcal{Q}_\alpha(u)$ holds for all $u \in \mathbb{R}^N$. We shall bound the power function by inserting a suitable vector $u^{(\alpha)}$ into $\mathcal{Q}_\alpha$. For that we use local polynomial reproductions. These come in two variations: with and without derivatives. We shall treat these two cases separately in the next sections.

4 Error Bounds without Derivatives

Here we use the following local polynomial reproduction from [1].

Theorem 1 *Let $\Omega := W(x_0,R) = \{x \in \mathbb{R}^d | \left\|x - x_0\right\|_\infty \leq R\}$ be a cube in $\mathbb{R}^d$. Then there exist constants $c_0, c_2 > 0$ such that for all $\ell \in \mathbb{N}$ and all $X = \{x_1,\dots,x_N\} \subset W(x_0,R)$ satisfying $h_{X,\Omega} \leq c_0/\ell$ there exist functions $u_j : \Omega \to \mathbb{R}$ such that*

1. $\sum_{j=1}^N p(x_j)u_j(x) = p(x)$ *for all $x \in \Omega$ and all polynomials $p \in \pi_\ell(\Omega)$,*
2. $\sum_{j=1}^N |u_j(x)| \leq e^{2d\gamma_d(\ell+1)}$ *for all $x \in \Omega$,*
3. $u_j(x) = 0,$ *if $\|x - x_j\|_2 > c_2 \ell h_{X,\Omega}$ and $x \in \Omega$.*

The constant γ_d is defined recursively by $\gamma_1 = 2$ and $\gamma_n = 2n(1+\gamma_{n-1})$ for $n = 2,3,\dots$.

Following the proof of [5, Theorem 11.21] the constants can be bounded by

$$c_0 := \frac{R}{6\gamma_d} \quad \text{and} \quad c_2 = \frac{3\sqrt{d}\gamma_d(\ell+1)}{\ell} \leq 6\sqrt{d}\gamma_d \quad . \tag{4}$$

The following theorem shows that interpolation with infinitely smooth kernels leads to arbitrarily high algebraic approximation orders for functions from the associated native spaces on any compact set contained in $(-1,1)^d$. We shall need this technical refinement of [5, Theorem 11.13] in order to derive exponential convergence orders for interpolation with power series kernels.

Theorem 2 *Let $\Omega = (-1,1)^d$ and $k \in \mathbb{N}$ be arbitrary. For $f \in \mathcal{N}_K$ we denote the interpolant based on $X = \{x_1, \ldots, x_N\}$ by $s_{f,X}$. Then there are constants $c_0, C > 0$, which depend only on d, such that for all data sets X with fill distance $h := h_{X,\Omega} \leq c_0/(2k)$ we have for all $f \in \mathcal{N}_K$*

$$\|f - s_{f,X}\|_{L_\infty\left((-1,1)^d\right)} \leq \left(\tilde{C}_K^{(2k)}\right)^{1/2} \frac{(Ch)^k}{\sqrt[4]{k}} \|f\|_{\mathcal{N}_K} . \tag{5}$$

The number $\tilde{C}_K^{(2k)}$ is defined by

$$\tilde{C}_K^{(2k)} := \max_{\substack{\beta \in \mathbb{N}_0^d \\ |\beta| = 2k}} \sup_{x,y \in \Omega} \left| D_2^\beta K(x,y) \right| ,$$

where D_2^β denotes the β-th derivative with respect to the second argument.

Proof: For fixed $x \in (-1,1)^d$ we choose a cube $W := W(0,r) \subset (-1,1)^d$ such that $(X \cup \{x\}) \subset \overset{o}{W}$ and $r \geq 1/2$. If we set $u := (u_1(x), \ldots, u_N(x))^T$ with the functions u_j from Theorem 1 for $\ell = 2k-1$, the power function can be bounded by [5, Proof of Theorem 11.13]

$$\left[P_{K,X}^{(0)}(x)\right]^2 \leq \mathcal{Q}(u) = -\sum_j u_j \left(R(x,x_j) + R(x_j,x) + \sum_i u_i R(x_i,x_j) \right)$$

with the Taylor remainder

$$R(w,z) := \sum_{|\beta|=2k} \frac{D_2^\beta K(w,\xi_{w,z})}{\beta!}(z-w)^\beta \ ,$$

where the point $\xi_{w,z}$ belongs to the segment $[w,z]$. By the choice of the vector u the inequalities $\sum_j |u_j| \le e^{2d\gamma_d 2k}$, $\|x-x_j\|_2 < c_2 2kh$ and $\|x_i - x_j\|_2 \le 4c_2 kh$ hold since otherwise $u_j = 0$. Using Stirling's formula and the identity $\sum_{|\beta|_d=k} \frac{1}{\beta!} = \frac{d^k}{k!}$ the first two terms can be bounded by

$$\begin{aligned}\left|\sum_j u_j R(x,x_j)\right| &\le \sum_j |u_j|\, \tilde{C}_K^{(2k)}\, \|x-x_j\|_2^{2k}\, \frac{d^{2k}}{(2k)!} \\ &\le e^{4d\gamma_d k} \tilde{C}_K^{(2k)}\, (2c_2kh)^{2k}\, \frac{(de)^{2k}}{(2k)^{2k}\sqrt{k}} \le \frac{1}{\sqrt{k}} E_1^{2k} \tilde{C}_K^{(2k)} h^{2k}\end{aligned}$$

and similarly

$$\left|\sum_j u_j R(x_j,x)\right| \le \frac{1}{\sqrt{k}} E_1^{2k} \tilde{C}_K^{(2k)} h^{2k}$$

with $E_1 := 12 d^{3/2}\gamma_d e^{2d\gamma_d+1}$. The last term can be bounded using the bound for the ℓ_1-norm of the coefficient vector twice. We get

$$\left|\sum_j u_j \sum_i u_i R(x_j,x_i)\right| \le \frac{1}{\sqrt{k}} E_2^{2k} \tilde{C}_K^{(2k)} h^{2k}$$

with $E_2 := 2e^{2d\gamma_d} E_1$. This finishes the proof with $C := E_2$. □

We shall now derive exponential error bounds. The actual bounds depend on the asymptotic behaviour of the number $\tilde{C}_K^{(2k)}$ for $k \to \infty$. These bounds holds for any infinitely smooth kernel and provide a direct way to derive exponential convergence orders.

Theorem 3 *If there is a constant $c \in \mathbb{R}$, such that $\tilde{C}_K^{(2k)} \le e^{ck}k^k$ for all $k \in \mathbb{N}$, then there exist constants $a, b, \tilde{h} > 0$, such that for all data sets X with fill distance*

$h \le \tilde{h}$ the error between any function $f \in \mathcal{N}_K$ and its interpolant $s_{f,X}$ can be bounded by

$$\|f - s_{f,X}\|_{L_\infty((-1,1))^d} \le e^{a \cdot log(bh)/h} \|f\|_{\mathcal{N}_K} \ .$$

If there is a constant $c \in \mathbb{R}$, such that $\tilde{C}_K^{(2k)} \le e^{ck} k^{2k}$ for all $k \in \mathbb{N}$, then there exist constants $A, \tilde{h} > 0$, such that for all sets X with $h \le \tilde{h}$ the error between any function $f \in \mathcal{N}_K$ and its interpolant $s_{f,X}$ can be bounded by

$$\|f - s_{f,X}\|_{L_\infty\left((-1,1)^d\right)} \le e^{-A/h} \|f\|_{\mathcal{N}_K} \ .$$

Proof: If $\tilde{C}_K^{(2k)} \le e^{ck} k^k$ for all $k \in \mathbb{N}$, Theorem 2 shows that the power function can be bounded by

$$\left[P_{K,X}^{(0)}\right]^2 \le \frac{1}{\sqrt{k}} C^{2k} \tilde{C}_K^{(2k)} h^{2k} \le \left(C_3 k h^2\right)^k$$

with $C_3 := C^2 e^c$, provided that $h \le \frac{c_0}{2k}$. If we define $\tilde{c} := \min\left\{\frac{c_0}{2}, \frac{1}{\sqrt{C_3}}\right\}$ we can for $h \le \tilde{c} =: \tilde{h}$ choose $k \in \mathbb{N}$ such that $\frac{\tilde{c}}{2k} \le h \le \frac{\tilde{c}}{k}$ holds. Then we find

$$\left[P_{K,X}^{(0)}\right]^2 \le k^{-k} \le e^{\tilde{c} \cdot log(2h/\tilde{c})/(2h)} \ ,$$

which establishes the claim with $a := \tilde{c}/4$ and $b := 2/\tilde{c}$.

In the case $\tilde{C}_K^{(2k)} \le e^{ck} k^{2k}$ we set $C_3 := C e^{c/2}$ and define $C_4 := \min\left\{\frac{c_0}{2}, \frac{1}{C_3 e}\right\}$. If we choose k such that $\frac{C_4}{2k} \le h \le \frac{C_4}{k}$ holds, we get

$$\left[P_{K,X}^{(0)}\right]^2 \le (C_3 k h)^{2k} \le e^{-2k} \le e^{-C_4/h} \ ,$$

which finishes the proof with $A := C_4/2$. □

We now want to analyze the asymptotic behaviour of the constant $\tilde{C}_K^{(2k)}$ for some typical choices of the weights w_α from (1). The estimates we derive here can also be applied to dot product kernels (3) since the asymptotic behaviour of the coefficients a_j can be expressed using $\tilde{w}_\alpha = a_{|\alpha|}\alpha!|\alpha|!$.

In general we have

$$\tilde{C}_K^{(2k)} = \max_{\substack{\beta\in\mathbb{N}_0^d \\ |\beta|=2k}} \sum_{\alpha\geq\beta} \frac{w_\alpha}{\alpha!(\alpha-\beta)!} \; .$$

Note that $w_\alpha \leq \tilde{w}_\alpha$ for (almost) all $\alpha \in \mathbb{N}_0^d$ implies $\tilde{C}_K^{(2k)} \leq \tilde{C}_{\tilde{K}}^{(2k)}$ for (almost) all $k \in \mathbb{N}$, if $K(x,y) = \sum \frac{w_\alpha}{\alpha!^2} x^\alpha y^\alpha$ and $\tilde{K}(x,y) = \sum \frac{\tilde{w}_\alpha}{\alpha!^2} x^\alpha y^\alpha$.

Lemma 1 *If $w_\alpha \leq C^{|\alpha|}\cdot\alpha!$ for all $\alpha \in \mathbb{N}_0^d$ with a constant $C > 0$, then the inequality $\tilde{C}_K^{(2k)} \leq C^{2k}e^{dC}$ holds for all $k \in \mathbb{N}$.*

If $w_\alpha = \alpha!^2 c^{|\alpha|}$ for all $\alpha \in \mathbb{N}_0^d$ with a constant $c < 1$, then we have $\tilde{C}_K^{(2k)} = \frac{c^{2k}(2k)!}{(1-c)^{2k+d}}$ for all $k \in \mathbb{N}$.

Proof: For $w_\alpha \leq C^{|\alpha|}\cdot\alpha!$ we have $\tilde{C}_K^{(2k)} \leq C^{2k}\cdot \sum_{\alpha\in\mathbb{N}_0^d} \frac{C^{|\alpha|}}{\alpha!} = C^{2k}\cdot e^{dC}$.

If $w_\alpha = \alpha!^2 c^{|\alpha|}$ for all $\alpha \in \mathbb{N}_0^d$ we have $K(x,y) = \prod_{i=1}^d \frac{1}{1-cx_iy_i}$. Thus we can compute the number $\tilde{C}_K^{(2k)}$ explicitely by taking the derivatives and find

$$\tilde{C}_K^{(2k)} = \max_{\substack{\beta\in\mathbb{N}_0^d \\ |\beta|=2k}} \sup_{x,y\in(-1,1)^d} \prod_{i=1}^d \left| \frac{c^{\beta_i}\beta_i! x_i^{\beta_i}}{(1-cx_iy_i)^{\beta_i+1}} \right| = \frac{c^{2k}(2k)!}{(1-c)^{2k+d}} \; .$$

□

Combining Lemma 1 and Theorem 3 yields the following result.

Corollary 1 *If there exists a constant $C > 0$, such that $w_\alpha \leq C^{|\alpha|}\cdot\alpha!$ for all $\alpha \in \mathbb{N}_0^d$, then there are constants $a, b > 0$, such that for all disrete sets X with sufficiently small fill distance $h \leq \tilde{h}$ the error between any function $f \in \mathcal{N}_K$ and* *its interpolant $s_{f,X}$ can be bounded by*

$$\|f - s_{f,X}\|_{L_\infty\left((-1,1)^d\right)} \le e^{a \cdot log(bh)/h} \|f\|_{\mathcal{N}_K} \ .$$

If there exists a constant $c < 1$, such that $w_\alpha \le \alpha!^2 c^{|\alpha|}$ for all $\alpha \in \mathbb{N}_0^d$, then there is a constant $A > 0$, such that for all sets X with $h \le \tilde{h}$ the error between any function $f \in \mathcal{N}_K$ and its interpolant $s_{f,X}$ can be bounded by

$$\|f - s_{f,X}\|_{L_\infty\left((-1,1)^d\right)} \le e^{-A/h} \|f\|_{\mathcal{N}_K} \ .$$

We point out that in these cases the derived error estimates hold even on the closure $[-1,1]^d$.

We can use the technique described above to derive error bounds for the native space interpolation problem for Gaussian kernels G. To bound the number $\tilde{C}_G^{(2k)}$ we use the decomposition

$$G(x,y) = e^{-c\|x-y\|_2^2} = e^{-c\|x\|_2^2} K(x,y) e^{-c\|y\|_2^2}$$

with the power series kernel $K(x,y) = \sum_{\alpha \in \mathbb{N}_0^d} \frac{(2c)^{|\alpha|}}{\alpha!} x^\alpha y^\alpha$. Then we can employ Leibniz' rule for higher order derivatives and Lemma 1 which leads to

$$\begin{aligned}
\tilde{C}_G^{(2k)} &= \max_{|\beta|=2k} \sup_{x,y} \left| D_2^\beta G\left(x,y\right) \right| \\
&= \max_{|\beta|=2k} \sup_{x,y} e^{\left(-c\|x\|^2\right)} \sum_{\alpha \le \beta} \binom{\beta}{\alpha} \left| D_2^{\beta-\alpha} K\left(x,y\right) \right| \cdot \left| D^\alpha e^{\left(-c\|y\|^2\right)} \right| \\
&\le C^k \max_{|\beta|=2k} \sum_{\alpha \le \beta} \binom{\beta}{\alpha} \sup_y \left| D^\alpha \exp\left(-c\left\|y\right\|_2^2\right) \right| \qquad (6)
\end{aligned}$$

with a constant $C > 0$, which depends only on the space dimension d and the scaling parameter c. To bound the remaining term $D^\alpha \exp\left(-c\|y\|_2^2\right)$ we use the identity

$$\frac{d^n}{dy^n}\exp\left(-cy^2\right) = c^{n/2} H_n\left(\sqrt{c}y\right)(-1)^n e^{-cy^2}$$

for $n \in \mathbb{N}$ and $y \in \mathbb{R}$ with the Hermite polynomials H_n [9]. The Hermite polynomials satisfy the recursion relation

$$\begin{pmatrix} H_{n+1}(x) \\ H_n(x) \end{pmatrix} = \prod_{j=1}^{n} \mathbf{M_x(j)} \begin{pmatrix} H_1(x) \\ H_0(x) \end{pmatrix}$$

with the matrices

$$\mathbf{M_x(j)} = \begin{pmatrix} 2x & -2j \\ 1 & 0 \end{pmatrix} \quad .$$

The eigenvalues of the matrices $\mathbf{M_x(j)}$ are given by $x \pm \sqrt{x^2 - 2j}$. The absolute values of the eigenvalues are $\sqrt{2j}$, independent of $x \in (-1,1)$. An upper bound for $H_{n+1}(x)$ is given by the product of the absolute values of the larger eigenvalues,

$$H_{n+1}(x) \le |H_1(x)| \prod_{j=1}^{n} \sqrt{2j} = 2 \cdot 2^{n/2} (n!)^{1/2} \le 2 \cdot 2^{n/2} n^{n/2} \,. \qquad (7)$$

Inserting this bound into (6) we get

$$\begin{aligned} \tilde{C}_G^{(2k)} &\le C^k \max_{|\beta|=2k} \sum_{\alpha\le\beta} \binom{\beta}{\alpha} \alpha^{\alpha/2} \le C^k \max_{|\beta|=2k} \prod_{j=1}^{d} \left(\beta_j^{1/2} + 1\right)^{\beta_j} \\ &\le C^k 2^{dk} \max_{|\beta|=2k} \left(|\beta| + d\right)^{|\beta|/2} \le C^k k^k \end{aligned}$$

where C denotes a generic positive constant depending only on d and the scaling parameter c. By Theorem 3 we have the approximation order $\|f - s_{f,X}\|_{L_\infty((-1,1))^d} \le e^{a\cdot log(bh)/h} \|f\|_{\mathcal{N}_G}$ for $f \in \mathcal{N}_G$ with positive constants a and b. Thus, for Gaussian kernels the procedure described above reproduces the well known results [5].

5 Error Bounds with Derivatives

Since both the generating function and its interpolant are smooth in the sense of $\mathcal{C}^\infty$, we now focus on estimates for the approximation error for the derivatives. To this end we proceed along the lines of the previous chapter, but here we use the following local polynomial reproduction from [5, Theorem 11.8].

Theorem 4 *Suppose that* $\Omega = (-1,1)^d,\ \ell \in \mathbb{N}_0$ *and* $\alpha \in \mathbb{N}_0^d$ *with* $|\alpha| \leq \ell$*. Then there exist constants* $c_0, c_1^{(\alpha)}, c_2^{(\alpha)} > 0$*, such that for all* $X = \{x_1, \ldots, x_N\} \subset \Omega$ *with* $h_{X,\Omega} \leq c_0/\ell^2$ *and every* $x \in \Omega$ *there exist numbers* $\tilde{u}_1^{(\alpha)}(x), \ldots, \tilde{u}_N^{(\alpha)}(x)$ *such that*

1. $\sum_{j=1}^N p(x_j)\tilde{u}_j^{(\alpha)}(x) = D^\alpha p(x)$ *for* $x \in \Omega$ *and* $p \in \pi_\ell(\Omega)$ *,*
2. $\sum_{j=1}^N |\tilde{u}_j^{(\alpha)}(x)| \leq c_1^{(\alpha)} h_{X,\Omega}^{-|\alpha|}$ *for all* $x \in \Omega$*,*
3. $\tilde{u}_j^{(\alpha)}(x) = 0$*, if* $\|x - x_j\|_2 > c_2^{(\alpha)} h_{X,\Omega}$ *and* $x \in \Omega$*.*

In this case the constants can be chosen as $c_1^{(\alpha)} \leq 2 \cdot 2^{-|\alpha|} \leq 2$ and $c_2^{(\alpha)} := M \cdot \ell^2/4$ with a positive constant M [5, Proposition 11.7].

With Theorem 4 we have the following algebraic error estimates for the derivatives for any infinitely smooth kernel. We shall again need this technical refinement of [5, Theorem 11.13] to derive explicit exponential orders for interpolation with power series kernels.

Theorem 5 *Let* $\Omega := (-1,1)^d$*,* $\alpha \in \mathbb{N}_0^d$ *be fixed and* $k \geq |\alpha|$*. Then there exist constants* $c_0, C > 0$*, such that for all sets* $X \subset \Omega$ *with* $h := h_{X,\Omega} \leq \min\left\{\frac{c_0}{2k^2}, 1\right\}$ *the error between a function* $f \in \mathcal{N}_K$ *and its interpolant* $s_{f,X}$ *can be bounded by*

$$\|D^\alpha f - D^\alpha s_{f,X}\|_{L_\infty} \leq \frac{1}{\sqrt[4]{2k-|\alpha|}} \left(C_K^{(2k)}\right)^{1/2} \left(\frac{Ck^2}{2k-|\alpha|}\right)^k h^{k-|\alpha|} \|f\|_{\mathcal{N}_K}\ .$$

The constant C depends on the space dimension d, but not on x, f, k and K, and the number $C_K^{(2k)}$ is defined by

$$C_K^{(2k)} := \max_{\substack{\beta,\nu\in\mathbb{N}_0^d \\ |\beta|+|\nu|=2k}} \sup_{x,y\in\Omega} \left| D_1^\beta D_2^\nu K(x,y) \right| ,$$

where D_j^γ denotes the γ-th derivative with respect to the j-th argument, $j = 1, 2$.

Proof: We fix $x \in (-1,1)^d$ and choose the vector $\tilde{u} := \left(\tilde{u}_1^{(\alpha)}(x), \ldots, \tilde{u}_N^{(\alpha)}(x)\right)^T$ with the functions $\tilde{u}_j^{(\alpha)}$ from Theorem 4 with $\ell+1 = 2k$. According to [5, Theorem 11.13] the power function can be bounded by

$$\left[P_{K,X}^{(\alpha)}(x)\right]^2 \leq -\sum_j \tilde{u}_j \left\{ R(x,x_j,\alpha) + S(x_j,x,\alpha) - \sum_i \tilde{u}_i R(x_i,x_j,0) \right\}$$

where

$$R(w,z,\nu) := \sum_{|\beta|=2k-|\nu|} \frac{D_2^\beta D_1^\nu K(w,\xi_{w,z}^\nu)}{\beta!}(z-w)^\beta \quad \text{and}$$
$$S(w,z,\nu) := \sum_{|\beta|=2k-|\nu|} \frac{D_2^{\beta+\nu} K(w,\eta_{w,z}^\nu)}{\beta!}(z-w)^\beta \; .$$

Here, $\xi_{w,z}^\nu$ and $\eta_{w,z}^\nu$ are points on the line segment between z and w.

By the choice of $\tilde{u}$ we have $\sum_j \left|\tilde{u}_j^{(\alpha)}\right| \leq c_1^{(\alpha)} h^{-|\alpha|}$ and $\|x - x_j\|_2 \leq c_2^{(\alpha)} h$. Using again Stirling's formula and the identity $\sum_{|\beta|_d=k} \frac{1}{\beta!} = \frac{d^k}{k!}$ we find with $c := 2deM + 1$

$$\begin{aligned} \left|\sum_j \tilde{u}_j R(x,x_j,\alpha)\right| &\leq \sum_j |\tilde{u}_j|\, C_K^{(2k)} \, \|x-x_j\|_2^{2k-|\alpha|} \, \frac{d^{2k-|\alpha|}}{(2k-|\alpha|)!} \\ &\leq c_1^{(\alpha)} h^{-|\alpha|} C_K^{(2k)} \frac{(deMk^2)^{2k-|\alpha|} \cdot h^{2k-|\alpha|}}{(2k-|\alpha|)^{2k-|\alpha|}\sqrt{2k-|\alpha|}} \\ &\leq \frac{1}{\sqrt{2k-|\alpha|}} C_K^{(2k)} \cdot \left(\frac{ck^2}{2k-|\alpha|}\right)^{2k} \cdot h^{2k-2|\alpha|} \end{aligned}$$

and similarly

$$\left| \sum_j \tilde{u}_j S\left(x_j, x, \alpha\right) \right| \leq \frac{1}{\sqrt{2k - |\alpha|}} C_K^{(2k)} \cdot \left(\frac{ck^2}{2k - |\alpha|} \right)^{2k} \cdot h^{2k-2|\alpha|} .$$

Finally, using the bound for the ℓ_1-norm of the coefficient vector twice we get

$$\begin{aligned} \left| \sum_j \tilde{u}_j \sum_i \tilde{u}_i R(x_i, x_j, 0) \right| &\leq \left(c_1^{(\alpha)} h^{-|\alpha|} \right)^2 C_K^{(2k)} \frac{(de)^{2k}}{\sqrt{k}(2k)^{2k}} \left(2Mk^2 h \right)^{2k} \\ &\leq \frac{1}{\sqrt{2k - |\alpha|}} C_K^{(2k)} \cdot \left(\frac{\tilde{c}k^2}{2k - |\alpha|} \right)^{2k} h^{2k-2|\alpha|} \end{aligned}$$

with $\tilde{c} := 4deM$. Setting $C := \max\{c, \tilde{c}\}$ finishes the proof. □

Theorem 5 shows that interpolation with infinitely smooth kernels leads to arbitrarily high algebraic approximation orders for derivatives of functions from the associated native spaces on any compact set in $(-1, 1)^d$. The philosophy of Theorem 3 applies analogously to derivatives. The following theorem holds for all infinitely smooth kernels and provides a direct way to derive exponential convergence orders for all derivatives.

Theorem 6 *Suppose $\alpha \in \mathbb{N}_0^d$ is an arbitrary but fixed multi-index.*

If there is a constant $F \geq 0$ such that $C_K^{(2k)} \leq e^{Fk} k^k$ holds for all $k \in \mathbb{N}$, then there are constants $A, B, \tilde{h} > 0$, which may depend on d and α, such that the approximation error between any function $f \in \mathcal{N}_K$ and its interpolant $s_{f,X}$ can be bounded by

$$\|D^\alpha f - D^\alpha s_{f,X}\|_{L_\infty} \leq e^{A \log(Bh)/\sqrt{h}} \|f\|_{\mathcal{N}_K}$$

for all data sets X with fill distance $h \leq \tilde{h}$.

If there is a constant $G > 0$ such that $C_K^{(2k)} \leq e^{Gk} k^{2k}$ holds for all $k \in \mathbb{N}$, then *there are constants $a, \tilde{h} > 0$, which may depend on d and α, such that for all data sets X with fill distance $h \leq \tilde{h}$ the approximation error between any function $f \in \mathcal{N}_K$ and its interpolant $s_{f,X}$ can be bounded by*

$$\|D^\alpha f - D^\alpha s_{f,X}\|_{L_\infty} \leq e^{-a/\sqrt{h}} \|f\|_{\mathcal{N}_K} \ .$$

Proof: Following Theorem 5 for $C_K^{(2k)} \leq e^{Fk}k^k$ the α-th power function is for $h \leq \frac{c_0}{2k^2}$ bounded by

$$\left\|P_{K,X}^{(\alpha)}\right\|_{L_\infty} \leq C^k k^{k/2} h^{k-|\alpha|} k^k \leq h^{-|\alpha|} \left(Chk^{3/2}\right)^k$$

for all $k \in \mathbb{N}$ with appropriate constants $C > 0$ which may depend on d and α but not on k or h. We set $\tilde{c} := \min\left\{\frac{c_0}{2}, \frac{1}{C}\right\}$ and choose $k \in \mathbb{N}$ such that $\frac{\tilde{c}}{2k^2} \leq h \leq \frac{\tilde{c}}{k^2}$. Here we assume the fill distance h to be sufficiently small to ensure that $k \geq |\alpha|$. With this choice we find with suitable constants $A, B > 0$

$$\left\|P_{K,X}^{(\alpha)}\right\|_{L_\infty} \leq h^{-|\alpha|} k^{-k/2} \leq e^{A\log(Bh)/\sqrt{h}} \ .$$

To prove the second part of the theorem one can proceed as in the proof of Theorem 3. □

We now have to study the asymptotic behaviour of the number $C_K^{(2k)}$ in the limit $k \to \infty$. By definition we have for power series kernels

$$C_K^{(2k)} = \max_{\substack{\beta,\nu \in \mathbb{N}_0^d \\ |\beta|+|\nu|=2k}} \sum_{\alpha \geq \beta,\nu} \frac{w_\alpha}{(\alpha-\beta)!(\alpha-\nu)!} \ .$$

Note that again $w_\alpha \leq \tilde{w}_\alpha$ for (almost) all $\alpha \in \mathbb{N}_0^d$ implies $C_K^{(2k)} \leq C_{\tilde{K}}^{(2k)}$ for (almost) all $k \in \mathbb{N}$. Here we will consider only the cases $w_\alpha = C\alpha!$ and $w_\alpha = \alpha!^2 c^{|\alpha|}$ with $c < 1$ where the number $C_K^{(2k)}$ can be computed in a simpler form.

Lemma 2 *If $w_\alpha = C\alpha!$ holds for all $\alpha \in \mathbb{N}_0^d$, the number $C_K^{(2k)}$ assumes the 'symmetric' form*

$$C_K^{(2k)} := \max_{\substack{\beta,\nu\in\mathbb{N}_0^d \\ |\beta|+|\nu|=2k}} \sup_{x,y\in\Omega} \left|D_1^\beta D_2^\nu K(x,y)\right| = \max_{\substack{\beta\in\mathbb{N}_0^d \\ |\beta|=k}} \sum_{\gamma\in\mathbb{N}_0^d} \frac{w_{\beta+\gamma}}{\gamma!^2} \ .$$

Proof: By symmetry of the problem we may assume $\beta_j \le \nu_j$ for all $j = 1, \ldots, d$. Without loss of generality we may assume $C = 1$. We use the factorization

$$C_K^{(2k)} = \max_{\substack{\beta,\nu\in\mathbb{N}_0^d \\ |\beta|+|\nu|=2k}} \prod_{j=1}^{d} \sum_{\alpha_j\ge\nu_j} \frac{\alpha_j!}{(\alpha_j-\beta_j)!(\alpha_j-\nu_j)!} \ .$$

Assuming the maximum value is achieved for $\beta \ne \nu,\ |\beta|+|\nu| = 2k$, there are two possibilities for the asymmetry. First, if there exists an index j such that $\nu_j \ge \beta_j+2$ holds, we define $\tilde\beta$ and $\tilde\nu$ by $\tilde\beta_i = \beta_i$ and $\tilde\nu_i = \nu_i$ for $i \ne j$, and $\tilde\nu_j = \nu_j - 1$ and $\tilde\beta_j = \beta_j + 1$. Then we have $\left|\tilde\beta\right| + |\tilde\nu| = 2k$ and since $\alpha_j - \beta_j \ge \alpha_j - \nu_j + 1$ for all $\alpha_j \in \mathbb{N}_0$ we find

$$\sum_{\alpha_j=\nu_j}^{\infty} \frac{\alpha_j!}{(\alpha_j-\nu_j)!(\alpha_j-\beta_j)!} \le \sum_{\alpha_j=\nu_j-1}^{\infty} \frac{\alpha_j!}{(\alpha_j-\tilde\nu_j)!(\alpha_j-\tilde\beta_j)!} \ .$$

Second, if there are indices $j < \ell$ such that $\nu_j = \beta_j + 1$ and $\nu_\ell = \beta_\ell + 1$ we define $\tilde\beta, \tilde\nu$ by $\tilde\nu_i = \nu_i,\ \tilde\beta_i = \beta_i$ for $i \ne j, \ell$, and $\tilde\nu_j = \tilde\beta_j = \beta_j + \beta_\ell + 1$ and $\tilde\beta_\ell = \tilde\nu_\ell = 0$. By induction in β_ℓ we see that

$$\frac{(\alpha_j+\beta_j+1)!}{\alpha_j!\,(\alpha_j+1)!} \cdot \frac{(\alpha_\ell+\beta_\ell+1)!}{\alpha_\ell!\,(\alpha_\ell+1)!} \le \frac{(\alpha_j+\beta_j+\beta_\ell+1)!}{\alpha_j!^2\alpha_\ell!}$$

holds for all $\alpha_j, \alpha_\ell, \beta_j, \beta_\ell \in \mathbb{N}$. This implies that after an index shift respective summands can be bounded by

$$\sum_{\alpha\ge\beta,\nu} \frac{\alpha!}{(\alpha-\beta)!\,(\alpha-\nu)!} \le \sum_{\alpha\ge\tilde\beta,\tilde\nu} \frac{\alpha!}{\left(\alpha-\tilde\beta\right)!\,(\alpha-\tilde\nu)!} \ .$$

Applying the described symmetrization operations to a pair (β, ν) we need only finitely many steps to construct a multi-index $\bar{\beta}$ with

$$\sum_{\alpha \geq \beta, \nu} \frac{w_\alpha}{(\alpha - \beta)!\,(\alpha - \nu)!} \leq \sum_{\alpha \geq \bar{\beta}} \frac{w_\alpha}{(\alpha - \bar{\beta})!^2} \, .$$

□

Using this representation of the number $C_K^{(2k)}$ we can determine its asymptotic behaviour.

Lemma 3 *If* $w_\alpha = \alpha!$ *for all* $\alpha \in \mathbb{N}_0^d$ *there exist positive constants* c_1, c_2*, such that*

$$e^{c_1 \sqrt{k}} k! \leq \; C_K^{(2k)} \leq \; e^{c_2 \sqrt{k}} k! \, .$$

Proof: In this case Lemma 2 gives

$$C_K^{(2k)} = \max_{\substack{\beta \in \mathbb{N}_0^d \\ |\beta| = k}} \prod_{i=1}^{d} \left(\sum_{\gamma_i \in \mathbb{N}_0} \frac{(\beta_i + \gamma_i)!}{\gamma_i!^2} \right) .$$

Therefore we shall compute the asymptotic behaviour of $\sum_{n \in \mathbb{N}_0} \frac{(b+n)!}{n!^2}$ as a function of $b \in \mathbb{N}$. For that we use the confluent hypergeometric functions and the Laguerre polynomials [10] to rewrite the sum as

$$\sum_{n \in \mathbb{N}_0} \frac{(b+n)!}{n!^2} = eb! \sum_{i=0}^{b} \frac{1}{i!} \binom{b}{i} = eb! L_b(-1) \, .$$

The Laguerre polynomials satisfy the recursion relation

$$\begin{pmatrix} L_{b+1}(-1) \\ \\ L_b(-1) \end{pmatrix} = \prod_{n=1}^{b} \mathbf{M}(\mathbf{n}) \begin{pmatrix} L_1(-1) \\ \\ L_0(-1) \end{pmatrix}$$

with the matrices

$$\mathbf{M}(\mathbf{n}) = \begin{pmatrix} 2 - \frac{n}{n+1} \\ \\ 1 \quad 0 \end{pmatrix} .$$

The eigenvalues of the matrices $\mathbf{M}(\mathbf{n})$ are given by $\lambda_\pm(n) = 1 \pm \frac{1}{\sqrt{n+1}}$. An upper bound for $L_{b+1}(-1)$ is given by the product of the larger eigenvalues,

$$L_b(-1) \leq \prod_{n=1}^{b-1} \lambda_+(n)\, L_1(-1) .$$

If we take the logarithm of this product we find

$$\sum_{n=1}^{b-1} \log\left(1 + \frac{1}{\sqrt{n+1}}\right) \leq \sum_{n=1}^{b-1} \frac{1}{\sqrt{n+1}} \leq \int_1^b dx \frac{1}{\sqrt{x}} = 2\sqrt{b} - 2 .$$

Thus we have

$$L_b(-1) \leq \prod_{n=1}^{b-1} \lambda_+(n)\, L_1(-1) \leq e^{2\sqrt{b}}\, L_1(-1) .$$

This shows that there is a constant $c_1 > 0$ such that $C_K^{(2k)} \leq e^{c_1\sqrt{k}} k!$ holds for all $k \in \mathbb{N}$. On the other hand side by Stirling's formula we have for all $b \in \mathbb{N}$

$$\sum_{n=0}^{b} \begin{pmatrix} b \\ \\ n \end{pmatrix} \frac{b!}{n!} \geq \begin{pmatrix} b \\ \\ \left[\sqrt{b}\right] \end{pmatrix} \frac{b!}{\left[\sqrt{b}\right]!} \geq \frac{\left(b - \left[\sqrt{b}\right]\right)^{\left[\sqrt{b}\right]-1}}{\left[\sqrt{b}\right]!^2} \cdot b!$$

$$\geq \frac{1}{2} b! \frac{\left(\left[\sqrt{b}\right] - 1\right)^{\left[\sqrt{b}\right]-1}}{\left(\left[\sqrt{b}\right] - 1\right)!} \geq c \cdot b! e^{c_1\sqrt{b}}$$

with appropriate constants $c, c_1 > 0$. Thus, there is a constant $c_2 > 0$ such that $C_K^{(2k)} \geq e^{c_2\sqrt{k}} k!$ holds for all $k \in \mathbb{N}$. □

We now consider the case $w_\alpha \leq c^{|\alpha|}\alpha!^2$ for all $\alpha \in \mathbb{N}_0^d$ with a positive parameter $c < 1$.

Lemma 4 *If $w_\alpha \le c^{|\alpha|}\alpha!^2$ for all $\alpha \in \mathbb{N}_0^d$ with a positive parameter $c<1$ we have*

$$C_K^{(2k)} := \max_{\substack{\beta,\nu\in\mathbb{N}_0^d\\|\beta|+|\nu|=2k}} \sup_{x,y\in\Omega} \left|D_1^\beta D_2^\nu K(x,y)\right| \le k^d \cdot \max_{\substack{\beta\in\mathbb{N}_0^d\\|\beta|=k}} \sum_{\alpha\in\mathbb{N}_0^d} \frac{w_{\alpha+\beta}}{\alpha!^2}\ .$$

Proof: As in the proof of Lemma 2 we may assume $\nu_j \ge \beta_j$ for all $j=1,\dots,N$. Again, there are two possibilities for an asymmetry. First, if there is an index j such that $\nu_j \ge \beta_j+2$ we define $\tilde\beta$ and $\tilde\nu$ by $\tilde\beta_i=\beta_i$ and $\tilde\nu_i=\nu_i$ for $i\ne j$, and $\tilde\beta_j=\beta_j+1$ and $\tilde\nu_j=\nu_j-1$. Since $\beta_j\le\nu_j-2$ implies

$$\frac{\alpha_j!^2c^{\alpha_j}}{(\alpha_j-\nu_j)!\,(\alpha_j-\beta_j)!} \le \frac{\alpha_j!^2c^{\alpha_j}}{(\alpha_j-\nu_j+1)!\,(\alpha_j-\beta_j-1)!}$$

for all $\alpha_j\ge\nu_j$, we find

$$\sum_{\alpha_j\ge\nu_j} \frac{\alpha_j!^2c^{\alpha_j}}{(\alpha_j-\nu_j)!\,(\alpha_j-\beta_j)!} \le \sum_{\alpha_j\ge\tilde\nu_j} \frac{\alpha_j!^2c^{\alpha_j}}{(\alpha_j-\tilde\nu_j)!\left(\alpha_j-\tilde\beta_j\right)!}\ .$$

Second, if there are indices $h<j$ such that $\nu_j=\beta_j+1$ and $\nu_h=\beta_h+1$ we use a different representation of $C_K^{(2k)}$, namely for $\Omega=(-1,1)^d$ we have

$$\begin{aligned}
C_K^{(2k)} &= \max_{\substack{\beta,\nu\in\mathbb{N}_0^d\\|\beta|+|\nu|=2k}} \sup_{x,y\in\Omega} \prod_{i=1}^d \left| \frac{\partial^{\beta_i}}{\partial x_i^{\beta_i}} \frac{\partial^{\nu_i}}{\partial y_i^{\nu_i}} \frac{1}{1-cx_iy_i}\right| \\
&= \max_{\substack{\beta,\nu\in\mathbb{N}_0^d\\|\beta|+|\nu|=2k}} \sup_{x,y\in\Omega} \prod_{i=1}^d \left| \frac{\partial^{\beta_i}}{\partial x_i^{\beta_i}} \frac{c^{\nu_i}\nu_i!x_i^{\nu_i}}{(1-cx_iy_i)^{\nu_i+1}}\right| \\
&= \max_{\substack{\beta,\nu\in\mathbb{N}_0^d\\|\beta|+|\nu|=2k}} \sup_{x,y\in\Omega} \prod_{i=1}^d \left| \sum_{\ell_i=0}^{\beta_i} \binom{\beta_i}{\ell_i}^2 \frac{\ell_i!\nu_i!^2}{(\nu_i-\beta_i+\ell_i)!} \frac{x_i^{\nu_i-\beta_i+\ell_i}c^{\nu_i+\ell_i}y_i^{\ell_i}}{(1-cx_iy_i)^{\nu_i+\beta_i+1}}\right| \\
&= \max_{\substack{\beta,\nu\in\mathbb{N}_0^d\\|\beta|+|\nu|=2k}} \prod_{i=1}^d \left| \frac{1}{(1-c)^{\nu_i+\beta_i+1}} \sum_{\ell_i=0}^{\beta_i} \binom{\beta_i}{\ell_i}^2 \frac{\nu_i!^2\,(\beta_i-\ell_i)!}{(\nu_i-\ell_i)!} c^{\nu_i+\beta_i-\ell_i}\right| \\
&= \frac{c^{2k}}{(1-c)^{2k+d}} \max_{\substack{\beta,\nu\in\mathbb{N}_0^d\\|\beta|+|\nu|=2k}} \prod_{i=1}^d \left| \sum_{\ell_i=0}^{\beta_i} \binom{\beta_i}{\ell_i}^2 \frac{\nu_i!^2\,(\beta_i-\ell_i)!}{(\nu_i-\ell_i)!} c^{-\ell_i}\right|\ .
\end{aligned} \tag{8}$$

We define the function Q for b, $n \in \mathbb{N}_0$ with $b \leq n$ by

$$Q(n,b) := \sum_{\ell=0}^{b} \binom{b}{\ell}^2 \frac{n!^2(b-\ell)!}{(n-\ell)!} c^{-\ell} \ .$$

Then for all $b \in \mathbb{N}_0$ we find

$$\begin{aligned} Q(b+1,b) &= \sum_{\ell=0}^{b} \binom{b}{\ell}^2 \frac{(b+1)!^2(b-\ell)!}{(b-\ell+1)!} c^{-\ell} = \sum_{\ell=0}^{b} \binom{b}{\ell}^2 \frac{(b+1)^2 b!^2}{(b-\ell+1)} c^{-\ell} \\ &\leq (b+1)^2 \sum_{\ell=0}^{b} \binom{b}{\ell}^2 b!^2 c^{-\ell} = (b+1)^2 Q(b,b) \ . \end{aligned} \tag{9}$$

Furthermore, since $\frac{1}{(b-\ell)!} \leq \frac{(b+1)^2}{(b+1-\ell)!}$ holds for all $\ell = 0, \ldots, b$, we have

$$\begin{aligned} Q(b+1,b) &= \sum_{\ell=0}^{b} \frac{b!^2(b+1)!^2}{\ell!^2(b-\ell+1)!(b-\ell)!} c^{-\ell} \\ &\leq \sum_{\ell=0}^{b} \frac{b!^2(b+1)!^2}{\ell!^2(b-\ell+1)!} \frac{(b+1)^2}{(b+1-\ell)!} c^{-\ell} \\ &= \sum_{\ell=0}^{b} \binom{b+1}{\ell}^2 (b+1)!^2 c^{-\ell} \leq Q(b+1,b+1) \ . \end{aligned} \tag{10}$$

If we now set $\tilde{\beta}_j = \tilde{\nu}_j = \beta_j$ and $\tilde{\beta}_h = \tilde{\nu_h} = \beta_h + 1$ we find using the estimates (9) and (10)

$$Q(\beta_j, \nu_j) \cdot Q(\beta_h, \nu_h) \leq (\beta_j + 1)^2 \cdot Q\left(\tilde{\beta}_j, \tilde{\nu}_j\right) \cdot Q\left(\tilde{\beta}_h, \tilde{\nu}_h\right) \ .$$

Since $\beta_j \leq 2k - \beta_j - 1$ we have $\beta_j + 1 \leq k$. To ‘symmetrize‘ a pair (β, ν) by the operations described above we have to apply the second step at most $\left[\frac{d}{2}\right]$ times. Thus, we constructively find a multi-index $\bar{\beta}$ such that

$$\sum_{\alpha \geq \beta, \nu} \frac{\alpha!^2 c^{|\alpha|}}{(\alpha-\beta)!\,(\alpha-\nu)!} \leq k^d \sum_{\alpha \geq \bar{\beta}} \frac{\alpha!^2 c^{|\alpha|}}{\left(\alpha-\bar{\beta}\right)!^2} \ .$$

□

Now we can determine the asymptotic behaviour of the number $C_K^{(2k)}$.

Lemma 5 *If* $w_\alpha \leq \alpha!^2 c^{|\alpha|}$ *holds for all* $\alpha \in \mathbb{N}_0^d$ *with a constant* $c < 1$*, then there exists a positive constant* $F > 0$*, such that*

$$C_K^{(2k)} \leq \ e^{Fk} k!^2$$

holds for all $k \in \mathbb{N}$*.*

Proof: Using Lemma 4 and the identity (8) we find

$$C_K^{(2k)} \leq \frac{k^d c^k k!^2}{(1-c)^{k+d}} \max_{|\beta|=k} \prod_{j=1}^{d} P_{\beta_j} \left(\frac{1+c}{1-c}\right)$$

with the Legendre polynomials P_n [9]. The Legendre polynomials satisfy the recursion relation

$$\begin{pmatrix} P_{n+1}(x) \\ \\ P_n(x) \end{pmatrix} = \prod_{j=1}^{n} \mathbf{M}(\mathbf{x}, \mathbf{j}) \begin{pmatrix} P_1(x) \\ \\ P_0(x) \end{pmatrix} ,$$

with the matrices

$$\mathbf{M}(\mathbf{x}, \mathbf{j}) = \begin{pmatrix} \frac{(2j+1)x}{j+1} & -\frac{j}{j+1} \\ \\ 1 & 0 \end{pmatrix} .$$

The larger eigenvalue of $\mathbf{M}(\mathbf{x}, \mathbf{j})$ is

$$\frac{x + 2jx + \sqrt{x^2 + 4x^2 j + 4x^2 j^2 - 4j - 4j^2}}{2(1+j)}$$
$$\leq \frac{2|x|(1+j)}{2(1+j)} + \frac{\sqrt{(4x^2+4)(j+1)^2}}{2(1+j)} \leq 2|x| + 1 .$$

Therefore, we have $P_n\left(\frac{1+c}{1-c}\right) \leq \left(2\frac{c+1}{1-c} + 1\right)^n$ for all $n \in \mathbb{N}$ which finishes the proof. □

Considering Theorem 6 and Lemmata 3 respectively 5 together we find the following approximation orders.

Corollary 2 *Suppose $\alpha \in \mathbb{N}_0^d$ is arbitrary but fixed.*

If $w_\beta \le C\beta!$ holds for all $\beta \in \mathbb{N}_0^d$ with a constant $C > 0$ then there are constants $A, B > 0$ such that for all data sets X with sufficiently small fill distance h the approximation error between any function $f \in \mathcal{N}_K$ and its interpolant $s_{f,X}$ is bounded by

$$\|D^\alpha f - D^\alpha s_{f,X}\|_{L_\infty(-1,1)^d} \le e^{A \log(Bh)/\sqrt{h}} \|f\|_{\mathcal{N}_K} \ .$$

If $w_\beta \le \beta!^2 c^{|\beta|}$ holds for all $\beta \in \mathbb{N}_0^d$ with a constant $c < 1$ then there is a constant $a > 0$ such that for all data sets X with sufficiently small fill distance h the approximation error between any function $f \in \mathcal{N}_K$ and its interpolant $s_{f,X}$ is bounded by

$$\|D^\alpha f - D^\alpha s_{f,X}\|_{L_\infty(-1,1)^d} \le e^{-a/\sqrt{h}} \|f\|_{\mathcal{N}_K} \ .$$

For the Gaussian kernel with $c = 1$ we can proceed along the lines of the previous chapter. Here we find using Leibniz' rule for higher order derivatives with the bounds on the Hermite polynomials (7) and Lemma 3

$$\begin{aligned}
C_G^{(2k)} &= \max_{|\beta|+|\gamma|=2k} \sup_{x,y} \left| D_x^\beta D_y^\gamma \left[\exp\left(-\|x\|_2^2\right) K(x,y) \exp\left(-\|y\|_2^2\right)\right]\right| \\
&\le C_K^{(2k)} \max_{|\beta|+|\gamma|=2k} \sum_{\alpha \le \gamma} \binom{\gamma}{\alpha} C^{|\alpha|} \alpha^{\alpha/2} \sum_{\delta \le \beta} \binom{\beta}{\delta} C^{|\beta|} \beta^{\delta/2} \\
&\le C^k C_K^{(2k)} \max_{|\beta|+|\gamma|=2k} \prod_{j=1}^d \left(\gamma_j^{1/2}+1\right)^{\gamma_j} \left(\beta_j^{1/2}+1\right)^{\beta_j} \\
&\le C^k C_K^{(2k)} \max_{|\beta|+|\gamma|=2k} \left((|\beta|+d)^{|\beta|/2} (|\gamma|+d)^{|\gamma|/2} \right) \le C^k k^k C_K^{(2k)} \le C^k k^{2k} \ ,
\end{aligned}$$

where C always denotes a positive constant. That means, for the Gaussian kernel for all $\alpha \in \mathbb{N}_0^d$ there is a positive constant c, which may depend on α, such that

$$\|D^\alpha f - D^\alpha s_{f,X}\|_{L_\infty\left((-1,1)^d\right)} \le e^{-c/\sqrt{h}} \|f\|_{\mathcal{N}_K}$$

holds for all $f \in \mathcal{N}_G$ and all data sets X with sufficiently small fill distance h.

6 Multivariate Polynomial Approximation

In this section we shall construct and analyze a polynomial approximant for a function from the native space of a power series kernel by truncating the interpolant studied in the previous sections. For that we define for $k \in \mathbb{N}$ the truncation operator $T^{(k)} : \mathcal{N}_K \to \mathcal{N}_K$ by

$$T^{(k)} \left(\sum_{\alpha \in \mathbb{N}_0^d} a_\alpha \, (\cdot)^\alpha \right) := \sum_{|\alpha| < k} a_\alpha \, (\cdot)^\alpha \ .$$

Lemma 6 *For all $f \in \mathcal{N}_K$ and all $k \in \mathbb{N}$ we have*

$$\left\| f - T^{(k)} f \right\|_{L_\infty([-1,1]^d)} \leq \left(A_K^{(k)} \right)^{1/2} \|f\|_{\mathcal{N}_K} \ ,$$

where the number $A_K^{(k)}$ is defined by $A_K^{(k)} := \sum_{|\alpha| \geq k} \frac{w_\alpha}{\alpha!^2}$.

Proof: For an arbitrary function $f(x) = \sum_\alpha a_\alpha x^\alpha$ in the native space of K, Hoelder's inequality shows

$$\left\| f - T^{(k)} f \right\|_{L_\infty} \leq \sum_{|\alpha| \geq k} |a_\alpha| \leq \left(\sum_{|\alpha| \geq k} a_\alpha^2 \frac{\alpha!^2}{w_\alpha} \right)^{1/2} \cdot \left(\sum_{|\alpha| \geq k} \frac{w_\alpha}{\alpha!^2} \right)^{1/2} \ .$$

□

Now we have to determine the asymptotic behaviour of the number $A_K^{(k)}$.

Lemma 7 *If $w_\alpha \leq \alpha!^2 c^{|\alpha|}$ for all $\alpha \in \mathbb{N}_0^d$ with a constant $c < 1$, then for almost all $k \in \mathbb{N}$ the number $A_K^{(k)}$ can be bounded by $A_K^{(k)} \leq e^{-Qk}$ with a constant $Q > 0$ that may depend on c but not on k.*

Proof: In this case we have $A_K^{(k)} = \sum_{|\alpha|\geq k} c^{|\alpha|}$. Thus, it suffices to show that $\sum_{|\alpha|\geq k} c^{|\alpha|} \leq c^k \cdot P_{d-1}(k)$ where P_{d-1} is a polynomial of degree less than d. This can be done by induction on d. For $d=1$ we have $\sum_{|\alpha|\geq k} c^{|\alpha|} = c^k/(1-c)$. For $d>1$ the induction hypothesis gives

$$\sum_{\substack{\alpha\in\mathbb{N}_0^d\\ |\alpha|_d\geq k}} c^{|\alpha|} = \sum_{\alpha_1=0}^{k-1} c^{\alpha_1} \sum_{|\beta|_{d-1}\geq k-\alpha_1} c^{|\beta|} + \sum_{\alpha_1=k}^{\infty} c^{\alpha_1} \sum_{|\beta|_{d-1}=0}^{\infty} c^{|\beta|}$$

$$\leq c^k \sum_{\alpha_1=0}^{k-1} P_{d-2}\left(k-\alpha_1\right) + \frac{c^k}{(1-c)^d} \ . \tag{11}$$

Since P_{d-2} is a polynomial of degree less than $d-1$ we can expand it in

$$P_{d-2}\left(k-\alpha_1\right) = \sum_{j=0}^{d-2} \mathfrak{b}_j \left(k-\alpha_1\right)^j = \sum_{j=0}^{d-2}\sum_{\ell=0}^{j} \mathfrak{b}_{j,\ell} k^{j-\ell}\alpha_1^{\ell}$$

with some real coefficients $\mathfrak{b}_{j,\ell}$. Using this representation, the first term of (11) can be rewritten as

$$c^k \sum_{\alpha_1=0}^{k-1} P_{d-2}\left(k-\alpha_1\right) = c^k \sum_{j=0}^{d-2}\sum_{\ell=0}^{j} \mathfrak{b}_{j,\ell} k^{j-\ell} \sum_{\alpha_1=0}^{k-1} \alpha_1^{\ell} \ .$$

Since $\sum_{\alpha_1=0}^{k-1} \alpha_1^{\ell} = \tilde{P}_{\ell+1}(k-1) = P_{\ell+1}(k)$ is a polynomial of degree $\ell+1$ there is a polynomial P_{d-1} such that

$$c^k \sum_{\alpha_1=0}^{k-1} P_{d-2}\left(k-\alpha_1\right) = c^k \sum_{j=0}^{d-2}\sum_{\ell=0}^{j} \mathfrak{b}_{j,\ell} P_{j+1}(k) = c^k P_{d-1}(k) \ .$$

□

Suppose a discrete set $X = \{x_1,\ldots,x_N\}$ and data $(f_1,\ldots,f_N)^T$ generated by an unknown function $f\in\mathcal{N}_K$. The Taylor polynomials of the analytic interpolant studied in the previous sections are polynomial approximants to the unknown generating function. By the triangle inequality we have

$$\left\|f - T^{(k)}\left(s_{f,X}\right)\right\|_{L_\infty} \leq \|f-s_{f,X}\|_{L_\infty} + \left\|s_{f,X} - T^{(k)}\left(s_{f,X}\right)\right\|_{L_\infty} \ .$$

Since $\|s_{f,X}\|_{\mathcal{N}_K} \leq \|f\|_{\mathcal{N}_K}$ [5] we can use Theorem 3 and Lemma 7 to bound the approximation error by

$$\left\| f - T^{(k)}\left(s_{f,X}\right) \right\|_{L_\infty([-1,1]^d)} \leq C\left(e^{-A/h} + e^{-Bk}\right) \|f\|_{\mathcal{N}_K} \;, \qquad (12)$$

where A, B and C denote positive constants. If we relate the limit processes $h \to 0$ and $k \to \infty$ we derive the following result.

Theorem 7 *Suppose weights $w_\alpha \leq \alpha!^2 c^{|\alpha|}$ with a constant $0 < c < 1$, $f \in \mathcal{N}_K$, $k \in \mathbb{N}$ and $X = \{x_1, \ldots, x_N\} \subset (-1,1)^d$, such that $h := h_X \leq 1/k$.*
Then there exist constants $L > 0$ and $0 < q < 1$ which may depend on f but not on k or h, such that for all $k \in \mathbb{N}$ the approximation error between f its polynomial approximant $T^{(k)} s_{f,X}$ can be bounded by

$$\left\| f - T^{(k)}\left(s_{f,X}\right) \right\|_{L_\infty([-1,1]^d)} < Lq^k \;.$$

Corollary 3 *We assume an analytic function $f(x) = \sum_{\alpha \in \mathbb{N}_0^d} a_\alpha x^\alpha$ with convergence radius $r > 1$, i.e., there exists $\epsilon > 0$ such that f converges absolutely and uniformly on $(-1-\epsilon, 1+\epsilon)^d$. Then there are constants $L > 0$ and $0 < q < 1$, such that there is a sequence of polynomials $(p_n)_n$ with $p_n \in \pi_n\left(\mathbb{R}^d\right)$, such that*

$$\|f - p_n\|_{L_\infty([-1,1]^d)} < Lq^n$$

holds for all $n \in \mathbb{N}$. The polynomials can be constructed explicitly from function values of f on a sufficiently dense discrete set.

Proof: To apply the construction described above it remains to show that there is a constant $C > 1$ such that $\sum_{\alpha \in \mathbb{N}_0^d} C^{|\alpha|} a_\alpha^2 < \infty$ holds, i.e., $f \in \mathcal{N}_K$ where

K has the weights $w_\alpha = \alpha!^2 C^{-|\alpha|}$ for $\alpha \in \mathbb{N}_0^d$. By the assumption that f has convergence radius $r \geq 1+\epsilon$ we have

$$\overline{\lim_{|\alpha|\to\infty}} \sqrt[|\alpha|]{|a_\alpha|} \leq \frac{1}{1+\epsilon} \, .$$

Since the sequence $\left(\sqrt[|\alpha|]{|a_\alpha|}\right)$ is bounded and $\sqrt[|\alpha|]{|a_\alpha|} \geq 0$ holds for all $\alpha \in \mathbb{N}_0^d$ we get

$$\overline{\lim_{|\alpha|\to\infty}} \sqrt[|\alpha|]{a_\alpha^2} \leq \left(\overline{\lim_{|\alpha|\to\infty}} \sqrt[|\alpha|]{|a_\alpha|}\right)^2 \leq \frac{1}{(1+\epsilon)^2} \, .$$

Therefore $\sum_{\alpha\in\mathbb{N}_0^d} a_\alpha^2 x^\alpha$ converges uniformly on $\left(-(1+\epsilon)^2, (1+\epsilon)^2\right)^d$ and if we choose $C := \frac{1+(1+\epsilon)^2}{2} > 1$ we have $\sum_{\alpha\in\mathbb{N}_0^d} c^{|\alpha|} a_\alpha^2 < \infty$. $\square$

This corollary states that an analytic function f which converges absolutely and uniformly on an open set Ω with $[-1,1]^d \subset \Omega$ can be uniformly approximated on the cube $[-1,1]^d$ by explicitly constructable polynomials of degree $\leq n$ at a geometric rate. The rate we derived here is exactly the one that a classical theorem by S. N. Bernstein [11, ch. IX] and its multivariate generalization [12] relate to the analyticity of functions on the interval $[-1,1]$ and the cube $[-1,1]^d$, respectively.

7 Application to Support Vector Regression

In this section we shall sketch how the polynomial approximation constructed in the previous section can be used to derive worst-case error estimates for several popular support vector machines (SVM) [7]. The typical problem considered in SV machines is not direct interpolation but e.g. regularized least squares (equivalent to ridge regression in statistics [13]). Suppose the values of an unknown function

$f \in \mathcal{N}_K$ are given on scattered locations $X = \{x_1, \ldots, x_N\} \subset (-1,1)^d =: \Omega$. The reconstruction Rf is the solution of a minimization problem

$$\min_{g \in \mathcal{N}_K} \sum_{j=1}^{N} V\left(g\left(x_j\right), f\left(x_j\right)\right) + \lambda \left\|g\right\|_{\mathcal{N}_K}^2 ,$$

where V denotes a nonnegative function with $V(t,t) = 0$ for all $t \in \mathbb{R}$, e.g. a loss function [14], and $\lambda > 0$ is a regularization parameter. Some results for approximation rates of SVM solutions by a kernel expansion with a limited number of terms can be found in [15], but we follow a different approach here. Following [16] good polynomial approximations lead to *sampling inequalities*. For any $k \in \mathbb{N}$ with $h_{X,\Omega} =: h \leq \frac{c_0}{2k^2}$ we can use the polynomial reproduction from Theorem 4 with $\tilde{u}_j := \tilde{u}_j^{(0)}$ to derive the inequality

$$\begin{aligned} |f(x)| &\leq |f(x) - p(x)| + |p(x)| \leq \|f-p\|_{L_\infty(\Omega)} + \sum_{j=1}^{k} \left|p\left(x_j\right)\right| \cdot |\tilde{u}_j(x)| \\ &\leq \|f-p\|_{L_\infty(\Omega)} + \|p|_X\|_{\ell_\infty(X)} \cdot \sum_{j=1}^{k} |\tilde{u}_j(x)| \\ &\leq \|f-p\|_{L_\infty(\Omega)} + 2 \cdot \left(\|f-p\|_{L_\infty(\Omega)} + \|f|_X\|_{\ell_\infty(X)} \right) \\ &\leq 3 \cdot \|f-p\|_{L_\infty(\Omega)} + 2 \cdot \|f|_X\|_{\ell_\infty(X)} \qquad (13) \end{aligned}$$

for any polynomial $p \in \pi_k\left(\mathbb{R}^d\right)$ and any point $x \in \Omega$. If we now choose $k \in \mathbb{N}$ such that $\frac{c_0}{4k^2} \leq h \leq \frac{c_0}{2k^2}$, equation (12) yields for all functions $f \in \mathcal{N}_K$ and the polynomial $T^{(k)} s_{f,X} \in \pi_k\left(\mathbb{R}^d\right)$

$$\left\| f - T^{(k)} s_{f,X} \right\|_{L_\infty(\Omega)} \leq C e^{-\frac{c}{\sqrt{h}}} \|f\|_{\mathcal{N}_K}$$

with positive constants C, c. If we insert this bound into estimate (13), we end up with

$$\|f\|_{L_\infty(\Omega)} \leq C e^{-\frac{c}{\sqrt{h}}} \|f\|_{\mathcal{N}_K} + 2 \cdot \|f|_X\|_{\ell_\infty(X)} . \qquad (14)$$

In order to derive error estimates for learning machines we shall apply this sampling inequality to a residual function $f - Rf$. For many popular support vector regression (SVR) algorithms, such as the ν- or ϵ-SVR [17], we know that a solution Rf exists and that it has the stability property [18]

$$\|f - Rf\|_{\mathcal{N}_K} \leq \|f\|_{\mathcal{N}_K} \ ,$$

which gives

$$\|f - Rf\|_{L_\infty(\Omega)} \leq Ce^{-\frac{c}{\sqrt{h}}} \|f\|_{\mathcal{N}_K} + 2 \cdot \|(f - Rf)\,|_X\|_{\ell_\infty(X)} \qquad (15)$$

with estimate (14). Usually the reconstruction is quite accurate on the given data, which implies that also the second term of (15) is small. Explicit bounds for the ϵ- and ν-SVR can be found in [18]. Therefore, the estimate (15) supports the good behaviour of power series kernels in those learning machines.

8 Future Work

Even in the special case of Gaussians, the optimal convergence rate of interpolants is not known. Inverse theorems are only available for non-analytic kernels [19], so far. Furthermore, the two polynomial reproduction scenarios used here need refinement and alignment.

For applications to multivariate polynomial interpolation and approximation, the connections to power series kernels should be investigated further, e.g. in order to derive spectral convergence orders for the de Boor/Ron multivariate polynomial interpolation.

Acknowledgements

First and foremost, I wish to thank Professor Robert Schaback for his continued support and encouragement. This work greatly benefitted from discussions with Professor Holger Wendland and Christian Rieger for which I would like to thank them. I am grateful to the German National Academic Foundation (Studienstiftung des deutschen Volkes) for their support.

References

1. Madych, W., Nelson, S.: Bounds on multivariate polynomials and exponential error estimates for multiquadric interpolation. J. Approx. Theory **70** (1992) 94–114
2. de Boor, C., Ron, A.: On multivariate polynomial interpolation. Constructive Approximation **6** (1990) 287–302
3. Schaback, R.: Interpolation by polynomials and radial basis functions. Constructive Approximation **21** (2005) 293–317
4. Driscoll, T., Fornberg, B.: Interpolation in the limit of increasingly flat radial basis functions. Computers and Mathematics with Applications **43** (2002) 413–422
5. Wendland, H.: Scattered Data Approximation. Cambridge Monographs on Applied and Computational Mathematics. Cambridge University Press, Cambridge (2005)
6. Christianini, N., Shawe-Taylor, J.: Kernel methods for pattern recognition. Cambridge University Press, Cambridge (2004)
7. Schölkopf, B., Smola, A.: Learning with kernels - Support Vector Machines, Regularisation, and Beyond. MIT Press, Cambridge, Massachussetts (2002)
8. Lu, F., Sun, H.: Positive definite dot product kernels in learning theory. Advances in Computational Mathematics **22** (2005) 181–198

9. Abramowitz, M.: Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Table. Dover, New York (1970)
10. Erdelyi, A., et al.: Higher Transcendental Functions, vol. II. McGraw-Hill, New York (1953-1955)
11. Natanson, I.: Constructive function theory, I. Frederick Ungar Publishing Co., New York (1965)
12. Baouendi, M., Goulaouic, C.: Approximation polynômiale de fonctions C^∞ et analytique. Ann. Inst. Fourier (Grenoble) **21** (1971) 149–173
13. Hoerl, A., Kennrad, R.: Ridge regression: Biased estimation for nonorthogonal problems. Technometrics **12** (1970) 55–67
14. Vapnik, V.: The nature of statistical learning theory. Springer, New York (1995)
15. Kurkova, V., Sanguineti, M.: Learning with generalization capability by kernel methods of bounded complexity. J. of Complexity **21** (2005) 350–367
16. Wendland, H., Rieger, C.: Approximate interpolation. Numerische Mathematik **101** (2005) 643–662
17. Schölkopf, B., Williamson, R., Bartlett, W.: New support vector algorithms. Neural Computation **12** (2000) 1207–1245
18. Rieger, C., Zwicknagl, B.: Deterministic error analysis of support vector machines and related regularized kernel methods. MPI-MIS preprint 152/2006 (2006)
19. Schaback, R., Wendland, H.: Inverse and saturation theorems for radial basis function interpolation. Mathematics of Computation **71** (2002) 669–681