Approximation in Hilbert spaces of the Gaussian and related analytic kernels

Toni Karvonen[1,2] — Yuya Suzuki[3]

[1]School of Engineering Sciences, Lappeenranta–Lahti University of Technology LUT, Finland
[2]Department of Mathematics and Statistics, University of Helsinki, Finland
[3]Department of Mathematics and Systems Analysis, Aalto University, Finland

May 10, 2025

Abstract

We consider linear approximation based on function evaluations in reproducing kernel Hilbert spaces of certain analytic weighted power series kernels and stationary kernels on the interval $[-1, 1]$. Both classes contain the popular Gaussian kernel $K(x, y) = \exp(-\frac{1}{2}\varepsilon^2(x - y)^2)$. For weighted power series kernels we derive almost matching upper and lower bounds on the worst-case error. When applied to the Gaussian kernel, our results state that, up to a sub-exponential factor, the nth minimal error decays as $(\varepsilon/n)^n (n!)^{-1/2}$. The proofs are based on weighted polynomial interpolation and classical polynomial coefficient estimates that we use to bound the Hilbert space norm of a weighted polynomial fooling function.

1 Introduction

Let $K \colon [-1, 1] \times [-1, 1] \to \mathbb{R}$ be a positive-semidefinite kernel on the interval $[-1, 1]$ and $H(K)$ its reproducing kernel Hilbert space equipped with inner product $\langle \cdot, \cdot \rangle_K$ and norm $\|\cdot\|_K$. This space consists of certain functions defined on $[-1, 1]$. Properties of the kernel determine which functions are contained in $H(K)$. For example, if K is m times differentiable in both arguments, all functions in $H(K)$ are m times differentiable; if K is analytic, $H(K)$ consists of analytic functions [39, Sec. 2.1]. In this paper we study approximation in $H(K)$ when the reproducing kernel is analytic and obtain reasonably tight upper and lower bounds on three types of worst-case errors.

Let $\|\cdot\|_p$ denote the $L^p = L^p([-1, 1])$-norm and let $x_1, \ldots, x_n \in [-1, 1]$ be any points. We assume that we have access to function evaluations at these points, a type of data known as standard information. Firstly, we study the minimal worst-case error

$$e_p^{\min}(x_1, \ldots, x_n) = \inf_{\psi_1, \ldots, \psi_n \in L^p} \sup_{0 \neq f \in H(K)} \frac{\|f - \sum_{k=1}^n f(x_k)\psi_k\|_p}{\|f\|_K} \tag{1.1}$$

of linear approximation given standard information at the points. Secondly, we consider the nth minimal error

$$e_p^{\min}(n) = \inf_{x_1, \ldots, x_n} e_p(x_1, \ldots, x_n) = \inf_{\substack{x_1, \ldots, x_n \\ \psi_1, \ldots, \psi_n \in L^p}} \sup_{0 \neq f \in H(K)} \frac{\|f - \sum_{k=1}^n f(x_k)\psi_k\|_p}{\|f\|_K}, \tag{1.2}$$

which is the smallest approximation error that can be attained with n function evaluations. Finally, we also study the pointwise worst-case error

$$e^{\min}(x \mid x_1, \ldots, x_n) = \inf_{u_1, \ldots, u_n \in \mathbb{R}} \sup_{0 \neq f \in H(K)} \frac{|f(x) - \sum_{k=1}^{n} f(x_k) u_k|}{\|f\|_K}, \qquad (1.3)$$

which only tells us about the error at a single fixed point $x \in [-1, 1]$. The pointwise worst-case error equals the power function in the radial basis function interpolation [40, 48] and the posterior or kriging variance in Gaussian process modelling [35, 41, 43]; see also Section 4.2. The original motivation for this work was to obtain bounds on the pointwise worst-case error that could be used to derive asymptotics for covariance kernel parameter estimators in Gaussian process modelling in a manner similar to [15].

We are in particular interested in the Gaussian kernel $K(x, y) = \exp(-\frac{1}{2}\varepsilon^2(x - y)^2)$, where $\varepsilon > 0$ is a shape parameter, and its Hilbert space. Although strong arguments have been made against its use in applications [43, p. 30], the Gaussian kernel remains popular, especially in machine learning [35, p. 83]. We study two classes of kernels that both contain the Gaussian kernel: *weighted power series kernels* given by

$$K(x, y) = \varphi(x)\varphi(y) \sum_{k=0}^{\infty} \alpha_k^{-1} x^k y^k \qquad (1.4)$$

for a suitable function $\varphi \colon [-1, 1] \to \mathbb{R}$ and positive coefficients α_k, and *stationary kernels* defined by a positive-semidefinite function $\Phi \colon \mathbb{R} \to \mathbb{R}$ via

$$K(x, y) = \Phi(x - y). \qquad (1.5)$$

The Gaussian kernel is obtained by setting $\alpha_k = \varepsilon^{-2k} k!$ and $\varphi(x) = e^{-\frac{1}{2}\varepsilon^2 x^2}$ in (1.4) or $\Phi(z) = \exp(-\frac{1}{2}\varepsilon^2 z^2)$ in (1.5). How fast the coefficients α_k blow up or how fast the Fourier transform of Φ tends to zero at infinity determines how smooth these kernels and their Hilbert spaces are. It would be natural to consider these kernels and approximation in their Hilbert spaces on the entire real line, rather than on a bounded interval as we do. However, the techniques that we use exploit the properties of polynomials on bounded intervals and do not extend easily to other cases.

In Section 3 we prove that the worst-case errors in (1.1)–(1.3) decay as $\alpha_n^{-1/2}$, up to an exponential factor, when K is a weighted power series kernel in (1.4) whose coefficients blow up factorially or faster. In Section 4 we prove analogous lower bounds when K is a stationary kernel in (1.5) defined by Φ whose Fourier transform tends to zero sufficiently fast. Our main results are Theorems 3.1 and 4.1. By applying Theorem 3.1 to the Gaussian kernel we obtain the following result.

Theorem 1.1. *Let $K(x, y) = \exp(-\frac{1}{2}\varepsilon^2(x - y)^2)$ be the Gaussian kernel, $p \in \{2, \infty\}$, and c_L and m_L the functions defined in (3.30) and (3.31). If $x_1, \ldots, x_n \in [-1, 1]$ are distinct, then*

$$c_{1,p} \left(\frac{\varepsilon}{2}\right)^n (n!)^{-1/2} \leq e_p^{\min}(x_1, \ldots, x_n) \leq c_{2,p}\, n^{-1/8 - 1/p} e^{\varepsilon\sqrt{n}} (2\varepsilon)^n (n!)^{-1/2}, \qquad (1.6)$$

$$c_{1,p} \left(\frac{\varepsilon}{2}\right)^n (n!)^{-1/2} \leq e_p^{\min}(n) \leq 2c_{2,p}\, n^{-1/8} e^{\varepsilon\sqrt{n}} \left(\frac{\varepsilon}{2}\right)^n (n!)^{-1/2}, \qquad (1.7)$$

$$c_{1,\infty} \left(\frac{\varepsilon}{4}\right)^n (n!)^{-1/2} \leq \frac{e^{\min}(x \mid x_1, \ldots, x_n)}{|\prod_{k=1}^{n}(x - x_k)|} \leq c_{2,p}\, n^{-1/8} e^{\varepsilon\sqrt{n}} \varepsilon^n (n!)^{-1/2}, \qquad (1.8)$$

where the lower bounds hold when $n \geq 1$ and the upper bounds when $n \geq m_L(\varepsilon^2)$, the constants $c_{1,\infty}$ and $c_{1,2}$ are obtained by setting $\varphi_{\min} = e^{-\frac{1}{2}\varepsilon^2}$ and $\lambda = \varepsilon^2$ in (3.9), and $c_{2,\infty} = \sqrt{c_L(\varepsilon^2)}$ and $c_{2,2} = \sqrt{2c_L(\varepsilon^2)}$.

The constant coefficients in the bounds could be slightly improved but we have opted for simplicity over optimality. The most remarkable feature of these bounds is that (1.7) shows that the rate of decay of the nth minimal error is controlled by the shape parameter ε. That is,

$$\frac{e_{p,\varepsilon}^{\min}(n)}{e_{p,\varepsilon'}^{\min}(n)} \to 0 \ \text{ if } \ \varepsilon < \varepsilon' \quad \text{and} \quad \frac{e_{p,\varepsilon}^{\min}(n)}{e_{p,\varepsilon'}^{\min}(n)} \to \infty \ \text{ if } \ \varepsilon > \varepsilon', \tag{1.9}$$

where the additional subscript indicates dependency on the shape parameter of the Gaussian kernel. To understand how estimators of the parameter ε behave in Gaussian process regression it is crucial to have results such as (1.9). For example, the maximum likelihood estimate of ε given data (i.e., function evaluations) at points $x_1, \ldots, x_n$ is

$$\hat{\varepsilon}_n = \arg\min_{\varepsilon > 0} \big\{ \mathrm{DF}(\varepsilon \mid x_1, \ldots, x_n) + \mathrm{MC}(\varepsilon \mid x_1, \ldots, x_n) \big\},$$

where the former term is a data-fit term that depends on the function evaluations and the latter a model complexity term that depends only on ε and the points. The model complexity (we do not discuss the data-fit here) can be written as

$$\mathrm{MC}(\varepsilon \mid x_1, \ldots, x_n) = 2 \sum_{i=1}^{n} \log e_{\varepsilon}^{\min}(x_i \mid x_1, \ldots, x_{i-1}),$$

where the first summand is the logarithm of the pointwise initial error at x_1 and the ε-dependency is made explicit [15, Sec. 2.3]. Precise control of the pointwise worst-case error with respect to ε is therefore essential in order to figure out how $\hat{\varepsilon}_n$ behaves as $n \to \infty$. To the best of our knowledge, no result resembling (1.9) has been proved for the Gaussian kernel or its relatives that we study. In a paper that has regrettably gone unnoticed outside of Gaussian process literature, Yarotsky [49] has proved a lower bound on the pointwise worst-case error that is slightly better than (1.8). However, his upper bound is looser than ours. We review Yarotsky's result in more detail in Section 4.2.

Upper bounds on the Gaussian and other analytic kernels abound in the scattered data approximation literature. These are usually expressed in terms of the power function, which is nothing but the pointwise worst-case error $e^{\min}(x \mid x_1, \ldots, x_n)$. Typical results for the Gaussian kernel state that, for any $1 \le p \le \infty$ and some constants c and γ,

$$\big\| e^{\min}(\cdot \mid x_1, \ldots x_n) \big\|_p \le c\, e^{-\gamma n \log n} = c\, n^{-\gamma n} \tag{1.10}$$

if the points $x_1, \ldots, x_n$ are sufficiently uniform on $[-1, 1]$; see [48, Thm. 11.22] and [37, Thm. 6.1]. Bounds for various weighted power series kernels with $\varphi \equiv 1$ may be found in [52, 53]. Because (1.10) is a special case of a bound that holds for the Gaussian kernel defined in any dimension and is proved by means of local polynomial reproduction [48, Ch. 3], the constant γ is rather complicated. For equispaced points on $[-1, 1]$ one may compute that $\gamma \le 1/48$ [19, Sec. 4.2], which gives an extremely sub-optimal rate since Stirling's formula shows that all bounds in Theorem 1.1 are, up to exponential factors, of order $n^{-n/2}$. Although apparently absent from the literature with the exception of [19], it is difficult to believe that upper bounds of order $n^{-n/2}$ are not part of the folklore of the field because, as we shall see in Section 3.4, proving them in dimension one via polynomial interpolation is a simple exercise. Connections between polynomial interpolation and interpolation with translates of the Gaussian kernel have been explored in the fascinating papers [32, 33]. Moreover, recent years have seen a host of results on approximation and integration in the Hilbert space of the Gaussian kernel when error is measured in the Gaussian-weighted L^2-norm over $\mathbb{R}^d$ [4, 6, 7, 17, 18, 20, 21, 25, 26, 42]. Results on covering numbers and regression rates for the Gaussian kernel may be found in [5, 24, 45, 47, 50, 51]. Closely related Hermite spaces have been studied in [10, 12, 13].

See also [1] for optimal rates of integration in the Hardy space of analytic functions that is obtained by setting $\varphi \equiv 1$ and $\alpha_k = r^{2k}$ for $r \neq 0$ in (1.4). Chapter VIII in [31] reviews various results on n-widths of spaces of analytic functions.

Our upper bounds are based on weighted polynomial interpolation, bounds on Laguerre polynomials, and the standard error estimate for polynomial interpolation. We use the standard fooling function technique to obtain the lower bounds (see [23] for some commentary on this technique). That is, for a set of n points we construct a function $f_n \in H(K)$ that vanishes at these points, so that the approximations in (1.1)–(1.3) are identically zero and the worst-case errors are bounded from below by $\|f_n\|_p/\|f_n\|_K$ or $|f_n(x)|/\|f_n\|_K$. The difficulty is then in estimating the $H(K)$-norm of f_n from above. If $H(K)$ were a Sobolev space, this would be relatively easy as f_n could be constructed out of bump functions whose Sobolev norms are straightforward to estimate. Because the spaces we consider consist of analytic functions, bump functions are out of question. Instead, we select f_n as a weighted polynomial interpolant, a more or less trivial construction. However, to estimate the $H(K)$-norm of such an f_n is not straightfoward. To do this we use the classical Markov inequality on the maxima of the derivatives and coefficients of a polynomial from 1892 and its L^2-version by Labelle [27]. To the best of our knowledge these results have not been used to bound worst-case errors before. We begin by illustrating this technique in the particularly tractable setting of Korobov spaces.

2 Prelude: Korobov spaces

Suppose that the reproducing kernel on $[-1, 1]$ is given by

$$K(x, y) = \sum_{k \in \mathbb{Z}} \lambda_k^{-1} \frac{e^{\pi \mathrm{i} k x}}{\sqrt{2}} \, \overline{\frac{e^{\pi \mathrm{i} k y}}{\sqrt{2}}} = \sum_{k \in \mathbb{Z}} \lambda_k^{-1} \frac{e^{\pi \mathrm{i} k x}}{\sqrt{2}} \, \frac{e^{-\pi \mathrm{i} k y}}{\sqrt{2}}$$

for positive λ_k. Here $\overline{z}$ stands for the complex conjugate of $z \in \mathbb{C}$. The reproducing kernel Hilbert space $H(K)$ is then [30, Sec. 2.1]

$$H(K) = \left\{ f(x) = \sum_{k \in \mathbb{Z}} c_k \frac{e^{\pi \mathrm{i} k x}}{\sqrt{2}} \: : \: \|f\|_K^2 = \sum_{k \in \mathbb{Z}} \lambda_k |c_k|^2 < \infty \right\}. \tag{2.1}$$

Because $\langle 2^{-1/2} e^{\pi \mathrm{i} \ell x}, 2^{-1/2} e^{\pi \mathrm{i} k x} \rangle_2 = \frac{1}{2} \int_{-1}^1 e^{\pi \mathrm{i}(\ell-k)x} \, \mathrm{d}x = \delta_{\ell k}$, where $\delta_{\ell k}$ is Kronecker's delta, the function $2^{-1/2} e^{\pi \mathrm{i} k x}$ is the kth eigenfunction and λ_k^{-1} the corresponding eigenvalue of the integral operator $T_K \colon H(K) \to H(K)$ given by

$$(T_K f)(x) = \int_{-1}^1 K(x, y) f(y) \, \mathrm{d}y. \tag{2.2}$$

If λ_k blows up as $|k|^{2s}$, the Hilbert space is a Sobolev space of periodic functions on $[-1, 1]$ with smoothness s, also known as a Korobov space. If λ_k grow exponentially fast, $H(K)$ consists of certain periodic analytic functions. The proof of the following proposition is analogous to the proofs for lower bounds in Sections 3 and 4.

Proposition 2.1. *If n is even and $x_1, \ldots, x_n \in [-1, 1]$ are any points, then*

$$e_2^{\min}(x_1, \ldots, x_n) \geq \left(\sum_{k=-n/2}^{n/2} \lambda_k \right)^{-1/2}. \tag{2.3}$$

4

Proof. Let $a \in (-1, 1)$ be not equal to any of $x_1, \ldots, x_n$. It is clear that the function

$$f_n(x) = \prod_{k=1}^{n} \frac{\sin(\frac{1}{2}\pi(x - x_k))}{\sin(\frac{1}{2}\pi(a - x_k))}$$

vanishes at $x_1, \ldots, x_n$ and equals one at a. This function is a trigonometric polynomial of order $n/2$ and has the form

$$f_n(x) = \sum_{k=-n/2}^{n/2} c_k \frac{e^{\pi i k x}}{\sqrt{2}} \tag{2.4}$$

for some coefficients $c_k \in \mathbb{C}$; see [54, Sec. 1 in Ch. X]. Parseval's identity tells us that

$$\|f_n\|_2 = \left(\sum_{\ell=-n/2}^{n/2} |c_\ell|^2 \right)^{1/2} \geq |c_k| \tag{2.5}$$

for $-n/2 \leq k \leq n/2$. Consequently, the Hilbert space characterisation in (2.1) gives

$$\|f_n\|_K^2 = \sum_{k=-n/2}^{n/2} \lambda_k |c_k|^2 \leq \|f_n\|_2^2 \sum_{k=-n/2}^{n/2} \lambda_k. \tag{2.6}$$

Because f_n vanishes at $x_1, \ldots, x_n$, the function $\sum_{k=1}^{n} f_n(x_k)\psi_k$ is identically zero for any $\psi_k \in L^2$. Therefore (1.1) and (2.6) yield

$$e_2^{\min}(x_1, \ldots, x_n) \geq \frac{\|f_n\|_2}{\|f_n\|_K} \geq \frac{\|f_n\|_2}{\|f_n\|_2 (\sum_{k=-n/2}^{n/2} \lambda_k)^{1/2}} = \left(\sum_{k=-n/2}^{n/2} \lambda_k \right)^{-1/2}. \qquad \square$$

The purpose of Proposition 2.1 is only to illustrate our proof technique. It is well known that square root of the nth largest eigenvalue of the integral operator (2.2) equals the Kolmogorov n-width in L^2, which is the smallest possible worst-case error in L^2 given n pieces of linear information [28, Cor. 4.12]. If $\lambda_k = \lambda_{-k}$ and $\lambda_{|k|}$ is non-decreasing in $|k|$, this implies the lower bound

$$e_2^{\min}(x_1, \ldots, x_n) \geq e_2^{\min}(n) \geq \lambda_{n/2}^{-1/2},$$

which is tighter than (2.3). In Sections 3 and 4 no such trivial lower bound is available because we do not have access to the eigenvalues of T_K and the polynomial basis functions we work with are not L^2-orthonormal. Moreover, polynomial coefficient bounds of the form (2.5) involve additional factors on the right-hand side, which complicates estimation of the sum in (2.3).

It is not surprising that the above technique yields sub-optimal lower bounds: the coefficient estimate (2.5) holds for *all* trigonometric polynomials, not just those that vanish at $x_1, \ldots, x_n$. That is, not all properties of the fooling function are exploited in the proof and it would be unrealistic to expect a tight lower bound. However, what is somewhat surprising is that such a crude technique nevertheless suffices for the almost tight lower bound in (1.6) for the nth minimal error.

3 Weighted power series kernels

Let $\varphi \colon [-1, 1] \to \mathbb{R}$ be a non-vanishing Lebesgue measurable function and $(\alpha_k)_{k=0}^{\infty}$ a positive sequence such that $\sum_{k=0}^{\infty} \alpha_k^{-1} < \infty$. It is easy to see that

$$K(x, y) = \varphi(x)\varphi(y) \sum_{k=0}^{\infty} \alpha_k^{-1} x^k y^k \tag{3.1}$$

5

is a well-defined strictly positive-definite kernel on $[-1, 1] \times [-1, 1]$. When $\varphi \equiv 1$, these kernels are called power series kernels [52, 53]. Accordingly, we call any kernel of the form (3.1) a weighted power series kernel. From (3.1) it follows that [30, Sec. 2.1] the Hilbert space in which K is reproducing is

$$H(K) = \left\{ f(x) = \varphi(x) \sum_{k=0}^{\infty} c_k x^k \; : \; \|f\|_K^2 = \sum_{k=0}^{\infty} \alpha_k c_k^2 < \infty \right\}. \tag{3.2}$$

That is, the rate of growth of α_k determines how large $H(K)$ is. We shall make two assumptions. Firstly, we assume that there are positive constants $\varphi_{\min}$ and $\varphi_{\max}$ such that

$$0 < \varphi_{\min} \leq |\varphi(x)| \leq \varphi_{\max} < \infty \tag{3.3}$$

for all $x \in [-1, 1]$. This is a very mild assumption. Our second assumption is that there exists a positive constant λ such that

$$\frac{\alpha_k}{\alpha_n} \leq \lambda^{n-k} \frac{k!}{n!} \tag{3.4}$$

for all $n \geq 0$ and $0 \leq k \leq n$. Implying that $\alpha_k \geq \alpha_0 \lambda^{-k} k!$ for every $k \geq 0$, this assumption is far from necessary to ensure that $H(K)$ consists of analytic functions. Why we employ (3.4) is related to deficiencies in the technique we use to obtain lower bounds and is discussed in more detail in Remarks 3.3 and 3.7. Note that we have not imposed any smoothness assumptions on the function φ, the qualitative properties of which play no role in our results. While a non-smooth φ yields a Hilbert space of non-smooth functions, these functions are "morally" smooth, being products of a smooth function and a known non-smooth function. It is necessary to include φ if our results are to apply to the Gaussian kernel.

We are aware of four weighted power series kernels that have closed form expressions, satisfy (3.3) and (3.4), and have appeared in the literature. Let $\varepsilon > 0$ be a shape parameter. The Gaussian kernel

$$K(x, y) = \exp\left(-\frac{1}{2} \varepsilon^2 (x - y)^2 \right) \tag{3.5}$$

is obtained by selecting $\alpha_k = \varepsilon^{-2k} k!$ and $\varphi(x) = e^{-\frac{1}{2} \varepsilon^2 x^2}$. The exponential kernel $K(x, y) = e^{\varepsilon x y}$ is obtained by selecting $\alpha_k = \varepsilon^{-k} k!$ and $\varphi \equiv 1$. Let $\tau > 0$. The analytic Hermite kernel (or Mehler kernel)

$$K(x, y) = \exp\left(-\frac{1}{2} \varepsilon^2 (x^2 + y^2) + \tau^2 \varepsilon^2 xy \right)$$

is obtained by selecting $\alpha_k = (\tau \varepsilon)^{-2k} k!$ and $\varphi(x) = e^{-\frac{1}{2} \varepsilon^2 x^2}$. If $\tau = 1$, this kernel reduces to the Gaussian kernel. But taking any $r \in (0, 1)$ and setting $\varepsilon^2 = r^2 / (1 - r^2)$ and $\tau^2 = 1/r$ yields, via Mehler's formula,

$$K(x, y) = \exp\left(-\frac{r^2 (x^2 + y^2) - 2rxy}{2(1 - r^2)} \right) = \sqrt{1 - r^2} \sum_{k=0}^{\infty} \frac{r^k}{k!} H_k(x) H_k(y),$$

where H_k are the probabilist's Hermite polynomials; see [14, Sec. 3.1.2] and [16, Sec. 5.2]. Finally, let $\mathcal{I}_0$ be the modified Bessel function of the first kind of order zero. The Bessel kernel $K(x, y) = \mathcal{I}_0(2\varepsilon \sqrt{xy})$ is obtained by selecting $\alpha_k = \varepsilon^{-k} k!^2$ and $\varphi \equiv 1$ [52, p. 64].

The bulk of this section is taken up by the proof of the following theorem that provides upper and lower bounds on the worst-case errors in (1.1)–(1.3).

Theorem 3.1. *Let K be a weighted power series kernel in (3.1) and $p \in \{2, \infty\}$. Suppose that φ and $(\alpha_k)_{k=0}^{\infty}$ satisfy (3.3) and (3.4). Let c_L and m_L be the functions defined in (3.30) and (3.31). If $x_1, \ldots, x_n \in [-1, 1]$ are distinct, then*

$$c_{1,p} \, 2^{-n} \alpha_n^{-1/2} \le e_p^{\min}(x_1, \ldots, x_n) \le c_{2,p} \, \varphi_{\max} \, n^{-1/8 - 1/p} e^{\sqrt{\lambda n}} 2^n \alpha_n^{-1/2}, \quad (3.6)$$

$$c_{1,p} \, 2^{-n} \alpha_n^{-1/2} \le e_p^{\min}(n) \le 2c_{2,p} \, \varphi_{\max} \, n^{-1/8} e^{\sqrt{\lambda n}} 2^{-n} \alpha_n^{-1/2}, \quad (3.7)$$

$$c_{1,\infty} \, 2^{-2n} \alpha_n^{-1/2} \le \frac{e^{\min}(x \mid x_1, \ldots, x_n)}{|\prod_{k=1}^{n}(x - x_k)|} \le c_{2,\infty} \, \varphi_{\max} \, n^{-1/8} e^{\sqrt{\lambda n}} \alpha_n^{-1/2}, \quad (3.8)$$

where the lower bounds hold when $n \ge 1$ and the upper bounds when $n \ge m_L(\lambda)$,

$$c_{1,\infty} = \sqrt{\frac{2}{1 + \lambda}} \, e^{-\frac{1}{32}\lambda^2} \varphi_{\min} \quad and \quad c_{1,2} = \frac{2\sqrt{2\pi}}{3e^3\sqrt{3(4 + \lambda)}} \, e^{-\frac{1}{32}\lambda^2} \varphi_{\min}, \quad (3.9)$$

and $c_{2,\infty} = \sqrt{c_L(\lambda)}$ and $c_{2,2} = \sqrt{2c_L(\lambda)}$.

The lower bounds in Theorem 3.1 are proved in Sections 3.1 and 3.2 and the upper bounds in Section 3.3. Section 3.4 applies the above theorem to the weighted power series kernels mentioned above. Bounds on worst-case errors given derivative information are derived in Section 3.5. Formally this corresponds to setting $x_1 = \cdots = x_n = a \in (-1, 1)$ in (3.6). Some remarks on extensions to higher dimensions can be found in Section 3.6.

3.1 Lower bounds for approximation in $L^{\infty}([-1, 1])$

In this section we prove lower bounds on $e_{\infty}^{\min}(x_1, \ldots, x_n)$, $e_{\infty}^{\min}(n)$, and $e^{\min}(x \mid x_1, \ldots, x_n)$. The Chebyshev polynomial of the first kind of degree n is

$$T_n(x) = \sum_{k=0}^{n} t_{n,k} x^k,$$

where the non-zero coefficients are

$$t_{n,n-2k} = (-1)^k \frac{2^{n-2k-1}n(n - k - 1)!}{k!(n - 2k)!} \quad \text{for} \quad 0 \le k \le \left\lfloor \frac{n}{2} \right\rfloor. \quad (3.10)$$

Note that the leading coefficient, $t_{n,n}$, equals 2^{n-1}. The roots of T_n are the Chebyshev nodes

$$x_{n,k} = \cos\left(\frac{\pi(k - 1/2)}{n}\right) \in (-1, 1) \quad \text{for} \quad 1 \le k \le n. \quad (3.11)$$

The extrema, which are either 1 or -1, of T_n on $[-1, 1]$ are located at

$$\tilde{x}_{n,k} = \cos(\pi k/n) \quad \text{for} \quad 0 \le k \le n. \quad (3.12)$$

That is, $T_n(x) \in [-1, 1]$ for all $x \in [-1, 1]$. For our purposes the importance of Chebyshev polynomials lies in a classical result by V. A. Markov from 1892 (for a brief history, see [34, p. 679]) and its generalisations [34, Sec. 16.3] that T_n has the maximal coefficients among polynomials of degree n that take values in $[-1, 1]$. Specifically [34, Thm. 16.3.3], if $P_n(x) = \sum_{k=0}^{n} a_k x^k$ is a polynomial of degree n with real coefficients a_k and $|P_n(\tilde{x}_{n,k})| \le 1$ for every $0 \le k \le n$, then

$$|a_{n-2k}| + |a_{n-2k-1}| \le |t_{n,n-2k}| \quad \text{for every} \quad 0 \le k \le \left\lfloor \frac{n - 1}{2} \right\rfloor. \quad (3.13)$$

The lemmas below use this result to estimate the $H(K)$-norm of the weighted polynomial $f_n(x) = \varphi(x)P_n(x)$.

7

Lemma 3.2. *Suppose that assumption* (3.4) *holds. Then for every $n \geq 1$ we have*

$$\tau_1 2^{2n} \alpha_n \leq \sum_{k=0}^{\lfloor n/2 \rfloor} \alpha_{n-2k} t_{n,n-2k}^2 \leq \tau_2 2^{2n} \alpha_n, \tag{3.14}$$

where

$$\tau_1 = \frac{1}{4} \quad and \quad \tau_2 = \frac{1}{2} e^{\frac{1}{16}\lambda^2}. \tag{3.15}$$

Proof. Equation (3.10) gives

$$\sum_{k=0}^{\lfloor n/2 \rfloor} \alpha_{n-2k} t_{n,n-2k}^2 = \alpha_n t_{n,n}^2 \sum_{k=0}^{\lfloor n/2 \rfloor} \frac{\alpha_{n-2k}}{\alpha_n} \left(\frac{t_{n,n-2k}}{t_{n,n}} \right)^2$$

$$= 2^{2n-2} \alpha_n \sum_{k=0}^{\lfloor n/2 \rfloor} \frac{\alpha_{n-2k}}{\alpha_n} \left(\frac{n(n-k-1)!}{(n-2k)!} \right)^2 \frac{1}{2^{4k} k!^2}.$$

Denote the sum on the last line by s_n and use assumption (3.4) to obtain the estimate

$$s_n \leq \sum_{k=0}^{\lfloor n/2 \rfloor} \lambda^{2k} \frac{(n-2k)!}{n!} \left(\frac{n(n-k-1)!}{(n-2k)!} \right)^2 \frac{1}{2^{4k} k!^2} = \sum_{k=0}^{\lfloor n/2 \rfloor} b_{n,k} \left(\frac{\lambda^2}{16} \right)^k \frac{1}{k!^2},$$

where

$$b_{n,k} = \frac{n(n-k-1)!^2}{(n-1)!(n-2k)!}.$$

It is straightforward to compute that $b_{n,0} = 1$ for any $n \geq 1$ and $b_{n,1} = n/(n-1) \leq 2$ for any $n \geq 2$. Furthermore,

$$\frac{b_{n,k}}{b_{n,k+1}} = \frac{(n-k-1)^2}{(n-2k)(n-2k-1)} > 1$$

for any $n \geq 2$ if $k \geq 1$. Therefore

$$s_n \leq 2 \sum_{k=0}^{\lfloor n/2 \rfloor} \left(\frac{\lambda^2}{16} \right)^k \frac{1}{k!^2} \leq 2 e^{\frac{1}{16}\lambda^2},$$

which gives the claimed upper bound. The lower bound follows from $s_n \geq 1$. $\qquad\square$

Remark 3.3. From the proof of Lemma 3.2 we see that the purpose of assumption (3.4), which states that

$$\frac{\alpha_k}{\alpha_n} \leq \lambda^{n-k} \frac{k!}{n!} \tag{3.16}$$

for every $n \geq 0$ and $0 \leq k \leq n$, is to ensure that the sum in (3.14) is dominated by the highest order term $\alpha_n t_{n,n}^2$. Suppose that $\alpha_k = k!^\beta$ for some $\beta > 0$ and note that (3.16) is satisfied if and only if $\beta \geq 1$. Then

$$\frac{\alpha_{n-2} t_{n,n-2}^2}{\alpha_n t_{n,n}^2} = \frac{1}{16} \cdot \frac{n^2}{n^\beta (n-1)^\beta}$$

is bounded from above if and only if $\beta \geq 1$, which shows that for $\beta \in (0,1)$ the sum in (3.14) is *not* dominated by $\alpha_n t_{n,n}^2$.

Lemma 3.4. *The function $f_n(x) = \varphi(x)T_n(x)$ is an element of $H(K)$. If assumption* (3.4) *holds, then*

$$\tau_1 2^{2n}\alpha_n \leq \|f_n\|_K^2 \leq \tau_2 2^{2n}\alpha_n$$

for every $n \geq 1$, where the constants τ_1 and τ_2 are given in (3.15).

Proof. The claim follows from Lemma 3.2 and the Hilbert space characterisation in (3.2) that gives $\|f_n\|_K^2 = \sum_{k=0}^{\lfloor n/2 \rfloor} \alpha_{n-2k}t_{n,n-2k}^2$. $\qquad\square$

Lemma 3.5. *Let $P_n(x) = \sum_{k=0}^n a_k x^k$ for $a_k \in \mathbb{R}$ be a polynomial of degree n. The function $f_n(x) = \varphi(x)P_n(x)$ is an element of $H(K)$. If assumption* (3.4) *holds, then*

$$\|f_n\|_K^2 = \sum_{k=0}^n \alpha_k a_k^2 \leq (1+\lambda)\,\tau_2 2^{2n}\alpha_n \|P_n\|_\infty^2$$

for every $n \geq 1$, where the constant τ_2 is given in (3.15).

Proof. Because the polynomial $Q_n = P_n/\|P_n\|_\infty$ has the coefficients $a_k/\|P_n\|_\infty$, we may assume that $\|P_n\|_\infty \leq 1$. In particular, $|P_n(\tilde{x}_{n,k})| \leq 1$ for every $0 \leq k \leq n$, where $\tilde{x}_{n,k}$ are extremal points of T_n in (3.12). We get from (3.13) that

$$|a_{n-2k}| + |a_{n-2k-1}| \leq |t_{n,n-2k}| \quad \text{for every} \quad 0 \leq k \leq \left\lfloor \frac{n-1}{2} \right\rfloor.$$

Furthermore, it is obvious that $|a_0| \leq \|P_n\|_\infty \leq t_{n,0} = 1$ when n is even. These inequalities and the characterisation (3.2) give

$$\|f_n\|_K^2 = \sum_{k=0}^n \alpha_k a_k^2 = \sum_{k=0}^{\lfloor n/2 \rfloor} \alpha_{n-2k}a_{n-2k}^2 + \sum_{k=0}^{\lfloor (n-1)/2 \rfloor} \alpha_{n-2k-1}a_{n-2k-1}^2$$

$$\leq \sum_{k=0}^{\lfloor n/2 \rfloor} \alpha_{n-2k}t_{n,n-2k}^2 + \sum_{k=0}^{\lfloor (n-1)/2 \rfloor} \alpha_{n-2k-1}t_{n,n-2k}^2.$$

From assumption (3.4) we get

$$\sum_{k=0}^{\lfloor (n-1)/2 \rfloor} \alpha_{n-2k-1}t_{n,n-2k}^2 \leq \lambda \sum_{k=0}^{\lfloor (n-1)/2 \rfloor} \frac{1}{n-2k}\alpha_{n-2k}t_{n,n-2k}^2 \leq \lambda \sum_{k=0}^{\lfloor n/2 \rfloor} \alpha_{n-2k}t_{n,n-2k}^2,$$

so that the claim follows from Lemma 3.2. $\qquad\square$

With these tools at hand we are ready to bound the worst-case errors from below.

Proofs of lower bounds in Theorem 3.1 for $p = \infty$. Let $P_n(x) = (x - x_1) \cdots (x - x_n)$. This is a polynomial of degree n that vanishes at the points $x_1, \ldots, x_n \in [-1,1]$. The function $f_n(x) = \varphi(x)P_n(x)$ is an element of $H(K)$ that vanishes at the same points. Therefore $\sum_{k=1}^n f_n(x_k)\psi_k \equiv 0$ and thus

$$e_\infty^{\min}(x_1, \ldots, x_n) = \inf_{\psi_1, \ldots, \psi_n \in L^\infty} \sup_{0 \neq f \in H(K)} \frac{\|f - \sum_{k=1}^n f(x_k)\psi_k\|_\infty}{\|f\|_K} \geq \frac{\|f_n\|_\infty}{\|f_n\|_K}.$$

The lower bounds in (3.6) and (3.7) for $p = \infty$ now follow from Lemma 3.5 and

$$\|f_n\|_\infty = \sup_{x \in [-1,1]} |\varphi(x)P_n(x)| \geq \varphi_{\min} \|P_n\|_\infty.$$

Observe that $\|P_n\|_\infty \leq 2^n$ since $|x - y| \leq 2$ for any $x, y \in [-1,1]$. Therefore

$$e^{\min}(x \mid x_1, \ldots, x_n) \geq \frac{|f_n(x)|}{\|f_n\|_K} \geq \frac{\varphi_{\min}}{\|f_n\|_K}\left|\prod_{k=1}^n (x - x_n)\right| \geq \frac{\varphi_{\min}\|P_n\|_\infty}{2^n\|f_n\|_K}\left|\prod_{k=1}^n (x - x_n)\right|,$$

so that the lower bound in (3.8) follows from Lemma 3.5. $\qquad\square$

3.2 Lower bounds for approximation in $L^2([-1, 1])$

In this section we prove lower bounds on $e_2^{\min}(x_1, \ldots, x_n)$ and $e_2^{\min}(n)$. The technique is identical to the one that we used for $p = \infty$. Define

$$l_{n,k} = \frac{(2k-1)!!}{k!} \sqrt{k + \tfrac{1}{2}} \binom{\lfloor (n-k)/2 \rfloor + k + \tfrac{1}{2}}{\lfloor (n-k)/2 \rfloor} \tag{3.17}$$

for any $0 \le k \le n$. Here $(2k-1)!! = (2k-1)(2k-3)\cdots 3 \cdot 1$ is the double factorial and

$$\binom{r}{k} = \frac{r(r-1)\cdots(r-k+1)}{k!}$$

the generalised binomial coefficient defined for any $r \in \mathbb{R}$ and $k \in \mathbb{N}_0$. In this section the constants $l_{n,k}$ play an analogous role to that played by $t_{n,n-2k}$ in Section 3.1 because Labelle [27] has proved an L^2-version of the polynomial coefficient bound (3.13) that involves $l_{n,k}$. Namely, if $P_n(x) = \sum_{k=0}^n a_k x^k$ is a polynomial of degree n with real coefficients a_k, then

$$|a_k| \le l_{n,k} \|P_n\|_2 \quad \text{for every} \quad 0 \le k \le n. \tag{3.18}$$

Equality in (3.18) holds only for certain sums of Legendre polynomials, but we shall not use this fact. See also [34, pp. 676–7].

Lemma 3.6. *Suppose that* (3.4) *holds. Then for every* $n \ge 1$ *we have*

$$\ell_1 2^{2n} \alpha_n \le \sum_{k=0}^n \alpha_k l_{n,k}^2 \le \ell_2 2^{2n} \alpha_n,$$

where

$$\ell_1 = \frac{1}{2\pi e^2} \quad \text{and} \quad \ell_2 = \frac{27e^6}{2\pi}\left(1 + \frac{\lambda}{4}\right) e^{\frac{1}{16}\lambda^2}. \tag{3.19}$$

Proof. For $k \ge 1$, the non-asymptotic Stirling's formula,

$$\sqrt{2\pi}\, n^{n+1/2} e^{-n} < n! < \sqrt{2\pi}\, n^{n+1/2} e^{-n+1}, \tag{3.20}$$

and the double factorial identity $(2k-1)!! = (2k)!/(2^k k!)$ yield

$$\left(\frac{(2k-1)!!}{k!}\right)^2 (k + \tfrac{1}{2}) = \frac{(2k)!^2}{2^{2k} k!^4}(k + \tfrac{1}{2}) \le 2^{2k} \frac{e^2}{\pi} \cdot \frac{k + \tfrac{1}{2}}{k} \le \frac{3e^2}{2\pi} \cdot 2^{2k} \tag{3.21}$$

and

$$\left(\frac{(2k-1)!!}{k!}\right)^2 (k + \tfrac{1}{2}) \ge 2^{2k} \frac{1}{\pi e^2} \cdot \frac{k + \tfrac{1}{2}}{k} \ge \frac{1}{2\pi e^2} \cdot 2^{2k}. \tag{3.22}$$

It is trivial to verify that these inequalities hold also for $k = 0$. The definition of $l_{n,k}$ in (3.17) thus gives

$$\frac{1}{2\pi e^2} \cdot 2^{2k} \binom{\lfloor (n-k)/2 \rfloor + k + \tfrac{1}{2}}{\lfloor (n-k)/2 \rfloor}^2 \le l_{n,k}^2 \le \frac{3e^2}{2\pi} \cdot 2^{2k} \binom{\lfloor (n-k)/2 \rfloor + k + \tfrac{1}{2}}{\lfloor (n-k)/2 \rfloor}^2.$$

Note that the generalised binomial coefficient equals one when $\lfloor (n-k)/2 \rfloor = 0$, so that

$$\frac{1}{2\pi e^2} \cdot 2^{2n} \le l_{n,n}^2 = \left(\frac{(2n-1)!!}{n!}\right)^2 (n + \tfrac{1}{2}) \le \frac{3e^2}{2\pi} \cdot 2^{2n}$$

10

by (3.21) and (3.22) and $k + \frac{1}{2}$ We now get

$$\sum_{k=0}^{n} \alpha_k l_{n,k}^2 \geq \alpha_n l_{n,n}^2 \geq \frac{1}{2\pi e^2} \cdot 2^{2n} \alpha_n,$$

which is the claimed lower bound, and

$$\sum_{k=0}^{n} \alpha_k l_{n,k}^2 = \alpha_n l_{n,n}^2 \sum_{k=0}^{n} \frac{\alpha_k}{\alpha_n} \left(\frac{l_{n,k}}{l_{n,n}} \right)^2$$

$$\leq \frac{3e^2}{2\pi} \cdot 2^{2n} \alpha_n \sum_{k=0}^{n} \frac{\alpha_k}{\alpha_n} \cdot \frac{3e^2}{2\pi} \cdot 2\pi e^2 \cdot 2^{-2(n-k)} \left(\binom{\lfloor (n-k)/2 \rfloor + k + \frac{1}{2}}{\lfloor (n-k)/2 \rfloor} \right)^2$$

$$= \frac{9e^6}{2\pi} \cdot 2^{2n} \alpha_n \sum_{k=0}^{n} \frac{\alpha_k}{\alpha_n} \cdot 2^{-2(n-k)} \left(\binom{\lfloor (n-k)/2 \rfloor + k + \frac{1}{2}}{\lfloor (n-k)/2 \rfloor} \right)^2.$$

Assumption (3.4) therefore gives

$$\sum_{k=0}^{n} \alpha_k l_{n,k}^2 \leq \frac{9e^6}{2\pi} \cdot 2^{2n} \alpha_n \sum_{k=0}^{n} \left(\frac{\lambda}{4} \right)^{n-k} \frac{k!}{n!} \left(\binom{\lfloor (n-k)/2 \rfloor + k + \frac{1}{2}}{\lfloor (n-k)/2 \rfloor} \right)^2. \tag{3.23}$$

We now show that the term

$$b_{n,k} = \frac{k!}{n!} \left(\binom{\lfloor (n-k)/2 \rfloor + k + \frac{1}{2}}{\lfloor (n-k)/2 \rfloor} \right)^2$$

is approximately $(n-k)!$. Denote $p_{n,k} = \lfloor (n-k)/2 \rfloor$ and suppose that $p_{n,k} \geq 1$. Then

$$b_{n,k} = \frac{k!}{n!} \cdot \frac{\prod_{\nu=1}^{p(n,k)} (\nu + k + \frac{1}{2})^2}{p(n,k)!^2}$$

$$= \frac{1}{p(n,k)!^2} \cdot \frac{\prod_{\nu=1}^{p(n,k)} (\nu + k + \frac{1}{2})^2}{\prod_{\nu=1}^{n-k} (\nu + k)}.$$

$$= \frac{1}{p(n,k)!^2} \cdot \frac{\prod_{\nu=1}^{p(n,k)} (\nu + k + \frac{1}{2})}{\prod_{\nu=1}^{p(n,k)} (\nu + k)} \cdot \frac{\prod_{\nu=1}^{p(n,k)} (\nu + k + \frac{1}{2})}{\prod_{\nu=p(n,k)+1}^{n-k} (\nu + k)}$$

$$= \frac{1}{p(n,k)!^2} \cdot \frac{\prod_{\nu=2}^{p(n,k)} (\nu + k - \frac{1}{2})}{\prod_{\nu=2}^{p(n,k)} (\nu + k)} \cdot \frac{p(n,k) + k + \frac{1}{2}}{k+1} \cdot \frac{\prod_{\nu=1}^{p(n,k)} (\nu + k + \frac{1}{2})}{\prod_{\nu=p(n,k)+1}^{n-k} (\nu + k)}$$

$$\leq \frac{1}{p(n,k)!^2} \cdot \frac{p(n,k) + k + \frac{1}{2}}{k+1} \cdot \frac{\prod_{\nu=1}^{p(n,k)} (\nu + k + \frac{1}{2})}{\prod_{\nu=p(n,k)+1}^{n-k} (\nu + k)}$$

$$= \frac{1}{p(n,k)!^2} \cdot \frac{p(n,k) + k + \frac{1}{2}}{k+1} \cdot \frac{k + \frac{3}{2}}{n} \cdot \frac{\prod_{\nu=2}^{p(n,k)} (\nu + k + \frac{1}{2})}{\prod_{\nu=p(n,k)+1}^{n-k-1} (\nu + k)}$$

$$\leq \frac{3}{p(n,k)!^2} \cdot \frac{\prod_{\nu=2}^{p(n,k)} (\nu + k + \frac{1}{2})}{\prod_{\nu=p(n,k)+1}^{n-k-1} (\nu + k)},$$

where the last inequality is a consequence of the inequalities $(p(n,k) + k + \frac{1}{2})/n \leq \frac{3}{2}$ and $(k + \frac{3}{2})/(k+1) \leq 2$. Now, in the second term on the last line, each term in the denominator is larger than any term in the numerator and there are at least as many terms in the denominator as there are in the numerator. Therefore,

$$b_{n,k} \leq \frac{3}{p(n,k)!^2} = \frac{3}{\lfloor (n-k)/2 \rfloor!^2} \tag{3.24}$$

11

if $n - k \geq 2$. By computing $b_{n,k}$ in the cases $k = n$ and $k = n - 1$ it is trivial to check that this inequality is valid for any $0 \leq k \leq n$. Using the estimate (3.24) in (3.23) yields

$$\sum_{k=0}^{n} \alpha_k l_{n,k}^2 \leq \frac{9e^6}{2\pi} \cdot 2^{2n} \alpha_n \sum_{k=0}^{n} \left(\frac{\lambda}{4}\right)^{n-k} b_{n,k} \leq \frac{27e^6}{2\pi} \cdot 2^{2n} \alpha_n \sum_{k=0}^{n} \left(\frac{\lambda}{4}\right)^{k} \frac{1}{\lfloor k/2 \rfloor!^2}$$

$$\leq \frac{27e^6}{2\pi} \cdot 2^{2n} \alpha_n \sum_{k=0}^{n} \left(\frac{\lambda}{4}\right)^{2k} \left(1 + \frac{\lambda}{4}\right) \frac{1}{k!^2}$$

$$\leq \frac{27e^6}{2\pi} \left(1 + \frac{\lambda}{4}\right) e^{\frac{1}{16}\lambda^2} \cdot 2^{2n} \alpha_n.$$

This completes the proof. $\qquad \square$

Remark 3.7. Just as the proof of Lemma 3.2 (see Remark 3.3), the proof of Lemma 3.6 is based on the fact that assumption (3.4) ensures that the term $\alpha_n l_{n,n}^2$ dominates the sum in (3.14). Suppose that $\alpha_k = k!^\beta$ for some $\beta > 0$ and note that (3.4) is satisfied if and only if $\beta \geq 1$. Then Stirling's formula yields

$$\frac{\alpha_{n-2} l_{n,n-2}^2}{\alpha_n l_{n,n}^2} = \left(\frac{(n-2)!}{n!}\right)^\beta \left(\frac{(2n-5)!!\, n!}{(2n-1)!!\, (n-2)!}\right)^2 \frac{n - \frac{3}{2}}{n + \frac{1}{2}} \left(\frac{n - \frac{1}{2}}{1}\right)^2$$

$$\sim \left(\frac{1}{n(n-1)}\right)^\beta \frac{1}{16} \cdot \frac{n}{n-2} \cdot \frac{n - \frac{3}{2}}{n + \frac{1}{2}} \cdot (n - \tfrac{1}{2})^2$$

$$\sim \frac{1}{16} \cdot n^{2(1-\beta)}$$

as $n \to \infty$. That is, the ratio is bounded from above if and only if $\beta \geq 1$, which shows that for $\beta \in (0, 1)$ the sum in (3.14) is *not* dominated by $\alpha_n l_{n,n}^2$.

Lemma 3.8. *Let $P_n(x) = \sum_{k=0}^{n} a_k x^k$ for $a_k \in \mathbb{R}$ be a polynomial of degree n. The function $f_n(x) = \varphi(x) P_n(x)$ is an element of $H(K)$. If assumption (3.4) holds, then*

$$\|f_n\|_K^2 = \sum_{k=0}^{n} \alpha_k a_k^2 \leq \ell_2 2^{2n} \alpha_n \|P_n\|_2^2$$

for every $n \geq 1$, where the constant ℓ_2 is given in (3.19).

Proof. The characterisation (3.2) and inequality (3.18) give

$$\|f_n\|_K^2 = \sum_{k=0}^{n} \alpha_k a_k^2 \leq \|P_n\|_2^2 \sum_{k=0}^{n} \alpha_k l_{n,k}^2.$$

The claim then follows from Lemma 3.6. $\qquad \square$

With these tools at hand we are ready to bound the worst-case errors from below.

Proofs of lower bounds in Theorem 3.1 for $p = 2$. The proof is identical to that for the case $p = \infty$, except that we now use Lemma 3.8 to estimate $\|f_n\|_K$. $\qquad \square$

3.3 Upper bounds

To bound the worst-case errors from above we shall use weighted polynomial interpolation. Upper bounds on derivatives of elements of $H(K)$ are needed for this purpose. The

following results closely resemble certain results in scattered data approximation literature, particularly in [52]. Define

$$C_K^{2n} = \sup_{x \in [-1,1]} \frac{\partial^{2n}}{\partial v^n \partial w^n} K(v,w) \bigg|_{\substack{v=x \\ w=x}}. \tag{3.25}$$

Lemma 3.9. *Let K be any positive-semidefinite kernel on $[-1,1] \times [-1,1]$ that is infinitely differentiable in both arguments. Then every $f \in H(K)$ is infinitely differentiable and*

$$\|f^{(n)}\|_\infty \leq \|f\|_K (C_K^{2n})^{1/2}$$

for every $n \geq 0$.

Proof. It is well known that, for a kernel K which is n times differentiable with respect to both of its arguments, every $f \in H(K)$ is n times differentiable and the reproducing property $\langle f, K(x,\cdot) \rangle_K = f(x)$ extends to differentiation [44, Cor. 4.36]. The Cauchy–Schwarz inequality thus gives

$$|f^{(n)}(x)| = \left| \left\langle f, \frac{\partial^n}{\partial x^n} K(x,\cdot) \right\rangle_K \right| \leq \|f\|_K \left\| \frac{\partial^n}{\partial x^n} K(x,\cdot) \right\|_K$$

for every $x \in [-1,1]$. Again by the derivative reproducing property,

$$\left\| \frac{\partial^n}{\partial x^n} K(x,\cdot) \right\|_K^2 = \frac{\partial^{2n}}{\partial v^n \partial w^n} K(v,w) \bigg|_{\substack{v=x \\ w=x}}.$$

The claim follows. $\qquad \square$

The Laguerre polynomial of degree $n \geq 0$ is defined as

$$L_n(x) = \sum_{k=0}^n \binom{n}{k} \frac{(-x)^k}{k!} = \frac{e^x}{n!} \cdot \frac{\mathrm{d}^n}{\mathrm{d}x^n} (e^{-x} x^n). \tag{3.26}$$

We make use of the following effective asymptotics for Laguerre polynomials from [3].

Proposition 3.10. *Let $x > 0$. Define*

$$\tilde{c}_L(x) = \frac{e^{-\frac{1}{2}x + 2\sqrt{x}}}{2\sqrt{\pi} x^{1/4}} \left(3 + \sqrt{\frac{\pi}{e}} + \frac{2}{\sqrt{\pi} x^{1/4}} + \frac{7}{16\sqrt{x}} + \sqrt{x}(2+x) \right)$$

and

$$\tilde{m}_L(x) = \max \left\{ 5x, x \left(1 + \frac{x}{2} \right)^2, \frac{9}{x} \right\}.$$

Then

$$L_n(-x) = \sum_{k=0}^n \binom{n}{k} \frac{x^k}{k!} \leq \tilde{c}_L(x) \, n^{-1/4} e^{2\sqrt{xn}} \tag{3.27}$$

for every $n \geq \tilde{m}_L(x)$.

Proof. Setting $a = 0$ in Section 6.3 of [3] gives

$$\begin{aligned}
L_n(-x) &\leq \frac{e^{-\frac{1}{2}x}}{2\sqrt{\pi}} \cdot \frac{e^{2\sqrt{x(n+1)}}}{x^{1/4}(n+1)^{1/4}} \left(1 + \frac{\overline{C}_1}{\sqrt{n+1}} + \mathcal{E}_1 + \mathcal{E}_{3,1} \right) \\
&\leq \frac{e^{-\frac{1}{2}x + 2\sqrt{x}}}{2\sqrt{\pi}} \cdot \frac{e^{2\sqrt{xn}}}{x^{1/4} n^{1/4}} (1 + \overline{C}_1 + \mathcal{E}_1 + \mathcal{E}_{3,1})
\end{aligned} \tag{3.28}$$

13

if $n \geq \max\{m_0, m_1, \ldots, m_7\}$, where $\overline{C}_1$, $\mathcal{E}_1$, $\mathcal{E}_{3,1}$ and $m_0, m_1, \ldots, m_7$ are certain non-negative constants. From the forms given for the constants m_i on p. 3303 we see that (3.28) holds if

$$n \geq \max\{m_0, m_1, \ldots, m_7\} = \max\left\{\frac{x}{4}, 5x, 0, -\frac{5}{4}x, 4x, x\left(1 + \frac{x}{2}\right)^2, \frac{9}{x}, 0\right\}$$

$$= \max\left\{5x, x\left(1 + \frac{x}{2}\right)^2, \frac{9}{x}\right\}.$$

The constants $\overline{C}_1$, $\mathcal{E}_1$ and $\mathcal{E}_{3,1}$ satisfy (see pp. 3306, 3304 and 3294, respectively)

$$\overline{C}_1 \leq \frac{7}{16\sqrt{x}} + 2\sqrt{x}(1 + x/2),$$

$$\mathcal{E}_1 \leq 2e^{-4\sqrt{x(n+1)}} + 2\sqrt{\pi}x^{1/4}(n+1)^{1/4}e^{-2\sqrt{x(n+1)}},$$

$$\mathcal{E}_{3,1} \leq \frac{2e^{-2\sqrt{x(n+1)}}}{\sqrt{\pi}x^{1/4}(n+1)^{1/4}}.$$

The last two of these we may further bound as

$$\mathcal{E}_1 \leq 2 + 2\sqrt{\pi}x^{1/4}(n+1)^{1/4}e^{-2\sqrt{x(n+1)}} \leq 2 + \sqrt{\frac{\pi}{e}} \quad \text{and} \quad \mathcal{E}_{3,1} \leq \frac{2}{\sqrt{\pi}x^{1/4}},$$

where we have used the inequality $2y^{1/4}e^{-2\sqrt{y}} \leq e^{-1/2}$ for $y \geq 0$. $\qquad\square$

Note that the rate in (3.27) cannot be improved because (3.28) is, in fact, not only an upper bound but an asymptotic equivalence for $L_n(-x)$. The following proposition is similar to the lemmas (in particular Lemma 3) in Section 5 of [52].[1] See also [53, Sec. 6.4]. Define the unweighted kernel

$$R(x, y) = \sum_{k=0}^{\infty} \alpha_k^{-1} x^k y^k. \tag{3.29}$$

Lemma 3.11. *Suppose that assumption* (3.4) *holds and let R be the unweighted kernel in* (3.29). *Then*

$$C_R^{2n} \leq c_L(\lambda)\, n^{-1/4} e^{2\sqrt{\lambda n}} \alpha_n^{-1} n!^2$$

for every $n \geq m_L(\lambda)$, where

$$c_L(\lambda) = \frac{e^{\frac{1}{2}\lambda + 2\sqrt{\lambda}}}{2\sqrt{\pi}\lambda^{1/4}}\left(3 + \sqrt{\frac{\pi}{e}} + \frac{2}{\sqrt{\pi}\lambda^{1/4}} + \frac{7}{16\sqrt{\lambda}} + \sqrt{\lambda}(2 + \lambda)\right) \tag{3.30}$$

and

$$m_L(\lambda) = \max\left\{5\lambda, \lambda\left(1 + \frac{\lambda}{2}\right)^2, \frac{9}{\lambda}\right\}. \tag{3.31}$$

---

[1]It seems that the eigenvalue bound used in the proof of Lemma 3 in [52] to estimate $L_n(-1)$ is erroneous because a more careful argument for bounding the log-sum gives

$$\sum_{k=1}^{n-1} \log\left(1 + \frac{1}{\sqrt{k+1}}\right) \leq \int_1^n \log\left(1 + \frac{1}{\sqrt{x}}\right) \mathrm{d}x = \sqrt{n} + (n-1)\log\left(1 + \frac{1}{n}\right) - \frac{1}{2}\log n - 1,$$

which implies that $L_n(-1) = O(n^{-1/2}e^{2\sqrt{n}})$. This contradicts the Laguerre asymptotics in [3]. Although its proof is erroneous, the upper bound $L_n(-1) \leq 2e^{2\sqrt{n}}$ derived in the proof of the lemma in question is correct (at least for sufficiently large n) in the light of [3] and our Proposition 3.10.

Proof. We compute

$$C_R^{2n} = \sup_{x=y\in[-1,1]} \sum_{k=n}^{\infty} \alpha_k^{-1} \left(\frac{k!}{(k-n)!} \right)^2 x^{k-n} y^{k-n} = \sum_{k=n}^{\infty} \alpha_k^{-1} \left(\frac{k!}{(k-n)!} \right)^2.$$

Assumption (3.4) yields

$$\begin{aligned}
C_R^{2n} = \alpha_n^{-1} \sum_{k=n}^{\infty} \frac{\alpha_n}{\alpha_k} \left(\frac{k!}{(k-n)!} \right)^2 &= \alpha_n^{-1} \sum_{k=0}^{\infty} \frac{\alpha_n}{\alpha_{n+k}} \left(\frac{(n+k)!}{k!} \right)^2 \\
&\leq \alpha_n^{-1} \sum_{k=0}^{\infty} \lambda^k \frac{n!}{(n+k)!} \left(\frac{(n+k)!}{k!} \right)^2 \qquad (3.32) \\
&= \alpha_n^{-1} n! \sum_{k=0}^{\infty} \lambda^k \frac{(n+k)!}{k!^2}.
\end{aligned}$$

From these computations it is easy to see that (set $\alpha_k = \lambda^{-k} k!$ so that each inequality becomes an equality)

$$\frac{\partial^{2n}}{\partial x^n \partial y^n} e^{\lambda xy} \bigg|_{\substack{x=1 \\ y=1}} = \lambda^n \sum_{k=0}^{\infty} \lambda^k \frac{(n+k)!}{k!^2}. \qquad (3.33)$$

However, by (3.26) we have an alternative expression for the above derivative in terms of Laguerre polynomials:

$$\frac{\partial^{2n}}{\partial x^n \partial y^n} e^{\lambda xy} \bigg|_{\substack{x=1 \\ y=1}} = \frac{\mathrm{d}^n}{\mathrm{d}y^n} \lambda^n y^n e^{\lambda y} \bigg|_{y=1} = e^\lambda \lambda^n n! L_n(-\lambda). \qquad (3.34)$$

Combining Equations (3.32)–(3.34) therefore yields

$$C_R^{2n} \leq e^\lambda L_n(-\lambda) \alpha_n^{-1} n!^2.$$

The claim follows from Proposition 3.10. $\qquad \square$

It is useful, though not necessary for what follows, to observe that assumption (3.4) and the analyticity φ ensure that $H(K)$ consists of analytic functions. We refer to [38, pp. 40–43] and [39, 46] for general discussion and results on analyticity in the context of reproducing kernel Hilbert spaces. Results on analyticity and related properties for kernels defined via Hermite polynomials may be found in [9, Sec. 3.1] and references therein.

Proposition 3.12. *Suppose that assumption* (3.4) *holds. If φ is analytic, then every element of $H(K)$ is analytic.*

Proof. Without loss of generality we can assume that $\alpha_0 = 1$. Then assumption (3.4) implies that $\alpha_n \geq \lambda^{-n} n!$. If $f \in H(R)$ for the unweighted kernel R in (3.29), Lemmas 3.9 and 3.11 yield

$$\begin{aligned}
\|f^{(n)}\|_\infty \leq \|f\|_R C^n n! \alpha_n^{-1/2} \leq \|f\|_R C^n n! \sqrt{\frac{\lambda^n}{n!}} &\leq \|f\|_R (C\lambda^{1/2})^n \sqrt{n!} \\
&\leq \|f\|_R (C\lambda^{1/2})^n n!
\end{aligned}$$

for a certain positive C, which implies that f is analytic [22, Lem. 1.2.10]. Since the mapping $f \mapsto \varphi f$ is an isometric isomorphism between $H(R)$ and $H(K)$, we see that if φ is analytic, every element of $H(K)$ is analytic, being a product of two analytic functions. $\qquad \square$

Let $f \colon I \to \mathbb{R}$ be an n times continuously differentiable function on a closed interval I. The unique polynomial $S_n f$ of degree $n-1$ that interpolates f at some distinct points $x_1, \ldots, x_n \in I$ can be written in terms of the Lagrange basis functions as

$$(S_n f)(x) = \sum_{k=1}^{n} f(x_k) \prod_{i \neq k} \frac{x - x_i}{x_k - x_i}. \tag{3.35}$$

Define the *weighted polynomial interpolant*

$$(S_n^\varphi f)(x) = \varphi(x) \sum_{k=1}^{n} \frac{f(x_k)}{\varphi(x_k)} \prod_{i \neq k} \frac{x - x_i}{x_k - x_i}. \tag{3.36}$$

Both of these interpolants have the form $\sum_{k=1}^{n} f(x_k) \psi_k$ for certain functions ψ_k. It is a standard result [11, Sec. 2.6] that for each $x \in I$ there exists $\xi_x \in I$ such that

$$f(x) - (S_n f)(x) = \frac{f^{(n)}(\xi_x)}{n!} \prod_{k=1}^{n} (x - x_k). \tag{3.37}$$

Upper bounds on the worst-case errors will be straightforward corollaries of the following proposition.

Proposition 3.13. *Suppose that assumption* (3.3) *holds. If* $f \in H(K)$ *and the points* $x_1, \ldots, x_n \in [-1, 1]$ *are distinct, then*

$$|f(x) - (S_n^\varphi f)(x)| \leq \varphi_{\max} \|f\|_K \frac{(C_R^{2n})^{1/2}}{n!} \left| \prod_{k=1}^{n} (x - x_k) \right| \tag{3.38}$$

for every $x \in [-1, 1]$.

Proof. Since the mapping $f \mapsto \varphi f$ is an isometric isomorphism, for every $f \in H(K)$ there is $g \in H(R)$ such that $f = \varphi g$ and $\|f\|_K = \|g\|_R$. The definitions of S_n and S_n^φ and the assumption (3.3) that φ is bounded from above give

$$|f(x) - (S_n^\varphi f)(x)| = |\varphi(x)| \, |g(x) - (S_n g)(x)| \leq \varphi_{\max} |g(x) - (S_n g)(x)| \,.$$

By Equation (3.37) and Lemma 3.9,

$$|g(x) - (S_n g)(x)| \leq \frac{\|g^{(n)}\|_\infty}{n!} \left| \prod_{k=1}^{n} (x - x_k) \right| \leq \|g\|_R \frac{(C_R^{2n})^{1/2}}{n!} \left| \prod_{k=1}^{n} (x - x_k) \right|$$

$$= \|f\|_K \frac{(C_R^{2n})^{1/2}}{n!} \left| \prod_{k=1}^{n} (x - x_k) \right|,$$

which gives the claim. $\qquad\square$

Note that the above proposition does not require that φ be differentiable. If φ is assumed to be infinitely differentiable and C_K^{2n} can be computed or estimated, one can use the polynomial interpolant (3.35) and the estimate

$$|f(x) - (S_n f)(x)| \leq \|f\|_K \frac{(C_K^{2n})^{1/2}}{n!} \left| \prod_{k=1}^{n} (x - x_k) \right| \tag{3.39}$$

instead of (3.38). However, since the weighting by φ in S_n^φ yields an interpolant that, at least intuitively, more resembles the elements of $H(K)$ than a polynomial interpolant, it is to be expected that (3.38) is tighter than (3.39). In Section 3.4 we demonstrate that this is indeed the case if K is the Gaussian kernel. Weighted polynomial interpolation provides a linear algorithm that attains the upper bounds in Theorem 3.1.

Proofs of upper bounds in Theorem 3.1. Let us first consider $p = \infty$. Let $S_n^\varphi f$ be the weighted polynomial interpolant in (3.36). Proposition 3.13 and Lemma 3.11 yield

$$|f(x) - (S_n^\varphi f)(x)| \leq \varphi_{\max} \|f\|_K \frac{(C_R^{2n})^{1/2}}{n!} \left| \prod_{k=1}^n (x - x_k) \right|$$

$$\leq \varphi_{\max} \|f\|_K \sqrt{c_L(\lambda)} \, n^{-1/8} e^{\sqrt{\lambda n}} \alpha_n^{-1/2} \left| \prod_{k=1}^n (x - x_k) \right|$$

when $n \geq m_L(\lambda)$. The upper bound in (3.8) is immediate, while that in (3.6) follows from the crude estimate $|\prod_{k=1}^n (x - x_k)| \leq 2^n$. To obtain a tighter bound for the nth minimal error we can select $x_1, \ldots, x_n$ as the roots of the nth Chebyshev polynomial T_n, for which we refer to Section 3.1. Because the leading coefficient of T_n is 2^{n-1}, we have

$$\sup_{x \in [-1,1]} \left| \prod_{k=1}^n (x - x_k) \right| = 2^{-n+1} \|T_n\|_\infty, \tag{3.40}$$

so that the upper bound in (3.7) follows from $\|T_n\|_\infty = 1$.

The upper bounds for $p = 2$ are proved identically. We obtain

$$\|f - S_n^\varphi f\|_2 \leq \varphi_{\max} \|f\|_K \sqrt{c_L(\lambda)} \, n^{-1/8} e^{\sqrt{\lambda n}} \alpha_n^{-1/2} \left(\int_{-1}^1 \prod_{k=1}^n (x - x_k)^2 \, dx \right)^{1/2}.$$

Estimating the integral as

$$\int_{-1}^1 \prod_{k=1}^n (x - x_k)^2 \, dx = \int_{-1}^0 \prod_{k=1}^n (x - x_k)^2 \, dx + \int_0^1 \prod_{k=1}^n (x - x_k)^2 \, dx$$

$$\leq \int_{-1}^0 (x - 1)^{2n} \, dx + \int_0^1 (x + 1)^{2n} \, dx$$

$$= \frac{2(2^{2n+1} - 1)}{2n + 1}$$

$$\leq n^{-1} 2^{2n+1}$$

gives the upper bound in (3.6). The upper bound in (3.7) is obtained by again selecting $x_1, \ldots, x_n$ as the roots of the nth Chebyshev polynomial T_n. This improves the above integral estimate to

$$\left(\int_{-1}^1 \prod_{k=1}^n (x - x_k)^2 \, dx \right)^{1/2} = \left(2^{-2n+2} \int_{-1}^1 T_n(x) \, dx \right)^{1/2} \leq 2^{-n+3/2}, \tag{3.41}$$

from which the upper bound of (3.7) follows.[2] $\qquad\square$

3.4 Application to specific kernels

In this section we apply (3.7) with $p = \infty$ to the four weighted power series kernels mentioned in the beginning of the section. Derivations for $p = 2$ and other worst-case errors are entirely analogous and thus omitted. It is notable that for each kernel the rate of decay of the nth minimal error is controlled by the shape parameter ε as in (1.9).

---

[2]By extremal properties of monic orthogonal polynomials, the constant $2^{3/2}$ in (3.41) could be optimised by selecting $x_1, \ldots, x_n$ as the roots of the nth Legendre polynomial.

The Gaussian kernel

$$K(x, y) = \exp\left(-\frac{1}{2}\varepsilon^2(x-y)^2\right) \tag{3.42}$$

is obtained from (3.1) by selecting

$$\alpha_k = \varepsilon^{-2k}k! \quad \text{and} \quad \varphi(x) = \exp\left(-\frac{1}{2}\varepsilon^2 x^2\right).$$

Corollary 3.14. *Let K be the Gaussian kernel in* (3.42) *and consider the setting of Theorem 3.1. Then*

$$c_1\left(\frac{\varepsilon}{2}\right)^n (n!)^{-1/2} \le e_\infty^{\min}(n) \le c_2 n^{-1/8} e^{\varepsilon\sqrt{n}}\left(\frac{\varepsilon}{2}\right)^n (n!)^{-1/2}$$

where the lower bound holds when $n \ge 1$ and the upper bound when $n \ge m_L(\varepsilon^2)$ and

$$c_1 = \sqrt{\frac{2}{1+\varepsilon^2}}\, e^{-(\frac{1}{32}\varepsilon^4 + \frac{1}{2}\varepsilon^2)} \quad \text{and} \quad c_2 = 2\sqrt{c_L(\varepsilon^2)}.$$

The upper bound of Corollary 3.14 is proved using weighted polynomial interpolation and Proposition 3.13. For the Gaussian kernel it is however easy to use polynomial interpolation and the estimate (3.39). Straightforward differentation of the Taylor series of the Gaussian function yields

$$C_K^{2n} = \sup_{x \in [-1,1]} \frac{\partial^{2n}}{\partial v^n \partial w^n} K(v, w)\Big|_{\substack{v=x \\ w=x}} = (-1)^n \frac{\mathrm{d}^{2n}}{\mathrm{d}z^{2n}} e^{-\frac{1}{2}\varepsilon^2 z^2}\Big|_{z=0} = \varepsilon^{2n}\frac{(2n)!}{2^n n!}.$$

Plugging this in (3.39) and using Stirling's formula produces

$$\|f - S_n f\|_\infty \le c\, \|f\|_K\, n^{-1/2}(\sqrt{2e}\,\varepsilon)^n n^{-n/2} \sup_{x \in [-1,1]} \left|\prod_{k=1}^n (x - x_k)\right|$$

for a positive constant c if $f \in H(K)$. In contrast, in the proof at the end of Section 3.3 we saw that

$$\|f - S_n^\varphi f\|_\infty \le \varphi_{\max}\|f\|_K\, \sqrt{c_L(\varepsilon^2)}\, n^{-1/8}e^{\varepsilon\sqrt{n}}\alpha_n^{-1/2} \sup_{x \in [-1,1]}\left|\prod_{k=1}^n (x - x_k)\right|$$

$$\le \tilde{c}\,\|f\|_K\, n^{-1/8-1/4}e^{\varepsilon\sqrt{n}}(\sqrt{e}\,\varepsilon)^n n^{-n/2} \sup_{x \in [-1,1]}\left|\prod_{k=1}^n (x - x_k)\right|$$

for a positive constant $\tilde{c}$, where the second inequality uses Stirling's formula (3.20). These estimates show that by using weighted polynomial interpolation one obtains an improvement of order $2^{-n/2}$ over polynomial interpolation.

The exponential kernel

$$K(x, y) = e^{\varepsilon xy} \tag{3.43}$$

is obtained from (3.1) by selecting

$$\alpha_k = \varepsilon^{-k}k! \quad \text{and} \quad \varphi(x) \equiv 1.$$

Corollary 3.15. *Let K be the exponential kernel in* (3.43) *and consider the setting of Theorem 3.1. Then*

$$c_1\left(\frac{\varepsilon^{1/2}}{2}\right)^n (n!)^{-1/2} \le e_\infty^{\min}(n) \le c_2 n^{-1/8}e^{\sqrt{\varepsilon n}}\left(\frac{\varepsilon^{1/2}}{2}\right)^n (n!)^{-1/2}$$

where the lower bound holds when $n \geq 1$ and the upper bound when $n \geq m_L(\varepsilon)$ and

$$c_1 = \sqrt{\frac{2}{1+\varepsilon}}\, e^{-\frac{1}{32}\varepsilon^2} \quad and \quad c_2 = 2\sqrt{c_L(\varepsilon)}.$$

Let $\tau > 0$. The analytic Hermite kernel

$$K(x,y) = \exp\left(-\frac{1}{2}\varepsilon^2(x^2+y^2) + \tau^2\varepsilon^2 xy\right) \tag{3.44}$$

is obtained from (3.1) by selecting

$$\alpha_k = (\tau\varepsilon)^{-2k} k! \quad and \quad \varphi(x) = \exp\left(-\frac{1}{2}\varepsilon^2 x^2\right).$$

Corollary 3.16. *Let K be the analytic Hermite kernel in (3.44) and consider the setting of Theorem 3.1. Then*

$$c_1\left(\frac{\tau\varepsilon}{2}\right)^n (n!)^{-1/2} \leq e_\infty^{\min}(n) \leq c_2 n^{-1/8} e^{\tau\varepsilon\sqrt{n}}\left(\frac{\tau\varepsilon}{2}\right)^n (n!)^{-1/2}$$

where the lower bound holds when $n \geq 1$ and the upper bound when $n \geq m_L(\tau^2\varepsilon^2)$ and

$$c_1 = \sqrt{\frac{2}{1+\tau^2\varepsilon^2}}\, e^{-(\frac{1}{32}\tau^4\varepsilon^4 + \frac{1}{2}\varepsilon^2)} \quad and \quad c_2 = 2\sqrt{c_L(\tau^2\varepsilon^2)}.$$

The Bessel kernel

$$K(x,y) = \mathcal{I}_0(2\varepsilon\sqrt{xy}) \tag{3.45}$$

is obtained from (3.1) by selecting

$$\alpha_k = \varepsilon^{-k} k!^2 \quad and \quad \varphi \equiv 1.$$

Corollary 3.17. *Let K be the Bessel kernel in (3.45) and consider the setting of Theorem 3.1. Then*

$$c_1\left(\frac{\varepsilon^{1/2}}{2}\right)^n (n!)^{-1} \leq e_\infty^{\min}(n) \leq c_2 n^{-1/8} e^{\sqrt{\varepsilon n}}\left(\frac{\varepsilon^{1/2}}{2}\right)^n (n!)^{-1}$$

where the lower bound holds when $n \geq 1$ and the upper bound when $n \geq m_L(\varepsilon)$ and

$$c_1 = \sqrt{\frac{2}{1+\varepsilon}}\, e^{-\frac{1}{32}\varepsilon^2} \quad and \quad c_2 = 2\sqrt{c_L(\varepsilon)}.$$

3.5 Derivative information

It is interesting to compare (3.6) to bounds for the worst-case error when information consists of derivative evaluations at a single point $a \in (-1,1)$. Here we consider the minimal worst-case error

$$e_\infty^{\min}(\mathrm{D}_a, n) = \inf_{\psi_1,\ldots,\psi_n \in L^\infty} \sup_{0 \neq f \in H(K)} \frac{\|f - \sum_{k=0}^{n-1} f^{(k)}(a)\psi_k\|_\infty}{\|f\|_K}.$$

For brevity we omit the case $p = 2$.

Theorem 3.18. *Suppose that assumptions* (3.3) *and* (3.4) *hold. Let c_L and m_L be the functions defined in* (3.30) *and* (3.31) *and* $a \in (-1, 1)$. *Then*

$$\varphi_{\min}\left(\frac{1 + |a|}{(1 + \lambda a^2)^{1/2}}\right)^n \alpha_n^{-1/2} \leq e_\infty^{\min}(\mathrm{D}_a, n) \leq \sqrt{c_L(\lambda)}\, \varphi_{\max}\, n^{-1/8} e^{\sqrt{\lambda n}}(1 + |a|)^n \alpha_n^{-1/2},$$

where the lower bound holds when $n \geq 1$ and the upper bound when $n \geq m_L(\lambda)$.

Proof. We first prove the upper bound. Consider the weighted Taylor approximation $T_a^n f$ given by

$$(T_a^n f)(x) = \varphi(x) \sum_{k=0}^{n-1} \frac{g^{(k)}(a)}{k!}(x - a)^k,$$

where $g = \varphi^{-1} f$. Taylor's theorem yields

$$\|f - T_a^n f\|_\infty \leq \varphi_{\max} \frac{\|g^{(n)}\|_\infty}{n!} \sup_{x \in [-1, 1]} |x - a|^n = \varphi_{\max} \frac{\|g^{(n)}\|_\infty}{n!}(1 + |a|)^n.$$

The upper bound follows from Lemmas 3.9 and 3.11 and $\|f\|_K = \|g\|_R$, with R the unweighted kernel in (3.29).

To prove the lower bound, let $g_n(x) = (x - a)^n$ and $f_n = \varphi g_n \in H(K)$. From the Hilbert space characterisation (3.2) and the binomial theorem we compute

$$
\begin{aligned}
\|f_n\|_K^2 = \sum_{k=0}^n \alpha_{n-k}\binom{n}{k}^2 a^{2k} = \alpha_n \sum_{k=0}^n \frac{\alpha_{n-k}}{\alpha_n}\binom{n}{k}^2 a^{2k} &\leq \alpha_n \sum_{k=0}^n \lambda^k \frac{(n-k)!}{n!}\binom{n}{k}^2 a^{2k} \\
&= \alpha_n \sum_{k=0}^n \binom{n}{k}(\lambda a^2)^k \frac{1}{k!} \\
&\leq \alpha_n \sum_{k=0}^n \binom{n}{k}(\lambda a^2)^k \\
&= \alpha_n (1 + \lambda a^2)^n.
\end{aligned}
$$

Clearly,

$$\|f_n\|_\infty = \|\varphi g_n\|_\infty \geq \varphi_{\min} \max\{g_n(-1), g_n(1)\} = \varphi_{\min}(1 + |a|)^n. \tag{3.46}$$

The derivatives up to order $n - 1$ of g_n vanish at $x = a$. This implies that also the derivatives of f_n vanish at $x = a$. Thus $\sum_{k=0}^{n-1} f^{(k)}(a)\psi_k \equiv 0$. Consequently,

$$e_\infty^{\min}(\mathrm{D}_a, n) = \inf_{\psi_1, \ldots, \psi_n \in L^\infty(\Omega)} \sup_{0 \neq f \in H(K)} \frac{\|f - \sum_{k=0}^{n-1} f^{(k)}(a)\psi_k\|_\infty}{\|f\|_K} \geq \frac{\|f_n\|_\infty}{\|f_n\|_K}.$$

Using $\|f_n\|_K^2 \leq \alpha_n(1 + \lambda a^2)^n$ and (3.46) yields the claimed lower bound. $\qquad\square$

Note that, up to a constant, the upper bound from Theorem 3.18 coincides with the upper bound on $\sup_{x \in [-1, 1]} e^{\min}(x \mid x_1, \ldots, x_n)$ that is obtained by formally setting $x_1 = \cdots = x_n = a$ in (3.8).

3.6 Some remarks on higher dimensions

Let us conclude with a few short remarks about how the results in this section could and could not be generalised to higher dimensions. The natural generalisation of the kernel (3.1) to dimension $d \in \mathbb{N}$ is

$$K_d(\mathbf{x}, \mathbf{y}) = \varphi(\mathbf{x})\varphi(\mathbf{y}) \sum_{\mathbf{k} \in \mathbb{N}_0^d} \alpha_{\mathbf{k}}^{-1} \mathbf{x}^{\mathbf{k}} \mathbf{y}^{\mathbf{k}} \quad \text{for} \quad \mathbf{x}, \mathbf{y} \in [-1, 1]^d, \tag{3.47}$$

with summation over d-dimensional non-negative multi-indices. This kernel is well-defined and strictly positive-definite if $\varphi\colon [-1,1]^d \to \mathbb{R}$ is non-vanishing and $\sum_{\mathbf{k}\in\mathbb{N}_0^d} \alpha_{\mathbf{k}}^{-1} < \infty$. The definitions of worst-case errors and the Hilbert space characterisations in Section 1 generalise naturally. Due to non-uniqueness of polynomial interpolation in dimensions higher than one it is no surprise that the worst-case error no longer tends to zero for an arbitrary sequence of points.

Proposition 3.19. *Suppose that $d \geq 2$ and $p \in [0,\infty]$. Let K_d be any kernel of the form (3.47). Then there exists a positive constant c such that for every $n \geq 1$ there are distinct points $\mathbf{x}_1, \ldots, \mathbf{x}_n \in [-1,1]^d$ such that*

$$e_p^{\min}(\mathbf{x}_1, \ldots, \mathbf{x}_n) \geq c.$$

Proof. Let $\mathbf{x}_1, \ldots, \mathbf{x}_n$ be any pairwise distinct points on the unit circle. The polynomial $P(\mathbf{x}) = x_1^2 + \cdots + x_d^2 - 1$, where $\mathbf{x} = (x_1, \ldots, x_d)$, vanishes on the unit circle and the function defined as $f(\mathbf{x}) = \varphi(\mathbf{x})P(\mathbf{x})$ is a non-zero element of $H(K_d)$. Therefore $\sum_{k=1}^n f(\mathbf{x}_k)\psi_k \equiv 0$. It follows that $e_p^{\min}(\mathbf{x}_1, \ldots, \mathbf{x}_n) \geq \|f\|_p / \|f\|_{K_d} > 0$, where the lower bound does not depend on n. $\square$

It would be straightforward to use tensor grids to prove limited generalisations to higher dimensions of Theorem 3.1. We also note that Bernstein [2] has proved a certain multivariate extension of the Markov inequality used in Section 3.1. See [8] for a somewhat more accessible source.

4 Stationary kernels

Let $\Phi\colon \mathbb{R} \to \mathbb{R}$ be a continuous and integrable positive-semidefinite function and consider the positive-semidefinite kernel K given by $K(x,y) = \Phi(x-y)$ for all $x,y \in \mathbb{R}$. Such kernels are variously called stationary or translation-invariant. Bochner's theorem [48, Thm. 6.6] ensures that the Fourier transform

$$\widehat{\Phi}(\omega) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \Phi(x)e^{ix\omega} \, \mathrm{d}x$$

of Φ is real and non-negative. The Hilbert space $H(K)$ consists of square-integrable functions whose Fourier transforms decay sufficiently fast [48, Thm. 10.12]:

$$H(K) = \left\{ f \in L^2(\mathbb{R}) \,:\, \|f\|_K^2 = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \frac{|\widehat{f}(\omega)|^2}{\widehat{\Phi}(\omega)} \, \mathrm{d}\omega < \infty \right\}. \tag{4.1}$$

The Gaussian kernel $K(x,y) = \exp(-\frac{1}{2}\varepsilon^2(x-y)^2)$ is one of the most straightforward examples of a stationary kernel, having

$$\Phi(z) = \exp\left(-\frac{1}{2}\varepsilon^2 z^2 \right) \quad \text{and} \quad \widehat{\Phi}(\omega) = \frac{1}{\varepsilon}\exp\left(-\frac{1}{2\varepsilon^2}\omega^2 \right). \tag{4.2}$$

By (4.1) the Hilbert space of the Gaussian kernel therefore consists of functions with super-exponentially decaying Fourier transforms. Note that exponential decay of the Fourier transform suffices for analyticity by the Paley–Wiener theorem [36, Thm. IX.13].

In this section we prove lower bounds on the worst-case errors in (1.1)–(1.3) under assumptions analogous to (3.3) and (3.4). Firstly, we assume that there is a positive constant $\Phi_{\min}$ such that

$$0 < \Phi_{\min} \leq |\Phi(x)| \tag{4.3}$$

for all $x \in [-1,1]$. We note that Φ is uniformly bounded from above on $[-1,1]$ by continuity. Secondly, define

$$\beta_k = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \frac{\widehat{\Phi}^{(k)}(\omega)^2}{\widehat{\Phi}(\omega)}\, \mathrm{d}\omega \quad \text{for} \quad k \geq 0. \tag{4.4}$$

We assume that $\widehat{\Phi}$ is infinitely differentiable and that there exists a positive constant λ such that

$$\beta_n < \infty \quad \text{and} \quad \frac{\beta_k}{\beta_n} \leq \lambda^{n-k} \frac{k!}{n!} \tag{4.5}$$

for all $n \geq 0$ and $0 \leq k \leq n$. Observe that in Section 3 the coefficient α_k is the squared Hilbert space norm of the weighted monomial $f(x) = \varphi(x)x^k$ by (3.2). Similarly, for $g(x) = \Phi(x)x^k$ we get

$$\|g\|_K^2 = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \frac{|\widehat{g}(\omega)|^2}{\widehat{\Phi}(\omega)}\, \mathrm{d}\omega = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \frac{|\mathrm{i}^k \widehat{\Phi}^{(k)}(\omega)|^2}{\widehat{\Phi}(\omega)}\, \mathrm{d}\omega = \beta_k. \tag{4.6}$$

The constants β_k and assumption (4.5) play the same role as α_k and (3.4) in Section 3.

Before proceeding, let us verify (4.5) for the Gaussian kernel. Since $\varphi = \Phi$ for the Gaussian kernel, it follows from (4.6) that $\beta_k = \alpha_k = \varepsilon^{-2k}k!$, but it is nevertheless useful to use Fourier transforms to check this. Equation (4.2) and the Rodrigues' formula for probabilist's Hermite polynomials give

$$\widehat{\Phi}^{(k)}(\omega) = \frac{(-1)^k}{\varepsilon^{k+1}} H_k\left(\frac{\omega}{\varepsilon}\right) \exp\left(-\frac{1}{2\varepsilon^2}\omega^2\right) = \frac{(-1)^k}{\varepsilon^k} H_k\left(\frac{\omega}{\varepsilon}\right) \widehat{\Phi}(\omega),$$

where H_k is the kth probabilist's Hermite polynomial. From the weighted L^2 norm of Hermite polynomials we obtain

$$\begin{aligned}
\beta_k = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \frac{\widehat{\Phi}^{(k)}(\omega)^2}{\widehat{\Phi}(\omega)}\, \mathrm{d}\omega &= \varepsilon^{-2k} \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} H_k\left(\frac{\omega}{\varepsilon}\right)^2 \frac{1}{\varepsilon} \exp\left(-\frac{1}{2\varepsilon^2}\omega^2\right) \mathrm{d}\omega \\
&= \varepsilon^{-2k} \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} H_k(\omega)^2 \exp\left(-\frac{1}{2}\omega^2\right) \mathrm{d}\omega \\
&= \varepsilon^{-2k} k!,
\end{aligned}$$

which satisfies (4.5) with $\lambda = \varepsilon^2$. We prove the following theorem in Section 4.1.

Theorem 4.1. *Let K be a stationary positive-semidefinite kernel and $p \in \{2,\infty\}$. Suppose that Φ and $(\beta_k)_{k=0}^{\infty}$ satisfy (4.3) and (4.5). Then*

$$\tilde{c}_{1,p}\,(n+1)^{-1/2}\, 2^{-n}\beta_n^{-1/2} \leq e_p^{\min}(x_1, \ldots, x_n), \tag{4.7}$$

$$\tilde{c}_{1,p}\,(n+1)^{-1/2}\, 2^{-n}\beta_n^{-1/2} \leq e_p^{\min}(n), \tag{4.8}$$

$$\tilde{c}_{1,\infty}\,(n+1)^{-1/2}\, 2^{-2n}\beta_n^{-1/2} \leq \frac{e^{\min}(x \mid x_1, \ldots, x_n)}{|\prod_{k=1}^{n}(x - x_k)|} \tag{4.9}$$

for every $n \geq 1$, where

$$\tilde{c}_{1,\infty} = \sqrt{\frac{2}{1+\lambda}}\, e^{-\frac{1}{32}\lambda^2} \Phi_{\min} \quad \text{and} \quad \tilde{c}_{1,2} = \frac{2\sqrt{2\pi}}{3e^3\sqrt{3(4+\lambda)}}\, e^{-\frac{1}{32}\lambda^2} \Phi_{\min}. \tag{4.10}$$

The only difference to the lower bounds of Theorem 3.1 is the additional factor $(n+1)^{-1/2}$ and the substitution of $\Phi_{\min}$ for $\varphi_{\min}$ in the constant. We omit upper bounds

as we have not found a convenient way to express them in terms of β_n. Note that polynomial interpolation (3.37) and Lemma 3.9, together with stationarity, yield the bound

$$
\begin{aligned}
|f(x) - (S_n f)(x)| &\leq \|f\|_K \frac{(C_K^{2n})^{1/2}}{n!} \left| \prod_{k=1}^n (x - x_k) \right| \\
&= \|f\|_K \left| \prod_{k=1}^n (x - x_k) \right| \frac{1}{n!} \sqrt{\Phi^{(2n)}(0)} \\
&= \|f\|_K \left| \prod_{k=1}^n (x - x_k) \right| \frac{1}{n!} \left(\int_{\mathbb{R}} \omega^{2n} \widehat{\Phi}(\omega) \, d\omega \right)^{1/2},
\end{aligned}
$$

which is in terms of the moments of $\widehat{\Phi}$ rather than of β_n. The upper bound in [49] (to be discussed in Section 4.2) is derived from the above estimate.

4.1 Proof of Theorem 4.1

To prove Theorem 4.1 we use the following lemma, a cousin of Lemmas 3.5 and 3.8.

Lemma 4.2. *Let $P_n(x) = \sum_{k=0}^n a_k x^k$ for $a_k \in \mathbb{R}$ be a polynomial of degree n. If assumption (4.5) holds, then the function $f_n(x) = \Phi(x) P_n(x)$ is an element of $H(K)$ and its norm satisfies*

$$
\|f_n\|_K^2 \leq (1 + \lambda)\, \tau_2 \,(n + 1)\, 2^{2n} \alpha_n \, \|P_n\|_\infty^2 \quad \text{and} \quad \|f_n\|_K^2 \leq \ell_2 \,(n + 1)\, 2^{2n} \alpha_n \, \|P_n\|_2^2
$$

for every $n \geq 1$, where the constants τ_2 and ℓ_2 are given in (3.15) and (3.19).

Proof. By the properties of the Fourier transform,

$$
\widehat{f_n}(\omega) = \sum_{k=0}^n a_k \mathrm{i}^k \widehat{\Phi}^{(k)}(\omega).
$$

From (4.1), Jensen's inequality and (4.4) we get

$$
\begin{aligned}
\|f_n\|_K^2 = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \frac{|\widehat{f_n}(\omega)|^2}{\widehat{\Phi}(\omega)} \, d\omega &= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \frac{1}{\widehat{\Phi}(\omega)} \left| \sum_{k=0}^n a_k \mathrm{i}^k \widehat{\Phi}^{(k)}(\omega) \right|^2 d\omega \\
&\leq \frac{n+1}{\sqrt{2\pi}} \int_{\mathbb{R}} \frac{1}{\widehat{\Phi}(\omega)} \sum_{k=0}^n a_k^2 \, \widehat{\Phi}^{(k)}(\omega)^2 \, d\omega \\
&= (n + 1) \sum_{k=0}^n \beta_k a_k^2.
\end{aligned}
$$

Lemmas 3.5 and 3.8 provide bounds on $\sum_{k=0}^n \alpha_k a_k^2$ if α_k satisfy (3.4). Because assumption (4.5) on β_k is identical to (3.4), the claimed bounds follow from these lemmas. $\square$

Proofs of lower bounds in Theorem 4.1. Select $f_n(x) = \Phi(x)(x - x_1) \cdots (x - x_n)$ and proceed as in Sections 3.1 and 3.2. Use Lemma 4.2 in the place of Lemmas 3.5 and 3.8 to estimate the norm of f_n, which yields lower bounds that differ from those in Theorem 3.1 by the factor $(n + 1)^{-1/2}$. $\square$

4.2 Bounds by Yarotsky

Yarotsky [49] has derived pointwise upper and lower bounds on the posterior variance in Gaussian process interpolation for a certain class of analytic stationary covariance kernels. It is well known that the Gaussian process posterior variance equals the squared worst-case error in the Hilbert space of the covariance kernel, a fact that one can verify by noting that [e.g., 29, Sec. 10.2]

$$
\begin{aligned}
e^{\min}(x \mid x_1, \ldots, x_n)^2 &= \inf_{u_1, \ldots, u_n \in \mathbb{R}} \sup_{0 \neq f \in H(K)} \frac{|f(x) - \sum_{k=1}^{n} f(x_k) u_k|^2}{\|f\|_K^2} \\
&= \inf_{u_1, \ldots, u_n \in \mathbb{R}} \left\| K(\cdot, x) - \sum_{k=1}^{n} u_k K(\cdot, x_k) \right\|_K^2 \\
&= \inf_{u_1, \ldots, u_n \in \mathbb{R}} \left(K(x, x) - 2 \sum_{k=1}^{n} u_k K(x, x_k) + \sum_{k,l=1}^{n} u_k u_l K(x_k, x_l) \right) \\
&= K(x, x) - \mathsf{k}_n(x)^\mathsf{T} \mathsf{K}_n^{-1} \mathsf{k}_n(x),
\end{aligned}
$$

where $\mathsf{k}_n(x)$ is an n-vector with elements $K(x, x_i)$ and K_n an $n \times n$ matrix with elements $K(x_i, x_j)$. The last expression, which requires assuming that the kernel is strictly positive-definite, equals the posterior variance in Equation (6) of [49].

Theorem 4.3 (Special case of Theorem 2 in [49]). *Let K be a strictly positive-definite stationary kernel such that*

$$
\widehat{\Phi}(\omega) = \exp(-a|\omega|^b) \quad for \quad a > 0 \ and \ b > 1.
$$

If n is sufficiently large and $x_1, \ldots, x_n \in [-1, 1]$ are distinct, then

$$
c\, e^{-n/2} (abe)^{-n/b} \left(\frac{(2n+1)^{1/b}}{n} \right)^{n+1/2}
$$
$$
\leq \frac{e^{\min}(x \mid x_1, \ldots, x_n)}{|\prod_{k=1}^{n}(x - x_k)|} \leq c\, e^n (abe)^{-n/b} \left(\frac{(2n+1)^{1/b}}{n} \right)^{n+1/2},
$$

where $c = (abe)^{-1/2b}$.

Yarotsky uses polynomial interpolation to prove the upper bound. His proof of the lower bound differs considerably from our proofs based on estimation of polynomial coefficients. Rather than construct a fooling function, he writes the worst-case error as an integral of the product of an exponential sum and the Fourier transform of Φ which he then truncates and bounds from below, a feat which involves an explicit expression for the distance between $(1, z, \ldots, z^n) \in \mathbb{C}^{n+1}$ and the span of $((1, z_k, \ldots, z_k^n))_{k=1}^{n} \subset \mathbb{C}^{n+1}$.

The assumption $b > 1$ ensures that $H(K)$ consists of analytic functions. Setting $b = 2$ and $a = 1/(2\varepsilon^2)$ gives a constant multiple of the Gaussian kernel by (4.2). In this case the lower and upper bounds in Theorem 4.3 are asymptotically constant multiples of

$$
n^{-1/4} \left(\frac{\sqrt{2}\, \varepsilon}{e} \right)^n n^{-n/2} \quad \text{and} \quad n^{-1/4} (\sqrt{2e}\, \varepsilon)^n n^{-n/2},
$$

respectively. For the Gaussian kernel our corresponding bounds in (3.8) are constant multiples of

$$
\left(\frac{\varepsilon}{4} \right)^n (n!)^{-1/2} \quad \text{and} \quad n^{-1/8} e^{\varepsilon\sqrt{n}} \varepsilon^n (n!)^{-1/2}. \tag{4.11}
$$

By Stirling's formula, $n! \sim \sqrt{2\pi}\, n^{n+1/2} e^{-n}$, so that the bounds (4.11) are asymptotically constant multiples of

$$n^{-1/4}\left(\frac{\sqrt{e}\,\varepsilon}{4}\right)^n n^{-n/2} \quad \text{and} \quad n^{-3/8} e^{\varepsilon\sqrt{n}} (\sqrt{e}\,\varepsilon)^n n^{-n/2}.$$

Because $\sqrt{2}/e \approx 0.52 > \sqrt{e}/4 \approx 0.41$ and $\sqrt{e} < \sqrt{2e}$, the lower bound in Theorem 4.3 is tighter than that in (3.8) but the upper bound is looser. That our upper bound is tighter is due to our use of weighted polynomial interpolation; Yarotsky uses polynomial interpolation, which is expected to yield worse upper bounds as discussed in Sections 3.3 and 3.4.

Acknowledgements

References

[1] ANDERSSON, J.-E., AND BOJANOV, B. D. A note on the optimal quadrature in H^p. *Numerische Mathematik 44* (1984), 301–308.

[2] BERNSTEIN, S. N. On certain elementary extremal properties of polynomials in several variables. *Doklady Akademii Nauk SSSR 59*, 5 (1948), 833–836. In Russian.

[3] BORWEIN, D., BORWEIN, J. M., AND CRANDALL, R. E. Effective Laguerre asymptotics. *SIAM Journal on Numerical Analysis 46*, 6 (2008), 3285–3312.

[4] CHEN, J., AND WANG, H. Average case tractability of multivariate approximation with Gaussian kernels. *Journal of Approximation Theory 239* (2019), 51–71.

[5] EBERTS, M., AND STEINWART, I. Optimal regression rates for SVMs using Gaussian kernels. *Electronic Journal of Statistics 7* (2013), 1–42.

[6] FASSHAUER, G., HICKERNELL, F., AND WOŹNIAKOWSKI, H. Average case approximation: Convergence and tractability of Gaussian kernels. In *Monte Carlo and Quasi-Monte Carlo Methods 2010* (2010), Springer Verlag, pp. 329–344.

[7] FASSHAUER, G., HICKERNELL, F., AND WOŹNIAKOWSKI, H. On dimension-independent rates of convergence for function approximation with Gaussian kernels. *SIAM Journal on Numerical Analysis 50*, 1 (2012), 247–271.

[8] GANZBURG, M. I. Sharp constants of approximation theory. V. An asymptotic equality related to polynomials with given Newton polyhedra. *Journal of Mathematical Analysis and Applications 499*, 1 (2021), 125026.

[9] GNEWUCH, M., HEFTER, M., HINRICHS, A., AND RITTER, K. Countable tensor products of Hermite spaces and spaces of Gaussian kernels. *Journal of Complexity 71* (2022), 101654.

[10] GNEWUCH, M., HINRICHS, A., RITTER, K., AND RÜSSMANN. Infinite-dimensional integration and L^2-approximation on Hermite spaces. *Journal of Approximation Theory 300* (2024), 106027.

[11] HILDEBRAND, F. B. *Introduction to Numerical Analysis*. Courier Corporation, 1987.

[12] IRRGEHER, C., KRITZER, P., LEOBACHER, G., AND PILLICHSHAMMER, F. Integration in Hermite spaces of analytic functions. *Journal of Complexity 31*, 3 (2015), 380–404.

[13] IRRGEHER, C., KRITZER, P., PILLICHSHAMMER, F., AND WOŹNIAKOWSKI, H. Approximation in Hermite spaces of smooth functions. *Journal of Approximation Theory 207* (2016), 98–126.

[14] IRRGEHER, C., AND LEOBACHER, G. High-dimensional integration on $\mathbb{R}^d$, weighted Hermite spaces, and orthogonal transforms. *Journal of Complexity 31*, 2 (2015), 174–205.

[15] KARVONEN, T. Asymptotic bounds for smoothness parameter estimates in Gaussian process interpolation. *SIAM/ASA Journal on Uncertainty Quantification 11*, 4 (2023), 1225–1257.

[16] KARVONEN, T. Small sample spaces for Gaussian processes. *Bernoulli 29*, 2 (2023), 875–900.

[17] KARVONEN, T., OATES, C. J., AND GIROLAMI, M. Integration in reproducing kernel Hilbert spaces of Gaussian kernels. *Mathematics of Computation 90*, 331 (2021), 2209–2233.

[18] KARVONEN, T., AND SÄRKKÄ, S. Gaussian kernel quadrature at scaled Gauss–Hermite nodes. *BIT Numerical Mathematics 59*, 4 (2019), 877–902.

[19] KARVONEN, T., TANAKA, K., AND SÄRKKÄ, S. Kernel-based interpolation at approximate Fekete points. *Numerical Algorithms 87* (2021), 445–468.

[20] KHARTOV, A. A., AND LIMAR, I. A. Asymptotic analysis in multivariate average case approximation with Gaussian kernels. *Journal of Complexity 70* (2022), 101631.

[21] KHARTOV, A. A., AND LIMAR, I. A. Asymptotic analysis in multivariate worst case approximation with Gaussian kernels. *Journal of Complexity 82* (2024), 101838.

[22] KRANTZ, S. G., AND PARKS, H. R. *A Primer of Real Analytic Functions*, 2nd ed. Birkhäuser, 2002.

[23] KRIEG, D., AND VYBÍRAL, J. New lower bounds for the integration of periodic functions. *Journal of Fourier Analysis and Applications 29* (2024), 41.

[24] KÜHN, T. Covering numbers of Gaussian reproducing kernel Hilbert spaces. *Journal of Complexity 27*, 5 (2011), 489–499.

[25] KUO, F. Y., SLOAN, I. H., AND WOŹNIAKOWSKI, H. Multivariate integration for analytic functions with Gaussian kernels. *Mathematics of Computation 86*, 304 (2017), 829–853.

[26] KUO, F. Y., AND WOŹNIAKOWSKI, H. Gauss–Hermite quadratures for functions from Hilbert spaces with Gaussian reproducing kernels. *BIT Numerical Mathematics 52*, 2 (2012), 425–436.

[27] LABELLE, G. Concerning polynomials on the unit interval. *Proceedings of the American Mathematical Society 20*, 2 (1969), 321–326.

[28] NOVAK, E., AND WOŹNIAKOWSKI, H. *Tractability of Multivariate Problems. Volume I: Linear Information*. No. 6 in EMS Tracts in Mathematics. European Mathematical Society, 2008.

[29] NOVAK, E., AND WOŹNIAKOWSKI, H. *Tractability of Multivariate Problems. Volume II: Standard Information for Functionals*. No. 12 in EMS Tracts in Mathematics. European Mathematical Society, 2010.

[30] PAULSEN, V. I., AND RAGHUPATHI, M. *An Introduction to the Theory of Reproducing Kernel Hilbert Spaces*. No. 152 in Cambridge Studies in Advanced Mathematics. Cambridge University Press, 2016.

[31] PINKUS, A. *n-Widths in Approximation Theory*. No. 7 in Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge. Springer-Verlag, 1985.

[32] PLATTE, R. B. How fast do radial basis function interpolants of analytic functions converge? *IMA Journal of Numerical Analysis 31*, 4 (2011), 1578–1597.

[33] PLATTE, R. B., AND DRISCOLL, T. A. Polynomials and potential theory for Gaussian radial basis function interpolation. *SIAM Journal on Numerical Analysis 43*, 2 (2005), 750–766.

[34] RAHMAN, Q. I., AND SCHMEISSER, G. *Analytic Theory of Polynomials.* No. 26 in London Mathematical Society Monographs New Series. Clarendon Press, 2002.

[35] RASMUSSEN, C. E., AND WILLIAMS, C. K. I. *Gaussian Processes for Machine Learning.* Adaptive Computation and Machine Learning. MIT Press, 2006.

[36] REED, M., AND SIMON, B. *Methods of Modern Mathematical Physics. II: Fourier Analysis, Self-Adjointness.* Elsevier, 1975.

[37] RIEGER, C., AND ZWICKNAGL, B. Sampling inequalities for infinitely smooth functions, with applications to interpolation and machine learning. *Advances in Computational Mathematics 32* (2010), 103–129.

[38] SAITOH, S. *Integral Transforms, Reproducing Kernels and Their Applications.* Chapman and Hall, 1997.

[39] SAITOH, S., AND SAWANO, Y. *Theory of Reproducing Kernels and Applications.* No. 44 in Developments in Mathematics. Springer, 2016.

[40] SCHABACK, R. Comparison of radial basis function interpolants. In *Multivariate Approximation: From CAGD to Wavelets.* World Scientific, 1993, pp. 293–305.

[41] SCHEUERER, M., SCHABACK, R., AND SCHLATHER, M. Interpolation of spatial data – A stochastic or a deterministic problem? *European Journal of Applied Mathematics 24*, 4 (2013), 601–629.

[42] SLOAN, I. H., AND WOŹNIAKOWSKI, H. Multivariate approximation for analytic functions with Gaussian kernels. *Journal of Complexity 45* (2018), 1–21.

[43] STEIN, M. L. *Interpolation of Spatial Data: Some Theory for Kriging.* Springer Series in Statistics. Springer, 1999.

[44] STEINWART, I., AND CHRISTMANN, A. *Support Vector Machines.* Information Science and Statistics. Springer, 2008.

[45] STEINWART, I., AND FISCHER, S. A closer look at covering number bounds for Gaussian kernels. *Journal of Complexity 62* (2021), 101513.

[46] SUN, H.-W., AND ZHOU, D.-X. Reproducing kernel Hilbert spaces associated with analytic translation-invariant Mercer kernels. *Journal of Fourier Analysis and Applications 14*, 1 (2008), 89–101.

[47] VAN DER VAART, A. W., AND VAN ZANTEN, J. H. Adaptive Bayesian estimation using a Gaussian random field with inverse Gamma bandwidth. *The Annals of Statistics 37*, 5B (2009), 2655–2675.

[48] WENDLAND, H. *Scattered Data Approximation.* No. 17 in Cambridge Monographs on Applied and Computational Mathematics. Cambridge University Press, 2005.

[49] YAROTSKY, D. Examples of inconsistency in optimization by expected improvement. *Journal of Global Optimization 56* (2013), 1773–1790.

[50] ZHOU, D.-X. The covering number in learning theory. *Journal of Complexity 18*, 3 (2002), 739–767.

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

[52] ZWICKNAGL, B. Power series kernels. *Constructive Approximation 29*, 1 (2009), 61–84.

[53] ZWICKNAGL, B., AND SCHABACK, R. Interpolation and approximation in Taylor spaces. *Journal of Approximation Theory 171* (2013), 65–83.

[54] ZYGMUND, A. *Trigonometric Series. Volumes I & II Combined,* 3rd ed. Cambridge Mathematical Library. Cambridge University Press, 2003.