Kernel-based interpolation at approximate Fekete points

Toni Karvonen,* Simo Särkkä,* and Ken’ichiro Tanaka†

December 17, 2019

Abstract

We construct approximate Fekete point sets for kernel-based interpolation by maximising the determinant of a kernel Gram matrix obtained via truncation of an orthonormal expansion of the kernel. Uniform error estimates are proved for kernel interpolants at the resulting points. If the kernel is Gaussian we show that the approximate Fekete points in one dimension are the solution to a convex optimisation problem and that the interpolants convergence with a super-exponential rate. A numerical experiment is provided for the Gaussian kernel.

1 Introduction

Kernel-based methods are widely used in interpolation and approximation of functions (Wendland, 2005; Fasshauer, 2007; Fasshauer and McCourt, 2015). Let $d \in \mathbb{N}$ and $\Omega \subset \mathbb{R}^d$ be a compact set with a non-empty interior. Given evaluations of a function $f: \Omega \to \mathbb{R}$ at a scattered set of distinct points $X_n = \{x_1, \ldots, x_n\} \subset \Omega$ and a continuous positive-definite kernel $K: \Omega \times \Omega \to \mathbb{R}$, the kernel interpolant $s_f$ is

$$s_f(x) = \sum_{k=1}^n c_k K(x, x_k),$$

where the coefficients $c_k$ are uniquely determined by the interpolation conditions $s_f(x_k) = f(x_k)$ for every $k = 1, \ldots, n$. The choice of the evaluation points $X_n$ can have a significant effect on the accuracy of the approximation $s_f(x) \approx f(x)$ at $x \notin X_n$. Popular methods for constructing “good” point sets include different types of greedy algorithms (Schaback and Wendland, 2000; De Marchi et al., 2005; Müller, 2009; Wirtz and Haasdonk, 2013; Santin and Haasdonk, 2017) that construct the next point $x_{n+1}$ by maximising the power function. An alternative approach is to select $n$ points concurrently by maximising

$$\det K_{X_n} = \det(K(x_k, x_m))_{k,m=1}^n,$$

the determinant of the kernel Gram matrix, over all sets of $n$ points $X_n \subset \Omega$. The resulting points are called Fekete points in an analogue to the classical Fekete points that maximise the Vandermonde determinant (Bos et al., 2010; Briani et al., 2012). The asymptotic distribution of these points for kernel-based interpolation in one dimension has been studied by Bos and Maier (2002) and Bos and De Marchi (2011).

*Department of Electrical Engineering and Automation, Aalto University, Espoo, Finland
†Department of Mathematical Informatics, Graduate School of Information Science and Technology, University of Tokyo, Japan

arXiv:1912.07316v1 [math.NA] 16 Dec 2019
Because maximisation of $\det K_{X_n}$ is typically intractable, in this article we study approximate Fekete points that are obtained by maximising the determinant of the kernel matrix of a truncated version of the kernel. Let \( \{ \varphi_\ell \}_{\ell=1}^\infty \) be an orthonormal basis of \( \mathcal{H}_K(\Omega) \), the reproducing kernel Hilbert space (RKHS) of \( K \). Then the kernel can be written as

\[
K(x,y) = \sum_{\ell=1}^\infty \varphi_\ell(x)\varphi_\ell(y).
\]

The approximate Fekete points \( X_n^* \) are then defined as any set of \( n \) points maximising

\[
\det \hat{K}_{X_n} = \det \left( \sum_{\ell=1}^n \varphi_\ell(x_k)\varphi_\ell(x_m) \right)_{k,m=1}^n.
\]

This and related constructions have been recently suggested by [Tanaka 2019] and in the context of numerical integration and sampling from determinantal point processes by [Belhadji et al. 2019] and [Gautier et al. 2019]. Our construction differs slightly from the prior work in that we do not require the basis functions \( \{ \varphi_\ell \}_{\ell=1}^\infty \) to arise from Mercer’s theorem, which significantly simplifies analysis and construction of the points, at least when the kernel is Gaussian. This article contains two main contributions:

- Let \( f \in \mathcal{H}_K(\Omega) \). In Section 3 we use a bound on the Lebesgue constant for interpolation with \( \{ \varphi_\ell \}_{\ell=1}^n \) to prove that

\[
\sup_{x \in \Omega} |f(x) - s_f(x)| \leq 2 \|f\|_{\mathcal{H}_K(\Omega)} (1 + n) \sup_{x \in \Omega} \left( \sum_{\ell=n+1}^\infty \varphi_\ell(x)^2 \right)^{1/2}
\]

for interpolation at any approximate Fekete points.

- In Section 4 we show that for a certain simple orthonormal expansion [Minh 2010] of the univariate Gaussian kernel

\[
K(x,y) = \exp \left( -\varepsilon^2 (x - y)^2 \right)
\]

with a scale parameter \( \varepsilon > 0 \) the objective function (1.1) is convex and has a unique maximiser. We then specialise the uniform error estimate (1.2) and some other results from Section 3 for the Gaussian kernel.

A numerical experiment for the Gaussian kernel is given in Section 5. We also discuss improved error estimates in subspaces of \( \mathcal{H}_K(\Omega) \) and tensor product extensions of the univariate approximate Fekete points for anisotropic multivariate Gaussian kernels.

## 2 Background

This section reviews basic properties of kernel interpolants and defines the approximate Fekete points studied in the remainder of the article.

### 2.1 Kernel-based interpolation

Every positive-definite \( K : \Omega \times \Omega \to \mathbb{R}^d \) on a general domain \( \Omega \subset \mathbb{R}^d \) induces a unique reproducing kernel Hilbert space \( \mathcal{H}_K(\Omega) \), which is a Hilbert space consisting of real-valued functions defined on \( \Omega \). The RKHS is characterised by the properties that \( K(\cdot,x) \in \mathcal{H}_K(\Omega) \) for every \( x \in \Omega \) and \( \langle f, K(\cdot,x) \rangle_{\mathcal{H}_K(\Omega)} = f(x) \) for every \( f \in \mathcal{H}_K(\Omega) \) and \( x \in \Omega \), the latter of which is known as the reproducing property.
Given a set of \( n \) distinct points, \( \mathcal{X}_n = \{ \mathbf{x}_1, \ldots, \mathbf{x}_n \} \subset \Omega \), the kernel interpolant \( s_f \) is the minimum-norm interpolant to a function \( f: \Omega \to \mathbb{R} \) at these points:

\[
s_f = \arg \min \{ \|g\|_{\mathcal{H}_K(\Omega)} : g \in \mathcal{H}_K(\Omega) \text{ s.t. } g(\mathbf{x}_k) = f(\mathbf{x}_k) \text{ for every } \mathbf{x}_k \in \mathcal{X}_n \}. \tag{2.1}
\]

This definition implies that \( \|s_f\|_{\mathcal{H}_K(\Omega)} \leq \|f\|_{\mathcal{H}_K(\Omega)} \). The main advantage in working in an RKHS as opposed to some different function space is that the minimum-norm interpolant has a simple algebraic form:

\[
s_f(x) = \sum_{k=1}^{n} c_k K(\mathbf{x}, \mathbf{x}_k) = \mathbf{c}^T \mathbf{k}_{\mathcal{X}_n}(\mathbf{x}), \tag{2.2}
\]

where we denote \( \mathbf{c} = (c_1, \ldots, c_n) \in \mathbb{R}^n \) and \( \mathbf{k}_{\mathcal{X}_n}(\mathbf{x}) = (K(\mathbf{x}, \mathbf{x}_k))_{k=1}^{n} \in \mathbb{R}^n \). The coefficients \( \mathbf{c} \) are

\[
\begin{bmatrix}
c_1 \\ \vdots \\ c_n
\end{bmatrix} = (K(\mathbf{x}_1, \mathbf{x}_1) \cdots K(\mathbf{x}_1, \mathbf{x}_n))^{-1} \begin{bmatrix} f(\mathbf{x}_1) \\ \vdots \\ f(\mathbf{x}_n) \end{bmatrix},
\]

where \( \mathbf{k}_{\mathcal{X}_n} = (K(\mathbf{x}_k, \mathbf{x}_m))_{k,m=1}^{n} \) is the positive-definite kernel Gram matrix. From this it follows that \( s_f \) is the unique interpolant to \( f \) at \( \mathcal{X}_n \) in the span of \( \{K(\cdot, \mathbf{x}_k)\}_{k=1}^{n} \).

The interpolant can be written using the cardinal functions \( u_k \in \text{span}\{K(\cdot, \mathbf{x}_k)\}_{k=1}^{n} \) satisfying \( u_k(\mathbf{x}_m) = \delta_{km} \) as \( s_f = \sum_{k=1}^{n} f(\mathbf{x}_k) u_k \). From the reproducing property and the Cauchy–Schwarz inequality it now follows that for any \( f \in \mathcal{H}_K(\Omega) \) the interpolation error admits the bound

\[
|f(x) - s_f(x)| = \left\langle f, K(\cdot, \mathbf{x}) - \sum_{k=1}^{n} K(\cdot, \mathbf{x}_k) u_k(\mathbf{x}) \right\rangle_{\mathcal{H}_K(\Omega)} \leq \|f\|_{\mathcal{H}_K(\Omega)} \left\| K(\cdot, \mathbf{x}) - \sum_{k=1}^{n} K(\cdot, \mathbf{x}_k) u_k(\mathbf{x}) \right\|_{\mathcal{H}_K(\Omega)} =: \|f\|_{\mathcal{H}_K(\Omega)} P_{\mathcal{X}_n}(x). \tag{2.3}
\]

The latter form is the point-wise worst-case approximation error. The power function can be also written in a determinantal form (e.g. Schaback 2005, Lemma 3)

\[
P_{\mathcal{X}_n}(x) = \sqrt{K(\mathbf{x}, \mathbf{x}) - \mathbf{k}_{\mathcal{X}_n}(\mathbf{x})^T \mathbf{K}^{-1}_{\mathcal{X}_n} \mathbf{k}_{\mathcal{X}_n}(\mathbf{x})} = \sup_{\|g\|_{\mathcal{H}_K(\Omega)} \leq 1} |f(x) - s_f(x)|. \tag{2.4}
\]

which suggests, via \( (\ref{2.3}) \), that points \( \mathcal{X}_n \) that maximise \( \det \mathbf{K}_{\mathcal{X}_n} \) ought to provide to small approximation error. Numerous explicit bounds on the error \( f - s_f \) in different norms and for different classes of kernels and functions within and without the RKHS can be found in Wendland (2005, Chapter 11); Wendland and Rieger (2005); Narcowich et al. (2006); Arcangéli et al. (2007) and Rieger and Zwicknagl (2010).

### 2.2 Approximate Fekete points

For the remainder of this article we assume that \( \Omega \) is a compact subset of \( \mathbb{R}^d \) with a non-empty interior and that the positive-definite kernel \( K: \Omega \times \Omega \to \mathbb{R} \) is continuous.
Let \( \{\varphi_{\ell}\}_{\ell=1}^{\infty} \) be an orthonormal basis of \( \mathcal{H}_K(\Omega) \). Then the kernel can be written as

\[
K(x, y) = \sum_{\ell=1}^{\infty} \varphi_{\ell}(x) \varphi_{\ell}(y)
\]

for all \( x, y \in \Omega \). It is easy to verify that \( K \) in (2.5) is the reproducing kernel: Any \( f \in \mathcal{H}_K(\Omega) \) has the expansion \( f = \sum_{\ell=1}^{\infty} \langle f, \varphi_{\ell} \rangle_{\mathcal{H}_K(\Omega)} \varphi_{\ell} \) so that

\[
\langle f, K(\cdot, x) \rangle_{\mathcal{H}_K(\Omega)} = \sum_{\ell, k=1}^{\infty} \langle \varphi_{\ell}, \varphi_k \rangle_{\mathcal{H}_K(\Omega)} (f, \varphi_{\ell})_{\mathcal{H}_K(\Omega)} \varphi_k(x) = \sum_{\ell=1}^{\infty} \langle \varphi_{\ell}, \varphi_{\ell} \rangle_{\mathcal{H}_K(\Omega)} \varphi_{\ell}(x) = f(x).
\]

The Fekete points for interpolation with the kernel (2.5) are the points that maximise the determinant

\[
det K_{X_n} = \det \begin{bmatrix} K(x_1, x_1) & \cdots & K(x_1, x_n) \\ \vdots & \ddots & \vdots \\ K(x_n, x_1) & \cdots & K(x_n, x_n) \end{bmatrix}
\]

of the kernel matrix. As exact computation of the Fekete points is typically challenging, we fix an orthonormal basis \( \{\varphi_{\ell}\}_{\ell=1}^{\infty} \) of \( \mathcal{H}_K(\Omega) \), truncate the expansion (2.5) after \( n \) terms and consider maximisation of the resulting approximation of the objective function (2.6).

Define the truncated kernel

\[
\hat{K}(x, y) = \sum_{\ell=1}^{n} \varphi_{\ell}(x) \varphi_{\ell}(y)
\]

and its kernel matrix \( \hat{K}_{X_n} = (K(x_k, x_m))_{k,m=1}^{n} \in \mathbb{R}^{n \times n} \). From (2.7) it is easy to see that

\[
\hat{K}_{X_n} = \Phi_{X_n} \Phi_{X_n}^T, \quad \text{where} \quad \Phi_{X_n} = \begin{bmatrix} \varphi_1(x_1) & \cdots & \varphi_n(x_1) \\ \vdots & \ddots & \vdots \\ \varphi_1(x_n) & \cdots & \varphi_n(x_n) \end{bmatrix}.
\]

The approximate Fekete points \( X^*_n = \{x^*_1, \ldots, x^*_n\} \subset \Omega \) are then any points such that

\[
X^*_n = \{x^*_1, \ldots, x^*_n\} \in \arg\max_{X_n = \{x_1, \ldots, x_n\} \subset \Omega} \det \hat{K}_{X_n} = \arg\max_{X_n = \{x_1, \ldots, x_n\} \subset \Omega} \det \Phi_{X_n}.
\]

Note that because \( \{\varphi_{\ell}\}_{\ell=1}^{n} \) are linearly independent, there exists \( X_n \subset \Omega \) such that \( \det \Phi_{X_n} > 0 \). As \( \Omega \) is compact and the continuity of \( K \) implies the continuity of the basis functions, there exist points \( X^*_n \) at which \( \det \Phi_{X_n} \) attains a maximal value.

Given a set \( X_n \) of \( n \) previously selected points, the popular \( P \)-greedy algorithm (De Marchi et al., 2005; Santin and Haasdonk, 2017) selects \( x_{n+1} \) such that

\[
x_{n+1} \in \arg\max_{x \in \Omega} P_{X_n}(x),
\]

which, using the block determinant identity and (2.4), can be written equivalently as

\[
x_{n+1} \in \arg\max_{x \in \Omega} \det \begin{bmatrix} K_{X_n} & K_{X_n}(x) \\ K(x, x) & K(x, x) \end{bmatrix} = \arg\max_{x \in \Omega} \det K_{X_n \cup \{x\}}.
\]

That is, the \( P \)-greedy points can be interpreted as greedily computed Fekete points. Because it is known (Santin and Haasdonk, 2017) that the interpolation error of the \( P \)-greedy algorithm decays fast (in some cases with an optimal rate), it is reasonable to expect that these rates are inherited by interpolation at the Fekete points, and by extension perhaps by interpolation at the approximate Fekete points.
3 Error estimates

This section provides upper bounds on the error of approximating \( f \in \mathcal{H}_K(\Omega) \) with the kernel interpolant \( s_f \) when the interpolation points are the approximate Fekete points from Section 2.2.

3.1 Interpolation with basis functions and Lebesgue constants

For any \( f: \Omega \to \mathbb{R} \) and any points \( X_n = \{x_1, \ldots, x_n\} \subset \Omega \) such that the matrix \( \Phi_{X_n} = (\varphi_m(x_k))_{k,m=1}^n \) is invertible there exists a unique interpolant \( s_f^\varphi \) such that

(i) \( s_f^\varphi(x_k) = f(x_k) \) for every \( k = 1, \ldots, n \);

(ii) \( s_f^\varphi \in \text{span}\{\varphi_\ell\}_{\ell=1}^n \).

From these requirements it follows that

\[
    s_f^\varphi = \sum_{k=1}^n c_k \varphi_k, \tag{3.1}
\]

where the coefficients are

\[
    \begin{bmatrix}
        c_1 \\
        \vdots \\
        c_n
    \end{bmatrix} = \begin{bmatrix}
        \varphi_1(x_1) & \cdots & \varphi_n(x_1) \\
        \vdots & \ddots & \vdots \\
        \varphi_1(x_n) & \cdots & \varphi_n(x_n)
    \end{bmatrix}^{-1} \begin{bmatrix}
        f(x_1) \\
        \vdots \\
        f(x_n)
    \end{bmatrix}.
\]

Alternatively, the interpolant can be written in the Lagrange form

\[
    s_f^\varphi = \sum_{k=1}^n f(x_k) u_k^\varphi, \tag{3.2}
\]

where \( u_k^\varphi \) are the Lagrange basis functions solved from

\[
    \begin{bmatrix}
        \varphi_1(x_1) & \cdots & \varphi_1(x_n) \\
        \vdots & \ddots & \vdots \\
        \varphi_n(x_1) & \cdots & \varphi_n(x_n)
    \end{bmatrix} \begin{bmatrix}
        u_1^\varphi(x) \\
        \vdots \\
        u_n^\varphi(x)
    \end{bmatrix} = \begin{bmatrix}
        \varphi_1(x) \\
        \vdots \\
        \varphi_n(x)
    \end{bmatrix}, \tag{3.3}
\]

for every \( x \in \Omega \). Define the Lebesgue constant

\[
    \Lambda_\varphi(X_n) = \sup_{x \in \Omega} \sum_{k=1}^n |u_k^\varphi(x)|. \tag{3.4}
\]

A standard argument yields a conservative upper bound on the Lebesgue constant at approximate Fekete points \((\text{Bos et al., 2010})\).

**Proposition 3.1.** If \( X_n^\varphi \) are any approximate Fekete points \((2.8)\), then the Lebesgue constant \((3.4)\) satisfies

\[
    \Lambda_\varphi(X_n^\varphi) \leq n. \tag{3.5}
\]

**Proof.** Cramer’s rule applied to \((3.3)\) gives

\[
    u_k^\varphi(x) = \frac{\det \Phi_{X_n}^k(x)}{\det \Phi_{X_n}}, \tag{3.6}
\]
Then for any \( \Theta \), Theorem 3.3.

Let \( \Phi \) be among all sets of \( n \) points within \( \Omega \) and \( \Phi^k \) with \( \Lambda_k(x) = \{x_1, \ldots, x_k, \ldots, x_n\} \),
\[
\det \Phi_{X_n}^k = \det \Phi_{X_n}^k(x) = \det \Phi_{X_n}^k(x).
\]

From (3.6) we thus get
\[
\Lambda_\phi(X_n^\ast) = \sup_{x \in \Omega} \sum_{k=1}^n |u_k^2(x)| = \sup_{x \in \Omega} \sum_{k=1}^n \left| \det \Phi_{X_n}^k(x) \right| \leq \sup_{x \in \Omega} \sum_{k=1}^n 1 = n.
\]

See De Marchi and Schaback (2010) for bounds on the Lebesgue constant \( \sup_{x \in \Omega} \sum_{k=1}^n |u_k(x)| \) for kernel interpolation when the RKHS is a Sobolev space.

### 3.2 Uniform error estimates

In this section we consider the uniform interpolation error \( \sup_{x \in \Omega} |f(x) - s_f(x)| \). Recall that since \( \{|\phi_\ell\}_{\ell=1}^\infty \) is an orthonormal basis of \( \mathcal{H}_K(\Omega) \), any \( f \in \mathcal{H}_K(\Omega) \) can be written as
\[
f = \sum_{\ell=1}^n f_\ell \phi_\ell \quad (3.7)
\]
for a square-summable sequence of coefficients \( f_\ell = \langle f, \phi_\ell \rangle_{\mathcal{H}_K(\Omega)} \in \mathbb{R} \). The RKHS norm of \( f \) in (3.7) is
\[
\|f\|_{\mathcal{H}_K(\Omega)}^2 = \sum_{\ell=1}^\infty f_\ell^2.
\]

That is, \( \mathcal{H}_K(\Omega) \) consists of functions having the form (3.7) such that their norm in (3.8) is finite.

**Lemma 3.2.** If \( f = \sum_{\ell=1}^\infty f_\ell \phi_\ell \in \mathcal{H}_K(\Omega) \), then
\[
\left| f(x) - \sum_{\ell=1}^n f_\ell \phi_\ell(x) \right| \leq \|f\|_{\mathcal{H}_K(\Omega)} \left( \sum_{\ell=n+1}^\infty \phi_\ell(x)^2 \right)^{1/2}
\]
for every \( x \in \Omega \).

**Proof.** The claim follows from the Cauchy–Schwarz inequality and (3.8):
\[
\left| f(x) - \sum_{\ell=1}^n f_\ell \phi_\ell(x) \right| = \left| \sum_{\ell=n+1}^\infty f_\ell \phi_\ell(x) \right| \leq \left( \sum_{\ell=n+1}^\infty f_\ell^2 \right)^{1/2} \left( \sum_{\ell=n+1}^\infty \phi_\ell(x)^2 \right)^{1/2}
\]
\[
\leq \|f\|_{\mathcal{H}_K(\Omega)} \left( \sum_{\ell=n+1}^\infty \phi_\ell(x)^2 \right)^{1/2}.
\]

**Theorem 3.3.** Let \( X_n = \{x_1, \ldots, x_n\} \subset \Omega \) be any points such that \( \Phi_{X_n} \) is invertible. Then for any \( f \in \mathcal{H}_K(\Omega) \),
\[
\sup_{x \in \Omega} |f(x) - s_f(x)| \leq 2 \|f\|_{\mathcal{H}_K(\Omega)} (1 + \Lambda_\phi(X_n)) \sup_{x \in \Omega} \left( \sum_{\ell=n+1}^\infty \phi_\ell(x)^2 \right)^{1/2}. \quad (3.9)
\]
To obtain a bound on $s\phi(x)$, where because $g \in \text{span}\{\phi\}_{k=1}^n$ and $s\phi(x)$ being the unique interpolant to $g$ in $	ext{span}\{\phi\}_{k=1}^n$ imply that $s\phi = g$. Finally, the Lagrange form (3.2) and Lemma 3.2 yield a bound on the third term:

$$\left| s\phi(x) - s\phi(x) \right| = \left| \sum_{k=1}^n [g(x_k) - f(x_k)] u\phi_k(x) \right|$$

$$\leq \| f \|_{H_K(\Omega)} \sum_{k=1}^n |u\phi_k(x)| \left( \sum_{\ell=n+1}^{\infty} \phi(\Omega)^2 \right) \frac{1}{2}$$

$$\leq \| f \|_{H_K(\Omega)} A\phi(\Omega) \sup_{x \in \Omega} \left( \sum_{\ell=n+1}^{\infty} \phi(\Omega)^2 \right) \frac{1}{2}.$$

Therefore,

$$\| f(x) - s\phi(x) \| \leq \| f \|_{H_K(\Omega)} (1 + A\phi(\Omega)) \sup_{x \in \Omega} \left( \sum_{\ell=n+1}^{\infty} \phi(\Omega)^2 \right) \frac{1}{2}.$$  

(3.10)

To obtain a bound on $\| f(x) - s\phi(x) \|$ observe that

$$\| f(x) - s\phi(x) \| \leq \| f(x) - s\phi(x) \| + \| s\phi(x) - s\phi(x) \|,$$

where, because $s\phi(x_k) = s\phi(x_k) = f(x_k)$ for $k = 1, \ldots, n$ and $\| s\phi \|_{H_K(\Omega)} \leq \| f \|_{H_K(\Omega)}$ by the norm-minimality property (2.1), both terms on the right-hand side obey the bound (3.10). The claim follows.

Proposition 3.4 immediately yields an error estimate for any approximate Fekete points.

**Corollary 3.4.** Suppose that $f \in H_K(\Omega)$ is interpolated at any approximate Fekete points (2.8). Then

$$\sup_{x \in \Omega} \| f(x) - s\phi(x) \| \leq 2 \| f \|_{H_K(\Omega)} (1 + n) \sup_{x \in \Omega} \left( \sum_{\ell=n+1}^{\infty} \phi(\Omega)^2 \right) \frac{1}{2}.$$  

(3.11)

Due to the presence of a supremum on the right-hand side of (3.9) and (3.11) it is difficult to make the bounds explicitly dependent on, for example, smoothness of the kernel as is usual in the error analysis of radial basis function interpolants (Wendland 2005, Chapter 11).

### 3.3 Improved error estimates in subspaces

It is known that the rate of convergence of kernel interpolation can be improved if the function being interpolated lives in a subset of the RKHS. The existing results in Schaback (1999, 2000, 2018) and Wendland (2005, Section 11.5) are particularly interesting when the kernel is finitely smooth\footnote{Wendland (2005, p. 192) goes as far as describing these results “almost pointless” for kernels, such as the Gaussian, that are associated with exponential rates of convergence.} Roughly speaking, in this case a typical
algebraic rate of convergence is “doubled” for sufficiently smooth elements of the RKHS. Specifically, let \( \mu \) be a Borel measure on \( \Omega \) that assigns positive measure to every open set and let \( \{ \psi_\ell \}_{\ell=1}^\infty \) and \( (\lambda_\ell)_{\ell=1}^\infty \) be the eigenfunctions and the positive decreasing eigenvalues of the integral operator \( Tf(x) = \int_\Omega K(x,y)f(y) \, dy \). By Mercer’s theorem (e.g. Sun [2005]),

\[
\mathcal{H}_K(\Omega) = \left\{ f \in L^2(\mu) : \| f \|_{\mathcal{H}_K(\Omega)}^2 = \sum_{\ell=1}^\infty \frac{\langle f, \psi_\ell \rangle^2_{L^2(\mu)}}{\lambda_\ell} < \infty \right\}.
\]

The standard improved error estimate states that for \( f \in \mathcal{H}_K(\Omega) \) such that \( f = Tf \) for some \( v \in L^2(\mu) \) the bound (2.3) is improved to

\[
|f(x) - s_f(x)| \leq \| v \|_{L^2(\mu)} \| P_{X_n}(x) \| P_{X_n} \|_{L^2(\mu)}.
\]

(3.12)

Because the range of \( T \) is

\[
T(L^2(\mu)) = \left\{ f \in L^2(\mu) : \| f \|_{\mathcal{H}_K(\Omega)}^2 = \sum_{\ell=1}^\infty \frac{\langle f, \psi_\ell \rangle^2_{L^2(\mu)}}{\lambda_\ell} < \infty \right\} \subset \mathcal{H}_K(\Omega),
\]

the collection of functions for which (3.12) holds is a subset of the RKHS. Theorem 3.5 below is significantly more flexible than this result and does not require that the Mercer expansion be used.

Let \( (\alpha_\ell)_{\ell=1}^\infty \) be a positive, increasing, and divergent sequence and define the subspace

\[
\mathcal{H}_K^{\alpha}(\Omega) = \left\{ f = \sum_{\ell=1}^\infty f_\ell \varphi_\ell : \| f \|_{\mathcal{H}_K^{\alpha}(\Omega)}^2 = \sum_{\ell=1}^\infty \alpha_\ell^2 f_\ell^2 < \infty \right\} \subset \mathcal{H}_K(\Omega).
\]

It is easy to verify that \( \mathcal{H}_K^{\alpha}(\Omega) \) is an RKHS and that its reproducing kernel is

\[
K^{\alpha}(x,y) = \sum_{\ell=1}^\infty \frac{1}{\alpha_\ell^2} \varphi_\ell(x) \varphi_\ell(y).
\]

**Theorem 3.5.** If \( f \in \mathcal{H}_K^{\alpha}(\Omega) \) is interpolated at any approximate Fekete points (2.8), then

\[
\sup_{x \in \Omega} |f(x) - s_f(x)| \leq 2 \| f \|_{\mathcal{H}_K^{\alpha}(\Omega)} (1 + n) \alpha_n^{-1} \sup_{x \in \Omega} \left( \sum_{\ell=n+1}^\infty \varphi_\ell(x)^2 \right)^{1/2}.
\]

**Proof.** When \( f \in \mathcal{H}_K^{\alpha}(\Omega) \), we replace the estimate of Lemma 3.2 with the following estimate:

\[
\left| f(x) - \sum_{\ell=1}^n f_\ell \varphi_\ell(x) \right|^2 = \left| \sum_{\ell=n+1}^\infty \alpha_\ell f_\ell \varphi_\ell(x) \right|^2 \leq \left( \sum_{\ell=n+1}^\infty \alpha_\ell^2 f_\ell^2 \right) \left( \sum_{\ell=n+1}^\infty \alpha_\ell^{-2} \varphi_\ell(x)^2 \right) \leq \| f \|_{\mathcal{H}_K^{\alpha}(\Omega)}^2 \alpha_n^{-2} \sum_{\ell=n+1}^\infty \varphi_\ell(x)^2.
\]

The proof of Theorem 3.5 then yields claimed uniform bound. \( \square \)
4 Gaussian kernel

For the most of this section we set $d = 1$ and consider the Gaussian kernel

$$K(x, y) = \exp \left(-\varepsilon^2 (x - y)^2\right)$$

(4.1)

with a scale parameter $\varepsilon > 0$. This kernel has the orthonormal expansion\(^2\)

$$K(x, y) = \sum_{\ell=0}^{\infty} \left(\sqrt{\frac{2\ell!}{\varepsilon^2}} x^\ell \exp(-\varepsilon^2 x^2)\right) \left(\sqrt{\frac{2\ell!}{\varepsilon^2}} y^\ell \exp(-\varepsilon^2 y^2)\right) =: \sum_{\ell=0}^{\infty} \varphi_\ell(x) \varphi_\ell(y)$$

(4.2)

and its RKHS is thus

$$H_K(\Omega) = \left\{ f(x) = \sum_{\ell=0}^{\infty} f_\ell \sqrt{\frac{2\ell!}{\varepsilon^2}} x^\ell \exp(-\varepsilon^2 x^2) : \sum_{\ell=0}^{\infty} f_\ell^2 < \infty \right\}.$$  

(4.3)

These results and other properties of the Gaussian kernel and its RKHS are studied in more detail in Steinwart et al. (2006) and Minh (2010). In Section 4.1 we show that, owing to the special structure of the above basis functions, the approximate Fekete points for the one-dimensional Gaussian kernel are solved from a convex optimisation problem. The rest of this section is then devoted to specialising the general error estimates of the previous section.

Note that most prior work, such as Tanaka (2019) and Belhadji et al. (2019), uses a well-known Mercer expansion of the Gaussian kernel instead of (4.2). This expansion is

$$K(x, y) = \sum_{\ell=0}^{\infty} \lambda_\ell \sigma_\ell \psi_\ell(x) \psi_\ell(y),$$

where the eigenfunctions are orthonormal with respect to the Gaussian measure with variance $\sigma^2$:

$$\frac{1}{\sqrt{2\pi}\sigma^2} \int_{\mathbb{R}} \psi^*_\ell(x) \psi_\ell(x) \exp \left(-\frac{x^2}{2\sigma^2}\right) \, dx = \delta_{\ell k}.$$

The eigenfunctions and values are (Fasshauer and McCourt, 2012)

$$\psi_\ell(x) = \sqrt{\beta \ell!} \exp(-\alpha^2 x^2) H_\ell(\sqrt{2\alpha \sigma x}) \quad \text{and} \quad \lambda_\ell = \sqrt{\frac{\alpha^2}{\alpha^2 + \delta^2 + \varepsilon^2}} \left(\frac{\varepsilon^2}{\alpha^2 + \delta^2 + \varepsilon^2}\right)^\ell,$$

where $H_\ell$ is the $\ell$th probabilists’ Hermite polynomial and the constants are

$$\alpha = \frac{1}{\sqrt{2\sigma}}, \quad \beta = (1 + 8\varepsilon^2 \sigma^2)^{1/4} \quad \text{and} \quad \delta^2 = \frac{1}{4\sigma^2}(\beta^2 - 1).$$

The expansion (4.2) used in this article is evidently much simpler to work with.

4.1 Approximate Fekete points via convex optimisation

Let

$$\hat{K}(x, y) = \sum_{\ell=0}^{n-1} \left(\sqrt{\frac{2\ell!}{\varepsilon^2}} x^\ell \exp(-\varepsilon^2 x^2)\right) \left(\sqrt{\frac{2\ell!}{\varepsilon^2}} y^\ell \exp(-\varepsilon^2 y^2)\right) = \sum_{\ell=0}^{n-1} \varphi_\ell(x) \varphi_\ell(y)$$

\(^2\)Observe that in this section we begin indexing of the expansion from zero to simplify notation.
be the truncation of the Gaussian kernel \((4.1)\) and \(\hat{K}_{X_n} = (\hat{K}(x_i, x_m))_{k,m=1}^n \in \mathbb{R}^{n \times n}\). Define the matrices

\[
\Phi_{X_n} = \begin{bmatrix}
\varphi_0(x_1) & \cdots & \varphi_{n-1}(x_1) \\
\vdots & \ddots & \vdots \\
\varphi_0(x_n) & \cdots & \varphi_{n-1}(x_n)
\end{bmatrix} \quad \text{and} \quad \mathcal{V}_{X_n} = \begin{bmatrix}
1 & x_1 & \cdots & x_1^{n-1} \\
\vdots & \vdots & \ddots & \vdots \\
1 & x_n & \cdots & x_n^{n-1}
\end{bmatrix},
\]

the latter of which is the classical Vandermonde matrix. Because \(\hat{K}_{X_n} = \Phi_{X_n} \Phi_{X_n}^T\) and the \(k\)th row of \(\Phi_{X_n}\) is that of \(\mathcal{V}_{X_n}\) multiplied by \((2^{k-1} \varepsilon^2 (k-1)/(k-1)!)^{1/2} \exp(-\varepsilon^2 x_k^2)\), we have

\[
(\det \hat{K}_{X_n})^{1/2} = |\det \Phi_{X_n}| = \left( \prod_{\ell=0}^{n-1} \frac{2^{\ell} \varepsilon^2}{\ell!} \right)^{1/2} \exp \left( -\varepsilon^2 \sum_{k=1}^{n} x_k^2 \right) |\det \mathcal{V}_{X_n}| = \left( \prod_{\ell=0}^{n-1} \frac{2^{\ell} \varepsilon^2}{\ell!} \right)^{1/2} \exp \left( -\varepsilon^2 \sum_{k=1}^{n} x_k^2 \right) \prod_{1 \leq i < j \leq n} (x_i - x_j),
\]

where the last equation uses the standard explicit expression for the Vandermonde determinant. This expression verifies that \(\hat{K}_{X_n}\) and \(\Phi_{X_n}\) are invertible whenever the points are distinct. Define

\[
W(x_1, \ldots, x_n) = \exp \left( -\varepsilon^2 \sum_{k=1}^{n} x_k^2 \right) \prod_{1 \leq i < j \leq n} (x_i - x_j). \quad (4.4)
\]

The approximate Fekete points \((2.8)\) for the Gaussian kernel are thus seen to be

\[
X_n^* = \left\{ x_1^*, \ldots, x_n^* \right\} \in \arg \max_{X_n=\{x_1, \ldots, x_n\} \subset \Omega} \det \hat{K}_X = \arg \max_{X_n=\{x_1, \ldots, x_n\} \subset \Omega} W(x_1, \ldots, x_n). \quad (4.5)
\]

Maximisation of \(W(x_1, \ldots, x_n)\) is equivalent to minimisation of the energy

\[
I(x_1, \ldots, x_n) = -\log W(x_1, \ldots, x_n) = \varepsilon^2 \sum_{k=1}^{n} x_k^2 + \sum_{1 \leq i < j \leq n} \log \frac{1}{|x_i - x_j|}
= \sum_{k=1}^{n} Q_\varepsilon(x_k) + \sum_{1 \leq i < j \leq n} N(x_i - x_j),
\]

where \(Q_\varepsilon(x) = \varepsilon^2 x\) and \(N(x) = 1/\log |x|\). To ensure that \(I\) is well-defined and non-unique in arising from ordering of the points is eliminated, define the simplex

\[
\mathcal{R}_n = \left\{ (x_1, \ldots, x_n) \in \Omega^n : x_1 < x_2 < \cdots < x_{n-1} < x_n \right\} \subset \Omega^n
\]
and consider \(I\) as a function on \(\mathcal{R}_n\). Adaptation of the proof of Theorem 3.3 in \cite{Tanaka2019} shows that the objective function \(I\) is convex and that there exists a unique minimiser \(X_n^* \in \mathcal{R}_n\).

**Proposition 4.1.** If \(\Omega \subset \mathbb{R}\) is a closed interval, then the energy function \(I : \mathcal{R}_n \to \mathbb{R}\) is convex and has a unique minimiser.

**Proof.** The Hessian matrix \(\nabla^2 I\) of \(I\) is

\[
(\nabla^2 I)_{ij} = \frac{\partial^2 I}{\partial x_i \partial x_j} = \begin{cases}
Q_\varepsilon''(x_i) + \sum_{k \neq i} N''(x_i - x_k) & (i = j), \\
-N''(x_i - x_j) & (i \neq j).
\end{cases}
\]
Because both
\[ N(x) = \log \frac{1}{|x|} \quad \text{and} \quad Q_\varepsilon(x) = \varepsilon^2 x^2 \]
are strictly convex on \( \mathbb{R} \setminus \{0\} \) and \( \mathbb{R} \), respectively, we have \( N'' > 0 \) and \( Q'' > 0 \). Therefore all diagonal elements of \( \nabla^2 I \) are always positive. Moreover,
\[
\sum_{k \neq i}|-N''(x_i - x_k)| = \sum_{k \neq i}N''(x_i - x_k) < \sum_{k \neq i}N''(x_i - x_k) + Q''_\varepsilon(x_i), \tag{4.6}
\]
which verifies that the Hessian is diagonally dominant and hence positive-definite. That is, the energy function \( I \) is convex on \( \mathcal{R}_n \).

To verify that there is a unique minimiser in the non-closed set \( \mathcal{R}_n \), consider the function \( J(\mathcal{X}_n) = \exp(-I(\mathcal{X}_n)) \) which is continuous on the closure of \( \mathcal{R}_n \) if we set \( J(\mathcal{X}_n) = 0 \) for every \( \mathcal{X}_n = \{x_1, \ldots, x_n\} \in \Omega^n \) such that \( x_i = x_{i+1} \) for some \( i \). Because it is positive on \( \mathcal{R}_n \), any maximiser of \( J \) is in \( \mathcal{R}_n \). As a maximiser of \( J \) is a minimiser of \( I \) and \( I \) is convex it follows that \( I \) must have a unique minimiser in \( \mathcal{R}_n \).

**Remark 4.2.** If we set \( \varepsilon = 0 \), the above optimisation problem becomes that of finding the Fekete points for polynomial interpolation. However, in this case the objective function \( I \) is no longer convex because \( Q''_\varepsilon(x_i) = 0 \) in (4.6). Our optimisation problem can be thus viewed as a regularised version of the standard Fekete problem. Based on this and the well-known convergence of kernel interpolants to polynomial interpolants at the so-called flat limit (Schaback, 2005; Lee et al., 2007; Karvonen and Särkkä, 2019) it may be expected that \( \mathcal{X}_n \) converge to the polynomial Fekete points as \( \varepsilon \to 0 \). We do not attempt to prove this.

### 4.2 Error estimates

In this section we denote \( c_\Omega = \sup_{x \in \Omega} |x| < \infty \).

**Lemma 4.3.** Consider the basis functions \( \varphi_\ell \) \( \ell \geq 4 \) and assume that \( n \geq 2\varepsilon^2 c_\Omega^2 \). Then
\[
\sup_{x \in \Omega} \left( \sum_{\ell = n}^{\infty} \varphi_\ell(x)^2 \right)^{1/2} \leq \left( \frac{\sqrt{2}\varepsilon c_\Omega}{\sqrt{n!}} \right)^n.
\]

**Proof.** By differentiation it is easy to see that \( \varphi_\ell(x)^2 \) attains its maximal value on \( \mathbb{R} \) at \( x_0 = (\ell/(2\varepsilon^2))^{1/2} \) and that \( \varphi_\ell^2 \) is decreasing on \( \left[-(\ell/(2\varepsilon^2))^{1/2}, 0\right] \) and increasing on \( [0,(\ell/(2\varepsilon^2))^{1/2}] \). It follows that
\[
\sup_{x \in \Omega} \varphi_\ell(x)^2 = \varphi_\ell(c_\Omega)^2 = \frac{2\ell \varepsilon^{2\ell}}{\ell!} c_\Omega^{2\ell} \exp(-2\varepsilon^2 c_\Omega^2)
\]
for every \( \ell \geq n \) if \( n \geq 2\varepsilon^2 c_\Omega^2 \). By Taylor’s theorem there is \( \xi \in [0,2\varepsilon^2 c_\Omega^2] \) such that
\[
\sup_{x \in \Omega} \sum_{\ell = n}^{\infty} \varphi_\ell(x)^2 \leq \exp(-2\varepsilon^2 c_\Omega^2) \sum_{\ell = n}^{\infty} \frac{(2\varepsilon^2 c_\Omega^2)^\ell}{\ell!} = \exp(-2\varepsilon^2 c_\Omega^2) \frac{\exp(\xi)}{n!} (2\varepsilon^2 c_\Omega^2)^n \leq \exp(-2\varepsilon^2 c_\Omega^2) \frac{\exp(2\varepsilon^2 c_\Omega^2)}{n!} (2\varepsilon^2 c_\Omega^2)^n = \frac{(2\varepsilon^2 c_\Omega^2)^n}{n!}.
\]
This proves the claim. \( \square \)
Using the estimate of Lemma 4.3 in Corollary 3.4 yields an explicit error estimate for interpolation with the Gaussian kernel.

**Theorem 4.4.** Consider the Gaussian kernel with the orthonormal expansion (4.2) and suppose that \( \Omega \subset \mathbb{R} \) is a closed interval. If \( f \in H_K(\Omega) \) is interpolated at the unique approximate Fekete points \( X_n^* \) defined in (4.5) and \( n \geq 2e^2c_\Omega^2 \), then

\[
\sup_{x \in \Omega} |f(x) - s_f(x)| \leq C_1 \| f \|_{H_K(\Omega)} n^{3/4} \exp \left( -n \left( \frac{1}{2} \log n - \log C_2 \right) \right),
\]

where \( C_1 = (128/\pi)^{1/4} \approx 2.53 \) and \( C_2 = \sqrt{2e} \epsilon \epsilon_{\Omega} \).

**Proof.** The claim follows from Corollary 3.4, Lemma 4.3 and the lower bound \( n! \geq \sqrt{2\pi} n^{n+1/2} e^{-n} \) of Stirling’s approximation (Robbins 1955):

\[
\sup_{x \in \Omega} |f(x) - s_f(x)| \leq 2 \| f \|_{H_K(\Omega)} \left(1 + n\right) \left( \sqrt{2\epsilon \epsilon_{\Omega}} \right)^n \frac{n}{\sqrt{n!}}
\leq 4 \| f \|_{H_K(\Omega)} \left( \sqrt{2\epsilon \epsilon_{\Omega}} \right)^n \frac{e^{n/2} n}{n^{n/2+1/4}}
\leq \left( \frac{128}{\pi} \right)^{1/4} \| f \|_{H_K(\Omega)} \left( \sqrt{2\epsilon \epsilon_{\Omega}} \right)^n \frac{e^{n/2} n}{n^{n/2+1/4}}
\leq \left( \frac{128}{\pi} \right)^{1/4} n^{3/4} \| f \|_{H_K(\Omega)} \left( \sqrt{2\epsilon \epsilon_{\Omega}} \right)^n \frac{e^{n/2} n}{n^{n/2+1/4}}
= C_1 \| f \|_{H_K(\Omega)} n^{3/4} \exp \left( -n \left( \frac{1}{2} \log n - \log C_2 \right) \right).
\]

\( \square \)

Also Theorem 3.5 can be specialised, and in some cases the kernel of the subspace \( H_K^0(\Omega) \) has an explicit form. For instance, set \( \alpha_\ell = \sqrt{\ell!} 2^{\ell/2} \). Then

\[
K^\alpha(x, y) = \exp \left( -\varepsilon^2(x^2 + y^2) \right) \sum_{\ell=0}^{\infty} \frac{1}{\alpha_\ell^2} \frac{2^{\ell/2} (xy)^\ell}{\ell!} = \exp \left( -\varepsilon^2(x^2 + y^2) \right) \sum_{\ell=0}^{\infty} \frac{1}{(\ell!)^2} (xy)^\ell,
\]

which can be written in terms of \( I_0 \), the modified Bessel function of the first kind:

\[
K^\alpha(x, y) = \exp \left( -\varepsilon^2(x^2 + y^2) \right) I_0(2\sqrt{xy}).
\]

**Theorem 4.5.** Consider the Gaussian kernel with the orthonormal expansion (4.2) and suppose that \( \Omega \subset \mathbb{R} \) is a closed interval. If \( f \in H_K^0(\Omega) \) is interpolated at the unique approximate Fekete points \( X_n^* \) defined in (4.5) and \( n \geq 2e^2c_\Omega^2 \), then

\[
\sup_{x \in \Omega} |f(x) - s_f(x)| \leq C_1 \| f \|_{H_K^0(\Omega)} n^{3/4} \alpha_n^{-1} \exp \left( -n \left( \frac{1}{2} \log n - \log C_2 \right) \right),
\]

where \( C_1 = (128/\pi)^{1/4} \approx 2.53 \) and \( C_2 = \sqrt{2e} \epsilon \epsilon_{\Omega} \).

If \( \Omega = [a, b] \subset \mathbb{R} \) is a closed interval, the standard fill-distance based bound (Rieger and Zwicknagl 2010 Theorem 6.1) for interpolation error is

\[
\sup_{x \in \Omega} |f(x) - s_f(x)| \leq 2 \| f \|_{H_K(\Omega)} \exp \left( C \log(h_{X_n, \Omega}) / h_{X_n, \Omega} \right)
\]

when the fill-distance

\[
h_{X_n, \Omega} = \sup_{x \in \Omega} \min_{x_k \in X_n} |x - x_k|
\]
is sufficiently small. The constant satisfies $C \leq \frac{1}{8} \min\{(b - a)/6, 1\}$ \footnote{This is the constant $C$ in Theorem 6.1 of \cite{RZ10}. To derive the claimed bound, observe that this constant is given as $C = \epsilon B/4$ for $B \leq \min\{(b - a)/6, 1\}$ in their proof of Theorem 4.5. On p. 120 they show that $\epsilon = 1/2$ if the kernel is Gaussian.} For the equispaced points

$$X_n = \left\{ a, a + \frac{b - a}{n}, \ldots, b - \frac{b - a}{n}, b \right\},$$

which have the minimal fill-distance $h_{X_n, \Omega} = (b - a)/n$, the bound \eqref{eq:4.8} becomes

$$\sup_{x \in \Omega} |f(x) - s_{\mathcal{F}}(x)| \leq 2\|f\|_{H_K(\Omega)} \exp\left( -C \frac{b - a}{b - a} n \left( \log n - \log(b - a) \right) \right),$$

where $C/(b - a) \leq 1/8$. Our bound \eqref{eq:4.7} for points $X_n^*$, being essentially of order $\exp(-1/2 n \log n)$, is thus tighter when $n$ is sufficiently large. However, a significant advantage of bounds of the type \eqref{eq:4.8} is that they apply to nested point sets (i.e., $X_n \subset X_{n+1}$ for every $n \geq 1$). It cannot be expected that the approximate Fekete point sets are nested. Further error estimates for Chebyshev-type nodes that cluster near the boundary are provided in \cite{RZ14}.

\textbf{Remark 4.6.} It is easy to see that in the Gaussian case the Lagrange basis functions in \eqref{eq:3.3} can be expressed in terms of the classical polynomial Lagrange functions:

$$u^\epsilon_k(x) = \exp(\epsilon^2 x_k^2) \exp(-\epsilon^2 x^2) l_k(x), \quad (4.9)$$

where

$$l_k(x) = \prod_{\ell \neq k} \frac{x - x_\ell}{x_k - x_\ell}.$$

Let $\Lambda_{\text{pol}}(\mathcal{X}_n) = \sup_{x \in \Omega} \sum_{k=1}^n |l_k(x)|$ be the Lebesgue constant for polynomial interpolation. It follows easily from \eqref{eq:4.9} and the boundedness of $\Omega$ that there exist $C_1, C_2 > 0$ such that

$$C_1 \Lambda_{\text{pol}}(\mathcal{X}_n) \leq \Lambda_{\mathcal{F}}(\mathcal{X}_n) \leq C_2 \Lambda_{\text{pol}}(\mathcal{X}_n)$$

for any $\mathcal{X}_n \subset \Omega$. This implies that in Theorem \ref{thm:3.3} the coefficient $1 + \Lambda_{\mathcal{F}}(\mathcal{X}_n)$ can be replaced with $1 + C_2 \Lambda_{\text{pol}}(\mathcal{X}_n)$, which means that convergence results are available if polynomial Lebesgue constants can be controlled (e.g., if $\mathcal{X}_n$ are the Chebyshev points).

\subsection*{4.3 Tensor product algorithms}

In this section we provide error estimates for interpolation with anisotropic Gaussian kernels in higher dimensions when the evaluation points are constructed as tensor products of the approximate Fekete points \eqref{eq:4.5}. Besides \cite{B10} \footnote{Besides\cite{B10} there does not appear to be much work on error estimates for general anisotropic kernels. \cite{Faietal12} and \cite{SW18} analyse the $L^2$-error of general linear algorithms for functions in the RKHS of an anisotropic Gaussian.} there does not appear to be much work on error estimates for general anisotropic kernels. \cite{Faietal12} and \cite{SW18} analyse the $L^2$-error of general linear algorithms for functions in the RKHS of an anisotropic Gaussian.

Let

$$\Omega = \Omega_1 \times \cdots \times \Omega_d \subset \mathbb{R}^d \quad \text{for} \quad \Omega_i = [a_i, b_i] \neq \emptyset$$

be a hyper-rectangle and

$$K(x, y) = \exp\left( -\sum_{i=1}^d \epsilon_i^2 (x_i - y_i)^2 \right) = \prod_{i=1}^d K_i(x_i, y_i) \quad (4.11)$$
with \( \varepsilon_i > 0 \) an anisotropic Gaussian kernel on \( \Omega \). Because the kernel (4.11) is a product of one-dimensional Gaussian kernels, its RKHS is the tensor product of their RKHSs:

\[
\mathcal{H}_K(\Omega) = \mathcal{H}_{K_1}(\Omega_1) \otimes \cdots \otimes \mathcal{H}_{K_d}(\Omega_d)
\]

where

\[
\mathcal{H}_{K_i}(\Omega_i) = \{ f(x) = f_1(x_1) \times \cdots \times f_d(x_d) : f_1 \in \mathcal{H}_{K_1}(\Omega_1), \ldots, f_d \in \mathcal{H}_{K_d}(\Omega_d) \}
\]

and

\[
\|f\|_{\mathcal{H}_K(\Omega)} = \|f_1\|_{\mathcal{H}_{K_1}(\Omega_1)} \times \cdots \times \|f_d\|_{\mathcal{H}_{K_d}(\Omega_d)} .
\]

Let \( n_1, \ldots, n_d \in \mathbb{N} \) and denote \( N = n_1 \times \cdots \times n_d \). We take the point set to be a tensor product of approximate Fekete point sets (4.15) for Gaussian kernels \( K_i \) on \( \Omega_i \):

\[
\mathcal{X}_N^* = \mathcal{X}_{n_1}^* \times \cdots \times \mathcal{X}_{n_d}^* \subset \Omega,
\]

where \( \mathcal{X}_{n_i}^* \subset \Omega_i \) stands for the set of \( n_i \) approximate Fekete points for kernel \( K_i \) on \( \Omega_i \). Due to the tensor product structure of the point set and the RKHS, the kernel interpolant \( s_f \) to \( f \in \mathcal{H}_K(\Omega) \) can be written as

\[
s_f(x) = s_{1,f_1}(x_1) \times \cdots \times s_{d,f_d}(x_d),
\]

where \( s_{i,f_i} \) is the kernel interpolant, based on \( K_i \), of \( f_i \in \mathcal{H}_{K_i}(\Omega_i) \) at the points \( \mathcal{X}_{n_i}^* \).

**Theorem 4.7.** Consider the multi-dimensional Gaussian kernel (4.11) and suppose that \( \Omega \subset \mathbb{R}^d \) is a hyper-rectangle of the form (4.10). If \( f \in \mathcal{H}_K(\Omega) \) is interpolated at the tensor product points \( \mathcal{X}_N^* \) defined in (4.12) and \( n_i \geq 2 \varepsilon_i c_{\Omega_i} \) for every \( i = 1, \ldots, d \), then

\[
\sup_{x \in \Omega} |f(x) - s_f(x)| \leq C_1 \|f\|_{\mathcal{H}_K(\Omega)} \sum_{i=1}^d n_i^{3/4} \exp \left( -n_i \left( \frac{1}{2} \log n_i - \log C_{i,2} \right) \right),
\]

where \( C_1 = (128/\pi)^{1/4} \approx 2.53 \) and \( C_{i,2} = \sqrt{2} \varepsilon_i c_{\Omega_i} \).

**Proof.** The argument is standard. Let \( f(x) = f_1(x_1) \times \cdots \times f_d(x_d) \) for \( f_i \in \mathcal{H}_{K_i}(\Omega_i) \) and denote \( f_{i,j}(x) = f_i(x_i) \times \cdots \times f_j(x_j) \) for \( 1 \leq i \leq j \leq d \). Then

\[
f(x) - s_f(x) = f_{2,d}(x)f_1(x_1) - s_{f_{2,d}}(x)s_{f_1}(x_1) = f_{2,d}(x)[f_1(x_1) - s_{1,f_1}(x_1)] + [f_{2,d}(x) - s_{f_{2,d}}(x)]s_{1,f_1}(x_1)
\]

\[
= \sum_{i=1}^d [f_i(x_i) - s_{i,f_i}(x_i)] f_{i+1:d}(x) \prod_{j=1}^{i-1} s_{j,f_j}(x_j),
\]

where the notational convention \( f_{d+1:d}(x) = 1 \) is used. By the reproducing property and the minimum-norm property (2.1),

\[
|f_i(x_i)| = \|f_i, K_i(\cdot, x_i)\|_{\mathcal{H}_{K_i}(\Omega_i)} \leq \|f_i\|_{\mathcal{H}_{K_i}(\Omega_i)} \quad \text{and} \quad |s_{i,f_i}(x_i)| \leq \|s_i\|_{\mathcal{H}_{K_i}(\Omega_i)}
\]

for any \( i \leq d \) and \( x_i \in \Omega_i \). Inserting these estimates and the bound (4.7) into (4.14) yields

\[
|f(x) - s_f(x)| \leq \sum_{i=1}^d |f_i(x_i) - s_{i,f_i}(x_i)| \prod_{j \neq i} \|f_j\|_{\mathcal{H}_{K_j}(\Omega_j)}
\]

\[
\leq C_1 \|f\|_{\mathcal{H}_K(\Omega)} \sum_{i=1}^d n_i^{3/4} \exp \left( -n_i \left( \frac{1}{2} \log n_i - \log C_{i,2} \right) \right).
\]

\(\square\)
In particular, if $n_1 = \cdots = n_d = n$ (so that $N = n^d$) and all $\Omega_i$ and $\epsilon_i$ are equal, the bound of Theorem 1.7 becomes

$$\sup_{x \in \Omega} |f(x) - s_f(x)| \leq C_1 \|f\|_{H^k(\Omega)} d N^{3/(4d)} \exp \left( - N^{1/d} \left( \frac{1}{2d} \log N - \log C_2 \right) \right).$$

It would be straightforward to generalise Theorem 4.5 to the tensor product setting.

![Figure 1: Maximal errors in interpolation of the function $f_m$ in $[0,1]$ on $\Omega = [-1,1]$ for three sequences of point sets. Left column: Kernel interpolants $[2,2]$ based on the Gaussian kernel $[4,1]$. Right column: Auxiliary interpolants $[3,1]$ constructed out of the basis functions $[4,2]$. Also displayed is the theoretical rate $[5,2]$ from Theorem 4.4.](image)
5 Numerical example

We use the kernel interpolant (2.2) based on the Gaussian kernel (4.1) with $\varepsilon = 1/\sqrt{2}$ to approximate the functions

$$f_m(x) = x^m \exp \left( x - \frac{x^2}{2} \right)$$

(5.1)

for $m \in \{5, 10, 15\}$ on $\Omega = [-1, 1]$. Using (4.3) and the expansion

$$f_m(x) = x^m \sum_{\ell=0}^{\infty} \frac{1}{\ell!} x^\ell \exp \left( -\frac{x^2}{2} \right)$$

we compute that

$$\|f_m\|_{\mathcal{H}_K(\Omega)}^2 = \sum_{\ell=m}^{\infty} \frac{\ell!}{((\ell - m)!)^2} = \sum_{\ell=0}^{\infty} \frac{(\ell + m)!}{(\ell!)^2} = \sum_{\ell=0}^{\infty} (\ell + 1) \cdots (\ell + m) \frac{1}{\ell!} < \infty,$$

which verifies that $f_m \in \mathcal{H}_K(\Omega)$ for every $m \in \mathbb{N}$.

The results are displayed in Figure 1 for both the kernel interpolant $s_{f_m}$ and the interpolant $s^\phi_{f_m}$ based on the basis functions $\{\phi_\ell\}_{\ell=0}^{n-1}$. Three different sequences of point sets were used for $n = 1, \ldots, 20$: (i) the approximate Fekete points (4.5), (ii) equispaced points, and (iii) Chebyshev points. We also plot the theoretical rate

$$n^{3/4} \exp \left( -\frac{1}{2} n (\log n - 1) \right)$$

(5.2)

of Theorem 4.4 (note that $C_2 = \sqrt{e}$ because $c_\Omega = \sup_{x \in [-1, 1]} |x| = 1$ and $\varepsilon = 1/\sqrt{2}$). The maximal errors $\sup_{x \in [-1, 1]} |f_m(x) - s_{f_m}(x)|$ and $\sup_{x \in [-1, 1]} |f_m(x) - s^\phi_{f_m}(x)|$ were substituted with $\max_{1 \leq i \leq 1000} |f_m(x_i) - s_{f_m}(x_i)|$ and $\max_{1 \leq i \leq 1000} |f_m(x_i) - s^\phi_{f_m}(x_i)|$, where $x_i$ are equispaced points dividing $[-1, 1]$ into 999 segments. The errors appear to behave roughly as predicted by the theoretical results. Note that construction of the kernel interpolants suffers from numerical instability when $n > 11$.

Acknowledgements

T. Karvonen was supported by the Aalto ELEC Doctoral School. This research was partially carried out while he was visiting the University of Tokyo, funded by the Finnish Foundation for Technology Promotion and Oskar Öflunds Stiftelse. S. Särkkä was supported by the Academy of Finland. K. Tanaka was supported by the grant-in-aid of Japan Society of the Promotion of Science with KAKENHI Grant Number 17K14241.

References

Arcangéli, R., de Silanes, M. C. L., and Torrens, J. J. (2007). An extension of a bound for functions in Sobolev spaces, with applications to $(m, s)$-spline interpolation and smoothing. Numerische Mathematik, 107(2):181–211.

Beatson, R. (2010). Error bounds for anisotropic RBF interpolation. Journal of Approximation Theory, 162(3):512–527.

Belhadji, A., Bardenet, R., and Chainais, P. (2019). Kernel quadrature with DPPs. In Advances in Neural Information Processing Systems, volume 32, pages 12907–12917.
Bos, L. and De Marchi, S. (2011). On optimal points for interpolation by univariate exponential functions. *Dolomites Research Notes on Approximation*, 4:8–12.

Bos, L., De Marchi, S., Sommariva, A., and Vianello, M. (2010). Computing multivariate Fekete and Leja points by numerical linear algebra. *SIAM Journal on Numerical Analysis*, 48(5):1984–1999.

Bos, L. P. and Maier, U. (2002). On the asymptotics of Fekete-type points for univariate radial basis interpolation. *Journal of Approximation Theory*, 119(2):252–270.

Briani, M., Sommariva, A., and Vianello, M. (2012). Computing Fekete and Lebesgue points: Simplex, square, disk. *Journal of Computational and Applied Mathematics*, 236(9):2477–2486.

De Marchi, S. and Schaback, R. (2010). Stability of kernel-based interpolation. *Advances in Computational Mathematics*, 32:155–161.

De Marchi, S., Schaback, R., and Wendland, H. (2005). Near-optimal data-independent point locations for radial basis function interpolation. *Advances in Computational Mathematics*, 23(3):317–330.

Fasshauer, G., Hickernell, F., and Wóźniakowski, H. (2012). On dimension-independent rates of convergence for function approximation with Gaussian kernels. *SIAM Journal on Numerical Analysis*, 50(1):247–271.

Fasshauer, G. and McCourt, M. (2015). *Kernel-based Approximation Methods Using MATLAB*. Number 19 in Interdisciplinary Mathematical Sciences. World Scientific Publishing.

Fasshauer, G. E. (2007). *Meshfree Approximation Methods with MATLAB*. Number 6 in Interdisciplinary Mathematical Sciences. World Scientific Publishing.

Fasshauer, G. E. and McCourt, M. J. (2012). Stable evaluation of Gaussian radial basis function interpolants. *SIAM Journal on Scientific Computing*, 34(2):A737–A762.

Gautier, G., Bardenet, R., and Valko, M. (2019). On two ways to use determinantal point processes for Monte Carlo integration. In *Advances in Neural Information Processing Systems*, volume 32, pages 7768–7777.

Karvonen, T. and Särkkä, S. (2019). Worst-case optimal approximation with increasingly flat Gaussian kernels. *arXiv:1906.02096v1*.

Lee, Y. J., Yoon, G. J., and Yoon, J. (2007). Convergence of increasingly flat radial basis interpolants to polynomial interpolants. *SIAM Journal on Mathematical Analysis*, 39(2):537–553.

Minh, H. Q. (2010). Some properties of Gaussian reproducing kernel Hilbert spaces and their implications for function approximation and learning theory. *Constructive Approximation*, 32(2):307–338.

Müller, S. (2009). *Komplexität und Stabilität von kernbasierten Rekonstruktionsmethoden*. PhD thesis, Institut für Numerische und Angewandte Mathematik, Georg-August-Universität Göttingen.

Narcowich, F. J., Ward, J. D., and Wendland, H. (2006). Sobolev error estimates and a Bernstein inequality for scattered data interpolation via radial basis functions. *Constructive Approximation*, 24(2):175–186.
Rieger, C. and Zwicknagl, B. (2010). Sampling inequalities for infinitely smooth functions, with applications to interpolation and machine learning. *Advances in Computational Mathematics*, 32:103–129.

Rieger, C. and Zwicknagl, B. (2014). Improved exponential convergence rates by oversampling near the boundary. *Constructive Approximation*, 39(2):323–341.

Robbins, H. (1955). A remark on Stirling’s formula. *The American Mathematical Monthly*, 62(1):26–29.

Santin, G. and Haasdonk, B. (2017). Convergence rate of the data-independent $P$-greedy algorithm in kernel-based approximation. *Dolomites Research Notes on Approximation*, 10:68–78.

Schaback, R. (1999). Improved error bounds for scattered data interpolation by radial basis functions. *Mathematics of Computation*, 68(225):201–216.

Schaback, R. (2000). A unified theory of radial basis functions: Native Hilbert spaces for radial basis functions II. *Journal of Computational and Applied Mathematics*, 121(1–2):165–177.

Schaback, R. (2005). Multivariate interpolation by polynomials and radial basis functions. *Constructive Approximation*, 21(3):293–317.

Schaback, R. (2018). Superconvergence of kernel-based interpolation. *Journal of Approximation Theory*, 235:1–19.

Schaback, R. and Wendland, H. (2000). Adaptive greedy techniques for approximate solution of large RBF systems. *Numerical Algorithms*, 24(3):239–254.

Sloan, I. H. and Woźniakowski, H. (2018). Multivariate approximation for analytic functions with Gaussian kernels. *Journal of Complexity*, 45:1–21.

Steinwart, I., Hush, D., and Scovel, C. (2006). An explicit description of the reproducing kernel Hilbert spaces of Gaussian RBF kernels. *IEEE Transactions on Information Theory*, 52(10):4635–4643.

Sun, H. (2005). Mercer theorem for RKHS on noncompact sets. *Journal of Complexity*, 21(3):337–349.

Tanaka, K. (2019). Generation of point sets by convex optimization for interpolation in reproducing kernel Hilbert spaces. *Numerical Algorithms*. Published online (doi:10.1007/s11075-019-00792-w).

Tanaka, K. and Sugihara, M. (2019). Design of accurate formulas for approximating functions in weighted Hardy spaces by discrete energy minimization. *IMA Journal of Numerical Analysis*, 39(4):1957–1984.

Wendland, H. (2005). *Scattered Data Approximation*. Number 17 in Cambridge Monographs on Applied and Computational Mathematics. Cambridge University Press.

Wendland, H. and Rieger, C. (2005). Approximate interpolation with applications to selecting smoothing parameters. *Numerische Mathematik*, 101(4):729–748.

Wirtz, D. and Haasdonk, B. (2013). A vectorial kernel orthogonal greedy algorithm. *Dolomites Research Notes on Approximation*, 6:83–100.