Error Bounds for Kernel-Based Numerical Differentiation

Oleg Davydov* Robert Schaback†

May 1, 2013

Abstract

The literature on meshless methods observed that kernel-based numerical differentiation formulae are robust and provide high accuracy at low cost. This paper analyzes the error of such formulas, using the new technique of *growth functions*. It allows to bypass certain technical assumptions that were needed to prove the standard error bounds on interpolants and their derivatives. Since differentiation formulas based on polynomials also have error bounds in terms of growth functions, we have a convenient way to compare kernel–based and polynomial–based formulas. It follows that kernel-based formulas are comparable in accuracy to the best possible polynomial–based formulas. A variety of examples is provided.

1 Introduction

We consider a linear differential operator D of order k in d real variables in the notation

$$Df = \sum_{|\alpha|\le k} a_\alpha \partial^\alpha f, \qquad \partial^\alpha := \frac{\partial^{|\alpha|}}{\partial \mathbf{x}^\alpha} = \frac{\partial^{|\alpha|}}{\partial x_1^{\alpha_1} \cdots \partial x_d^{\alpha_d}}, \qquad |\alpha| = \alpha_1 + \cdots + \alpha_d, \tag{1}$$

where a_α are real functions of the independent variable such that

$$\sum_{|\alpha|=k} |a_\alpha(\mathbf{z})| \neq 0 \tag{2}$$

for a point $\mathbf{z} \in \mathbb{R}^d$ that will be kept fixed as well as D throughout. We shall consider formulas

$$Df(\mathbf{z}) \approx \sum_{j=1}^{N} w_j \, f(\mathbf{x}_j) \tag{3}$$

that allow to evaluate $Df(\mathbf{z})$ approximately using only function values at the nodes.

Given a point set $\mathbf{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$, a standard way of obtaining numerical differentiation formulas is by requiring that the formula should be exact for polynomials of certain order. We denote by Π_q^d the space of all d-variate polynomials of order at most q, i.e. of total degree less than q, and $\Pi_0^d := \{0\}$.

*Department of Mathematics, University of Strathclyde, 26 Richmond Street, Glasgow G1 1XH, Scotland, oleg.davydov@strath.ac.uk

†Institut für Numerische und Angewandte Mathematik, Universität Göttingen, Lotzestraße 16–18, 37083 Göttingen, Germany, schaback@math.uni-goettingen.de

Definition 1. Let D be a linear differential operator of order k. A numerical differentiation formula (3) based on a set $\mathbf{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_N\} \subset \mathbb{R}^d$ of centres is said to be *polynomially consistent of order* $m \geq 1$ if it is exact for any polynomial $p \in \Pi^d_{m+k}$.

Polynomially consistent formulas are extensively used in the finite difference methods for partial differential equations, where $\mathbf{X}$ typically is a scaled and translated version of a particular configuration of centres, such as the five point star used for the discretisation of the Laplacian, and the weights are obtained by scaling a single *stencil*. This approach is indeed efficient and effective if the centres used for the discretisation of a problem are regularly distributed. In contrast to this, we consider the setting where the weights w_j need to be generated for each unique scattered constellation of centres $\mathbf{x}_1, \ldots, \mathbf{x}_N$ that arises for example in the cause of adaptive refinement or relocation.

A polynomially consistent formula of order m can be obtained by numerically solving the equations

$$Dp(\mathbf{z}) = \sum_{j=1}^{N} w_j \, p(\mathbf{x}_j) \text{ for all } p \in \Pi^d_{m+k} \tag{4}$$

for weights w_j. The achievable order of polynomial consistency on a set $\mathbf{X}$ is limited by solvability of (4) and will depend crucially on the geometry of $\mathbf{X}$ and the differential operator D. For example, if all points are on a line, only derivatives along that line can be approximated to reasonable order.

Numerical differentiation formulas on scattered centres are needed in *meshless methods* where the trial functions are often written "entirely in terms of nodes" [2], in particular in *generalized finite difference methods* [4]. For example, *shape functions* $u_1, \ldots, u_N$ can be parametrized in terms of their values at scattered nodes $x_1, \ldots, x_N$ to yield interpolatory *trial functions* [6]

$$u_f(\mathbf{x}) := \sum_{j=1}^{N} u_j(\mathbf{x}) \, f(\mathbf{x}_j)$$

and formulas that are exact on these functions take the form

$$Df(\mathbf{z}) \approx Du_f(\mathbf{z}) = \sum_{j=1}^{N} Du_j(\mathbf{z}) \, f(\mathbf{x}_j)$$

which is of the required form with $w_j = Du_j(\mathbf{z})$.

Similarly, the weights of a *kernel based differentiation formula* are obtained by requiring that

$$Dr_{\mathbf{X},K,f}(\mathbf{z}) = \sum_{j=1}^{N} w_j \, f(\mathbf{x}_j),$$

where $r_{\mathbf{X},K,f}$ is a *kernel interpolant* of the data $f(\mathbf{x}_j)$, $j = 1, \ldots, N$, defined by a positive definite or conditionally positive definite kernel K, for example a radial basis function interpolant. This means that there are no geometric restrictions on $\mathbf{X}$ and the weights are computed by solving a single and simple linear system (see (18)) and lead to a formula (3) that often works nicely without having a positive polynomial consistency

order. The kernel based approach gains a growing popularity in meshless methods in both weak and strong form [11, Sections 12–14]. The high approximation quality and robustness of the method have been observed in many numerical experiments despite the lack of polynomial consistency.

This paper provides a theoretical justification for these observations. We show (Theorem 9) that the error of a kernel based differentiation formula is bounded by

$$\inf_{m\geq 1} \rho_{m+k,D}(\mathbf{z},\mathbf{X})\, C_{K,f,m}, \tag{5}$$

where $C_{K,f,m}$ is a measure of smoothness of K and f, independent of the geometry of $\mathbf{X}$, and $\rho_{q,D}(\mathbf{z},\mathbf{X})$ is the *growth function* defined as

$$\rho_{q,D}(\mathbf{z},\mathbf{X}) = \sup\big\{Dp(\mathbf{z}) : p\in\Pi_q^d,\ |p(\mathbf{x}_i)|\leq \|\mathbf{x}_i-\mathbf{z}\|_2^q,\ i=1,\dots,N\big\}.$$

It turns out that the error of polynomially consistent formulas of order m is also governed by the expressions of the type

$$\rho_{m+k,D}(\mathbf{z},\mathbf{X})\, C_{f,m},$$

where $C_{f,m}$ accounts for the smoothness of f, see (38) and (39) for precise formulations of the lower and upper bounds. Therefore any data set $\mathbf{X}$ allowing some polynomial consistency leads to a kernel–based differentiation formula that shares the error behaviour with the differentiation formula of the optimal polynomial consistency order on that set, without knowing or calculating that order. This explains the robustness of kernel-based formulas in numerical applications.

As an illustration, we consider (see Section 4) the standard regular five–point star approximation for the Laplacian in 2 variables. It has a polynomial consistency order $m=2$ and is exact on bivariate polynomials of order at most 4. Its error is $\mathcal{O}(h^2)$ under scaling. Our results show that for all sufficiently smooth kernels, the kernel–based differentiation formulas for this special geometry have the same $\mathcal{O}(h^2)$ error without being polynomially consistent at all.

Note that the standard theory of error bounds for kernel interpolation, including the error of derivatives, see e.g. [13], accounts for the geometry of $\mathbf{X}$ by resorting to the *fill distance*, a measure of density of centers. This however presumes that the centers of the same density should be equally well suited for interpolation or numerical differentiation, and misses the fact that kernel based approximation actually does the best a polynomially consistent method can do on a given point set. As a results, the fill distance type estimates turn out to be either not applicable or utterly pessimistic for the numerical differentiation formulas because of their requirement that the centres are "sufficiently dense." In the proof of the bound (5) we follow the classical scheme involving a power function estimate and local polynomial reproduction [14], but instead of the density assumption we use a duality argument relating the error to the polynomial growth, see also [3, 1]. Note that although the density assumption was not used in [3, 1], the estimates obtained there are not optimal when applied to the RBF numerical differentiation formulas.

The paper is organized as follows. In Section 2 we introduce kernel-based numerical differention formulas and discuss their existence and computation. Section 3 is devoted

to our main results giving error bounds in terms of the growth function. In Section 4 we derive lower and upper bounds for polynomially consistent formulas and compare them to the kernel-based formulas. Finally, in Section 5 we provide numerical examples that illustrate the theoretical results of the paper and compare the errors of numerical differention formulas generated by Gaussian, Wendland and Matérn kernels, as well as the polynomially consistent formulas of different orders. In particular, we provide examples where the highest polynomial consistency order is not optimal, and demonstrate that the kernel-based formulas are 'aware' of this, in line with the prediction given by (5).

2 Kernel-based numerical differentiation

Let

$$K : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$$

be a symmetric *kernel* on $\mathbb{R}^d$. We will often require existence of certain partial derivatives of K. Since it is important to distinguish both d-dimensional arguments of K, we will use the following notation:

$$\partial^{\alpha,\beta} K(\mathbf{x},\mathbf{y}) := \frac{\partial^{|\alpha|}}{\partial \mathbf{x}^\alpha}\Big(\frac{\partial^{|\beta|}}{\partial \mathbf{y}^\beta} K(\mathbf{x},\mathbf{y})\Big).$$

The kernel K is assumed to be *conditionally positive definite (cpd)* of some order $s = 0, 1, 2, \ldots$. This means that for any distinct points $\mathbf{x}_i \in \mathbb{R}^d$, $i = 1, \ldots, N$ the *kernel matrix*

$$[K(\mathbf{x}_i, \mathbf{x}_j)]_{i,j=1}^N$$

is positive definite on the subspace of $\mathbb{R}^N$ of vectors $a \in \mathbb{R}^N$ satisfying *zero moment conditions*

$$\sum_{j=1}^N a_j\, p(\mathbf{x}_j) = 0, \qquad \text{for all } p \in \Pi_s^d.$$

If $s = 0$, the kernel is called (unconditionally) *positive definite*. For later use, we set $M := \dim(\Pi_s^d) = \binom{d+s-1}{d} \geq 0$ and fix a basis $\{p_1, \ldots, p_M\}$ for Π_s^d.

Obviously, a cpd-kernel of order s is also a cpd-kernel of any order higher than s. In particular, any positive definite kernel can be treated as a cpd-kernel of some nonzero order and a polynomial term can be added to ensure that the interpolant (7) and any numerical differentiation formula derived from it reproduce polynomials of order s.

A *translation-invariant* kernel K has the form $K(\mathbf{x},\mathbf{y}) = \Phi(\mathbf{x}-\mathbf{y})$, where $\Phi : \mathbb{R}^d \to \mathbb{R}$ is a real-valued function on $\mathbb{R}^d$. A *radial basis function* (RBF) is a translation- and rotation-invariant kernel $K(\mathbf{x},\mathbf{y}) = \phi(\|\mathbf{x}-\mathbf{y}\|_2)$ with a scalar function $\phi : [0,\infty) \to \mathbb{R}$.

Since the kernel-based differentiation formulas are derived from the derivatives of interpolants, we now consider interpolation using a conditionally positive kernel K of order s on sets of distinct points $\mathbf{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_N\} \subset \mathbb{R}^d$ using the values of a real function f at the points $\mathbf{x}_1, \ldots, \mathbf{x}_N$. The kernel interpolants $r_{\mathbf{X},K,f}$ have the form

$$r_{\mathbf{X},K,f} = \sigma_{\mathbf{a},\mathbf{X},\mathbf{b}} := \sum_{j=1}^N a_j K(\cdot, \mathbf{x}_j) + \sum_{j=1}^M b_j p_j, \tag{6}$$

where the coefficients $\mathbf{a} = \{a_j\} \in \mathbb{R}^N$ and $\mathbf{b} = \{b_j\} \in \mathbb{R}^M$ in (6) are determined from the conditions

$$r_{\mathbf{X},K,f}(\mathbf{x}_k) = \sum_{j=1}^{N} a_j K(\mathbf{x}_k, \mathbf{x}_j) + \sum_{j=1}^{M} b_j p_j(\mathbf{x}_k) = f_k, \qquad k = 1, \ldots, N, \tag{7}$$

and

$$\sum_{j=1}^{N} a_j \, p_i(\mathbf{x}_j) = 0, \qquad 1 \le i \le M. \tag{8}$$

This is uniquely solvable (see e.g. [13]) under the assumptions that $N \ge M = \dim \Pi_s^d$ and $\mathbf{X}$ is a *unisolvent set* for Π_s^d, that is $p|_{\mathbf{X}} = 0$ implies $p = 0$.

The representation (6) of functions is a direct sum, as can be easily shown, and we view it as a split into a *kernel* part and a *polynomial* part. Now conditional positive definiteness implies that

$$(\sigma_{\mathbf{a},\mathbf{X},\mathbf{b}}, \sigma_{\mathbf{c},\mathbf{Y},\mathbf{d}})_K := \sum_{\mathbf{x}_j \in \mathbf{X}} \sum_{\mathbf{y}_k \in \mathbf{Y}} a_j K(\mathbf{x}_j, \mathbf{y}_k) c_k$$

is a semi-inner product which is zero on the polynomial part (if present) and an inner product on the kernel part of (6). Going over to the completion turns the kernel part into a Hilbert space, and thus there is a *native* space $\mathcal{F}_K$ of functions on Ω which is the direct sum of a Hilbert space and the polynomial space Π_s^d, arising as completion of the space of functions of the form (6). It is not straightforward to interpret the abstract completion as a space of functions, but we leave details on this to [9].

Since we want to discuss differentiation formulas, we have to consider continuous linear functionals on the native space. This starts with functionals of the form

$$\lambda_{\mathbf{a},\mathbf{X}} \; : \; f \mapsto \sum_{j=1}^{N} a_j f(\mathbf{x}_j) \tag{9}$$

for all sets $\mathbf{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$ that are Π_s^d–unisolvent and all vectors $\mathbf{a} \in \mathbb{R}^N$ that satisfy zero moment conditions on $\mathbf{X}$ in the sense of (8), that is $\lambda_{\mathbf{a},\mathbf{X}} p = 0$ for each $p \in \Pi_s^d$. We can define an inner product on these functionals by

$$(\lambda_{\mathbf{a},\mathbf{X}}, \lambda_{\mathbf{c},\mathbf{Y}})_K := \sum_{\mathbf{x}_j \in \mathbf{X}} \sum_{\mathbf{y}_k \in \mathbf{Y}} a_j K(\mathbf{x}_j, \mathbf{y}_k) c_k = \lambda_{\mathbf{a},\mathbf{X}} \sigma_{\mathbf{c},\mathbf{Y},0} = (\sigma_{\mathbf{a},\mathbf{X},0}, \sigma_{\mathbf{c},\mathbf{Y},0})_K \tag{10}$$

and this implies that on the functionals of the above form we have the Riesz map

$$R(\lambda_{\mathbf{a},\mathbf{X}}) = \sigma_{\mathbf{a},\mathbf{X},0}$$

mapping functionals into functions arising as kernel parts of (6). This carries over to the completions of the pre–Hilbert spaces, and we see that continuous linear functionals on the native space arise as limits of functionals of the form (9) under the topology induced by the inner product (10). Furthermore, (10) implies that

$$|\lambda_{\mathbf{a},\mathbf{X}} f| = |(\sigma_{\mathbf{a},\mathbf{X},0}, f)_K| \le \|\sigma_{\mathbf{a},\mathbf{X},0}\|_K \|f\|_K = \|\lambda_{\mathbf{a},\mathbf{X}}\|_K \|f\|_K$$

holds for all functions $f \in \mathcal{F}_K$, and

$$\|\lambda_{\mathbf{a},\mathbf{X}}\|_K^2 = \sum_{j,k=1}^{N} a_j a_k K(\mathbf{x}_j, \mathbf{x}_k) = \lambda^{\mathbf{x}}_{\mathbf{a},\mathbf{X}} \lambda^{\mathbf{y}}_{\mathbf{a},\mathbf{X}} K(\mathbf{x}, \mathbf{y}) \tag{11}$$

where the upper index shows the variable that the functional acts on. This proves that all functionals λ will be admissible that arise as limits of functionals of the form $\lambda_{\mathbf{a},\mathbf{X}}$ and give the right–hand side of the above equation a finite limit. Moreover, for any such functional we have $\lambda^{\mathbf{y}} K(\cdot, \mathbf{y}) \in \mathcal{F}_K$, and (11) remains valid in the form

$$\|\lambda\|_K^2 = \|\lambda^{\mathbf{y}} K(\cdot, \mathbf{y})\|_K^2 = \lambda^{\mathbf{x}} \lambda^{\mathbf{y}} K(\mathbf{x}, \mathbf{y}). \tag{12}$$

Lemma 2. *The linear functional $\partial^\alpha_{\mathbf{z}} : f \mapsto \partial^\alpha f(\mathbf{z})$ for a fixed $\mathbf{z} \in \Omega$ and $|\alpha| \geq s$ is continuous on $\mathcal{F}_K$ if $\partial^{\alpha,\alpha} K(\mathbf{z}, \mathbf{z})$ exists.*

Proof. We take finite–difference approximations to $\partial^\alpha_{\mathbf{z}}$ of the form $\lambda_{\mathbf{a},\mathbf{X}}$ with Π^d_s–unisolvent sets $\mathbf{X}$ and let all points move towards $\mathbf{z}$ to produce the derivative in the limit. The functionals $\lambda_{\mathbf{a},\mathbf{X}}$ vanish on Π^d_s because $\lambda_{\mathbf{a},\mathbf{X}} p$ reproduces $\partial^\alpha_{\mathbf{z}} p$ exactly for all $p \in \Pi^d_s$. □

The above argument fails for lower–order derivatives, since we need that the functionals $\lambda_{\mathbf{a},\mathbf{X}}$ vanish on Π^d_s. But there is a workaround:

Lemma 3. *All functionals of the form*

$$\mu^\alpha_{\mathbf{b},\mathbf{Y},\mathbf{z}} \; : \; f \mapsto \partial^\alpha f(\mathbf{z}) - \sum_{j=1}^{M} b_j f(\mathbf{y}_j) \tag{13}$$

that vanish on Π^d_s and are based on Π^d_s–unisolvent sets $\mathbf{Y}$ are continuous on $\mathcal{F}_K$ if $\partial^{\alpha,\alpha} K(\mathbf{z}, \mathbf{z})$ exists.

Proof. We take finite–difference approximations to $\partial^\alpha_{\mathbf{z}}$ of the form $\lambda_{\mathbf{a},\mathbf{X}}$ with Π^d_s–unisolvent sets $\mathbf{X}$ not containing $\mathbf{z}$ or points of $\mathbf{Y}$, but we do not assume zero moment conditions for the coefficient vectors $\mathbf{a}$ on Π^d_s. Then we consider the functional

$$f \mapsto \lambda_{\mathbf{a},\mathbf{X}} f - \sum_{j=1}^{M} b_j f(\mathbf{y}_j) = \lambda_{\mathbf{a},\mathbf{X}} f - \partial^\alpha f(\mathbf{z}) + \mu^\alpha_{\mathbf{b},\mathbf{Y},\mathbf{z}}$$

which vanishes on Π^d_s and is based on the unisolvent set $\mathbf{X} \cup \mathbf{Y}$. This functional vanishes on Π^d_s because $\lambda_{\mathbf{a},\mathbf{X}} p$ reproduces $\partial^\alpha_{\mathbf{z}} p$ exactly for all $p \in \Pi^d_s$. Thus the functional can be used in the standard argument, and we can let all points of $\mathbf{X}$ move towards $\mathbf{z}$ to produce the limit functional $\mu^\alpha_{\mathbf{b},\mathbf{Y},\mathbf{z}}$. □

Of course, the above results generalize to values of linear differential operators:

Lemma 4. *Let D be a linear differential operator of order k. All functionals of the form*

$$\mu_{\mathbf{b},\mathbf{Y},\mathbf{z}} \; : \; f \mapsto Df(\mathbf{z}) - \sum_{j=1}^{M} b_j f(\mathbf{y}_j) \tag{14}$$

that vanish on Π_s^d *and are based on* Π_s^d*–unisolvent sets* $\mathbf{Y}$ *are continuous on* $\mathcal{F}_K$ *if* $\partial^{\alpha,\alpha} K(\mathbf{z},\mathbf{z})$ *exists for all* $|\alpha| \le k$*. Moreover,*

$$\|\mu_{\mathbf{b},\mathbf{Y},\mathbf{z}}\|_K^2 = \mu^{\mathbf{x}}_{\mathbf{b},\mathbf{Y},\mathbf{z}} \mu^{\mathbf{y}}_{\mathbf{b},\mathbf{Y},\mathbf{z}} K(\mathbf{x},\mathbf{y}). \tag{15}$$

For the goals of this paper, we shall focus on functionals of the form (14), but note that more general functionals are admissible as well, e.g. local integrals over derivatives.

Concerning approximations of derivatives, we should add a practical recipe. The linear system

$$\begin{array}{rcrcl} \sum_{j=1}^N u_j(\mathbf{x}) K(\mathbf{x}_k,\mathbf{x}_j) & + & \sum_{j=1}^M v_j(\mathbf{x}) p_j(x_k) & = & K(\mathbf{x},\mathbf{x}_k),\ 1 \le k \le N, \\ \sum_{j=1}^N u_j(\mathbf{x}) p_i(\mathbf{x}_j) & + & 0 & = & p_i(\mathbf{x}),\ 1 \le i \le M, \end{array} \tag{16}$$

is uniquely solvable as well as (7) and (8), having the same coefficient matrix. The solution functions $u_1,\dots,u_N$ are a Lagrange basis, i.e. they satisfy $u_j(\mathbf{x}_k) = \delta_{jk},\ 1 \le j,k \le N$, and they are a linear combination of the functions $p_1,\dots,p_M$ as well as $K(\cdot,\mathbf{x}_1),\dots,K(\cdot,\mathbf{x}_N)$, but, as can be shown by some linear algebra, their coefficients in terms of the latter satisfy zero moment conditions on $\mathbf{X}$. Some more details on the Lagrange basis are in [11, Section 3.3].

Thus the interpolant $r_{\mathbf{X},K,f}$ to some function f can be written as

$$r_{\mathbf{X},K,f} = \sum_{j=1}^N u_j f(\mathbf{x}_j),$$

and each derivative thereof takes the form

$$\partial^\alpha r_{\mathbf{X},K,f} = \sum_{j=1}^N \underbrace{\partial^\alpha u_j}_{\text{weights}} f(\mathbf{x}_j) \approx \partial^\alpha f$$

fitting into (3) if evaluated at $\mathbf{z}$. These derivatives exist if the derivatives $\partial^\alpha K(\cdot,\mathbf{x}_j)$ exist, since they are obtained from them via the linear system. Note that the above equation directly gives the approximation of a derivative via kernel methods. In particular, by applying $\lambda := \delta_{\mathbf{z}} D$ to both sides of (16), we see that the weight vector $\mathbf{w}$ of the numerical differentiation formula

$$Df(\mathbf{z}) \approx Dr_{\mathbf{X},K,f}(\mathbf{z}) = \sum_{j=1}^N w_j\, f(\mathbf{x}_j), \tag{17}$$

where $w_j = Du_j(\mathbf{z})$, can be obtained by solving the linear system

$$\begin{array}{rcrcll} \sum_{j=1}^N w_j K(\mathbf{x}_k,\mathbf{x}_j) & + & \sum_{j=1}^M c_j p_j(x_k) & = & \delta_{\mathbf{z}} DK(\cdot,\mathbf{x}_k), & 1 \le k \le N, \\ \sum_{j=1}^N w_j p_i(\mathbf{x}_j) & + & 0 & = & \delta_{\mathbf{z}} Dp_i, & 1 \le i \le M. \end{array} \tag{18}$$

Note that the components of the "dummy" vector $\mathbf{c}$ are not needed for the numerical differentiation.

Note further that we have polynomial reproduction

$$r_{\mathbf{X},K,p} = p \quad \text{for all } \; p \in \Pi_s^d$$

and hence

$$Dp(\mathbf{z}) = \sum_{j=1}^{N} w_j \, p(\mathbf{x}_j) \quad \text{for all } \; p \in \Pi_s^d$$

in the conditionally positive definite case of order s due to the second set of equations in (16).

We summarize the results of this section in the following statement.

Proposition 5. *Let D be a linear differential operator of order k and let K be a cpd-kernel of order s. Kernel–based numerical differentiation formula (17) is well defined as long as $\partial^{\alpha,\beta}K(\mathbf{z},\mathbf{z})$ exists for all $|\alpha|, |\beta| \le k$ and $\mathbf{X}$ is a unisolvent set for Π_s^d. The formula (17) is exact for $p \in \Pi_s^d$ and its weight vector $\mathbf{w}$ can be found by solving the linear system* (18). □

In the translation-invariant case $K(\mathbf{x},\mathbf{y}) = \Phi(\mathbf{x}-\mathbf{y})$ on $\mathbb{R}^d$, the *native space* $\mathcal{F}_K$ generated by the kernel K can be described with the help of the *generalised Fourier transform* $\hat{f}$ as

$$\mathcal{F}_K = \{f \in L_2(\mathbb{R}^d) \, : \, \|f\|_K < \infty\},$$

where the (semi-) norm is

$$\|f\|_K := (2\pi)^{-d/4} \left\| \hat{f} / \sqrt{\widehat{\Phi}} \right\|_{L_2(\mathbb{R}^d)}, \qquad f \in L_2(\mathbb{R}^d),$$

see [13, Theorem 10.21]. Recall that $\hat{f}$ is given for any $f \in L_1(\mathbb{R}^d)$ by the usual formula

$$\hat{f}(\mathbf{x}) = (2\pi)^{-d/2} \int_{\mathbb{R}^d} e^{-i\mathbf{x}^T\boldsymbol{\omega}} f(\boldsymbol{\omega}) \, \mathbf{d}\boldsymbol{\omega}, \qquad \mathbf{x} \in \mathbb{R}^d,$$

and it is defined in a distributional sense for certain functions $f \notin L_1(\mathbb{R}^d)$, see e.g. [13, Section 8.2].

In the Fourier case, functionals λ have a complex–valued formal Fourier transform $\hat{\lambda}$ defined via

$$\lambda(f) = (2\pi)^{-d/2} \int_{\mathbb{R}^d} \hat{f}(\boldsymbol{\omega}) \, \hat{\lambda}(\boldsymbol{\omega}) \, \mathbf{d}\boldsymbol{\omega}$$

on functions f for which this integral exists. By Cauchy–Schwarz inequality, continuity on the native space of Φ holds if

$$\|\lambda\|_K^2 = (2\pi)^{-d/2} \int_{\mathbb{R}^d} \hat{\Phi}(\boldsymbol{\omega}) \, |\hat{\lambda}(\boldsymbol{\omega})|^2 \, \mathbf{d}\boldsymbol{\omega} < \infty$$

This is $\lambda^{\mathbf{y}} \overline{\lambda^{\mathbf{x}} \Phi(\mathbf{x}-\mathbf{y})}$, illustrating Lemmas 2 and 3.

To shed some light on the hidden peculiarities here, we consider the example of thin–plate splines $\Phi(\mathbf{x}-\mathbf{y}) = \|\mathbf{x}-\mathbf{y}\|_2^2 \log \|\mathbf{x}-\mathbf{y}\|_2$ in $\mathbb{R}^2$, which is a conditionally

positive definite radial kernel of order 2. It is the fundamental solution of the biharmonic equation, and since the Fourier transform of $-\Delta$ is $\|\boldsymbol{\omega}\|_2^2$, its generalized Fourier transform in $\mathbb{R}^2$ thus is $\hat{\Phi}(\boldsymbol{\omega}) = \|\boldsymbol{\omega}\|_2^{-4}$ up to a constant [13, Theorem 8.17]. In this case, continuous linear functionals are required to vanish on linear polynomials, and point evaluation functionals are not continuous. After quite some work, the native space can be identified with a Beppo–Levi space [13, Theorem 10.43], and we get continuous linear functionals if we compose L_2 functionals with second derivatives, e.g. $\lambda(f) = -\mu(\Delta f)$ with $\mu \in (L_2(\mathbb{R}^2))^*$. Their norms then are

$$\|\lambda\|_K^2 = (2\pi)^{-d/2} \int_{\mathbb{R}^d} \hat{\Phi}(\boldsymbol{\omega})\, |\hat{\Delta}(\boldsymbol{\omega})|^2 |\hat{\mu}(\boldsymbol{\omega})|^2 \, \mathbf{d}\boldsymbol{\omega} = (2\pi)^{-d/2} \int_{\mathbb{R}^d} |\hat{\mu}(\boldsymbol{\omega})|^2 \, \mathbf{d}\boldsymbol{\omega}.$$

Lemma 2 is not applicable because the kernel is not smooth enough, while Lemma 3 works for $|\alpha| \le 1$. On the other hand, Lemma 5 allows $|\alpha| = 2$, but not with continuity.

3 Error bounds in terms of a growth function

We now proceed towards error bounds for kernel–based approximations of derivative formulas. We start with a slight generalization of a standard result in kernel based approximation, see e.g. [13, Theorem 11.4 and 11.5]:

Lemma 6. *Let λ be a linear functional of the form $\lambda = \delta_{\mathbf{z}} D$, that is $\lambda f = Df(\mathbf{z})$, where D is a linear differential operator of order k. Assume that $\partial^{\alpha,\beta} K(\mathbf{z},\mathbf{z})$ exists for all $|\alpha|, |\beta| \le k$. Furthermore, let $\mathbf{X}$ be a unisolvent set for Π_s^d. Then the interpolant $r_{\mathbf{X},K,f}$ to a function $f \in \mathcal{F}_K$ satisfies*

$$|\lambda f - \lambda\, r_{\mathbf{X},K,f}| \le P_{\lambda,\mathbf{X}}\, \|f\|_K, \tag{19}$$

where $P_{\lambda,\mathbf{X}}$ is the power function *defined by*

$$P_{\lambda,\mathbf{X}}^2 = \min\Big\{ Q_{\lambda,\mathbf{X}}(\mathbf{w}) \,:\, \mathbf{w} \in \mathbb{R}^N,\, \lambda\, p = \sum_{j=1}^N w_j p(\mathbf{x}_j) \text{ for all } p \in \Pi_s^d \Big\}, \tag{20}$$

where

$$Q_{\lambda,\mathbf{X}}(\mathbf{w}) := \epsilon^{\mathbf{x}}_{\lambda,\mathbf{X},\mathbf{w}} \epsilon^{\mathbf{y}}_{\lambda,\mathbf{X},\mathbf{w}} K(\mathbf{x},\mathbf{y}), \qquad \text{with} \quad \epsilon_{\lambda,\mathbf{X},\mathbf{w}} := \lambda - \sum_{j=1}^N w_j\, \delta_{\mathbf{x}_j}. \tag{21}$$

Proof. We consider the functional

$$\mu \,:\, f \mapsto \lambda f - \lambda\, r_{\mathbf{X},K,f} = \lambda f - \sum_{j=1}^N f(\mathbf{x}_j) \lambda u_j$$

on the native space $\mathcal{F}_K$ using the Lagrange basis of (16). It vanishes on Π_s^d due to polynomial reproduction of interpolants, and it is a special case of the functionals $\epsilon_{\lambda,\mathbf{X},\mathbf{w}}$ in (21) for $w_j = \lambda u_j$. Due to our assumptions on λ, all functionals $\epsilon_{\lambda,\mathbf{X},\mathbf{w}}$ vanishing on Π_s^d satisfy Lemma 4, so that the quadratic form $Q_{\lambda,\mathbf{X}}(\mathbf{w})$ is well–defined and

$$\|\epsilon_{\lambda,\mathbf{X},\mathbf{w}}\|_K^2 = \epsilon^{\mathbf{x}}_{\lambda,\mathbf{X},\mathbf{w}} \epsilon^{\mathbf{y}}_{\lambda,\mathbf{X},\mathbf{w}} K(\mathbf{x},\mathbf{y}) = Q_{\lambda,\mathbf{X}}(\mathbf{w}),$$

which proves

$$|\epsilon_{\lambda,\mathbf{X},\mathbf{w}} f|^2 \le Q_{\lambda,\mathbf{X}}(\mathbf{w}) \|f\|_K^2$$

for all coefficient vectors $\mathbf{w}$ admissible in (20).

The minimization of the quadratic form $Q_{\lambda,\mathbf{X}}(\mathbf{w})$ with respect to $\mathbf{w}$ under the conditions of polynomial reproduction in (20) can be carried out via Lagrange multipliers in a standard way [13, Theorem 11.5], and the result is the system (16). This proves that the optimal $\mathbf{w}$ is given by $w_j = \lambda u_j$, and we get (19). □

The above result shows in particular that

$$Dr_{\mathbf{X},K,f}(\mathbf{z}) = \sum_{j=1}^{N} w_j^* f(x_j), \qquad w_j^* = \lambda u_j, \quad j = 1, \ldots, N,$$

provides the *optimal recovery* of $Df(\mathbf{z})$ for f in the native space $\mathcal{F}_K$ among all numerical differentiation formulas exact for polynomials in Π_s^d. Namely, under the hypotheses of Lemma 6, let $\mathbf{w}$ be any weight vector satisfying the condition of polynomial reproduction in (20). We have shown that

$$\sup_{\|f\|_K \le 1} |Df(\mathbf{z}) - \sum_{j=1}^{N} w_j f(x_j)| = \sqrt{Q_{\lambda,\mathbf{X}}(\mathbf{w})}, \tag{22}$$

whereas

$$\sup_{\|f\|_K \le 1} |Df(\mathbf{z}) - \sum_{j=1}^{N} w_j^* f(x_j)| = \sqrt{Q_{\lambda,\mathbf{X}}(\mathbf{w}^*)} = P_{\lambda,\mathbf{X}}, \tag{23}$$

which is the infimum of $\sqrt{Q_{\lambda,\mathbf{X}}(\mathbf{w})}$ over all such weights $\mathbf{w}$.

We now deduce an upper bound for $Q_{\lambda,\mathbf{X}}(\mathbf{w})$ by using the Boolean sum of Taylor polynomials of K with respect to $\mathbf{x}$ and $\mathbf{y}$ at $\mathbf{z}$.

Lemma 7. *Let $\lambda = \delta_{\mathbf{z}} D$, that is $\lambda f = Df(\mathbf{z})$, where D is a linear differential operator of order k, and let $\mathbf{X} = \{\mathbf{x}_1 \ldots, \mathbf{x}_N\} \subset \mathbb{R}^d$ be a Π_s^d-unisolvent set. Assume that for some $q \ge \max\{s, k+1\}$, the cpd kernel K of order s possesses continuous partial derivatives $\partial^{\alpha,\beta} K(\mathbf{x}, \mathbf{y})$, $|\alpha|, |\beta| \le q$ for all $(\mathbf{x}, \mathbf{y}) \in \Omega_{\mathbf{z}} \times \Omega_{\mathbf{z}}$, where $\Omega_{\mathbf{z}} \subset \mathbb{R}^d$ is any domain that contains $\{\mathbf{z}\} \cup \mathbf{X}$ and is star-shaped with respect to $\mathbf{z}$. Furthermore, let $\mathbf{w}$ satisfy $\lambda p = \sum_{j=1}^{N} w_j p(\mathbf{x}_j)$ for all $p \in \Pi_q^d$. Then*

$$Q_{\lambda,\mathbf{X}}(\mathbf{w}) \le \Big(\frac{1}{q!} \sum_{i=1}^{N} |w_i| \, \|\mathbf{x}_i - \mathbf{z}\|_2^q\Big)^2 \Big(\sum_{|\alpha|,|\beta|=q} \binom{q}{\alpha} \binom{q}{\beta} \max_{\mathbf{x},\mathbf{y} \in S_{\mathbf{z},\mathbf{X}}} |\partial^{\alpha,\beta} K(\mathbf{x}, \mathbf{y})|^2 \Big)^{1/2}, \tag{24}$$

where

$$S_{\mathbf{z},\mathbf{X}} := \bigcup_{i=1}^{N} [\mathbf{z}, \mathbf{x}_i] \subset \Omega_{\mathbf{z}}.$$

Proof. Let $T_{q,\mathbf{z}} f$ denote the Taylor polynomial of order q of a function f centered at $\mathbf{z}$

$$T_{q,\mathbf{z}} f(\mathbf{x}) = \sum_{|\alpha| < q} \frac{(\mathbf{x} - \mathbf{z})^\alpha}{\alpha!} \partial^\alpha f(\mathbf{z}). \tag{25}$$

Then, by a well-known remainder formula, for any $f \in C^q(\Omega_{\mathbf{z}})$,

$$f(\mathbf{x}) - T_{q,\mathbf{z}} f(\mathbf{x}) = q \sum_{|\alpha|=q} \frac{(\mathbf{x}-\mathbf{z})^\alpha}{\alpha!} \int_0^1 (1-t)^{q-1} \partial^\alpha f(\mathbf{z} + t(\mathbf{x}-\mathbf{z}))\, dt, \quad \mathbf{x} \in \Omega_{\mathbf{z}}. \quad (26)$$

Since $K(\mathbf{x}, \cdot) \in C^q(\Omega_{\mathbf{z}})$ for each $\mathbf{x} \in \Omega_{\mathbf{z}}$, its Taylor polynomial in $\mathbf{y}$,

$$T^{\mathbf{y}}_{q,\mathbf{z}} K(\mathbf{x}, \mathbf{y}) = \sum_{|\alpha|<q} \frac{(\mathbf{y}-\mathbf{z})^\alpha}{\alpha!} \partial^{0,\alpha} K(\mathbf{x}, \mathbf{z}),$$

satisfies

$$K(\mathbf{x}, \mathbf{y}) - T^{\mathbf{y}}_{q,\mathbf{z}} K(\mathbf{x}, \mathbf{y}) = q \sum_{|\alpha|=q} \frac{(\mathbf{y}-\mathbf{z})^\alpha}{\alpha!} \int_0^1 (1-t)^{q-1} \partial^{0,\alpha} K(\mathbf{x}, \mathbf{z} + t(\mathbf{y}-\mathbf{z}))\, dt$$

for all $\mathbf{x}, \mathbf{y} \in \Omega_{\mathbf{z}}$. Since $\partial^{0,\alpha} K(\cdot, \mathbf{y}) \in C^q(\Omega_{\mathbf{z}})$, $|\alpha| < q$, for any fixed $\mathbf{y} \in \Omega_{\mathbf{z}}$, we can apply the remainder formula to the Taylor polynomial of $S(\mathbf{x}, \mathbf{y}) := K(\mathbf{x}, \mathbf{y}) - T^{\mathbf{y}}_{q,\mathbf{z}} K(\mathbf{x}, \mathbf{y})$ in $\mathbf{x}$, which gives

$$S(\mathbf{x}, \mathbf{y}) - T^{\mathbf{x}}_{q,\mathbf{z}} S(\mathbf{x}, \mathbf{y}) = q \sum_{|\alpha|=q} \frac{(\mathbf{x}-\mathbf{z})^\alpha}{\alpha!} \int_0^1 (1-t)^{q-1} \partial^{0,\alpha} S(\mathbf{z} + t(\mathbf{x}-\mathbf{z}), \mathbf{y})\, dt.$$

We observe that

$$S(\mathbf{x}, \mathbf{y}) - T^{\mathbf{x}}_{q,\mathbf{z}} S(\mathbf{x}, \mathbf{y}) = K(\mathbf{x}, \mathbf{y}) - P(\mathbf{x}, \mathbf{y}),$$

where $P(\mathbf{x}, \mathbf{y})$ is the Boolean sum of Taylor polynomials,

$$\begin{aligned} P(\mathbf{x}, \mathbf{y}) &= T^{\mathbf{x}}_{q,\mathbf{z}} K(\mathbf{x}, \mathbf{y}) + T^{\mathbf{y}}_{q,\mathbf{z}} K(\mathbf{x}, \mathbf{y}) - T^{\mathbf{x}}_{q,\mathbf{z}} T^{\mathbf{y}}_{q,\mathbf{z}} K(\mathbf{x}, \mathbf{y}) \\ &= \sum_{|\alpha|<q} \frac{(\mathbf{x}-\mathbf{z})^\alpha}{\alpha!} \partial^{\alpha,0} K(\mathbf{z}, \mathbf{y}) + \sum_{|\alpha|<q} \frac{(\mathbf{y}-\mathbf{z})^\alpha}{\alpha!} \partial^{0,\alpha} K(\mathbf{x}, \mathbf{z}) \\ &\quad - \sum_{|\alpha|,|\beta|<q} \frac{(\mathbf{x}-\mathbf{z})^\alpha (\mathbf{y}-\mathbf{z})^\beta}{\alpha!\beta!} \partial^{\alpha,\beta} K(\mathbf{z}, \mathbf{z}). \end{aligned}$$

Since for all α, β with $|\alpha|, |\beta| < q$,

$$\begin{aligned} &\epsilon^{\mathbf{x}}_{\lambda,\mathbf{X},\mathbf{w}} \epsilon^{\mathbf{y}}_{\lambda,\mathbf{X},\mathbf{w}} \big((\mathbf{x}-\mathbf{z})^\alpha \partial^{\alpha,0} K(\mathbf{z}, \mathbf{y})\big) = \epsilon^{\mathbf{x}}_{\lambda,\mathbf{X},\mathbf{w}} \big((\mathbf{x}-\mathbf{z})^\alpha\big) \epsilon^{\mathbf{y}}_{\lambda,\mathbf{X},\mathbf{w}} \big(\partial^{\alpha,0} K(\mathbf{z}, \mathbf{y})\big) = 0, \\ &\epsilon^{\mathbf{x}}_{\lambda,\mathbf{X},\mathbf{w}} \epsilon^{\mathbf{y}}_{\lambda,\mathbf{X},\mathbf{w}} \big((\mathbf{y}-\mathbf{z})^\alpha \partial^{0,\alpha} K(\mathbf{x}, \mathbf{z})\big) = \epsilon^{\mathbf{x}}_{\lambda,\mathbf{X},\mathbf{w}} \big(\partial^{0,\alpha} K(\mathbf{x}, \mathbf{z})\big) \epsilon^{\mathbf{y}}_{\lambda,\mathbf{X},\mathbf{w}} \big((\mathbf{y}-\mathbf{z})^\alpha\big) = 0, \\ &\epsilon^{\mathbf{x}}_{\lambda,\mathbf{X},\mathbf{w}} \epsilon^{\mathbf{y}}_{\lambda,\mathbf{X},\mathbf{w}} \big((\mathbf{x}-\mathbf{z})^\alpha (\mathbf{y}-\mathbf{z})^\beta\big) = \epsilon^{\mathbf{x}}_{\lambda,\mathbf{X},\mathbf{w}} \big((\mathbf{x}-\mathbf{z})^\alpha\big) \epsilon^{\mathbf{y}}_{\lambda,\mathbf{X},\mathbf{w}} \big((\mathbf{y}-\mathbf{z})^\beta\big) = 0, \end{aligned}$$

we conclude that

$$\epsilon^{\mathbf{x}}_{\lambda,\mathbf{X},\mathbf{w}} \epsilon^{\mathbf{y}}_{\lambda,\mathbf{X},\mathbf{w}} \big(P(\mathbf{x}, \mathbf{y})\big) = 0,$$

and hence

$$\begin{aligned} Q_{\lambda,\mathbf{X}}(\mathbf{w}) &= \epsilon^{\mathbf{x}}_{\lambda,\mathbf{X},\mathbf{w}} \epsilon^{\mathbf{y}}_{\lambda,\mathbf{X},\mathbf{w}} \big(K(\mathbf{x}, \mathbf{y}) - P(\mathbf{x}, \mathbf{y})\big) \\ &= \lambda^{\mathbf{x}} \lambda^{\mathbf{y}} R(\mathbf{x}, \mathbf{y}) - 2 \sum_{i=1}^N w_i \lambda^{\mathbf{x}} R(\mathbf{x}, \mathbf{x}_i) + \sum_{i,j=1}^N w_i w_j R(\mathbf{x}_i, \mathbf{x}_j), \end{aligned}$$

where we set $R := K - P$ and used the fact that both $K(\mathbf{x},\mathbf{y})$ and $P(\mathbf{x},\mathbf{y})$ are symmetric.

Since $q > k$, we have

$$\lambda^{\mathbf{x}}\big((\mathbf{x}-\mathbf{z})^\alpha\big) = \sum_{|\gamma|\le k} a_\gamma(\mathbf{z})\Big(\frac{\partial^{|\gamma|}}{\partial \mathbf{x}^\gamma}(\mathbf{x}-\mathbf{z})^\alpha\Big)\Big|_{\mathbf{x}=\mathbf{z}} = 0, \qquad |\alpha| = q.$$

Hence for any $\mathbf{y}$

$$\lambda^{\mathbf{x}} R(\mathbf{x},\mathbf{y}) = \lambda^{\mathbf{x}}\Big(q\sum_{|\alpha|=q}\frac{(\mathbf{x}-\mathbf{z})^\alpha}{\alpha!}\int_0^1 (1-t)^{q-1}\partial^{\alpha,0}S(\mathbf{z}+t(\mathbf{x}-\mathbf{z}),\mathbf{y})\,dt\Big) = 0,$$

which implies $\lambda^{\mathbf{x}} R(\mathbf{x},\mathbf{x}_i) = 0$ and $\lambda^{\mathbf{x}}\lambda^{\mathbf{y}} R(\mathbf{x},\mathbf{y}) = \lambda^{\mathbf{y}}\lambda^{\mathbf{x}} R(\mathbf{x},\mathbf{y}) = 0$, thus reducing the above expression for $Q_{\lambda,\mathbf{X}}(\mathbf{w})$ to its last term,

$$Q_{\lambda,\mathbf{X}}(\mathbf{w}) = \sum_{i,j=1}^N w_i w_j R(\mathbf{x}_i,\mathbf{x}_j). \tag{27}$$

To estimate it, we use the remainder of the Taylor polynomial $T^{\mathbf{x}}_{q,\mathbf{z}} S(\mathbf{x},\mathbf{y})$ to write

$$R(\mathbf{x}_i,\mathbf{x}_j) = q\sum_{|\alpha|=q}\frac{(\mathbf{x}_i-\mathbf{z})^\alpha}{\alpha!}\int_0^1 (1-t)^{q-1}\partial^{\alpha,0}S(\mathbf{x}_i^t,\mathbf{x}_j)\,dt,$$

where $\mathbf{x}_i^t := \mathbf{z} + t(\mathbf{x}_i - \mathbf{z})$, and

$$\partial^{\alpha,0}S(\mathbf{x}_i^t,\mathbf{x}_j) = q\sum_{|\beta|=q}\frac{(\mathbf{x}_j-\mathbf{z})^\beta}{\beta!}\int_0^1 (1-s)^{q-1}\partial^{\alpha,\beta}K(\mathbf{x}_i^t,\mathbf{x}_j^s)\,ds.$$

By the multinomial theorem it is easy to see that

$$\sum_{|\alpha|=q}\frac{(\mathbf{x}_j-\mathbf{z})^{2\alpha}}{\alpha!} = \frac{1}{q!}\|\mathbf{x}_j-\mathbf{z}\|_2^{2q}. \tag{28}$$

Hence, the weighted Cauchy-Schwarz inequality gives

$$\begin{aligned}
|R(\mathbf{x}_i,\mathbf{x}_j)|^2 &\le \Big(\sum_{|\alpha|=q}\frac{|(\mathbf{x}_i-\mathbf{z})^\alpha|}{\alpha!}\max_{0\le t\le 1}\big|\partial^{\alpha,0}S(\mathbf{x}_i^t,\mathbf{x}_j)\big|\Big)^2 \\
&\le \sum_{|\alpha|=q}\frac{(\mathbf{x}_i-\mathbf{z})^{2\alpha}}{\alpha!}\sum_{|\alpha|=q}\frac{1}{\alpha!}\max_{0\le t\le 1}\big|\partial^{\alpha,0}S(\mathbf{x}_i^t,\mathbf{x}_j)\big|^2 \\
&\le \frac{1}{q!}\|\mathbf{x}_i-\mathbf{z}\|_2^{2q}\sum_{|\alpha|=q}\frac{1}{\alpha!}\sum_{|\beta|=q}\frac{(\mathbf{x}_j-\mathbf{z})^{2\beta}}{\beta!}\sum_{|\beta|=q}\frac{1}{\beta!}\max_{0\le t,s\le 1}\big|\partial^{\alpha,\beta}K(\mathbf{x}_i^t,\mathbf{x}_j^s)\big|^2 \\
&= \frac{1}{(q!)^2}\|\mathbf{x}_i-\mathbf{z}\|_2^{2q}\|\mathbf{x}_j-\mathbf{z}\|_2^{2q}\sum_{|\alpha|,|\beta|=q}\frac{1}{\alpha!\beta!}\max_{0\le t,s\le 1}\big|\partial^{\alpha,\beta}K(\mathbf{x}_i^t,\mathbf{x}_j^s)\big|^2,
\end{aligned}$$

and (24) follows in view of (27). □

Lemmas 6 and 7 imply the error bound

$$|Df(\mathbf{z}) - Dr_{\mathbf{X},K,f}(\mathbf{z})| \le \frac{\sqrt{M_{K,q}}}{q!}\|f\|_K \sum_{i=1}^N |w_i|\,\|\mathbf{x}_i - \mathbf{z}\|_2^q, \qquad f \in \mathcal{F}_K, \tag{29}$$

where $q \ge \max\{s, k+1\}$, $\mathbf{w}$ is any vector satisfying $Dp(\mathbf{z}) = \sum_{j=1}^N w_j p(\mathbf{x}_j)$ for all $p \in \Pi_q^d$, and

$$M_{K,q} := \Big(\sum_{|\alpha|,|\beta|=q} \binom{q}{\alpha}\binom{q}{\beta} \|\partial^{\alpha,\beta} K\|^2_{C(\Omega\times\Omega)} \Big)^{1/2}, \tag{30}$$

with $\Omega \subset \mathbb{R}^d$ such that $S_{\mathbf{z},\mathbf{X}} \subset \Omega$ and $\partial^{\alpha,\beta}K(\mathbf{x},\mathbf{y}) \in C(\Omega\times\Omega)$ for all $|\alpha|, |\beta| \le q$. We stress that $\mathbf{w}$ in (29) is not the weight vector $\mathbf{w}$ of (3) determined by (18).

Clearly, q and $\mathbf{w}$ are parameters of the bound (29) rather than of the approximation method, and hence for given $D, \mathbf{z}, \mathbf{X}$ one can look for the optimal choice of $q, \mathbf{w}$ that minimize the right hand side of (29). Of course, the greater q is chosen, the more restrictive is the condition on $\mathbf{w}$, and for a sufficiently large q it is impossible to find any feasible vector $\mathbf{w}$.

For a fixed q, the optimal value of $\sum_{i=1}^N |w_i|\,\|\mathbf{x}_i - \mathbf{z}\|_2^q$ has a useful interpretation as the maximum attainable value of $Dp(\mathbf{z})$ subject to the conditions $|p(\mathbf{x}_i)| \le \|\mathbf{x}_i - \mathbf{z}\|_2^q$. More precisely, the *growth function* $\rho_{q,D}(\mathbf{z},\mathbf{X})$ is defined by

$$\rho_{q,D}(\mathbf{z},\mathbf{X}) = \sup\big\{ Dp(\mathbf{z}) : p \in \Pi_q^d,\ |p(\mathbf{x}_i)| \le \|\mathbf{x}_i - \mathbf{z}\|_2^q,\ i = 1,\dots,N \big\}. \tag{31}$$

Note that a slightly different growth function has been used in the RBF error bounds in [3, 1]. (The growth function is also related to the *norming constant* used in the error bounds introduced in [7].) The following statement is easily obtainable by the duality of linear functionals, see e.g. [1, Lemma 4].

Lemma 8. *The identity*

$$\inf\Big\{ \sum_{i=1}^N |w_i|\,\|\mathbf{x}_i - \mathbf{z}\|_2^q :\ \mathbf{w} \in \mathbb{R}^N,\ Dp(\mathbf{z}) = \sum_{j=1}^N w_j p(\mathbf{x}_j) \text{ for all } p \in \Pi_q^d \Big\} = \rho_{q,D}(\mathbf{z},\mathbf{X})$$

holds as soon as there exists an admissible vector $\mathbf{w}$ *in the left hand side. If there is no such vector, then* $\rho_{q,D}(\mathbf{z},\mathbf{X}) = \infty$.

From this and (29) we obtain the following result.

Theorem 9. *Let* K *be an cpd-kernel of order* s *on* Ω*, and let* D *be a linear differential operator of order* k*. Assume that* $\mathbf{X} = \{\mathbf{x}_1 \dots, \mathbf{x}_N\} \subset \Omega$ *is a* Π_s^d*-unisolvent set. Then, for any* $\mathbf{z} \in \Omega$*, such that* $S_{\mathbf{z},\mathbf{X}} \subset \Omega$*, and any* $q \ge \max\{s, k+1\}$ *such that* $\partial^{\alpha,\beta}K(\mathbf{x},\mathbf{y}) \in C(\Omega\times\Omega)$ *for all* $|\alpha|, |\beta| \le q$*,*

$$P_{\lambda,\mathbf{X}} \le \rho_{q,D}(\mathbf{z},\mathbf{X})\frac{\sqrt{M_{K,q}}}{q!}, \qquad \lambda f = Df(\mathbf{z}), \tag{32}$$

and hence

$$|Df(\mathbf{z}) - Dr_{\mathbf{X},K,f}(\mathbf{z})| \le \rho_{q,D}(\mathbf{z},\mathbf{X})\frac{\sqrt{M_{K,q}}}{q!}\|f\|_K, \qquad f \in \mathcal{F}_K. \tag{33}$$

Note that in the case of the identity operator $Df = f$, Theorem 9 gives a new error bound for the kernel based interpolation.

4 Comparison to polynomially consistent formulas

Recall that the auxiliary weights w_i in (29) satisfy $Dp(\mathbf{z}) = \sum_{j=1}^N w_j p(\mathbf{x}_j)$ for all $p \in \Pi_q^d$, that is, they are the weights of a numerical differentiation formula

$$Df(\mathbf{z}) \approx \sum_{j=1}^N w_j\, f(\mathbf{x}_j)$$

exact for polynomials of order q and hence polynomially consistent of order $q-k$ in the sense of Definition 1.

The following theorem gives an error bound for polynomially consistent formulas strikingly similar to (29).

Theorem 10. *Any differentiation formula (3) which is polynomially consistent of order $m \geq 1$ for a linear differential operator D of order k has an error bound*

$$|Df(\mathbf{z}) - \sum_{j=1}^N w_j\, f(\mathbf{x}_j)| \leq \frac{1}{(m+k)!} |f|_{\infty,m+k,\Omega} \sum_{j=1}^N |w_j| \|\mathbf{x}_j - \mathbf{z}\|_2^{m+k} \tag{34}$$

for all $f \in C^{m+k}(\Omega)$, where $\Omega \subset \mathbb{R}^d$ is any domain that contains $S_{\mathbf{z},\mathbf{X}}$. Here, we use the following semi–norm on $C^q(\Omega)$

$$|f|_{\infty,q,\Omega} := \Big(\sum_{|\alpha|=q} \binom{q}{\alpha} \|\partial^\alpha f\|_{\infty,\Omega}^2 \Big)^{1/2}.$$

Proof. Let $T_{q,\mathbf{z}} f$ denote the Taylor polynomial of order q of a function f centered at $\mathbf{z}$ as defined in (25). In view of the remainder formula (26), by setting $R = f - T_{q,\mathbf{z}} f$ for $q := m+k$ and using the fact that $DR(\mathbf{z}) = 0$, we can estimate the error of an order m polynomially consistent numerical differentiation formula as

$$|Df(\mathbf{z}) - \sum_{j=1}^N w_j\, f(\mathbf{x}_j)| = |DR(\mathbf{z}) - \sum_{j=1}^N w_j\, R(\mathbf{x}_j)| \leq \sum_{j=1}^N |w_j R(\mathbf{x}_j)|$$

By (26),

$$|R(\mathbf{x}_j)| \leq \sum_{|\alpha|=q} \frac{|(\mathbf{x}_j - \mathbf{z})^\alpha|}{\alpha!} \|\partial^\alpha f\|_{C(\Omega)}.$$

By (28) and Cauchy-Schwarz inequality,

$$\begin{aligned} \Big(\sum_{|\alpha|=q} \frac{|(\mathbf{x}_j - \mathbf{z})^\alpha|}{\alpha!} \|\partial^\alpha f\|_{C(\Omega)} \Big)^2 &\leq \sum_{|\alpha|=q} \frac{(\mathbf{x}_j - \mathbf{z})^{2\alpha}}{\alpha!} \sum_{|\alpha|=q} \frac{\|\partial^\alpha f\|_{C(\Omega)}^2}{\alpha!} \\ &= \frac{1}{q!} \|\mathbf{x}_j - \mathbf{z}\|_2^{2q} \sum_{|\alpha|=q} \frac{\|\partial^\alpha f\|_{C(\Omega)}^2}{\alpha!}, \end{aligned}$$

which implies

$$|R(\mathbf{x}_j)| \leq \|\mathbf{x}_j - \mathbf{z}\|_2^q \Big(\frac{1}{q!} \sum_{|\alpha|=q} \frac{1}{\alpha!} \|\partial^\alpha f\|_{\infty,\Omega}^2 \Big)^{1/2}, \tag{35}$$

and (34) follows. ∎

We now show that the bound (34) cannot be significantly improved for any numerical differentiation formula if the centres in $\mathbf{X}$ are well separated.

Theorem 11. *Given* $\mathbf{z}$ *and* $\mathbf{X}$*, let* $K \geq 1$ *be such that*

$$\|\mathbf{x}_j - \mathbf{z}\|_2 \leq K \operatorname{dist}(\mathbf{x}_j, \mathbf{X} \setminus \{\mathbf{x}_j\}), \quad j = 1, \ldots, N. \tag{36}$$

For any numerical differentiation formula (3) *there exists a function* $f \in C^\infty(\mathbb{R}^d)$ *such that*

$$|Df(\mathbf{z}) - \sum_{j=1}^N w_j\, f(\mathbf{x}_j)| \geq C|f|_{\infty,m+k,\Omega} \sum_{j=1}^N |w_j| \|\mathbf{x}_j - \mathbf{z}\|_2^{m+k}, \tag{37}$$

where C *depends only on* m, k, N, d *and* K*. Moreover, if* (3) *is polynomially consistent of order* m*, then*

$$|Df(\mathbf{z}) - \sum_{j=1}^N w_j\, f(\mathbf{x}_j)| \geq C\rho_{m+k,D}(\mathbf{z}, \mathbf{X})|f|_{\infty,m+k,\Omega}. \tag{38}$$

Proof. Denote by s_j, $j = 1, \ldots, N$, the distance from $\mathbf{x}_j$ to $\mathbf{X} \cup \{\mathbf{z}\} \setminus \{\mathbf{x}_j\}$ and consider the function

$$f(\mathbf{x}) = \sum_{\substack{j=1 \\ \mathbf{x}_j \neq \mathbf{z}}}^N \operatorname{sign}(w_j) \|\mathbf{x}_j - \mathbf{z}\|_2^{m+k} \phi\Big(\frac{\mathbf{x} - \mathbf{x}_j}{s_j}\Big),$$

where ϕ is a C^∞ 'bump function' supported in the unit disk in $\mathbb{R}^d$, for example

$$\phi(\mathbf{x}) = \begin{cases} e^{-1/(1-\|\mathbf{x}\|_2^2)} e, & \text{if } \|\mathbf{x}\|_2 \leq 1, \\ 0, & \text{otherwise.} \end{cases}$$

Then $Df(\mathbf{z}) = 0$ and

$$|Df(\mathbf{z}) - \sum_{j=1}^N w_j\, f(\mathbf{x}_j)| = \sum_{j=1}^N |w_j| \|\mathbf{x}_j - \mathbf{z}\|_2^{m+k}.$$

Since

$$\frac{\partial^{|\alpha|}}{\partial \mathbf{x}^\alpha} \phi\Big(\frac{\mathbf{x} - \mathbf{x}_j}{s_j}\Big) = \frac{1}{s_j^{m+k}} \partial^\alpha \phi\Big(\frac{\mathbf{x} - \mathbf{x}_j}{s_j}\Big), \quad |\alpha| = m + k,$$

it follows that

$$|f|_{\infty,m+k,\Omega} \leq \sum_{\substack{j=1 \\ \mathbf{x}_j \neq \mathbf{z}}}^N \Big(\frac{\|\mathbf{x}_j - \mathbf{z}\|_2}{s_j}\Big)^{m+k} |\phi|_{\infty,m+k,\mathbb{R}^d} \leq N K^{m+k} |\phi|_{\infty,m+k,\mathbb{R}^d},$$

which proves (37). The bound (38) follows immediately in view of Lemma 8. ∎

If the weight vector $\mathbf{w}$ is uniquely determined by the consistency conditions (4), then Theorem 10 and Lemma 8 lead to the upper bound in the form similar to (33),

$$|Df(\mathbf{z}) - \sum_{j=1}^{N} w_j\, f(\mathbf{x}_j)| \le \rho_{m+k,D}(\mathbf{z},\mathbf{X}) \frac{|f|_{\infty,m+k,\Omega}}{(m+k)!}, \tag{39}$$

which shows that up to the difference in the norm of f, the error in both cases is determined by the growth function.

Moreover, in the case when for given $\mathbf{z}, \mathbf{X}$ and m there are more than one polynomially consistent formula for the operator D, we consider the *minimal polynomially consistent formulas* which minimise the weighted ℓ_1-norm $\sum_{j=1}^{N} |w_j| \|\mathbf{x}_j - \mathbf{z}\|_2^{m+k}$ with respect to all weights $\mathbf{w}$ satisfying the consistency conditions (4). Note that such minimal formulas, with an addtional assumption of *positivity*, have been studied recently in [12]. The following statement follows immediately from Theorem 10 and Lemma 8.

Theorem 12. *Let (3) be a minimal polynomially consistent differentiation formula of order $m \ge 1$ for a linear differential operator D of order k. Then* (39) *holds for all $f \in C^{m+k}(\Omega)$, where $\Omega \subset \mathbb{R}^d$ is any domain that contains $S_{\mathbf{z},\mathbf{X}}$.*

As a consequence, we also obtain the following *sampling inequality.*

Corollary 13. *Let $\rho_{m+k,D}(\mathbf{z},\mathbf{X}) < \infty$ for a linear differential operator D of order k and some $m \ge 1$, and let $f \in C^{m+k}(\Omega)$ for some domain $\Omega \subset \mathbb{R}^d$ containing $S_{\mathbf{z},\mathbf{X}}$. If $f|_{\mathbf{X}} = 0$, then*

$$|Df(\mathbf{z})| \le \rho_{m+k,D}(\mathbf{z},\mathbf{X}) \frac{|f|_{\infty,m+k,\Omega}}{(m+k)!}. \tag{40}$$

Proof. Indeed, by Lemma 8 $\rho_{m+k,D}(\mathbf{z},\mathbf{X}) < \infty$ implies the existence of a polynomially consistent formula (3) of order m. By Theorem 12 a minimal formula satisfies (39), which implies (40) since $f|_{\mathbf{X}} = 0$. □

See [8] for a survey of sampling inequalities and their use.

If D is a *homogeneous* linear differential operator of order k, then any polynomially consistent numerical differentiation formula generates a whole family of polynomially consistent formulas by scaling. Indeed, assume that the formula (3) is consistent of order m. Then it is easy to see that for any $h > 0$ the formula

$$(Df)(\mathbf{z}) \approx \sum_{j=1}^{N} \frac{w_j}{h^k}\, f(\mathbf{x}_j^h), \qquad \mathbf{x}_j^h := \mathbf{z} + h(\mathbf{x}_j - \mathbf{z}) \tag{41}$$

is consistent of the same order, and the error bound (34) takes the form

$$|Df(\mathbf{z}) - \sum_{j=1}^{N} \frac{w_j}{h^k}\, f(\mathbf{x}_j^h)| \le h^m \frac{|f|_{\infty,m+k,\Omega^h}}{(m+k)!} \sum_{j=1}^{N} |w_j| \|\mathbf{x}_j - \mathbf{z}\|_2^{m+k},$$

where Ω^h contains $S_{\mathbf{z},\mathbf{X}^h}$, with $\mathbf{X}^h := \mathbf{z} + h(\mathbf{X} - \mathbf{z})$. Thus, for a minimal polynomially consistent formula,

$$|Df(\mathbf{z}) - \sum_{j=1}^{N} \frac{w_j}{h^k}\, f(\mathbf{x}_j^h)| \le h^m \rho_{m+k,D}(\mathbf{z},\mathbf{X}) \frac{|f|_{\infty,m+k,\Omega^h}}{(m+k)!}. \tag{42}$$

For a homogeneous D, we have

$$\rho_{q,D}(\mathbf{z}, \mathbf{X}^h) = h^{q-k}\rho_{q,D}(\mathbf{z}, \mathbf{X}). \tag{43}$$

Indeed, by setting $p_h(\mathbf{x}) = h^{-q}p(\mathbf{z} + h(\mathbf{x} - \mathbf{z}))$, we have $p(\mathbf{x}_j^h) = h^q p_h(\mathbf{x}_j)$, $\|\mathbf{x}_j^h - \mathbf{z}\|_2^q = h^q\|\mathbf{x}_j - \mathbf{z}\|_2^q$ and $Dp(\mathbf{z}) = h^{q-k}Dp_h(\mathbf{z})$, and hence

$$\begin{aligned}\rho_{q,D}(\mathbf{z}, \mathbf{X}^h) &= \sup\{Dp(\mathbf{z}) : p \in \Pi_q^d,\ |p(\mathbf{x}_j^h)| \le \|\mathbf{x}_j^h - \mathbf{z}\|_2^q\} \\ &= h^{q-k}\sup\{Dp_h(\mathbf{z}) : p_h \in \Pi_q^d,\ |p_h(\mathbf{x}_j)| \le \|\mathbf{x}_j - \mathbf{z}\|_2^q\}.\end{aligned}$$

Therefore we obtain from Theorem 9 for the scaled centres $\mathbf{X}^h$,

$$|Df(\mathbf{z}) - Dr_{\mathbf{X}^h,K,f}(\mathbf{z})| \le h^m \rho_{m+k,D}(\mathbf{z}, \mathbf{X})\frac{\sqrt{M_{K,m+k}}}{(m+k)!}\|f\|_K, \tag{44}$$

showing the same scaling of the error as in (42). However, the weights of the formulas generated by $r_{\mathbf{X}^h,K,f}$ will not be in general the scaled versions of the weights generated by $r_{\mathbf{X},K,f}$.

As an example consider the classical five point numerical differentiation formula for the Laplacian $D = \Delta$ in two variables. It is given by the centres

$$\mathbf{X} = \{\mathbf{x}_j\}_{j=1}^5 = \{\mathbf{z}, \mathbf{z} \pm (0,1), \mathbf{z} \pm (1,0)\}$$

and weights $w_1 = -4$, $w_2 = w_3 = w_4 = w_5 = 1$ and is exact for all cubic polynomials $p \in \Pi_4^2$. Then $\rho_{4,\Delta}(\mathbf{z}, \mathbf{X}) = 4$. Indeed, assuming $|p(\mathbf{x}_j)| \le \|\mathbf{x}_j - \mathbf{z}\|_2^4$, $j = 1, \ldots, 5$, we have $p(\mathbf{x}_1) = 0$, $|p(\mathbf{x}_j)| \le 1$, $j = 2, \ldots, 5$, which implies $\Delta p(\mathbf{z}) = \sum_{j=1}^5 w_j p(\mathbf{x}_j) \le 4$. On the other hand, $\tilde{p}(\mathbf{x}) = \|\mathbf{x}\|_2^2$ satisfies these conditions, and $\Delta\tilde{p}(\mathbf{z}) = 4$. Since $\sum_{j=1}^5 |w_j|\|\mathbf{x}_j - \mathbf{z}\|_2^4 = 4$, we see that this formula is minimal. Now (42) gives the following error bound for the scaled formulas

$$|\Delta f(\mathbf{z}) - \sum_{j=1}^5 \frac{w_j}{h^2} f(\mathbf{x}_j^h)| \le \frac{h^2}{6}|f|_{\infty,4,\Omega^h}.$$

Note that by using a standard error formula for the central difference approximation of the second derivative in one variable $h^{-2}\big(f(x+h) - 2f(x) + f(x-h)\big) - f''(x) = \frac{h^2}{12}f^{(4)}(\xi)$, for some $\xi \in (x-h, x+h)$, a slightly better estimate is obtained in the form

$$|\Delta f(\mathbf{z}) - \sum_{j=1}^5 \frac{w_j}{h^2} f(\mathbf{x}_j^h)| \le \frac{h^2}{12}\Big(\max_{|x_1 - z_1| \le h}\Big|\frac{\partial^4 f}{\partial x_1^4}(x_1, 0)\Big| + \max_{|x_2 - z_2| \le h}\Big|\frac{\partial^4 f}{\partial x_2^4}(0, x_2)\Big|\Big).$$

By (44) we have a similar bound for the error of the kernel based formulas on the same scaled centres assuming that $s \le 4$ and $M_{K,4} < \infty$,

$$|\Delta f(\mathbf{z}) - \Delta r_{\mathbf{X}^h,K,f}(\mathbf{z})| \le \frac{h^2}{6}\sqrt{M_{K,4}}\|f\|_K.$$

Note that this error bound cannot be obtained by the standard approach of estimating the error in terms of the fill distance [13, Section 11.2]. Indeed, the 5 centres are far

from being "unisolvent" for polynomials, or "sufficiently dense" in the sense of the hypothesis of a small fill distance arising in the standard literature on error bounds of kernel–based approximations. In fact, a simple calculation (see [3, Section 3]) shows that the arguments based on Markov inequality as in [13, Proposition 11.6] would predict the estimate $|\Delta f(\mathbf{z}) - \Delta r_{\mathbf{X}^h,K,f}(\mathbf{z})| = \mathcal{O}(h)$ and thus convergence of the approximate Laplacian as $h \to 0$ only if $\mathbf{X}$ consists of no less than hundreds of centers.

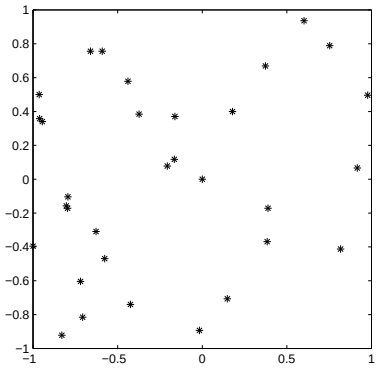
By comparing Theorem 9 with the estimates (39) and (38) we observe the following important feature of the kernel-based formulas: The estimate (33) is valid for *any* $q \geq \max\{s, k+1\}$ bounded above only by the smoothness class of the kernel. Therefore, the error is governed by $\rho_{q,D}(\mathbf{z}, \mathbf{X})$ with *optimal* q. Numerical examples in the next section illustrate that indeed the error of the kernel-based formula is comparable to the error of the minimal polynomial formula with optimal order of polynomial reproduction.

5 Numerical Examples

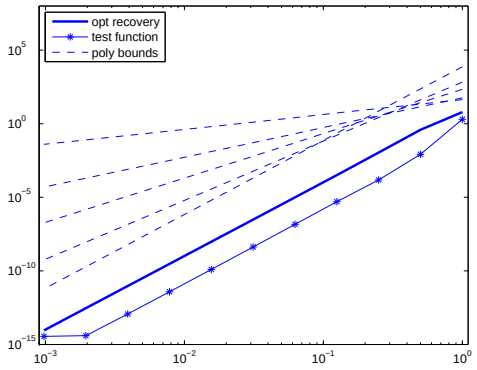
Our experiments are for the approximation of $\lambda f = \Delta f(\mathbf{z})$, where Δ is the Laplace operator in 2D and $\mathbf{z} \in \mathbb{R}^2$. We consider three sets $\hat{\mathbf{X}}_i \subset [-1,1]^2$, $i = 1,2,3$, of 32 points containing the origin, and generate numerical differentiation centres by scaling and translating these sets as $\mathbf{X}_i^h = \mathbf{z} + h\hat{\mathbf{X}}_i$, where $h = 2^{-n}$, $n = 0, \ldots, 10$. The set $\hat{\mathbf{X}}_1$ consists of the origin and 31 random points in $[-1,1]^2$ drawn from the uniform distribution, see Figure 1(a). The set $\hat{\mathbf{X}}_2$ includes 32 points on a straight line, a hyperbola and an ellipse, see Figures 1(c) and 2(a), perturbed (except of the point at the origin) randomly by at most 10^{-6} in both coordinate directions. The set $\hat{\mathbf{X}}_3$ includes 32 points on three parallel straight lines, see Figures 1(e) and 2(b), also perturbed in a similar way.

In the first example we consider kernel-based numerical differentiation using the Gaussian kernel $K(\mathbf{x}, \mathbf{y}) = e^{-\varepsilon^2\|\mathbf{x}-\mathbf{y}\|^2}$ with $\varepsilon = 2$. Figures 1(b)(d)(f) show the power function $P_{\lambda,\mathbf{X}}$, that is the error (23) of the optimal recovery on the unit ball of the Gaussian native space, the bounds (32) for $P_{\lambda,\mathbf{X}}$ obtained with various orders $q = 3, \ldots, 7$ (distinguishable by their slopes), and the maximum numerical differentiation error for the test function $f_1(x_1, x_2) = \frac{15^{1/4}}{2} e^{-5(x_1-0.1)^2-4x_2^2}$ over $\mathbf{z}$ in a uniform 21×21-grid in $\mathbf{z}_0 + [-0.1, 0.1]$, where $\mathbf{z}_0 = (0.23, 0.34)$. Note that $\|f\|_K = 1$ in the Gaussian ($\varepsilon = 2$) native space on $\mathbb{R}^2$, $\Delta f(\mathbf{z}_0) \approx -5.0748$ and the values of $\Delta f(\mathbf{z})$ vary between -11.066 and -0.11128 over the grid. Gaussian numerical differentiation formulas are obtained by solving (18) directly for $h = 1, \frac{1}{2}, \frac{1}{4}$, by the Gauss-QR method (see [5] and references therein) for $h = \frac{1}{8}, \frac{1}{16}, \frac{1}{32}$, and using the variable-precision arithmetic (between 32 and 64 digits) of MATLAB Symbolic Math Toolbox for the smaller values of h. For the evaluation of the power function we used the variable-precision arithmetic already for $h \leq \frac{1}{16}$ because of numerical instabilities due to cancellations inherent in (21). The constants $M_{K,q}$ in (32) are computed using $\Omega = \mathbf{z}_0 + h[-1.1, 1.1]^2$ in (30).

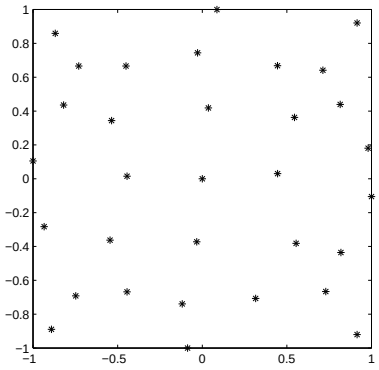
Since $\dim \Pi_7^2 = 28 < 32 < 36 = \dim \Pi_8^2$, the highest order we can expect for the growth function to be finite for 32 centres in general position is $q = 7$. The smallest q for which Theorem 9 applies to the Laplacian as a second order operator is $q = 3$. Therefore, we compute the bounds according to (32) for all $q = 3, \ldots, 7$. The corresponding dashed curves in Figure 1 are close to straight lines because the factor



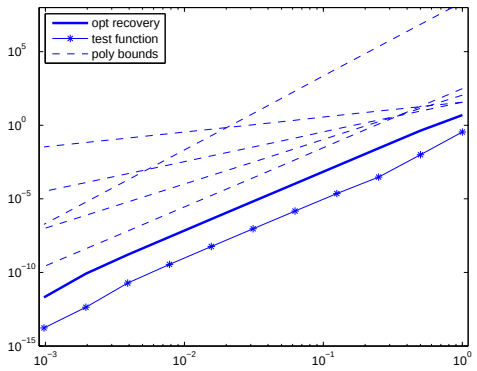
(a) $\hat{\mathbf{X}}_1$: random points



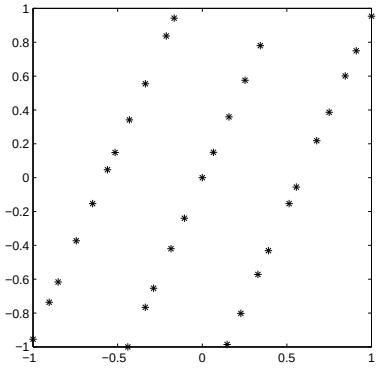
(b) Differentiation errors for $\mathbf{X}_1^h$



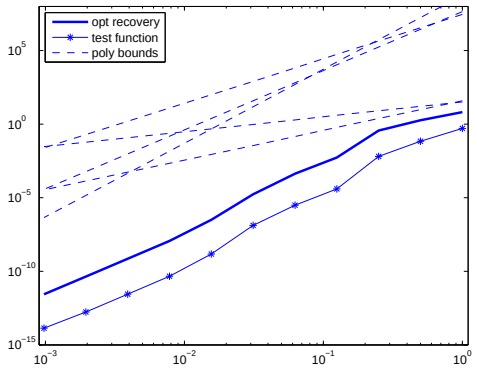
(c) $\hat{\mathbf{X}}_2$: near to a curve of degree 5



(d) Differentiation errors for $\mathbf{X}_2^h$

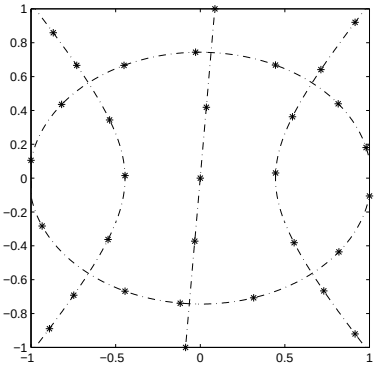


(e) $\hat{\mathbf{X}}_3$: near to a curve of degree 3

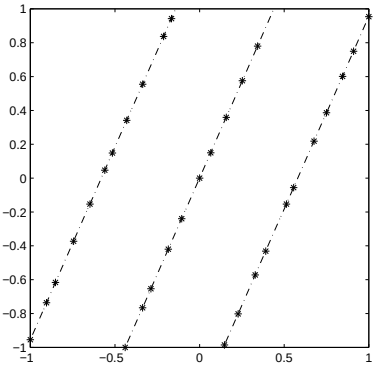


(f) Differentiation errors for $\mathbf{X}_3^h$

Figure 1: Numerical differentiation of the Laplacian with Gaussian formulas and bounds according to Theorem 9 for $\mathbf{X}_i^h$, $i = 1, 2, 3$. Left: Sets of centres. Right: Differentiation errors and bounds as functions of h, `opt recovery`: optimal recovery error (23); `test function`: differentiation error for the test function f_1; `poly bounds`: error bounds (32) for various q.



(a) $\hat{\mathbf{X}}_2$ and a curve of degree 5



(b) $\hat{\mathbf{X}}_3$ and a curve of degree 3

Figure 2: .

$M_{K,q}$ in (32) changes insignificantly with scaling whereas the growth function satisfies $\rho_{q,D}(\mathbf{z}, \mathbf{X}_i^h) = h^{q-2}\rho_{q,D}(\mathbf{z}, \hat{\mathbf{X}}_i)$, which generates distinctive slopes for the lines. We can see that the error curves in Figures 1(b)(d) generally follow the slope of the best bounding line which corresponds to $q = 7$ for $\hat{\mathbf{X}}_1$ and $q = 6$ for $\hat{\mathbf{X}}_2$. For the set $\hat{\mathbf{X}}_3$ in Figure 1(f) we can recognise a transition from the slope corresponding to the bound of order $q = 4$ for $h = 1, \frac{1}{2}, \frac{1}{4}$ to the slope of the bound of order $q = 6$ for $h \leq \frac{1}{128}$. The non-optimal behaviour for the sets $\hat{\mathbf{X}}_2$ and $\hat{\mathbf{X}}_3$ can be explained by their proximity to lower order algebraic curves, see Figure 2.

In the second example we compare the errors of the numerical differentiation of the Laplacian $\lambda f = \Delta f(\mathbf{z})$ by several kernel-based formulas and the minimal polynomially consistent formulas of various orders for functions in the Sobolev space $H^6(\mathbb{R}^2)$ on the above sets of centres $\mathbf{X}_i^h = \mathbf{z} + h\hat{\mathbf{X}}_i$. Recall that $H^s(\mathbb{R}^d) = W_2^s(\mathbb{R}^d)$ coincides with the native space of the *Matérn kernel*

$$M_{s,d}(\mathbf{x}, \mathbf{y}) := \frac{\mathcal{K}_{s-d/2}(\|\mathbf{x} - \mathbf{y}\|)\|\mathbf{x} - \mathbf{y}\|^{s-d/2}}{2^{s-1}\Gamma(s)}, \quad s > d/2,$$

where $\mathcal{K}_\nu(\mathbf{x})$ denotes the modified Bessel function of second kind. Since $M_{s,d}(\mathbf{x}, \mathbf{y}) = \Phi(\mathbf{x} - \mathbf{y})$ with $\widehat{\Phi}(\boldsymbol{\omega}) = (1 + \|\boldsymbol{\omega}\|^2)^{-s}$, it follows that

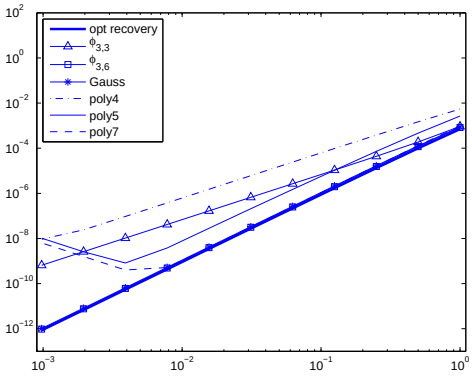
$$\|f\|^2_{H^s(\mathbb{R}^d)} := (2\pi)^{-d/2} \int_{\mathbb{R}^d} |\hat{f}(\boldsymbol{\omega})|^2 (1 + \|\boldsymbol{\omega}\|^2)^s \, \mathbf{d}\boldsymbol{\omega} = \|f\|^2_{M_{s,d}}.$$

Hence the kernel-based numerical differentiation formula (17) obtained using the Matérn kernel $M_{6,2}$ delivers the optimal recovery (23) on $H^6(\mathbb{R}^2)$. In addition to the Matérn kernel $M_{6,2}$ we consider kernel-based formulas generated by the Gaussian kernel $K(\mathbf{x}, \mathbf{y}) = e^{-\varepsilon^2\|\mathbf{x}-\mathbf{y}\|^2}$ with $\varepsilon = 0.5$, and by two Wendland kernels, $K(\mathbf{x}, \mathbf{y}) = \phi_{3,3}(\varepsilon\|\mathbf{x} - \mathbf{y}\|)$ and $K(\mathbf{x}, \mathbf{y}) = \phi_{3,6}(\varepsilon\|\mathbf{x} - \mathbf{y}\|)$ [13, Section 9.4],

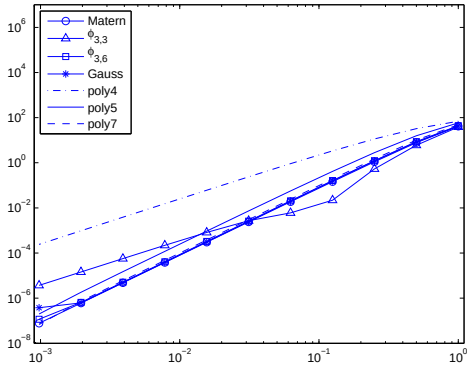
$$\phi_{3,3}(r) = (1 - r)_+^8(32r^3 + 25r^2 + 8r + 1),$$
$$\phi_{3,6}(r) = (1 - r)_+^{14}(46189r^6 + 73206r^5 + 54915r^4 + 24500r^3 + 6755r^2 + 1078r + 77),$$

both with $\varepsilon = 0.07$. Note that $\phi_{3,3}$ is 6, and $\phi_{3,6}$ 12 times continuously differentiable. Moreover, we also compute the error of the minimal polynomially consistent formulas of various orders considered in Section 4. Figures 3(a)(c)(e) present the maximum approximation error (22) of each numerical differentiation formula on the space $H^6(\mathbb{R}^2)$ computed using the functional $Q_{\lambda,\mathbf{X}}(\mathbf{w})$ of (21) for the Matérn kernel $M_{6,2}$. In particular, the weights obtained with the Matérn kernel itself provide the error of the optimal recovery on $H^6(\mathbb{R}^2)$. Note that $Q_{\lambda,\mathbf{X}}(\mathbf{w})$ is independent of the choice of $\mathbf{z}$ because the Matérn kernel is translation invariant. In Figures 3(b)(d)(f) we present the errors of the same formulas for the approximation of $\Delta f_2(0,0)$ for the test function $f_2(x_1,x_2) = \phi_{3,2}(\sqrt{x_1^2+x_2^2})$, where $\phi_{3,2}(r) = (1-r)^6(35r^2+18r+3)$ is another Wendland function. Note that $f_2 \in W_\infty^5(\mathbb{R}^2)$ and $H^6(\mathbb{R}^2)$ is continuously embedded in $W_\infty^5(\mathbb{R}^2)$. We used the variable-precision arithmetic for the weights of the Wendland and Matérn kernel-based differentiation formulas, and for the computation of the errors. However, the weights of the minimal polynomial formulas are computed in the standard double precision arithmetic, which explains the erroneous behaviour of some of the polynomial curves in Figures 3(a)(c)(e) for $h \leq \frac{1}{128}$.

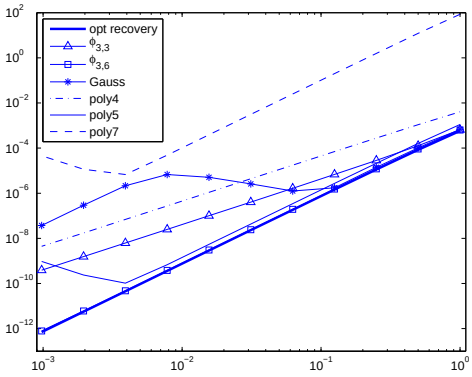
The results in Figure 3 indicate an h^3 order of the optimal recovery error for both $\mathbf{X}_1^h$ and $\mathbf{X}_2^h$ and h^2 for $\mathbf{X}_3^h$. Clearly, h^3 is indeed the best possible order for the recovery of the pointwise value of the Laplacian as a second order operator for functions in $H^6(\mathbb{R}^2)$. This is contrasted with the h^5 and h^4 orders achieved on $\mathbf{X}_1^h$ and $\mathbf{X}_2^h$, respectively, on the native space of the Gaussian in our first experiments, see Figures 1(b)(d). The order h^2 for $\mathbf{X}_3^h$ is explained by the proximity of this set to a curve of degree 3. In contrast to Figure 1(f), no improvement of the slope of the optimal error curve is seen for small h. Note that $\partial^{\alpha,\beta} M_{6,2}(\mathbf{x},\mathbf{y})$ is discontinuous when $\mathbf{x} = \mathbf{y}$ if $|\alpha|+|\beta| = 10$, and hence Theorem 9 only implies an estimate of order h^2 for the kernel-based numerical differentiation with $M_{6,2}$ for centres in a good position. For all formulas the error for the test function f_2 and the maximum error on $H^6(\mathbb{R}^2)$ behave strikingly similar, with the exception that Wendland function-based formulas seem advantageous for f_2 when h is relatively large, which might be related to the fact that f_2 itself belongs to the family of Wendland functions with the distinctive piecewise polynomial structure. Whereas the Matérn kernel-based differentiation formulas seem to be the best performers in all cases, the results delivered by the other methods strongly depend on the character of the centres. The best performing polynomial method seems to be the one which corresponds to the order for which the growth functions are the smallest: $q = 7$ for $\mathbf{X}_1^h$, $q = 6$ for $\mathbf{X}_2^h$ and $q = 4$ for $\mathbf{X}_3^h$. (Recall that the dashed lines in Figure 1(b)(d)(f) represent the growth functions of the orders $q = 3,\dots,7$ multiplied by the a factor slowly varying with h. Note that in the second experiment the curves for the polynomial methods with $q = 5$ and $q = 6$ are very close and we decided to only include $q = 5$ in Figure 3.) Although the Gaussian kernel with the chosen shape parameter $\varepsilon = 0.5$ gives almost optimal results for the random centres in $\mathbf{X}_1^h$, for the more problematic centres its error curve seems to 'gravitate' towards a higher order polynomial error curve ($q = 7$ for $\mathbf{X}_2^h$ and $q = 5$ for $\mathbf{X}_3^h$), giving rise to a particularly bizarre shape in Figures 3(c)(d) where this behaviour only manifests itself for $h \leq \frac{1}{128}$, whereas for $h \geq \frac{1}{8}$ the Gaussian formulas are almost optimal, and there is a transition region for $\frac{1}{128} \leq h \leq \frac{1}{8}$, where the error is actually larger for smaller h. There are similar patterns in the error curves for $\phi_{3,6}$ on $\mathbf{X}_3^h$ with a gravitaton towards the polynomial curve with $q = 5$. The slope of the error curves for



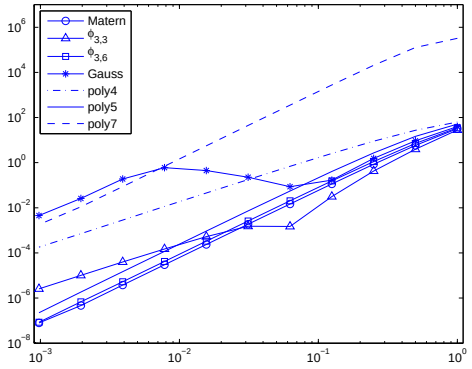
(a) $\mathbf{X}_1^h$: error on $H^6(\mathbb{R}^2)$



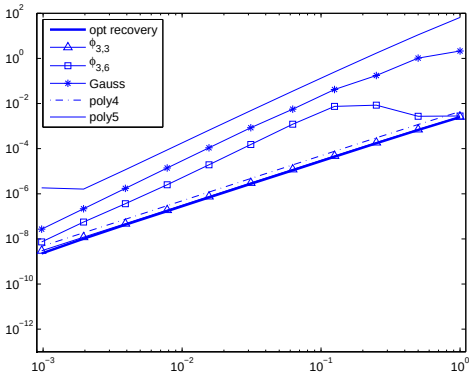
(b) $\mathbf{X}_1^h$: error for test function f_2



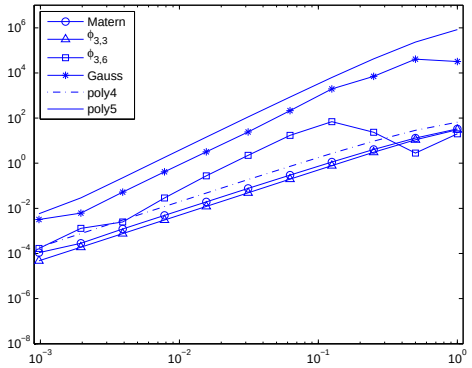
(c) $\mathbf{X}_2^h$: error on $H^6(\mathbb{R}^2)$



(d) $\mathbf{X}_2^h$: error for test function f_2



(e) $\mathbf{X}_3^h$: error on $H^6(\mathbb{R}^2)$



(f) $\mathbf{X}_3^h$: error for test function f_2

Figure 3: Error of numerical differentiation of the Laplacian on the Sobolev space $H^6(\mathbb{R}^2)$ (left) and for the test function f_2 using centres in $\mathbf{X}_i^h$, $i = 1, 2, 3$ (right), as function of h. The curves correspond to the weights obtained by different methods. `opt recovery`, `Matern`: Matérn kernel, which provides optimal recovery on $H^6(\mathbb{R}^2)$; $\phi_{3,3}, \phi_{3,6}$: Wendland kernels with $\varepsilon = 0.07$; `Gauss`: Gaussian kernel with $\varepsilon = 0.5$; `poly4, poly5, poly7`: minimal polynomial formulas of corresponding orders.

$\phi_{3,3}$ resembles that of the polynomial curves with $q = 4$, giving the approximation order h^2 on all sets. We see that in all cases one of the two Wendland kernel-based formulas with the shape parameter $\varepsilon = 0.07$ is almost optimal ($\phi_{3,6}$ for $\mathbf{X}_1^h$ and $\mathbf{X}_2^h$ and $\phi_{3,3}$ for $\mathbf{X}_3^h$).

Further examples along the lines of Figure 3 are in [10, Section 9, Table 1, Figures 3 to 6]. There, the error norms of various differentiation formulas, including those with polynomial consistency and optimized weights, are not only compared on Sobolev space, but also on Beppo–Levi spaces. The latter spaces are well–adapted to formulas with polynomial consistency.

Acknowledgement

The authors are grateful to the Alexander von Humboldt Foundation for the financial support that helped them to get together and work on this project for several days in September 2011.

References

[1] R. Beatson, O. Davydov, and J. Levesley. Error bounds for anisotropic RBF interpolation. *Journal of Approximation Theory*, 162:512–527, 2010.

[2] T. Belytschko, Y. Krongauz, D. Organ, M. Fleming, and P. Krysl. Meshless methods: an overview and recent developments. *Computer Methods in Applied Mechanics and Engineering, special issue*, 139:3–47, 1996.

[3] O. Davydov. Error bound for radial basis interpolation in terms of a growth function. In A. Cohen, J. L. Merrien, and L. L. Schumaker, editors, *Curve and Surface Fitting: Avignon 2006*, pages 121–130. Nashboro Press, Brentwood, 2007.

[4] O. Davydov and D. T. Oanh. Adaptive meshless centres and RBF stencils for Poisson equation. *J. Comput. Phys.*, 230:287–304, 2011.

[5] O. Davydov and D. T. Oanh. On the optimal shape parameter for Gaussian radial basis function finite difference approximation of the Poisson equation. *Comput. Math. Appl.*, 62:2143–2161, 2011.

[6] M. A. García-March, M. Arevalillo-Herráez, F. R. Villatoro, F. Giménez, and P. Fernández de Córdoba. A generalized finite difference method using Coatmèlec lattices. *Comput. Phys. Comm.*, 180(7):1125–1133, 2009.

[7] K. Jetter, J. Stöckler, and J. Ward. Error estimates for scattered data interpolation on spheres. *Mathematics of Computation*, 68:733–747, 1999.

[8] C. Rieger, B. Zwicknagl, and R. Schaback. Sampling and stability. In M. Dæhlen, M. Floater, T. Lyche, J.-L. Merrien, K. Mørken, and L. Schumaker, editors, *Mathematical Methods for Curves and Surfaces*, volume 5862 of *Lecture Notes in Computer Science*, pages 347–369, 2010.

[9] R. Schaback. Native Hilbert spaces for radial basis functions I. In M. Buhmann, D. H. Mache, M. Felten, and M. Müller, editors, *New Developments in Approximation Theory*, number 132 in International Series of Numerical Mathematics, pages 255–282. Birkhäuser Verlag, 1999.

[10] R. Schaback. Direct discretizations with applications to meshless methods for PDEs. submitted, http://www.num.math.uni-goettingen.de/schaback/research/group.html, 2013.

[11] R. Schaback and H. Wendland. Kernel techniques: from machine learning to meshless methods. *Acta Numerica*, 15:543–639, 2006.

[12] B. Seibold. Minimal positive stencils in meshfree finite difference methods for the Poisson equation. *Comput. Methods Appl. Mech. Eng.*, 198(3-4):592–601, 2008.

[13] H. Wendland. *Scattered Data Approximation*. Cambridge University Press, 2005.

[14] Z. M. Wu and R. Schaback. Local error estimates for radial basis function interpolation of scattered data. *IMA J. Numer. Anal.*, 13(1):13–27, 1993.