Interpolation with the polynomial kernels

Giacomo Elefante[a] · Wolfgang Erb[a] · Francesco Marchetti[a] · Emma Perracchione[b] · Davide Poggiali[c] · Gabriele Santin[d]

Abstract

The polynomial kernels are widely used in machine learning and they are one of the default choices to develop kernel-based classification and regression models. However, they are rarely used and considered in numerical analysis due to their lack of strict positive definiteness. In particular they do not enjoy the usual property of unisolvency for arbitrary point sets, which is one of the key properties used to build kernel-based interpolation methods.

This paper is devoted to establish some initial results for the study of these kernels, and their related interpolation algorithms, in the context of approximation theory. We will first prove necessary and sufficient conditions on point sets which guarantee the existence and uniqueness of an interpolant. We will then study the Reproducing Kernel Hilbert Spaces (or native spaces) of these kernels and their norms, and provide inclusion relations between spaces corresponding to different kernel parameters. With these spaces at hand, it will be further possible to derive generic error estimates which apply to sufficiently smooth functions, thus escaping the native space. Finally, we will show how to employ an efficient stable algorithm to these kernels to obtain accurate interpolants, and we will test them in some numerical experiment. After this analysis several computational and theoretical aspects remain open, and we will outline possible further research directions in a concluding section.

This work builds some bridges between kernel and polynomial interpolation, two topics to which the authors, to different extents, have been introduced under the supervision or through the work of Stefano De Marchi. For this reason, they wish to dedicate this work to him in the occasion of his 60th birthday.

1 Introduction

Positive definite kernels are widely used in a variety of problems ranging from numerical analysis to machine learning, including Gaussian process regression.

In different fields they come into play from different directions. In numerical analysis, they provided data-dependent bases that permit interpolation of scattered data [7, 15, 37]; in machine learning, they are usually the result of the application of a feature map on the input data, used to transform linear algorithms into nonlinear ones by means of an high dimensional space [33, 34, 35]; in Gaussian process regression, they represent the covariance function of a stochastic process [29].

Despite this remarkable variety, a large part of the success of kernel methods is due to the fact that they can be analyzed to some extent within the unified framework of Reproducing Kernel Hilbert Spaces (RKHS) [31], which provide a solid mathematical underpinning to different algorithmic approaches. This connection has the additional benefit that novel ideas and points of view may often spread from one field to another through this common perspective, see e.g. [1, 21, 23, 26, 32] for a few recent examples.

However, it is still the case that some requirements and conditions are peculiar to one or the other specific setting, and thus this spill over is not always possible. In particular, a major difference between the point of view of machine learning and Gaussian process on one hand, and the one of numerical analysis on the other, is the definiteness of the kernel in the following sense.

Definition 1.1 (Definiteness classes). Let $\Omega \neq \emptyset$ be a set and let $k : \Omega \times \Omega \to \mathbb{R}$ be symmetric. Then k is said to be positive definite on Ω if for all $N \in \mathbb{N}$ and for all sets $X := \{x_i\}_{i=1}^N \subset \Omega$ the kernel matrix $A := (k(x_i, x_j))_{i,j=1}^N \in \mathbb{R}^{N \times N}$ is positive semidefinite. The kernel is additionally said to be strictly positive definite if A is positive definite whenever the points in X are pairwise distinct.

In the analysis of stochastic processes kernels are used as covariance functions, which are in general only positive definite. In machine learning, data approximation models are usually defined as the solution of an optimization problem, which can be proven to be convex even if the employed kernel is just positive definite. The approximation models considered in numerical analysis are instead mostly interpolatory, and their existence is guaranteed for general distributions of the interpolation points only if the kernel is strictly positive definite. Namely, given an input space $\Omega \neq \emptyset$, a set $X := \{x_i\}_{i=1}^N \subset \Omega$ of pairwise distinct interpolation points, and a set $Y := \{y_i\}_{i=1}^N \subset \mathbb{R}$ of target values, a kernel is used to build an interpolatory model

$$s(x) := \sum_{i=1}^N c_i k(x, x_i), \ x \in \Omega, \qquad s(x_i) = y_i, \ \text{for all } 1 \le i \le N. \tag{1}$$

[a]Dipartimento di Matematica "Tullio Levi-Civita", Università degli Studi di Padova (Italy)
[b]Dipartimento di Scienze Matematiche "Giuseppe Luigi Lagrange", Politecnico di Torino (Italy)
[c]FAR Networks S.r.l., Cernusco S.N., Milano (Italy)
[d]Digital Society Center (DIGIS), Fondazione Bruno Kessler, Trento (Italy)

This model exists precisely when there exists a vector $c := (c_1, \ldots, c_N)^T \in \mathbb{R}^N$ which solves

$$Ac = y, \tag{2}$$

with $y := (y_1, \ldots, y_N)^T \in \mathbb{R}^N$ and A as in Definition 1.1. The solvability of this system is in turn guaranteed if A is positive definite and thus invertible for any set X, i.e., if k is strictly positive definite. In other words, strictly positive definite kernels are used in numerical analysis to construct data-dependent interpolation models of the form (1), thus enabling the interpolation of arbitrarily scattered data for arbitrary input space dimension. This capability in particular permits to overcome the limitations of classical techniques such as polynomial interpolation, which require instead precise geometrical conditions on the interpolation points.

This distinction has the effect that many kernels which are commonly employed in machine learning and Gaussian process regression are almost unknown in approximation theory and numerical analysis, since they are only positive definite. In particular, in this paper we consider the notable family of polynomial kernels $k_{a,p}(x, y) := (a + \langle x, y \rangle)^p$ defined for $a \geq 0$ and $p \in \mathbb{N}$ on a subset $\Omega \neq \emptyset$ of the Euclidean space $\mathbb{R}^d$. These kernels are widely used in machine learning, where they are often even considered the essential basic example of a positive definite kernel. For example, together with the Gaussian kernel, polynomial kernels are the only ones implemented by default in the widespread Scikit-Learn Python machine learning library [8, 28], as well as in the Matlab Statistics and Machine Learning Toolbox [36].

Due to their lack of strict positive definiteness, they have however received little attention in approximation theory. For this reason in this work we aim at establishing an interpolation theory for these polynomial kernels. In particular, after recalling some additional details on their definition and properties in Section 2, we characterize in Section 3 sets of points X which are unisolvent for the polynomial kernels, i.e., which allow unique interpolation. As proven in Theorem 3.5, we obtain the remarkable result that any set of pairwise distinct points in $\mathbb{R}^d$ is unisolvent provided p is chosen large enough. After the existence of an interpolant is established, we obtain in Section 4 an error bound for its approximation error. This in particular allows one to "escape the native space" in the sense of [25], i.e., approximating functions which are outside of the RKHS of the kernel. In the same section we also use an argument of [39] to provide a characterization of the RKHS, which turns out to be simply the space $\mathbb{P}_p^d(\Omega)$ of d variate polynomials of degree p over Ω or the corresponding homogeneous space $\mathbb{H}_p^d(\Omega)$, each equipped with a suitable inner product. Moreover, we study the stability of this interpolation process in Section 4.2, and show that, although the direct solution of the linear system (2) is possibly highly unstable, one can apply the celebrated RBF-QR algorithm [17, 18] to $k_{a,p}$, thus obtaining stable computations. From this analysis of existence, convergence, and stability it turns out that polynomial-kernel interpolation is strictly related to standard polynomial interpolation, perhaps unsurprisingly. In particular, we argue that good interpolation points for these kernels can be found from good interpolation points for polynomial interpolation, such as [3, 4, 5, 6, 9, 10]. Finally, we test our findings in a number of experiments in Section 6, and comment on some possible extensions in Section 7.

2 Background on polynomials and the polynomial kernels

2.1 Multivariate polynomial spaces

We start by recalling some notation and the necessary background results on multivariate polynomials.

Let $d \in \mathbb{N}$ and $p \in \mathbb{N}$. Given a multiindex $\zeta := (\zeta_1, \ldots, \zeta_d) \in \mathbb{N}_0^d$, we write $|\zeta| := \zeta_1 + \ldots + \zeta_d$ for its length and $\zeta! := \zeta_1! \cdot \ldots \cdot \zeta_d!$ for its factorial, and denote the monomial with degrees ζ as $x^\zeta := \prod_{i=1}^d x_i^{\zeta_i}$, $x \in \mathbb{R}^d$. For two multiindices $\zeta, \beta \in \mathbb{N}_0^d$, the term $\delta_{\zeta\beta}$ has value 1 when $\zeta_i = \beta_i$ for all $1 \leq i \leq d$, and otherwise it has value zero.

We denote as $\mathbb{P}_p^d$ the space of polynomials over $\mathbb{R}^d$ of total degree at most p, and as $\mathbb{H}_p^d$ the corresponding homogeneous space, i.e., the space of polynomials over $\mathbb{R}^d$ of total degree exactly p. The two spaces have dimension $M_p^d := \binom{d+p}{d} = \dim(\mathbb{P}_p^d)$ and $M_p^{d-1} = \dim(\mathbb{H}_p^d)$, respectively.

Using a notation that will be motivated in the next section, for any $a \geq 0$ we denote two sets of multi-indices that we will use repeatedly as

$$I_a(p, d) := \begin{cases} \left\{ \zeta \in \mathbb{N}_0^d, |\zeta| \leq p \right\}, & a > 0 \\ \left\{ \zeta \in \mathbb{N}_0^d, |\zeta| = p \right\}, & a = 0, \end{cases} \tag{3}$$

and set

$$M_a := \begin{cases} M_p^d = \dim(\mathbb{P}_p^d), & a > 0 \\ M_p^{d-1} = \dim(\mathbb{H}_p^d), & a = 0, \end{cases} \tag{4}$$

so that $M_a = \dim(I_a(p, d))$.

2.2 The polynomial kernels

With these notations in hand, we can now give a formal definition of the family of polynomial kernels and state some of their properties. The content of this section is a collection of classical results, for which we refer e.g. to [33, 34, 35].

Let $d \in \mathbb{N}$, $\Omega \subset \mathbb{R}^d$, and let $a \geq 0, p \in \mathbb{N}$ be fixed. The polynomial kernel $k := k_{a,p} : \Omega \times \Omega \to \mathbb{R}$ is defined as $k(x, y) := (a + \langle x, y \rangle)^p$, $x, y \in \Omega$, where $\langle x, y \rangle$ is the inner product on $\mathbb{R}^d$.

If $a > 0$, using the notation (3) and the multinomial expansion we have

$$k_{a,p}(x, y) = \left(a + \sum_{i=1}^d x_i y_i \right)^p = \sum_{\substack{(\zeta_0, \zeta) \in \mathbb{N}_0^{d+1} \\ \zeta_0 + |\zeta| = p}} \frac{p! a^{\zeta_0}}{\zeta_0! \zeta!} x^\zeta y^\zeta = \sum_{\zeta \in \mathbb{N}_0^d, |\zeta| \leq p} \frac{p! a^{p - |\zeta|}}{(p - |\zeta|)! \zeta!} x^\zeta y^\zeta = \sum_{\zeta \in I_a(p,d)} d_\zeta^a x^\zeta y^\zeta, \tag{5}$$

where for $a > 0$ we defined

$$d_\zeta^a := \frac{p! a^{p-|\zeta|}}{(p-|\zeta|)! \zeta!}, \quad \zeta \in I_a(p,d). \tag{6}$$

Considering an arbitrary enumeration $\{\zeta^{(1)}, \ldots, \zeta^{(M_a)}\}$ of the set $I_a(p,d)$, with M_a as defined in (4), the representation (5) shows that the function $\Phi_{a,p} : \mathbb{R}^d \to \mathbb{R}^{M_a}$ defined for $a > 0$ by

$$\Phi_{a,p}(x) := \left[\sqrt{d_{\zeta^{(1)}}^a} \, x^{\zeta^{(1)}}, \sqrt{d_{\zeta^{(2)}}^a} \, x^{\zeta^{(2)}}, \ldots, \sqrt{d_{\zeta^{(M_a)}}^a} \, x^{\zeta^{(M_a)}} \right]^T, \quad x \in \mathbb{R}^d, \tag{7}$$

is a feature map for $k_{a,p}$ on $\mathbb{R}^d$ with feature space $\mathbb{R}^{M_a}$, i.e., it holds $k_{a,p}(x,y) = \Phi_{a,p}(x)^T \Phi_{a,p}(y)$ for all $x, y \in \mathbb{R}^d$.

If instead $a = 0$, again with the notation (3) the same chain of reasoning as above proves that

$$k_{0,p}(x,y) = \sum_{\zeta \in I_0(p,d)} d_\zeta^0 x^\zeta y^\zeta, \qquad d_\zeta^0 := \frac{p!}{\zeta!}, \ \zeta \in I_0(p,d). \tag{8}$$

In this case, taking an arbitrary enumeration $\{\zeta^{(1)}, \ldots, \zeta^{(M_0)}\}$ of this set, with M_0 as in (4), we obtain as before that a feature map for $k_{0,p}$ is $\Phi_{0,p} : \mathbb{R}^d \to \mathbb{R}^{M_0}$ with

$$\Phi_{0,p}(x) := \left[\sqrt{d_{\zeta^{(1)}}^0} \, x^{\zeta^{(1)}}, \sqrt{d_{\zeta^{(2)}}^0} \, x^{\zeta^{(2)}}, \ldots, \sqrt{d_{\zeta^{(M_0)}}^0} \, x^{\zeta^{(M_0)}} \right]^T, \quad x \in \Omega. \tag{9}$$

Both for $a > 0$ and $a = 0$, the existence of a feature map implies that the polynomial kernel is positive definite, i.e., for each $X := \{x_i\}_{i=1}^N \subset \Omega$ the kernel matrix $A := (k_{a,p}(x_i, x_j))_{i,j=1}^N \in \mathbb{R}^{N \times N}$ is positive semidefinite. This is immediate from Definition 1.1 since A is the Gramian matrix of the N vectors $\{\Phi_{a,p}(x) : x \in X\}$. On the other hand, since the image of this feature map is an M_a-dimensional feature space, and $M_a < \infty$, the kernel $k_{a,p}$ is not strictly positive definite, i.e., the matrix A may be singular even for pairwise distinct points X. In particular, if $N > M_a$ there exists no set $X \in \mathbb{R}^d$ of N points such that the kernel matrix is non singular, since this would require the N vectors $\{\Phi_{a,p}(x) : x \in X\}$ to be linearly independent in an $M_a < N$ dimensional space.

We recall moreover (see e.g. [31]) that each positive definite kernel $k : \Omega \times \Omega \to \mathbb{R}$ is associated to a RKHS $(\mathcal{H}, \langle \cdot, \cdot \rangle_{\mathcal{H}})$, which is a Hilbert space of functions from Ω to $\mathbb{R}$, where the kernel acts as a reproducing kernel, i.e., it holds

- $k(\cdot, x) \in \Omega$ for all $x \in \Omega$
- $\langle k(\cdot, x), f \rangle_{\mathcal{H}} = f(x)$ for all $x \in \Omega$ and $f \in \mathcal{H}$,

and that this RKHS is unique given a positive definite kernel k and a set Ω.

The RKHS of a kernel is often called its native space in the approximation theory literature. We denote as $\mathcal{H}_{a,p} := \mathcal{H}_{a,p}(\Omega)$ the native space of $k_{a,p}$ on Ω, and we will discuss its characterization in Section 4.

Remark 1. We remark that there are possible extensions to the definition of the polynomial kernel that we use in this section. Most notably, one may consider more general sets Ω, not necessarily in the Euclidean space, and replace $\langle x, y \rangle$ with a corresponding inner product on Ω. Although this extension is of potential interest, we do not consider it in this paper.

3 Existence and characterization of unisolvent sets

We start by analyzing conditions on a set of interpolation points that guarantee the existence of a unique polynomial kernel interpolant. As recalled in Section 1, the kernel interpolant (1) exists and is unique whenever the linear system (2) has a unique solution, i.e., when the kernel matrix of $k_{a,p}$ on X is invertible, i.e., positive definite since it is positive semidefinite by construction. We will thus investigate conditions for the invertibility of this kernel matrix.

To derive our characterization we are going to use some relations with interpolation points for classical polynomials. To this end, we recall the following definition.

Definition 3.1 (Unisolvent set). Let $\Omega \subset \mathbb{R}^d$ and let $U \subset C(\Omega)$ be a finite dimensional linear space of continuous functions. A set of $N := \dim(U)$ points $X := \{x_i\}_{i=1}^N \subset \Omega$ is said to be U-unisolvent if one of the following equivalent conditions hold:

(i) For each $Y := \{y_i\}_{i=1}^N$ there exists a unique element $u \in U$ which interpolates Y on X, i.e, $u(x_i) = y_i$, $1 \le i \le N$.

(ii) If $u \in U$ and $u_{|X} = 0$, then $u = 0$.

(iii) If $\{u_1, \ldots, u_N\}$ is a basis of U, then the matrix $V := V\left(\{u_j\}_{j=1}^N, X\right) := [u_j(x_i)]_{i,j=1}^N \in \mathbb{R}^{N \times N}$ is invertible.

We furthermore recall that the matrix V defined in point (iii) is called a Vandermonde matrix if $U = \mathbb{P}_d^p$ for some $p, d \in \mathbb{N}$, and if $\{u_1, \ldots, u_N\}$ is any enumeration of the monomial basis of this space.

The fundamental step to derive the results of this section is the following simple lemma, which establishes a connection between the kernel matrix and a rectangular Vandermonde matrix.

Lemma 3.1. *Let $p \in \mathbb{N}$ and $a \ge 0$, and let $\left\{\zeta^{(i)}\right\}_{i=1}^{M_a}$ be an enumeration of $I_a(p,d)$. Let furthermore $X_N \subset \Omega$ be a set of $N \le M_a$ pairwise distinct points and let $V \in \mathbb{R}^{N \times M}$ be the Vandermonde matrix given by the evaluation of the monomials $\{x^{\zeta^{(i)}}\}_{i=1}^{M_a}$ on X_N, with columns ordered according to the chosen enumeration, i.e.,*

$$V := \begin{bmatrix} x_1^{\zeta^{(1)}} & \cdots & x_1^{\zeta^{(M_a)}} \\ \vdots & \ddots & \vdots \\ x_N^{\zeta^{(1)}} & \cdots & x_N^{\zeta^{(M_a)}} \end{bmatrix}. \tag{10}$$

Then for all $a > 0$ the kernel matrix A of $k_{a,p}$ on X_N satisfies

$$A = VDV^T,$$

where $D := \mathrm{diag}\left(d^a_{\zeta^{(1)}}, \ldots, d^a_{\zeta^{(M_a)}} \right)$ and d^a_ζ is defined as in (6) *for $a > 0$ and as in* (8) *for $a = 0$.*

Proof. We consider the case $a > 0$, since for $a = 0$ the argument is the same. The result follows by direct computation using the representation (5) of the kernel. Indeed, using the definition of V, D given in the statement we have for all $1 \le i, j \le N$ that

$$A_{ij} = k_{a,p}(x_i, x_j) = \sum_{\zeta \in I(p,d)} d^a_\zeta x_i^\zeta x_j^\zeta = \sum_{\ell=1}^{M_a} d^a_{\zeta^{(\ell)}} x_i^{\zeta^{(\ell)}} x_j^{\zeta^{(\ell)}} = \sum_{\ell=1}^{M_a} d^a_\ell V_{i\ell} V_{j\ell} = (VDV^T)_{ij},$$

which is the desired representation. □

We furthermore need the following result which shows that, under a certain rank condition, any set of points can be completed to a unisolvent set. Observe that for our purposes it is sufficient to prove that the set X_M defined in the lemma is in $\mathbb{R}^d$, so we do not put much care in constraining its location into a smaller compact subset of $\mathbb{R}^d$.

Lemma 3.2. *Let $M, N \in \mathbb{N}$, $N \le M$, $U := \mathrm{span}\{u_i\}_{i=1}^M$ be a linear space of functions in $\mathbb{R}^d$, and let $X_N \subset \mathbb{R}^d$ be such that the matrix $V := [u_j(x_i)]_{1 \le i \le N, 1 \le j \le M} \in \mathbb{R}^{N \times M}$ has full row rank. Then there exist an U-unisolvent set $X_M \subset \mathbb{R}^d$ with $X_N \subset X_M$.*

Proof. The proof is a simple extension of Lemma 1 in [2]. Since V has full row rank there exists $\{j_1, \ldots, j_N\} \subset \{1, \ldots, M\}$ such that $\tilde{V} := [V_{i,j_\ell}]_{1 \le i, \ell \le N} \in \mathbb{R}^{N \times N}$ is invertible, and in particular $\det(\tilde{V}) \ne 0$. We can then pick any $j_{N+1} \in \{1, \ldots, M\} \setminus \{j_1, \ldots, j_N\}$ and consider the matrix function $V'(x) := V(\{u_{j_\ell}\}_{\ell=1}^{N+1}, X_N \cup \{x\})$. Computing the determinant of $V'(x)$ by expanding the last column gives

$$g(x) := \det(V'(x))) = \sum_{j=1}^{N+1} c_j u_j(x),$$

where $c_{N+1} = \det(\tilde{V}) \ne 0$ by assumption. It follows that $g \ne 0$ since the u_j are linearly independent, and thus there exists an $x_{N+1} \in \mathbb{R}^d$ such that $g(x_{N+1}) \ne 0$ (clearly $x_{N+1} \notin X_N$). We can thus define $X_{N+1} := X_N \cup \{x_{N+1}\}$ and repeat the operation by picking another j_{N+2} until $\{j_1, \ldots, j_M\} = \{1, \ldots, M\}$. □

These lemmas immediately give the first characterization of unisolvency for polynomial kernel interpolation, that is stated in the following proposition. Observe that in this case we make a distinction between points in Ω, which is the given domain where the interpolation problem is defined, and point which instead may be in $\mathbb{R}^d \setminus \Omega$.

Proposition 3.3. *Let $p \in \mathbb{N}$ and $a \ge 0$, and let $X_N \subset \Omega$ be a set of $N \le M_a$ pairwise distinct points. Then the kernel matrix of $k_{a,p}$ on X_N is invertible (and thus positive definite) if and only if the Vandermonde matrix* (10) *has full row rank N.*

In particular, this is the case if and only if there exists a set $X_{M_a} \subset \Omega$ such that $X_N \subset X_{M_a}$ and X_{M_a} is $\mathbb{P}_p^d(\Omega)$-unisolvent if $a > 0$, or $\mathbb{H}_p^d(\Omega)$-unisolvent if $a = 0$.

Proof. For any $u \in \mathbb{R}^N \setminus \{0\}$ we have by Lemma 3.1 that $u^T A u = u^T V D V^T u = v^T D v$, with $v := V^T u \in \mathbb{R}^{M_a}$. Since the matrix D is invertible by construction, we have that $v^T D v = 0$ if and only if $v = 0$, i.e., if $V^T u = 0$, i.e., if and only if the columns of V^T - or the rows of V - are linearly dependent. This proves that in fact $u^T A u \ne 0$ for all $u \in \mathbb{R}^N \setminus \{0\}$, i.e., A is positive definite, if and only if the rows of V are linearly independent.

It remains to prove that this condition is equivalent to the existence of a set X_{M_a} that contains X_N and is unisolvent for the corresponding space of polynomials. If X_N is a subset of a set of unisolvent points X_{M_a} (either for $\mathbb{P}_p^d$ or $\mathbb{H}_p^d$), then the columns of V are clearly linearly independent, since V is obtained by selecting N rows from the full Vandermonde matrix V' given by the evaluation of the same monomials on the entire set of points X_{M_a}, and V' is invertible by definition since X_{M_a} is unisolvent for the corresponding space of polynomials. The converse implication follows instead from Lemma 3.2. □

As one may expect, the condition of the last proposition is related to the unisolvency of the interpolation set for standard polynomial interpolation. However, it is remarkable that it is sufficient (and necessary) to have $X_N \subset X_{M_a}$ with X_{M_a} a polynomially unisolvent set (either $\mathbb{P}_p^d$-unisolvent for $a > 0$, or $\mathbb{H}_p^d$-unisolvent for $a = 0$), since this opens the possibility to solve polynomial-like interpolation problems with an arbitrary number of interpolation points. This is in contrast with the case of standard polynomial interpolation, which requires $N = M_a$ and M_a takes only some specific values, depending on d and p. Moreover, the construction of Proposition 3.3 shows that the points $X_{M_a} \setminus X_N$ are not bounded to be in Ω, but can be chosen in the entire space $\mathbb{R}^d$.

As a consequence of Proposition 3.3, we show that the connection with polynomial interpolation is even stronger, i.e., kernel interpolation with minimal p is in fact plain polynomial interpolation, either in $\mathbb{P}_p^d$ or $\mathbb{H}_p^d$. We have the following.

Corollary 3.4. *Let $p \in \mathbb{N}$ and $a \ge 0$. Assume that $N = M_a$ and that the points X_N are $\mathbb{P}_p^d$-unisolvent if $a > 0$ or $\mathbb{H}_p^d$-unisolvent if $a = 0$. Then the polynomial kernel interpolant on X_N coincides with the polynomial interpolant from $\mathbb{P}_p^d$ if $a > 0$ or from $\mathbb{H}_p^d$ if $a = 0$.*

Proof. We have $A = VDV^T$ from Lemma 3.1, and since $M_a = N$ then V is a square invertible matrix. It follows that $A = VB$ with an invertible matrix of change of basis $B := DV^T$, and thus the monomial basis $\{x^{\zeta^{(i)}}\}_{i=1}^{M_a}$ spans the same space of the kernel basis $\{k(\cdot, x_i)\}_{i=1}^N$ of (1). In particular the two interpolants coincide by uniqueness (see point (i) of Definition 3.1). □

Finally, we combine Proposition 3.3 with a construction of polynomial-unisolvent sets in order to derive an explicit characterization of point sets for interpolation with the polynomial kernel in the case $a > 0$.

Theorem 3.5. *Let $a > 0$ and let $X_N \subset \Omega$ be a set of N pairwise distinct points. Then for any $p \geq d(N-1)$ there exists a set $X_{M_p^d} \subset \Omega$ of M_p^d points such that $X_N \subset X_{M_p^d}$ and $X_{M_p^d}$ is $\mathbb{P}_p^d$-unisolvent. In particular, the kernel matrix of $k_{a,p}$ on X_N is invertible if $p \geq d(N-1)$.*

Proof. By virtue of [11, Theorem 1], it is sufficient for $X_{M_p^d}$ to satisfy the following geometric property: For each $x_i \in X_{M_p^d}$, there exist p distinct hyperplanes $G_{i1}, \ldots, G_{ip}$ such that

1. x_i does not lie on any of these hyperplanes;

2. all the other nodes $X_{M_p^d} \setminus \{x_i\}$ lie on at least one of the hyperplanes.

In order to construct $X_{M_p^d}$, we can proceed as follows. For each $x_i \in X_N$, let $L_{i_1}, \ldots, L_{i_d}$ be d hyperplanes such that $L_{i_1} \cap \cdots \cap L_{i_d} = \{x_i\}$ and $x_j \notin L_{i_1} \cup \cdots \cup L_{i_d}$ for $j \neq i$ with $j, i = 1, \ldots, N$. Moreover, such Nd hyperplanes are chosen to be pairwise non-parallel. This construction is always feasible. To get a better intuition, it is possible to reason iteratively. With $x_i \in X_N$ fixed, consider d hyperplanes that do not intersect any point in X_N besides x_i, thus they are not parallel and moreover choose them such that they do not contain any line connecting x_i with another of the other points. Then, consider another point $x_j \in X_N$, $j \neq i$. We can again choose d hyperplanes that intersect at x_j and do not contain the lines connecting x_j with the other points in X_N and are not parallel with the other previously chosen. This is possible due to the infinitely many directions of the hyperplanes passing through x_j. Furthermore, we can surely select pairwise non-parallel hyperplanes (if two hyperplanes turn out to be parallel, we can simply rotate one of the two by using the corresponding node as center of rotation and avoid the parallelism).

Then, proceeding with the proof, the considered hyperplanes intersect at $\binom{Nd}{d}$ points that satisfy the mentioned geometric property. Then, by setting $p = Nd - d = d(N-1)$, we have $\binom{Nd}{d} = M_p^d$ and $X_{M_p^d}$ consists of such points.

Moreover, we observe that it is possible to construct $X_{M_p^d}$ for any $p = Nd - d + j$, $j \geq 1$, by adding to the Nd hyperplanes further j pairwise non-parallel ones that intersect the other hyperplanes in distinct new points, taking then the resulting intersection points. This proves that the same construction works for any $p \geq d(N-1)$.

Finally, Proposition 3.3 proves that X_N can be used for interpolation with the polynomial kernel $k_{a,p}$ if $p \geq d(N-1)$. $\qquad \square$

We remark that this theorem has the notable implication that, if $a > 0$, any set of pairwise distinct points can be used for interpolation with $k_{a,p}$ provided that p is large enough.

Remark 2. It remains open to investigate if this construction works also for $a = 0$, i.e., for the space of homogenous polynomials. Moreover, it would be interesting to investigate if the lower bound $p \geq d(N-1)$ provided by the theorem is sharp. For now, we observe that this is clearly optimal for $d = 1$, since in this case the theorem implies that any $p \geq N-1$ can be used, or in fact that no point needs to be added. This corresponds to the fact that any set of pairwise distinct points is unisolvent for polynomial interpolation in $\mathbb{R}$.

4 Interpolation, stability, and error estimation

Now that the existence of interpolants is established we want to understand the corresponding approximation behavior.

In the last section we dealt with interpolation problems in terms of generic target data $y \in \mathbb{R}^N$. From now on, we will additionally assume that there exists a continuous function $f \in C(\Omega)$ such that $y_i := f(x_i)$, $1 \leq i \leq N$. In this case, we will denote the $k_{a,p}$-interpolant (1) as $I_{X,a,p} f$. From the representation (1), it is immediate to see that $I_{X,a,p} f$ is an element of the linear subspace $V_{a,p}(X) := \text{span}\{k_{a,p}(\cdot, x) : x \in X\} \subset \mathcal{H}_{a,p}$, and thus $I_{X,a,p}$ can be understood as a map from $C(\Omega)$ (or $\mathcal{H}_{a,p}$) to $V_{a,p}(X)$. For this reason, we are interested in obtaining more explicit information on the native space $\mathcal{H}_{a,p}(\Omega)$ on $\Omega \subset \mathbb{R}^d$, which will connect it to suitable spaces of polynomials.

In any case, we underline that in order to obtain asymptotic stability or convergence estimates, one would like to consider an increasing number N of interpolation points. To ensure the existence of the interpolant, one thus needs to consider a kernel $k_{a,p}$ with an N (or X) dependent parameter p, and the design of an optimal choice of p to have stability and convergence is an interesting question. In the following sections we will try to connect the stability and accuracy of kernel interpolation to polynomial interpolation, which seems a promising way to address these issues.

Remark 3. In contrast with polynomial interpolation, the interpolation space $V_{a,p}(X)$ is depending on X. In the language of Definition 3.1, we should thus say that X allows unique interpolation by $k_{a,p}$ if and only if the points X are $V_{a,p}(X)$-unisolvent. To simplify the presentation, in the following we will instead simply write that the set X is $\mathcal{H}_{a,p}$-unisolvent.

4.1 Native space

We first derive some characterization of the space $\mathcal{H}_{a,p}$ for $a \geq 0$ and $p \in \mathbb{N}$, and to this end we introduce some additional notation. Given $\gamma \in \mathbb{N}_0^d$ with $|\gamma| \leq p$, and $f \in C^p(\Omega)$, we write

$$D^\gamma f(x) := \prod_{i=1}^d \left(\partial_{x^{(i)}}^{\gamma^{(i)}} f(x) \right)$$

for the derivative of f with multi-index γ. For all $a \geq 0$ and $\zeta \in I_a(p, d)$, we furthermore define the weights

$$w_\zeta^a := (\zeta!)^2 d_\zeta^a = \begin{cases} \frac{p! \zeta!}{(p-|\zeta|)!} a^{p-|\zeta|}, & a > 0, \\ p! \zeta!, & a = 0, \end{cases} \tag{11}$$

so that (5) and (8) give

$$k_{a,p}(x,y) = \sum_{\zeta \in I_a(p,d)} d_\zeta^a x^\zeta y^\zeta = \sum_{\zeta \in I_a(p,d)} w_\zeta^a \frac{x^\zeta}{\zeta!} \frac{y^\zeta}{\zeta!}.$$

For $\gamma \in I_a(p,d)$ and $x \in \Omega$ this implies that

$$\left(D_x^\gamma k_{a,p}(x,y) \right)_{x=0} = \sum_{\zeta \in I_a(p,d)} w_\zeta^a \left(D^\gamma \frac{x^\zeta}{\zeta!} \right)_{x=0} \frac{y^\zeta}{\zeta!} = w_\gamma^a \frac{y^\gamma}{\gamma!}, \tag{12}$$

since $\left(D^\gamma x^\zeta \right)_{x=0} = 0$ if $\zeta \neq \gamma$, while $D^\gamma x^\gamma / \gamma! = 1$.

With these observations we can prove the following result. The characterization of $\mathcal{H}_{a,p}$ is a special case of the argument of Section 2 in [39] that we repeat for completeness.

Theorem 4.1. *The native space of $k_{a,p}$ on Ω is $\mathcal{H}_{a,p}(\Omega) = \mathbb{P}_p^d(\Omega)$ if $a > 0$ and $\mathcal{H}_{0,p} = \mathbb{H}_p^d(\Omega)$ if $a = 0$, with the inner product*

$$\langle f, g \rangle_{\mathcal{H}_{a,p}} := \sum_{\gamma \in I_a(p,d)} \frac{1}{w_\gamma^a} D^\gamma f(0) D^\gamma g(0), \quad f, g \in \mathcal{H}_{a,p}. \tag{13}$$

Proof. We prove the result for $a > 0$, since if $a = 0$ the same reasoning applies. Namely, the expression (13) clearly defines a symmetric, bilinear, and positive definite form on $\mathbb{P}_p^d(\Omega)$, and thus an inner product. Since $\mathbb{P}_p^d(\Omega)$ is finite it follows that $\left(\mathbb{P}_p^d(\Omega), \langle \cdot, \cdot \rangle_{\mathcal{H}_{a,p}} \right)$ is a Hilbert space. If we prove that $k_{a,p}$ is a reproducing kernel on this space then it must hold that $\mathcal{H}_{a,p}(\Omega) = \mathbb{P}_p^d(\Omega)$ by uniqueness of the native space of a given kernel (see Section 2).

This is indeed the case since $k_{a,p}(\cdot, x) \in \mathbb{P}_p^d(\Omega)$ for all $x \in \Omega$ thanks to (5). Moreover, for all $f := \sum_{\beta \in I_a(p,d)} c_\beta x^\beta \in \mathbb{P}_p^d(\Omega)$ it holds $D^\gamma f(0) = \gamma! c_\gamma$, and thus for all $x \in \Omega$ using (12) we have

$$\langle f, k_{a,p}(\cdot, y) \rangle_{\mathcal{H}_{a,p}} = \sum_{\gamma \in I_a(p,d)} \frac{1}{w_\gamma^a} D^\gamma f(0) \left(D_x^\gamma k_{a,p}(x,y) \right)_{x=0} = \sum_{\gamma \in I_a(p,d)} \frac{1}{w_\gamma^a} \gamma! c_\gamma w_\gamma^a \frac{y^\gamma}{\gamma!} = \sum_{\gamma \in I_a(p,d)} c_\gamma y^\gamma = f(y),$$

which is the reproducing property of the kernel. We thus have that $k_{a,p}$ is a reproducing kernel and the first part of the theorem is proven. The same argument works for $a = 0$ and $\mathbb{H}_p^d$ using (8) in place of (5). $\quad\square$

This characterization makes it possible to study in an explicit manner the effect of the parameters a, p on the native spaces, and thus on the corresponding approximants.

Corollary 4.2. *For any $a \geq 0$, $p \in \mathbb{N}$ we have the following.*

 (i) *If $0 < a' \leq a$ then the native spaces of $k_{a,p}$ and $k_{a',p}$ are norm equivalent, i.e., $\mathcal{H}_{a,p} = \mathcal{H}_{a',p}$ as sets and*

$$\left(a'/a \right)^{p/2} \|f\|_{\mathcal{H}_{a',p}} \leq \|f\|_{\mathcal{H}_{a,p}} \leq \|f\|_{\mathcal{H}_{a',p}} \quad \text{for all } f \in \mathcal{H}_{a,p}, \tag{14}$$

 while for any $a > 0$ it holds $\mathcal{H}_{0,p} \subset \mathcal{H}_{a,p}$ with $\|f\|_{\mathcal{H}_{a,p}} = \|f\|_{\mathcal{H}_{0,p}}$ for all $f \in \mathcal{H}_{0,p}$.

 (ii) *If $a > 0$ and $p, q \in \mathbb{N}$ with $0 \leq q \leq p$, we have that $\mathcal{H}_{a,q} \subset \mathcal{H}_{a,p}$, and the norms of the two spaces are equivalent on $\mathcal{H}_{a,q}$ with*

$$a^{(q-p)/2} \|f\|_{\mathcal{H}_{a,p}} \leq \|f\|_{\mathcal{H}_{a,q}} \leq a^{(q-p)/2} \binom{p}{p-q}^{1/2} \|f\|_{\mathcal{H}_{a,p}} \quad \text{for all } f \in \mathcal{H}_{a,q}. \tag{15}$$

Proof. For the first point we have clearly $\mathcal{H}_{a,p} = \mathcal{H}_{a',p} = \mathbb{P}_p^d$ as sets. Moreover for all $\zeta \in I_a(p,d)$ we have

$$w_\zeta^a = \frac{p! \zeta!}{(p-|\zeta|)!} a^{p-|\zeta|} = \left(\frac{a}{a'} \right)^{p-|\zeta|} w_\zeta^{a'},$$

and thus

$$w_\zeta^{a'} \leq w_\zeta^a \leq \left(a/a' \right)^p w_\zeta^{a'}, \tag{16}$$

since $a/a' \geq 1$. Using this relation in the definition (13) of the norm implies that

$$\left(\frac{a'}{a} \right)^{p/2} \|f\|_{\mathcal{H}_{a',p}} \leq \|f\|_{\mathcal{H}_{a,p}} \leq \|f\|_{\mathcal{H}_{a',p}},$$

which is (14). Finally $\mathcal{H}_{0,p} = \mathbb{H}_p^d \subset \mathbb{P}_p^d = \mathcal{H}_{a,p}$ for all $a > 0$, and so it makes sense to compute the $\mathcal{H}_{a,p}$-norm of $f(x) = \sum_{\zeta \in I_0(p,d)} c_\zeta x^\zeta \in \mathcal{H}_{0,p}$. Since $D^\gamma f(0) = 0$ for all $\gamma \in \mathbb{N}_0^d$ with $|\gamma| < p$, and since $w_\gamma^a = w_\gamma^0$ if $|\gamma| = p$ (see (11)), we have

$$\|f\|_{\mathcal{H}_{a,p}}^2 = \sum_{\gamma \in I_a(p,d)} \frac{1}{w_\gamma^a} (D^\gamma f(0))^2 = \sum_{\gamma \in I_0(p,d)} \frac{1}{w_\gamma^a} (D^\gamma f(0))^2 = \sum_{\gamma \in I_0(p,d)} \frac{1}{w_\gamma^0} \frac{a^{p-|\gamma|}}{(p-|\gamma|)!} (D^\gamma f(0))^2 = \|f\|_{\mathcal{H}_{0,p}}^2,$$

and this concludes the proof of the first part.

In the second case the space inclusion is also clear, and to prove the norm equivalence we write $w_\zeta^a(p)$, $w_\zeta^a(q)$ with an explicit dependence on the polynomial degrees p, q. For $f \in \mathcal{H}_{a,q}$ we have by definition of the norm of $\mathcal{H}_{a,p}$ that

$$\|f\|_{\mathcal{H}_{a,p}}^2 = \sum_{\zeta \in I_a(p,d)} \frac{1}{w_\zeta^a(p)} \left(D^\zeta f(0)\right)^2 = \sum_{\zeta \in I_a(q,d)} \frac{1}{w_\zeta^a(p)} \left(D^\zeta f(0)\right)^2, \tag{17}$$

where the second equality follows from the fact that the derivatives vanish for $|\zeta| > q$. Moreover for all $\zeta \in I_a(q,d)$ we have

$$w_\zeta^a(p) = \frac{p! \zeta!}{(p-|\zeta|)!} a^{p-|\zeta|} = \frac{p \cdots (q+1) q! \zeta!}{(p-|\zeta|) \cdots (q+1-|\zeta|)(q-|\zeta|)!} a^{q-|\zeta|} a^{p-q} = \frac{p \cdots (q+1)}{(p-|\zeta|) \cdots (q+1-|\zeta|)} a^{p-q} w_\zeta^a(q),$$

and this quantity is minimized when $|\zeta| = 0$ and maximized when $|\zeta| = q$, giving

$$a^{p-q} w_\zeta^a(q) \leq w_\zeta^a(p) \leq a^{p-q} \binom{p}{p-q} w_\zeta^a(q).$$

Inserting these bounds in (17) gives

$$a^{q-p} \binom{p}{p-q}^{-1} \|f\|_{\mathcal{H}_{a,q}}^2 \leq \|f\|_{\mathcal{H}_{a,p}}^2 \leq a^{q-p} \|f\|_{\mathcal{H}_{a,q}}^2 \quad \text{for all } f \in \mathcal{H}_{a,q},$$

which can be rearranged to obtain (15). $\square$

Remark 4. The inclusion relation $\mathcal{H}_{a,p} \subset \mathcal{H}_{a,p'}$ for $p \leq p'$ and $a = 1$, and the equivalence between the corresponding norms, was already proven in [38] by other arguments (see Proposition 4.3 and Proposition 6.3). The other relations are instead new to the best of our knowledge. Moreover, the case (ii) cannot be extended to $a = 0$, since in this case the spaces $\mathcal{H}_{0,p}$ and $\mathcal{H}_{0,p-q}$ have empty intersection unless $q = p$. Finally, we remark that the general case $a \neq a'$, $p \neq q$ can be obtained by combining the two cases (i) and (ii) of Corollary 4.2.

We point out that (13) implies that the monomials are orthogonal in $\mathcal{H}_{a,p}$, i.e., for all $\zeta, \beta \in I_a(p,d)$ we have

$$\left\langle \frac{x^\zeta}{\zeta!}, \frac{x^\beta}{\beta!} \right\rangle_{\mathcal{H}_{a,p}} = \frac{1}{w_\zeta^a} \delta_{\zeta\beta}. \tag{18}$$

From this fact we may also deduce that the norm inequality (14) is sharp for all $0 < a' \leq a$. Indeed, the right inequality in (14) is an equality for $f(x) := x^\zeta$ with $|\zeta| = p$, since in this case we have by (18) that

$$\left\|x^\zeta\right\|_{\mathcal{H}_{a,p}}^2 = \frac{(\zeta!)^2}{w_\zeta^a} = \frac{\zeta!}{p!} = \frac{(\zeta!)^2}{w_\zeta^{a'}} = \left\|x^\zeta\right\|_{\mathcal{H}_{a',p}}^2,$$

where we used the definition (11) of w_ζ^a and $w_\zeta^{a'}$. Similarly, the left inequality in (14) is met for $|\zeta| = 0$, i.e. $x^\zeta = 1$ and $\zeta! = 1$, since in this case we have again by (18) that

$$\left\|x^\zeta\right\|_{\mathcal{H}_{a,p}}^2 = \frac{1}{w_\zeta^a} = \frac{1}{a^p} = \left(\frac{a'}{a}\right)^p \frac{1}{w_\zeta^{a'}} = \left(\frac{a'}{a}\right)^p \left\|x^\zeta\right\|_{\mathcal{H}_{a',p}}^2.$$

In particular, maximal-degree monomials have the same norm independently of $a > 0$, while lower degree monomials have an increasingly large norm as $a \to 0$, up to not even being elements of $\mathcal{H}_{a,p}$ in the limiting case, since indeed $\mathcal{H}_{0,p}$ is the homogeneous space. In this sense, the parameter $a \geq 0$ has a regularizing effect, promoting high degree components in a minimal norm solution of the interpolation problem.

For (15), similar arguments prove that the equality is obtained for $|\zeta| = 0$ (the lower bound), and $|\zeta| = q$ (the upper bound). It thus happens that if $p \geq q$, the elements of $\mathcal{H}_{a,q}$ have an $\mathcal{H}_{a,p}$-norm which increases with a factor between $a^{(q-p)/2}$ and $a^{(q-p)/2} \binom{p}{p-q}^{1/2}$. In particular, low-degree and high-degree monomials have norm that are increasingly separated, and thus also increasing p has a regularization effect, and the monomials x^ζ have minimal norm in $\mathcal{H}_{a,p}$ with $p = |\zeta|$.

4.2 Stability

We now derive a simple stability results for the interpolation map $I_{X,a,p} : C(\Omega) \to V_{a,p}(X)$, as a function of a set $X \subset \Omega$ of $\mathcal{H}_{a,p}$-unisolvent points. We recall that the Lebesgue function associated to the interpolation process is defined as

$$\lambda_{X,a,p}(x) := \sup_{0 \neq f \in V_{a,p}(X)} \frac{|f(x)|}{\|f_{|X}\|_\infty}, \quad x \in \Omega, \tag{19}$$

such that one obtains the stability bound

$$|I_{X,a,p} f(x)| \leq \lambda_{X,a,p}(x) \|f_{|X}\|_\infty \leq \lambda_{X,a,p}(x) \|f\|_{L_\infty(\Omega)} \quad \text{for all } f \in C(\Omega), x \in \Omega, \tag{20}$$

which can be also written in terms of the associated Lebesgue constant $\Lambda_{X,a,p} := \|\lambda_{X,a,p}\|_{L_\infty(\Omega)}$ as

$$\left\|I_{X,a,p} f\right\|_{L_\infty(\Omega)} \leq \Lambda_{X,a,p} \|f_{|X}\|_\infty \leq \Lambda_{X,a,p} \|f\|_{L_\infty(\Omega)} \quad \text{for all } f \in C(\Omega). \tag{21}$$

Moreover, the fact that X is unisolvent ensures the existence of a Lagrange basis $\{\ell_{i,a,p}\}_{i=1}^N$ of $V_{a,p}(X)$, which gives

$$I_{X,a,p}f(x) = \sum_{i=1}^N f(x_i)\ell_{i,a,p}(x), \quad x \in \Omega, \tag{22}$$

and $\lambda_{X,a,p}(x) = \sum_{i=1}^N |\ell_{i,a,p}(x)|$.

Although we are still not able to obtain explicit bounds on $\lambda_{X,a,p}$ and $\Lambda_{X,a,p}$, we can prove the following result.

Theorem 4.3. *Let $B \subset \mathbb{R}^d$ be a set and let $X \subset \Omega$ be a set of $\mathbb{P}_p^d$-unisolvent points. Let*

$$\Lambda_X^{pol} := \sup_{0 \neq f \in \mathbb{P}_p^d(B)} \frac{\|f\|_{L_\infty(B)}}{\|f_{|X}\|_\infty} \tag{23}$$

be the Lebesgue constant for polynomial interpolation of degree p on X.

Let $a \geq 0$, $p \in \mathbb{N}$, and let X be $\mathcal{H}_{a,p}$-unisolvent. For any $X_{M_a} \supset X$ which is $\mathbb{P}_p^d$-unisolvent, it holds

$$\left\|I_{X,a,p}f\right\|_{L_\infty(\Omega)} \leq \Lambda_{X_{M_a}}^{pol} \|f\|_{L_\infty(B)} \quad \text{for all } f \in C(B),$$

where $B \subset \mathbb{R}^d$ is any set which contains X_{M_a}.

Proof. The results simply follows from the fact that $I_{X,a,p}f \in \mathcal{H}_{a,p} = \mathbb{P}_p^d$ and by the definition of $\Lambda_{X_{M_a}}^{pol}$. Such X_{M_a} exists thanks to Proposition 3.3, but since it needs not to be contained in Ω we consider an enclosing set B. □

This results shows that polynomial kernel interpolation on X is at least as stable as polynomial interpolation on any $X_{M_a} \supset X$. This points to the fact that understanding how to complete a set of points to a set of polynomially unisolvent points of small Lebesgue constant may be highly relevant in this context. The fact is also related to the possible minimality of the construction in Theorem 3.5 (see also Remark 3). Moreover, although choosing $X_{M_a} \setminus X_N$ outside of Ω may possibly lead to a smaller Lebesgue constant, one pays the price of obtaining a bound in terms of the norm $\|f\|_{L_\infty(B)}$ computed on a larger set $B \supset \Omega$, which may possibly be significantly larger than $\|f\|_{L_\infty(\Omega)}$.

On the other hand, if one is free to choose N points to sample a function to construct a $k_{a,p}$ interpolant, this result suggests that it could be a good idea to select them from a $\mathbb{P}_p^d$-unisolvent set with small Lebesgue constant.

4.3 Error estimation

As it is typically the case in kernel interpolation, we start by assuming that $f \in \mathcal{H}_{a,p}$. In this case, we recall that the interpolant as a map $I_{X,a,p} : \mathcal{H}_{a,p} \to V_{a,p}(X)$ coincides with the $\mathcal{H}_{a,p}$-orthogonal projection onto $V_{a,p}(X)$. The norm of the associated error operator is the power function $P_{X,a,p}(x)$ defined by

$$P_{X,a,p}(x) := \sup_{0 \neq f \in \mathcal{H}_{a,p}} \frac{\left|f(x) - I_{X,a,p}f(x)\right|}{\|f\|_{\mathcal{H}_{a,p}}}. \tag{24}$$

By definition, the interpolation error can be controlled as

$$\left|f(x) - I_{X,a,p}f(x)\right| \leq P_{X,a,p}(x)\|f\|_{\mathcal{H}_{a,p}} \quad \text{for all } f \in \mathcal{H}_{a,p}. \tag{25}$$

This bound allows one to separate the error in a term depending only on f and one depending only on X, Ω and $k_{a,p}$.

In particular, worst-case error bounds in $\mathcal{H}_{a,p}$ can be derived by obtaining uniform bounds on $P_{X,a,p}$ in terms of the fill distance

$$h_X := \sup_{x \in \Omega} \min_{y \in X} \|x - y\|.$$

We refer to Chapter 11 in [37] for details on this approach.

We are not able to obtain bounds of this type yet, and we rather use the power function to outline a method to derive more general error bounds. Indeed, as mentioned in Section 4.2 the interpolant $I_{X,a,p}f$ is well defined also for any $f \in C(\Omega)$, even for $f \notin \mathcal{H}$, since its computation requires only the knowledge of $f_{|X}$, and it is of interest to study the resulting approximation error also in this case. We have the following result.

Proposition 4.4. *Let $X_N \subset \Omega$ be $\mathcal{H}_{a,p}$-unisolvent. Then for all $f \in C(\Omega)$ we have*

$$\left|(f - I_{X,a,p}f)(x)\right| \leq \left(1 + \lambda_{X,a,p}(x)\right)\left\|f - f_p^\star\right\|_{L_\infty(\Omega)} + P_{X,a,p}(x)\left\|f^\star - I_{X,a,p}f_p^\star\right\|_{\mathcal{H}_{a,p}}, \quad x \in \Omega, \tag{26}$$

where

$$f_p^\star := \inf_{g \in \mathbb{P}_p^d(\Omega)} \|f - g\|_{L_\infty(\Omega)}$$

is the uniform best polynomial approximant of f.

Proof. We have

$$f - I_{X,a,p} f = f - f_p^\star + f_p^\star - I_{X,a,p} f_p^\star + I_{X,a,p} f_p^\star - I_{X,a,p} f. \tag{27}$$

Using the form (22) of the interpolant, a standard argument gives

$$|(I_{X,a,p} f_p^\star - I_{X,a,p} f)(x)| = |I_{X,a,p}(f_p^\star - f)(x)| = \sum_{i=1}^{N} |\ell_{i,a,p}(x)| |(f_p^\star - f)(x_i)| \leq \lambda_{X,a,p}(x) \max_{1 \leq i \leq N} |(f_p^\star - f)(x_i)| \tag{28}$$

$$\leq \lambda_{X,a,p}(x) \|f^\star - f\|_{L_\infty(\Omega)}.$$

Moreover, since $f_p^\star \in \mathcal{H}_{a,p} = \mathbb{P}_p^d$ by definition, we can apply the standard power function bound (25) and get

$$|(f_p^\star - I_{X,a,p} f_p^\star)(x)| \leq P_{X,a,p}(x) \left\| f_p^\star - I_{X,a,p} f_p^\star \right\|_{\mathcal{H}_{a,p}}. \tag{29}$$

Inserting these two bounds in (27) we get

$$|(f - I_{X,a,p} f)(x)| \leq |(f - f_p^\star)(x)| + |(f_p^\star - I_{X,a,p} f_p^\star)(x)| + |(I_{X,a,p} f_p^\star - I_{X,a,p} f)(x)|$$

$$\leq \left\| f - f_p^\star \right\|_{L_\infty(\Omega)} + |(f_p^\star - I_{X,a,p} f_p^\star)(x)| + |(I_{X,a,p} f_p^\star - I_{X,a,p} f)(x)|$$

$$\leq \left(1 + \lambda_{X,a,p}(x)\right) \left\| f - f_p^\star \right\|_{L_\infty(\Omega)} + P_{X,a,p}(x) \left\| f_p^\star - I_{X,a,p} f_p^\star \right\|_{\mathcal{H}_{a,p}},$$

which is the bound of the statement.   □

This result shows that it is possible to obtain error bounds for interpolation of functions outside of the native space (thus escaping the native space [25], or working in the so-called misspecified setting [19, 20]).

Observe also that in a sense inequality (26) is sharp with respect to the relation between polynomial interpolation in $\mathbb{P}_p^d$ and kernel interpolation in $\mathcal{H}_{a,p}$. Indeed, on one hand if $f \in \mathcal{H}_{a,p} = \mathbb{P}_p^d$ then $f - f_p^\star = 0$, and thus (26) reduces to the usual power function bound (25) for kernel interpolation. On the other hand, if instead $N = M_a$ then $I_{X,a,p} f$ coincides with the polynomial interpolant of f (see Corollary 3.4). It follows that $f - I_{X,a,p} f = 0$ for all $f \in \mathcal{H}_{a,p} = \mathbb{P}_p^d$, and thus $P_{X,a,p} = 0$ (see (24)). In this case (26) reduces to the standard Lebesgue function bound for polynomial interpolation.

Between these two limits, the proposition suggests as well that one may try to optimize the set X_{M_a} and its subset X_N, in order to balance the contribution of the two terms in (26).

Moreover, we have the following.

Corollary 4.5. *Under the assumptions of Proposition 4.4, let X_{M_a} with $X_N \subset X_{M_a} \subset B \subset \mathbb{R}^d$ be any $\mathbb{P}_p^d$-unisolvent set. Then*

$$\left|(f - I_{X,a,p} f)(x)\right| \leq \left(1 + \lambda_{X_{M_a}}^{pol}(x)\right) \left(\left\| f - f_p^\star \right\|_{L_\infty(B)} + \max_{x \in X_{M_a}} P_{X,a,p}(x) \left\| f^\star - I_{X,a,p} f_p^\star \right\|_{\mathcal{H}_{a,p}} \right), \quad x \in B, \tag{30}$$

where $\lambda_{X_{M_a}}^{pol}(x)$ is the Lebesgue function for polynomial interpolation.

Proof. We follow the same steps as in the proofs of Proposition 4.4. Instead of (28), since $f_p^\star - I_{X,a,p} f_p^\star \in \mathbb{P}_p^d$ we have

$$|(I_{X,a,p} f_p^\star - I_{X,a,p} f)(x)| \leq \lambda_{X_{M_a},a,p}(x) \|f^\star - f\|_{L_\infty(B)},$$

with the same argument as in Theorem 4.3. Moreover, again because $f_p^\star - I_{X,a,p} f_p^\star \in \mathbb{P}_p^d$ we have

$$|(f_p^\star - I_{X,a,p} f_p^\star)(x)| \leq \lambda_{X_{M_a}}^{pol}(x) \left\| (f_p^\star - I_{X,a,p} f_p^\star)_{|X_{M_a}} \right\|_\infty,$$

and from this we can proceed to replace (29) with

$$|(f_p^\star - I_{X,a,p} f_p^\star)(x)| \leq \lambda_{X_{M_a}}^{pol}(x) \max_{x \in X_{M_a}} P_{X,a,p}(x) \left\| f_p^\star - I_{X,a,p} f_p^\star \right\|_{\mathcal{H}_{a,p}}.$$

Inserting these two bounds in (27) we get the result.   □

Without the term involving the power function, inequality (30) is the error bound for polynomial interpolation of f on X_{M_a}, and we thus have that the error of kernel interpolation on X is comparable with the error of polynomial interpolation on X_{M_a}. To quantify this relation it would be sufficient to prove bounds on $\max_{x \in X_{M_a}} P_{X,a,p}(x)$, which should be expected to be much easier to bound than $\left\| P_{X,a,p} \right\|_{L_\infty(\Omega)}$ and with a smaller value, provided the points X_{M_a} are not too far from Ω. Moreover, one may expect that such a bound depends on the relation between X_N and X_{M_a}, instead of on h_X.

Remark 5. Other approaches may be followed to obtain error bounds for interpolation with the polynomial kernels. Most notably, one can use the zero lemma of [24]. Namely, for $\tau := k + s$ with $k \in \mathbb{N}$, $k > d/2$, $s \in (0, 1]$, and $1 \leq t \leq \infty$, we denote as $W_t^\tau(\Omega)$ the $L_t(\Omega)$-Sobolev space of fractional smoothness τ. Theorem 2.12 in [24] proves that if Ω has a sufficiently smooth boundary, then there is a constant C_k depending on k such that for any $1 \leq r \leq \infty$ and for any $f \in W_t^\tau(\Omega)$ it holds

$$\left\| f - I_{X,a,p} f \right\|_{W_r^\tau(\Omega)} \leq C_k h_X^{\tau - d(1/t - 1/r)_+} \left| f - I_{X,a,p} f \right|_{W_t^\tau(\Omega)},$$

where $(x)_+ := \max(x, 0)$. Taking in particular $\tau = k + s > p$ gives thus

$$\left\| f - I_{X,a,p} f \right\|_{W_r^\tau(\Omega)} \leq C_k h_X^{\tau - d(1/t - 1/r)_+} \left| f \right|_{W_t^\tau(\Omega)},$$

where $\left| I_{X,a,p} f \right|_{W_r^\tau(\Omega)} = 0$ since $\tau > p$. However, p needs to increase with N (see the beginning of Section 4), and thus $\tau = k + s > p$ is possible only if k is itself increasing. Since C_k is increasing with k, to use this approach one would need to work out explicitly the growth of C_k in terms of p, with a small as possible p. A similar approach has already been followed in [39], but in that case the kernels are strictly positive definite, so one does not need to change the kernel depending on the points.

5 Stable computations

As we will demonstrate numerically in Section 6, the computation of a $k_{a,p}$-interpolant may be significantly unstable from a computational point of view, even if we proved in Section 4.2 that the Lebesgue constant of this interpolant can be growing quite slowly, for example with a rate comparable to that of polynomial interpolation (see Theorem 4.3).

The fact that the interpolation process is provably stable even if its actual computation is unstable has been observed and studied thoroughly in kernel interpolation [12, 13]. Moreover, this discrepancy has been attributed to the use of the so-called direct method, i.e., the inversion of the kernel matrix, and several stable algorithm has been introduced to overcome this problem.

In particular, given parameters $p \in \mathbb{N}$, $a \geq 0$ and a set of $\mathcal{H}_{a,p}$-unisolvent points $X \subset \Omega$, in this section we show how to apply the RBF-QR algorithm to the polynomial kernels to construct a stable basis $\{u_j\}_{j=1}^N$ of $V_{a,p}(X)$. We follow Section 4.2 of [18] and Section 4.1 in [17], where in our case the monomial basis plays the role of the Mercer basis of the kernel.

To this end we define an arbitrary but fixed ordering $\{\zeta^{(i)}\}_{i=1}^{M_a}$ of the multiindices $I_a(p, d)$ and construct the diagonal matrix $D \in \mathbb{R}^{M_a \times M_a}$ as in Lemma 3.1. Using the same ordering of the monomials, we split D as

$$D = \begin{bmatrix} D_1 & 0 \\ 0 & D_2 \end{bmatrix}, \tag{31}$$

now with $D_1 \in \mathbb{R}^{N \times N}$ and $D_2 \in \mathbb{R}^{(M_a - N) \times (M_a - N)}$. Observe in particular that the multi-indices may be sorted in such a way that the diagonal entries of D are sorted in non-increasing order, so that D_1 contains the large entries, and D_2 the small ones that are likely the cause of the instability that we observed in the numerical inversion of A. Observe that, in case of equality among values of the diagonal, multiple possible ordering are possible. We are not investigating this aspect here, and just assume that an arbitrary valid order is fixed.

We then assemble the Vandermonde matrix $V \in \mathbb{R}^{N \times M_a}$ associated to the same monomial ordering and compute the QR decomposition $V = QR$, with $Q \in \mathbb{R}^{N \times N}$ an orthogonal matrix, and where the matrix $R \in \mathbb{R}^{N \times M_a}$ is splitted as

$$R := [R_1 | R_2], \quad R_1 \in \mathbb{R}^{N \times N}, \quad R_2 \in \mathbb{R}^{N \times (M_a - N)}. \tag{32}$$

It follows from Lemma 3.1 that

$$A = V D V^T = V D R^T Q^T = V \begin{bmatrix} D_1 & 0 \\ 0 & D_2 \end{bmatrix} \begin{bmatrix} R_1^T \\ R_2^T \end{bmatrix} Q^T = V \begin{bmatrix} D_1 R_1^T \\ D_2 R_2^T \end{bmatrix} Q^T = V \begin{bmatrix} I \\ D_2 R_2^T R_1^{-T} D_1^{-1} \end{bmatrix} D_1 R_1^T Q^T, \tag{33}$$

where $D_1 R_1^T$ is invertible because D is invertible and V, and thus R, have full rank since the points X are unisolvent (see Proposition 3.3).

We now recall that for an arbitrary basis $\{u_j\}_{j=1}^N$ of $V_{a,p}(X)$, if $C_u \in \mathbb{R}^{N \times N}$ is the matrix of change of basis from $\{k_{a,p}(\cdot, x_i)\}_{i=1}^N$ to this new basis, and $V_u := [u_j(x_i)]_{i,j=1}^N$, then [27] shows that it holds $A = V_u C_u^{-1}$. We may interpret the decomposition (33) in these terms, and assume that $C_u^{-1} := D_1 R_1^T Q^T$ is the inverse of the matrix of change of basis from the stable basis to the kernel basis of translates. With this definition, observe that we also have from (33) that

$$V_u = A C_u = V \begin{bmatrix} I \\ D_2 R_2^T R_1^{-T} D_1^{-1} \end{bmatrix} = V C_u',$$

where now

$$C_u' := \begin{bmatrix} I \\ D_2 R_2^T R_1^{-T} D_1^{-1} \end{bmatrix} \in \mathbb{R}^{M_a \times N} \tag{34}$$

is a rectangular matrix that expresses the new basis in terms of the monomial basis. This can thus be used to express the stable basis without the need of passing through the unstable kernel basis. Observe moreover that $D_2 R_2^T R_1^{-T} D_1^{-1}$ is an $(M_a - N) \times N$ matrix, and that its computation (and thus the computation of C_u), requires only the inverses of the $N \times N$ matrices R_1 and D_1, which can be computed efficiently since R_1 is triangular and D_1 is diagonal.

Algorithm 1 Construction of the interpolant

1: Input: Parameters $p \in \mathbb{N}$, $a \geq 0$, $\mathcal{H}_{a,p}$-unisolvent interpolation points $X := \{x_i\}_{i=1}^N \subset \mathbb{R}^d$ and target values $y \in \mathbb{R}^N$.
2: Define an ordering $I_u := \{\zeta^{(i)}\}_{i=1}^{M_a}$ of the multiindices $I_a(p, d)$.
3: Evaluate the Vandermonde matrix $V \in \mathbb{R}^{N \times M_a}$ associated to the points X and the given ordering (see Lemma 3.1).
4: Compute the QR decomposition $QR = V$, and define R_1, R_2 as in (32).
5: Evaluate the diagonal matrix $D \in \mathbb{R}^{M_a \times M_a}$ (see Lemma 3.1) and define D_1, D_2 as in (31).
6: Compute C_u' as in (34).
7: Compute the coefficients $c_u := C_u^{-1} y$.
8: Output: C_u', c_u, I_u.

We can thus work directly in terms of this new stable basis, both to solve the linear system and to evaluate the interpolant as

$$I_{a,p,X}(x) = \sum_{j=1}^N (c_u)_j u_j(x), \ \ x \in \Omega,$$

for a suitable vector of coefficients $c_u \in \mathbb{R}^N$. The computation of the interpolant is summarized in Algorithm 1.

Once the matrix C_u', the vector c_u, and the ordering I_u are computed, they can be used to evaluate the interpolant on any set $X_{eval} := \{x_i'\}_{i=1}^{N_{eval}} \subset \mathbb{R}^d$ of $N_{eval} \in \mathbb{N}$ evaluation points, again by using the corresponding Vandermonde matrix. We describe this process in Algorithm 2, which returns the vector y_{eval} with $(y_{eval})_i := I_{a,p,X}(x_i')$.

Algorithm 2 Evaluation of the interpolant

1: Input: Evaluation points $X_{eval} := \{x_i'\}_{i=1}^{N_{eval}} \subset \mathbb{R}^d$, $N_{eval} \in \mathbb{N}$, matrix C_u', vector c_u, and indices I_u from Algorithm 1.
2: Evaluate the Vandermonde matrix $V \in \mathbb{R}^{N_{eval} \times M_a}$ (see Lemma 3.1) associated to the points X_{eval} and the ordering I_u.
3: Evaluate the stable basis on X_{eval} as $V_u := V C_u'$.
4: Evaluate the interpolant on X_{eval} as $y_{eval} = V_u c_u$.
5: Output: y_{eval}.

We remark that this approach works exactly in the same way to interpolate vector-valued functions $f : \mathbb{R}^d \to \mathbb{R}^{d'}$, $d' \in \mathbb{N}$. In this case $y \in \mathbb{R}^{N \times d'}$ collects the evaluations of f on X rowwise, and the output c_u of Algorithm 1 is a matrix $c_u \in \mathbb{R}^{N \times d'}$. In particular, this approach can also be used to compute the Lagrange basis functions (22) by a single run of Algorithm 1, simply by defining $d' := N$ and $y := I \in \mathbb{R}^{N \times N}$. The resulting output $y_{eval} \in \mathbb{R}^{N_{eval} \times N}$ of Algorithm 2 contains as columns the Lagrange functions evaluated on X_{eval} (see Remark 13.8 in [16]).

Remark 6. The RBF-QR algorithm includes also a series of further ad-hoc optimizations that we are not discussing here for simplicity. Moreover, the stable basis obtained by this process is expressed in terms of the monomial basis, and it is thus clearly a polynomial basis. Similar approaches based on QR decompositions are used for point selection in polynomial interpolation, such as the Approximate Fekete Points (AFP) of [4]. In particular, the matrices Q and V_u should be related to some sort of orthogonal polynomials, which could be interesting to investigate to further improve the stability of the algorithm.

6 Numerical experiments

We test now numerically some aspects that were discussed in the previous sections. For simplicity we restrict to $d = 1$ and set $\Omega = [-1, 1]$, so that (3) and (4) give $I_a(p, 1) = \{0, 1, \ldots, p\}$ and $M_a = p + 1$ if $a > 0$, while $I_0(p, 1) = \{p\}$ and $M_0 = 1$. Moreover, (5) and (8) simplify to

$$k_{a,p}(x, y) = \sum_{\zeta=0}^p \binom{p}{\zeta} a^{p-\zeta}(xy)^\zeta, \quad k_{0,p}(x, y) = (xy)^p.$$

Some examples of the values of $k_{a,p}(\cdot, 1/2)$ are visualized in Figure 1.

6.1 Convergence of the interpolant and stable computations

We start by comparing kernel interpolation with polynomial interpolation, and demonstrate the potential instability of the direct method and the effectiveness of the RBF-QR approach.

We recall that $M_a = p + 1$ for $d = 1$, and thus the constraint $N \leq M_a$ means that we need to require $p \geq N - 1$. Moreover, for any $p \in \mathbb{N}$ any set of pairwise distinct points is the subset of a set of $\mathbb{P}_p^1$-unisolvent points, and thus in view of Proposition 3.3 we can solve interpolation problems with any set of N pairwise distinct points provided that $p \geq N - 1$.

As an example we interpolate the smooth function $f(x) := \cos(10x)$ sampled at N Chebyshev points with $N = 5, \ldots, 50$. For each N we test polynomial interpolation, and kernel interpolation with $p = p(N) \in \{N - 1, N + 1, N + 3, N + 5\}$ and $a \in \{5, 10\}$.

We report in Figure 2 the corresponding maximum absolute errors with respect to the exact values of f, computed on a grid of $N_{eval} = 1000$ equally spaced points. It is remarkable to observe that the direct approach (first row of Figure 2) fails to compute a converging interpolant even for small values of N and even if the corresponding polynomial interpolant is stable and convergent. On the other hand, switching to RBF-QR (second row of Figure 2) resolves this instability, and for all the tested
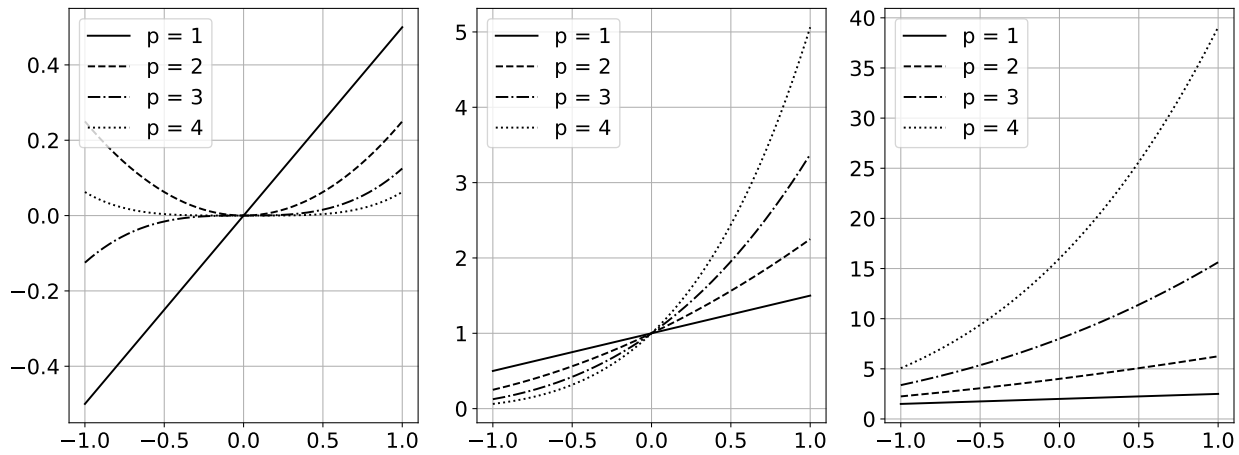
Figure 1: Values of the kernel $k_{a,p}(\cdot, 1/2)$ on $[-1, 1]$ for $p \in \{1, 2, 3, 4\}$ and for $a = 0$ (left), $a = 1$ (center), and $a = 2$ (right).

values of p and a the kernel interpolants converge at the same speed of the polynomial interpolant. However, also in this stable case the convergence saturates at an error of roughly 10^{-12} at $N = 30$, and from there on the stability seems to slightly degrade, up to some small oscillations for large N. It is possible that the adoption of further optimizations, as discussed in Remark 6, may lead to an even stronger stability.

6.2 Lagrange functions and Lebesgue constant

We now consider the stability aspects of the interpolant, analysing the associated Lagrange functions and Lebesgue constant.

We first show that RBF-QR is indeed effective also in computing these cardinal functions. As an example, Figure 3 shows the Lagrange functions of $k_{10,25}$ corresponding to $N = 15$ Chebyshev points. Even for this small number of points, it is immediately clear that a stable algorithm is needed to have an accurate computation.

We use these stably computed functions to evaluate the Lebesgue function and the Lebesgue constant and compare it with the ones of polynomial interpolation. As an example, for $N = 5$ Chebyshev points we consider the polynomial kernel $k_{a,p}$ with $p \in \{N-1, N+9, N+19, N+29\}$ and $a = 5$. The first row in Figure 4 shows the Lagrange functions for the different values of p, and it is clear that an increase of p has the effect of smoothing the oscillations in the interior of the domain and enlarging the oscillations close to the boundary. This reflects into a Lebesgue function (left panel in Figure 4) which is decreasing in the interior and increasing close to the boundary as p increases. This behavior causes the Lebesgue function (right panel in Figure 4) to be initially decreasing and then increasing. In particular, the minimum value for this set of points and parameter a is reached for $p > N - 1$, i.e., there exists a polynomial kernel with a Lebesgue constant which is strictly smaller than that of the polynomial interpolant.

To further investigate the stability of the interpolants we look at the asymptotic behavior of the Lebesgue constant. We compare the same kernels $k_{a,p}$ used in the previous section, but now testing both equally spaced and Chebyshev points, for $N = 5, \ldots, 45$. We restrict to a maximal $N = 45$ because the same instability in the computations observed in Section 6.1 appears here for large N, up to making the results completely unreliable for $N \approx 50$. The results are reported in Figure 5. It is clear that in all cases the growth of the Lebesgue constant coincides with that of polynomial interpolation (i.e., $p = N - 1$), and thus it has the well known logarithmic growth for Chebyshev points (left panels in Figure 5), and exponential growth for equally spaced points (right panels in Figure 5). It is important to notice that the growth seems to be not affected by the value of a, and especially not even by that of p. This latter fact is relevant because it implies that, at least for $d = 1$, the value of $\Lambda_{X,a,p}$ depends on X and not on X_{M_a}, and especially it seems that the bound of Theorem 4.3 is quite pessimistic. Moreover, it seems that the difference between the values of the Lebesgue constants of polynomial and kernel interpolation observed in Figure 4 (bottom right) are not so significant for growing N.

7 Conclusions and perspectives

In this paper we derived some initial results for the application and analysis of polynomial kernels for the solution of interpolation problems. We derived necessary and sufficient conditions for the existence of a unique interpolant and we provided an explicit description of the native spaces of the polynomial kernels. In particular, we analyzed in some detail the effect of the kernel parameters on these spaces. These results were further used to derive some first quantification of the stability and convergence of these interpolants, with particular attention to the connection with the corresponding results in polynomial interpolation. Finally, we have shown that a direct solution of the interpolation system leads to inaccurate computations, and that the use of the RBF-QR algorithm can significantly mitigate this issue.

Several points remain open and will be the subject of future research. In particular, we outlined in several occasions that a proper selection of the degree p may be crucial, and that its choice may balance between accuracy and stability. A better
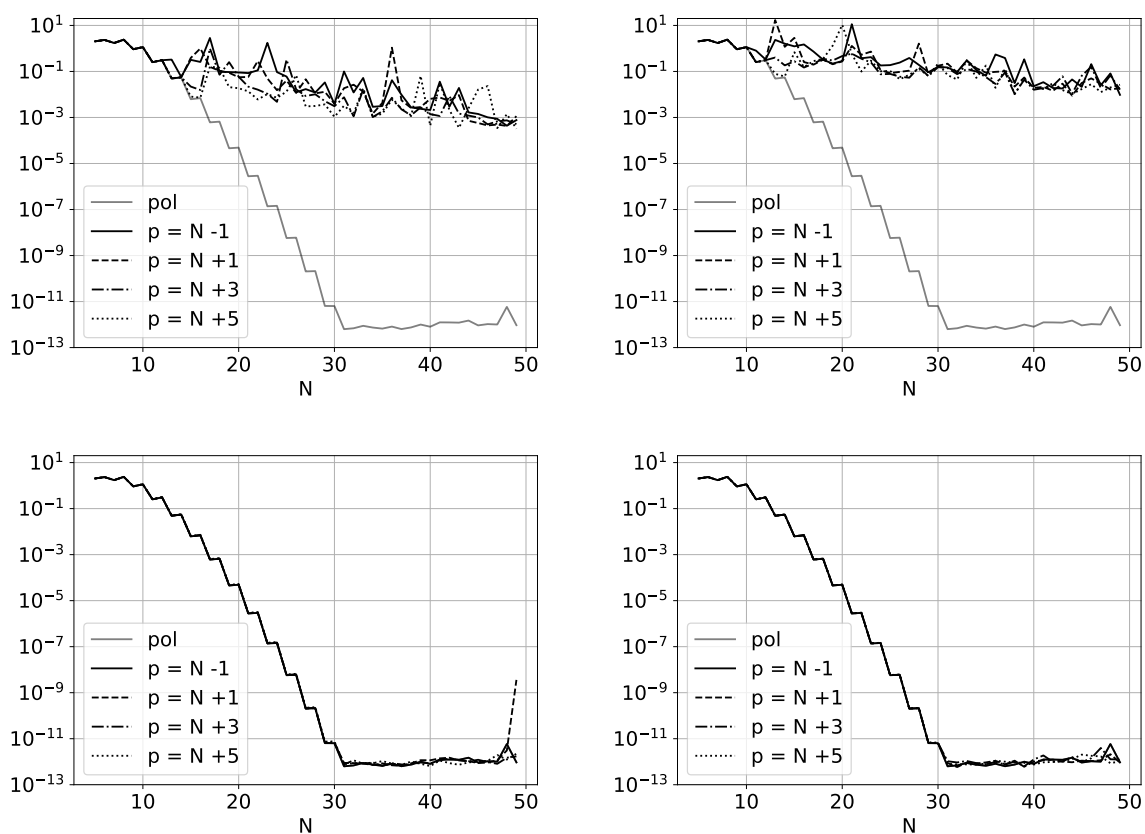
Figure 2: Convergence of the maximal absolute error of interpolation of the function $f(x) = \cos(10x)$ using $N = 5, \dots, 50$ Chebyshev points. For each figure, we test a polynomial interpolant (gray line), and kernel interpolants with various values of p, and $a = 5$ (left column) and $a = 10$ (right column). The kernel interpolants are computed with the direct method (first row) and with RBF-QR (second row).
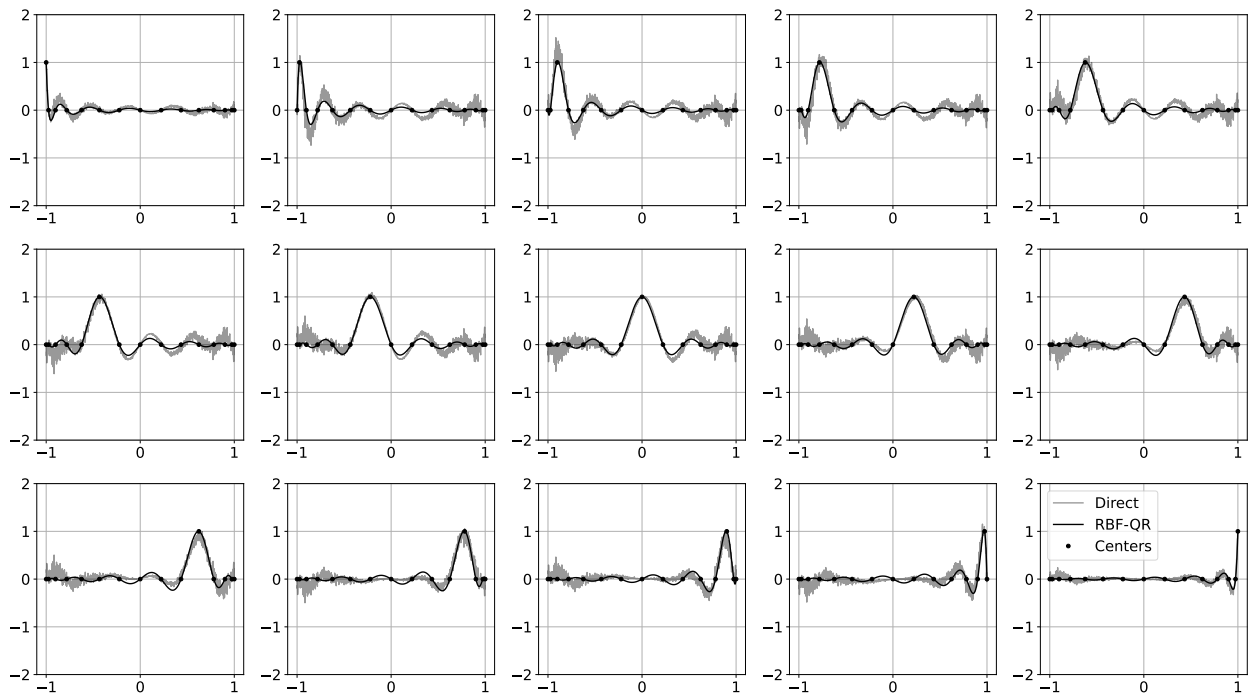
Figure 3: Lagrange functions for the polynomial kernel with $p = 25$ and $a = 10$, and corresponding to $N = 15$ Chebyshev points (black dots), computed with the direct method (gray lines) and with RBF-QR (black lines).
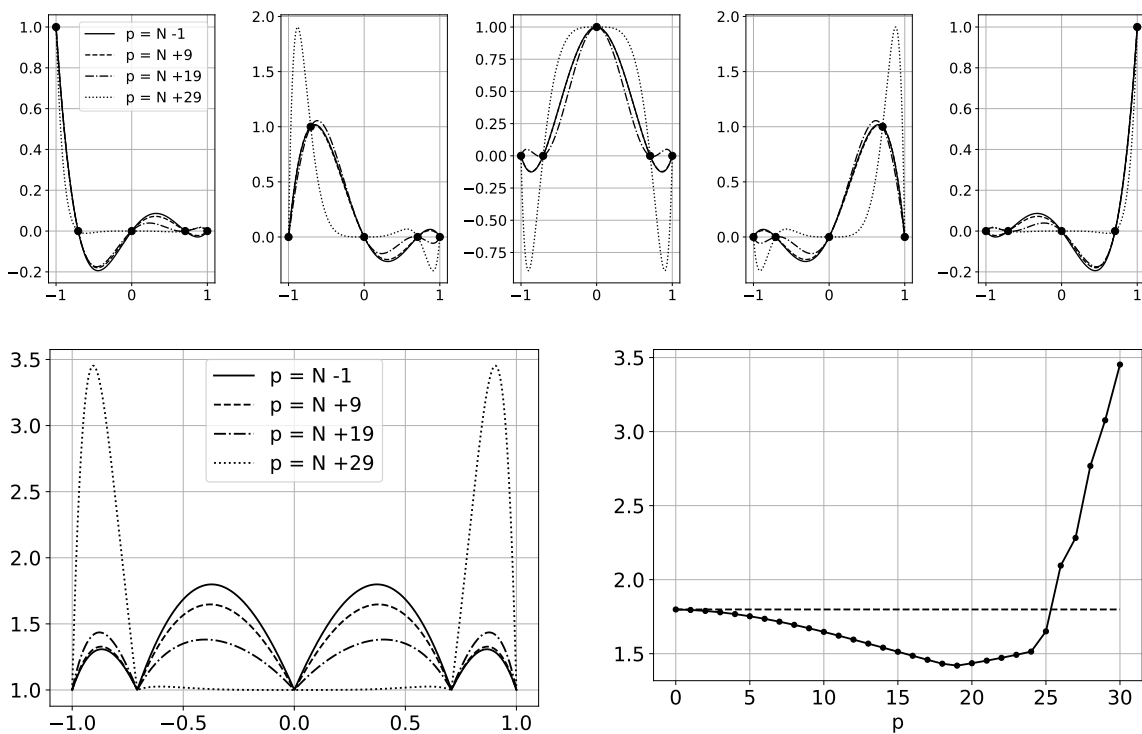
Figure 4: Lagrange functions (top), Lebesgue function (bottom left), and Lebesgue constant (bottom right) for interpolation with a polynomial kernel $k_{5,p}$ on $N = 5$ Chebyshev points and various values of p, as reported in each panel.
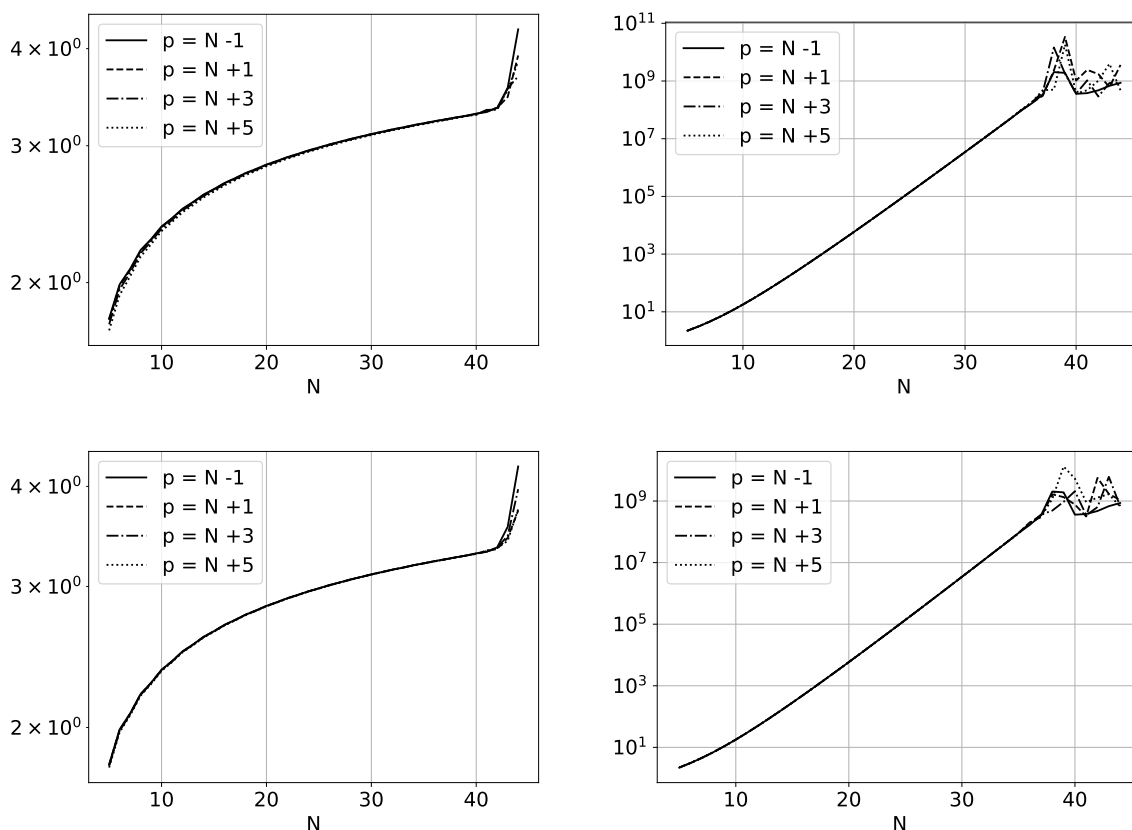
Figure 5: Growth of the Lebesgue constant associated to $N = 5, \ldots, 45$ Chebyshev (left) and equally spaced (right) points. We test kernel interpolants with various values of p, and $a = 5$ (top row) and $a = 10$ (bottom row). We remark that the results for $p = N - 1$ coincide with those of polynomial interpolation.

understanding of the role of this parameter and a systematic method for its determination are open problems. This aspect may be connected to the so-called overparameterized regime and to ridgeless regression in machine learning, since by increasing p one may aim at solving a data fitting problem by interpolation without regularization, and use the parameter a as an implicit regularizer (see e.g. [22, 26, 30]).

Moreover, some results of this paper point to the fact that the properties of an interpolation set X may be related to those of a superset X_{M_a} of polynomially unisolvent points. Also in this case a quantitative relation is missing, as well as suitable algorithms to select X from X_{M_a} or completing X to X_{M_a}. In both cases, it would be interesting to investigate processes related to Leja and approximate Fekete points in this context, as well as to P-greedy points [14].

References

[1] M. Belkin. Approximation beats concentration? An approximation view on inference with smooth radial kernels. In *Conference On Learning Theory, COLT 2018, Stockholm, Sweden, 6-9 July 2018.*, pages 1348–1361, 2018.

[2] L. Białas-Cież and J.-P. Calvi. Homogeneous minimal polynomials with prescribed interpolation conditions. *Transactions of the American Mathematical Society*, 368(12):8383–8402, 2016.

[3] L. Bos, M. Caliari, S. De Marchi, M. Vianello, and Y. Xu. Bivariate Lagrange interpolation at the Padua points: The generating curve approach. *Journal of Approximation Theory*, 143(1):15–25, 2006. Special Issue on Foundations of Computational Mathematics.

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

[5] L. Bos, S. De Marchi, A. Sommariva, and M. Vianello. Weakly admissible meshes and discrete extremal sets. *Numerical Mathematics: Theory, Methods and Applications*, 4(1):1–12, 2011.

[6] L. Bos, S. De Marchi, M. Vianello, and Y. Xu. Bivariate Lagrange interpolation at the Padua points: the ideal theory approach. *Numerische Mathematik*, 108(1):43–57, Nov 2007.

[7] M. D. Buhmann. *Radial Basis Functions: theory and implementations*, volume 12 of *Cambridge Monographs on Applied and Computational Mathematics*. Cambridge University Press, Cambridge, 2003.

[8] L. Buitinck, G. Louppe, M. Blondel, F. Pedregosa, A. Mueller, O. Grisel, V. Niculae, P. Prettenhofer, A. Gramfort, J. Grobler, R. Layton, J. VanderPlas, A. Joly, B. Holt, and G. Varoquaux. API design for machine learning software: experiences from the scikit-learn project. In *ECML PKDD Workshop: Languages for Data Mining and Machine Learning*, pages 108–122, 2013.

[9] M. Caliari, S. De Marchi, and M. Vianello. Bivariate polynomial interpolation on the square at new nodal sets. *Applied Mathematics and Computation*, 165(2):261–274, 2005.

[10] M. Caliari, S. Marchi, and M. Vianello. Algorithm 886: Padua2d—Lagrange interpolation at Padua points on bivariate domains. *ACM Trans. Math. Softw.*, 35(3), Oct 2008.

[11] K. C. Chung and T. H. Yao. On lattices admitting unique Lagrange interpolations. *SIAM J. Numer. Anal.*, 14(4):735–743, 1977.

[12] S. De Marchi and R. Schaback. Stability constants for kernel-based interpolation processes. Technical report, Dipartimento di Informatica, Università degli Studi di Verona, 2008.

[13] S. De Marchi and R. Schaback. Stability of kernel-based interpolation. *Adv. Comput. Math.*, 32(2):155–161, 2010.

[14] S. De Marchi, R. Schaback, and H. Wendland. Near-optimal data-independent point locations for Radial Basis Function interpolation. *Adv. Comput. Math.*, 23(3):317–330, 2005.

[15] G. E. Fasshauer. *Meshfree Approximation Methods with MATLAB*, volume 6 of *Interdisciplinary Mathematical Sciences*. World Scientific Publishing Co. Pte. Ltd., Hackensack, NJ, 2007.

[16] G. E. Fasshauer and M. McCourt. *Kernel-Based Approximation Methods Using MATLAB*, volume 19 of *Interdisciplinary Mathematical Sciences*. World Scientific Publishing Co. Pte. Ltd., Hackensack, NJ, 2015.

[17] G. E. Fasshauer and M. J. McCourt. Stable evaluation of Gaussian Radial Basis Function interpolants. *SIAM J. Sci. Comput.*, 34(2):A737–A762, 2012.

[18] B. Fornberg, E. Larsson, and N. Flyer. Stable computations with Gaussian Radial Basis Functions. *SIAM J. Sci. Comput.*, 33(2):869–892, 2011.

[19] M. Kanagawa, B. K. Sriperumbudur, and K. Fukumizu. Convergence guarantees for kernel-based quadrature rules in misspecified settings. *Advances in Neural Information Processing Systems*, 29, 2016.

[20] M. Kanagawa, B. K. Sriperumbudur, and K. Fukumizu. Convergence analysis of deterministic kernel-based quadrature rules in misspecified settings. *Foundations of Computational Mathematics*, Jan 2019.

[21] T. Karvonen, G. Wynne, F. Tronarp, C. Oates, and S. Särkkä. Maximum likelihood estimation and uncertainty quantification for gaussian process approximation of deterministic functions. *SIAM/ASA Journal on Uncertainty Quantification*, 8(3):926–958, 2020.

[22] T. Liang and A. Rakhlin. Just interpolate: Kernel "ridgeless" regression can generalize. *The Annals of Statistics*, 48(3):1329–1347, 2020.

[23] M. McCourt and G. E. Fasshauer. Stable likelihood computation for Gaussian random fields. In *Recent Applications of Harmonic Analysis to Function Spaces, Differential Equations, and Data Science*, pages 917–943. Springer, 2017.

[24] F. J. Narcowich, J. D. Ward, and H. Wendland. Sobolev bounds on functions with scattered zeros, with applications to Radial Basis Function surface fitting. *Mathematics of Computation*, 74(250):743–763, 2005.

[25] F. J. Narcowich, J. D. Ward, and H. Wendland. Sobolev error estimates and a Bernstein inequality for scattered data interpolation via Radial Basis Functions. *Constructive Approximation*, 24(2):175–186, Sep 2006.

[26] N. Pagliana, A. Rudi, E. De Vito, and L. Rosasco. Interpolation and learning with scale dependent kernels. *arXiv preprint arXiv:2006.09984*, 2020.

[27] M. Pazouki and R. Schaback. Bases for kernel-based spaces. *J. Comput. Appl. Math.*, 236(4):575–588, 2011.

[28] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[29] C. E. Rasmussen and C. K. I. Williams. *Gaussian Processes for Machine Learning*. The MIT Press, 2006.

[30] D. Richards, J. Mourtada, and L. Rosasco. Asymptotics of ridge (less) regression under general source condition. In *International Conference on Artificial Intelligence and Statistics*, pages 3889–3897. PMLR, 2021.

[31] S. Saitoh and Y. Sawano. *Theory of Reproducing Kernels and Applications*. Developments in Mathematics; 44. Springer, Singapore, 2016.

[32] M. Scheuerer, R. Schaback, and M. Schlather. Interpolation of spatial data – a stochastic or a deterministic problem? *European Journal of Applied Mathematics*, 24(4):601–629, 2013.

[33] B. Schölkopf and A. Smola. *Learning with Kernels*. The MIT Press, 2002.

[34] J. Shawe-Taylor and N. Cristianini. *Kernel Methods for Pattern Analysis*. Cambridge University Press, 2004.

[35] I. Steinwart and A. Christmann. *Support Vector Machines*. Science + Business Media. Springer, 2008.

[36] The MathWorks Inc. MATLAB R2021b Statistics and Machine Learning Toolbox, Natick, Massachusetts, USA.

[37] H. Wendland. *Scattered Data Approximation*, volume 17 of *Cambridge Monographs on Applied and Computational Mathematics*. Cambridge University Press, Cambridge, 2005.

[38] H. Zhang and L. Zhao. On the inclusion relation of Reproducing Kernel Hilbert Spaces. *Analysis and Applications*, 11(02):1350014, 2013.

[39] B. Zwicknagl. Power series kernels. *Constructive Approximation*, 29(1):61–84, Feb 2009.