

Approximation with Conditionally Positive Definite Kernels on Deficient Sets



Oleg Davydov

Abstract Interpolation and approximation of functionals with conditionally positive definite kernels is considered on sets of centers that are not determining for polynomials. It is shown that polynomial consistence is sufficient in order to define kernel-based numerical approximation of the functional with usual properties of optimal recovery. Application examples include generation of sparse kernel-based numerical differentiation formulas for the Laplacian on a grid and accurate approximation of a function on an ellipse.

Keywords Conditionally positive definite kernels · Numerical differentiation · Optimal recovery · Saddle point problem

1 Introduction

Let Ω be a set and P a finite dimensional space of functions on Ω . A function $K : \Omega \times \Omega \rightarrow \mathbb{R}$ is said to be a *conditionally positive definite kernel with respect to P* if for any finite set $X = \{x_1, \dots, x_n\} \subset \Omega$ the quadratic form $\sum_{i,j=1}^n c_i c_j K(x_i, x_j)$ is positive for all $c \in \mathbb{R}^n \setminus \{0\}$ such that $\sum_{i=1}^n c_i p(x_i) = 0$ for all $p \in P$ [9].

Given data (x_j, f_j) , $j = 1, \dots, n$, with $x_j \in \Omega$, $f_j \in \mathbb{R}$, a sum of the form

$$\sigma(x) = \sum_{j=1}^n c_j K(x, x_j) + \tilde{p}, \quad c_j \in \mathbb{R}, \quad \tilde{p} \in P \quad (1)$$

can be used to solve the interpolation problem

$$\sigma(x_i) = f_i, \quad i = 1, \dots, n. \quad (2)$$

O. Davydov (✉)
University of Giessen, Giessen, Germany
e-mail: oleg.davydov@math.uni-giessen.de

Moreover, a solution of (2) satisfying the condition

$$\sum_{j=1}^n c_j p(x_j) = 0 \quad \text{for all } p \in P, \quad (3)$$

can always be found [9, p. 117]. This solution is unique if X is a *determining set* for P , that is $p \in P$ and $p|_X = 0$ implies $p = 0$.

In meshless finite difference methods, conditionally positive definite kernels with respect to spaces of polynomials are often used to produce numerical approximations of linear functionals

$$\lambda f \approx \sum_{i=1}^n w_i f(x_i), \quad w_i \in \mathbb{R}, \quad (4)$$

such as the value $\lambda f = Df(x)$ of a differential operator D applied to a function f at a point $x \in \Omega$. If the interpolant $\sigma = \sigma_f$ satisfying (1)–(3) with $f_i = f(x_i)$ is uniquely defined, then the weights w_i of (4) can be obtained by the approximation $\lambda f \approx \lambda \sigma_f$, which leads to the conditions

$$\sum_{j=1}^n w_j K(x_i, x_j) + \tilde{p}(x_i) = \lambda' K(x_i), \quad i = 1, \dots, n, \quad \text{for some } \tilde{p} \in P, \quad (5)$$

$$\sum_{i=1}^n w_i p(x_i) = \lambda p \quad \text{for all } p \in P, \quad (6)$$

where $\lambda' K : \Omega \rightarrow \mathbb{R}$ is the function obtained by applying λ to the first argument of K . The weights w_i are in this case uniquely determined by the conditions (5)–(6). In particular, by introducing a basis for the space P , we can write both (1)–(3) and (5)–(6) as systems of linear equations with the same matrix which is non-singular as soon as X is a determining set for P . Solving this system is the standard way to obtain the weights w_i , see e.g. [7]. It is demonstrated in [1] that the weights satisfying (5)–(6) for a polyharmonic kernel K significantly improve the performance of meshless finite difference methods in comparison to the weights obtained by unconditionally positive definite kernels such as the Gaussian. In addition, these weights provide optimal recovery of λf from the data $f(x_i)$, $i = 1, \dots, n$, on spaces of functions of finite smoothness, see e.g. [9, Chapter 13].

An alternative interpretation of (5)–(6) is that the approximation (4) of λf is required to be exact for all $f = \sigma$ in the form (1) with coefficients c_j satisfying (3). Indeed, this can be easily shown with the help of the Fredholm alternative for matrices, see Theorem 1 below. In particular, (6) already expresses exactness of (4) for all elements of P . In the case when Ω is a domain in \mathbb{R}^d and P is a space of d -variate polynomials, (6) can be used to obtain error bounds for the numerical differentiation with weights w_i , see e.g. [5, 6].

However, exactness (6) for $p \in P$ is sometimes achievable without X being a determining set for P . We then say that X is P -consistent for λ . The best known examples are the Gauss quadrature when $\lambda f = \int_a^b f(x) dx$ and the five point stencil for the two-dimensional Laplacian. Moreover, P -consistent sets with n significantly smaller than the dimension of P often can be used for the numerical discretization of the Laplace operator on gridded nodes in irregular domains, leading to sparser differentiation matrices [3].

In this paper we study numerical approximation formulas (4) obtained by requiring exactness conditions (5)–(6) on “deficient” sets X that are not determining for P .

Our main result (Theorem 2 and Corollary 1) shows that a unique formula satisfying these conditions exists as soon as X is P -consistent. Another consequence is that the coefficients c_j of the interpolant (1) are uniquely defined for any X (Corollary 2). Numerical differentiation formulas obtained in this way provide optimal recovery on native spaces of the kernels. We also discuss computational methods for the weights of the formula (4) and coefficients of the interpolant (1). In particular, a null space method can be used for the saddle point problems (5)–(6) or (1)–(3) even if in the case of deficient sets they do not satisfy restrictions usually required in the literature [2].

In the last section we describe two types of deficient sets that arise naturally in applications. First, deficient subsets of a grid may be used for numerical differentiation of the Laplacian (Sect. 3.1). Second, function values and differential operators on algebraic surfaces, in this case an ellipse, may be approximated using data located on the manifold, which are necessarily deficient sets for polynomials in the ambient space of degree at least the order of the surface (Sect. 3.2). In both cases, numerical results demonstrate a robust performance of the suggested numerical methods, and a reasonable approximation quality of the polyharmonic kernels we employ in the experiments.

2 Approximation on Deficient Sets

We assume that $K : \Omega \times \Omega \rightarrow \mathbb{R}$ is a conditionally positive definite kernel with respect to a linear space P of functions on Ω , with $\dim P = m$. Let $\{p_1, \dots, p_m\}$ be a basis for P . By writing $\tilde{p} = \sum_{j=1}^m v_j p_j$, $v_j \in \mathbb{R}$, conditions (5)–(6) give rise to a linear system with respect to w_j and v_j , in block matrix form,

$$\begin{bmatrix} K_X & P_X \\ P_X^T & 0 \end{bmatrix} \cdot \begin{bmatrix} w \\ v \end{bmatrix} = \begin{bmatrix} a \\ b \end{bmatrix}, \quad (7)$$

where

$$K_X = [K(x_i, x_j)]_{i,j=1}^n, \quad P_X = [p_j(x_i)]_{i,j=1}^{n,m},$$

$$w = [w_j]_{j=1}^n, \quad v = [v_j]_{j=1}^m, \quad a = [\lambda' K(x_i)]_{i=1}^n, \quad b = [\lambda p_j]_{j=1}^m.$$

Condition (3) in matrix form is

$$P_X^T c = 0, \quad c = [c_j]_{j=1}^n,$$

that is c belongs to the null space $N(P_X^T)$ of P_X^T . Since

$$P_X v = [\tilde{p}(x_i)]_{i=1}^n, \quad \tilde{p} = \sum_{j=1}^m v_j p_j,$$

we see that the condition that X is a determining set for P is equivalent to $N(P_X) = 0$.

We show that the conditions (5)–(6) express the exactness of (4) for the sums σ conditional on (3), even when X is a *deficient set* for P , that is $N(P_X) \neq 0$. Recall that this condition is equivalent to $R(P_X^T) \neq \mathbb{R}^m$, where $R(A)$ denotes the range of a matrix A .

Theorem 1 *Let $X = \{x_1, \dots, x_n\} \subset \Omega$. An approximation formula (4) satisfies the exactness condition $\lambda\sigma = \sum_{i=1}^n w_i \sigma(x_i)$ for all sums σ in the form (1) with coefficients c_j satisfying (3) if and only if (5)–(6) holds for the weights w_i , $i = 1, \dots, n$.*

Proof The exactness condition is

$$\sum_{j=1}^n c_j \lambda' K(x_j) + \lambda p = \sum_{j=1}^n c_j \sum_{i=1}^n w_i K(x_i, x_j) + \sum_{i=1}^n w_i p(x_i)$$

for all $c = [c_j]_{j=1}^n$ satisfying (3) and all $p \in P$. In particular, for $c = 0$ we obtain (6), and rewrite the condition as

$$\sum_{j=1}^n c_j \left(\lambda' K(x_j) - \sum_{i=1}^n w_i K(x_i, x_j) \right) = 0 \quad \text{for all } c \in N(P_X^T).$$

By the Fredholm alternative for matrices this is equivalent to

$$\left[\lambda' K(x_j) - \sum_{i=1}^n w_i K(x_i, x_j) \right]_{j=1}^n \in R(P_X),$$

which is in turn equivalent to (5) in view of the symmetry of the kernel K . \square

Linear systems of the type (7) have been extensively studied under the name of equilibrium equations [8, Section 4.4.6] or saddle point problems [2] because they arise in many application areas. Our approach below is a variation of the null space

techniques described in [2, Section 6]. However, usual assumptions that $n \geq m$, K_X is positive semidefinite and P_X has full column rank are not satisfied in our case of interest.

As long as X is a deficient set, $R(P_X^T) \neq \mathbb{R}^m$ and hence the solvability of $P_X^T w = b$ cannot be guaranteed for all b . Nevertheless, should this last equation have a solution for w , there is a unique weight vector w satisfying (7).

Theorem 2 *There is a unique vector w satisfying (7) if and only if $b \in R(P_X^T)$.*

Proof The necessity of the condition $b \in R(P_X^T)$ is obvious. To show the sufficiency, assume that $P_X^T w_0 = b$ for some $w_0 \in \mathbb{R}^n$. Then the solution w must satisfy $P_X^T(w - w_0) = 0$ if it exists, so we look for w in the form

$$w = w_0 + \tilde{u}, \quad \tilde{u} \in N(P_X^T).$$

Let M be a matrix whose columns form a basis for $N(P_X^T)$. Then $\tilde{u} = Mu$ for some vector u , and we may write (7) equivalently as a linear system with respect to u and v ,

$$K_X Mu + P_X v = a - K_X w_0. \quad (8)$$

Since $M^T P_X = 0$, it follows that

$$M^T K_X Mu = M^T (a - K_X w_0). \quad (9)$$

Since K is conditionally positive definite, the matrix $M^T K_X M$ is positive definite, and hence there is a unique u determined by the last equation. The existence of some $v \in \mathbb{R}^m$ such that (8) holds is equivalent to the claim that $K_X Mu - a + K_X w_0 \in R(P_X)$. This claim follows from the Fredholm alternative since (9) implies that $K_X Mu - a + K_X w_0 \perp N(P_X^T)$. Thus, u and v satisfying (8) exist, and u is uniquely determined. Then $w = w_0 + Mu$ is a unique vector satisfying (7). \square

Remark 1 Theorem 2 is valid for any linear system (7) with arbitrary matrices A and B replacing K_X and P_X , respectively, and arbitrary a, b , as soon as A is definite on $N(B^T)$, that is $x^T A x \neq 0$ for all $x \in N(B^T) \setminus \{0\}$. Indeed, this condition implies that $M^T A M$ is non-singular and hence the argument in the proof goes through.

As long as the condition $b \in R(P_X^T)$ is satisfied, the weight vector w may be found by any solution method applicable to the system (7), for example via the pseudoinverse of its matrix when it is singular. Alternatively, we may use the null space matrix M of the above proof and find w from the linear system

$$\begin{bmatrix} M^T K_X \\ P_X^T \end{bmatrix} w = \begin{bmatrix} M^T a \\ b \end{bmatrix}, \quad (10)$$

which is in general overdetermined but has full rank because its solution w is unique. Indeed, any solution w of (10) satisfied (7) for some v since $M^T K_X w = M^T a$ implies $K_X w - a \perp N(P_X^T)$ and thus $K_X w - a \in R(P_X)$. We refer to [2, Section 6] for the computational methods for the null space matrix M . One obvious possibility is to employ the right singular vectors of P_X , see [8, Eq. (2.5.4)]. Should v be needed, it can be computed as a solution of the consistent linear system

$$P_X v = a - K_X w. \quad (11)$$

For example we can use

$$v = P_X^+(a - K_X w), \quad (12)$$

where P_X^+ denotes the Moore-Penrose pseudoinverse of P_X , is the unique v with the smallest 2-norm.

We formulate two immediate consequences of Theorem 2 for the numerical approximation of functionals and for the interpolation. Note that (1)–(3) can be written in the form (7) with w replaced by c , $a = [f_i]_{i=1}^n$, and $b = 0$. In particular, the condition $0 \in R(P_X^T)$ of Theorem 2 is trivially satisfied.

Corollary 1 *For any X and λ there is a unique numerical approximation formula (4) satisfying (5)–(6) as soon as (6) is solvable.*

Corollary 2 *For any data (x_j, f_j) , $j = 1, \dots, n$, one or more interpolants σ satisfying (1)–(3) exist and their coefficients c_j , $j = 1, \dots, n$, are uniquely determined.*

Thanks to Theorem 1 we also obtain the property known for the case of a determining set X that the approximation $\lambda f \approx \sum_{i=1}^n w_i f(x_i)$ can be found by requiring $\lambda f \approx \lambda \sigma$ for any interpolant σ of Corollary 2 with $f_i = f(x_i)$.

Looking specifically at numerical differentiation, consider the case when $\Omega = \mathbb{R}^d$, $\lambda f = Df(x)$ for a linear differential operator D of order k and $x \in \mathbb{R}^d$, and $P = \mathbb{P}_q^d$, the space of d -variate polynomials of total order at most q (that is, total degree at most $q - 1$) for some $q \in \mathbb{N}$. Any kernel K that is conditionally positive definite with respect to P generates a native semi-Hilbert space $F(K, P)$ of functions on Ω with null space P , see e.g. [9]. By inspecting the arguments in Section 2 and Lemma 6 of [4], we see that thanks to Corollary 1, the optimal recovery property of the weights w_i defined by (5)–(6) remains valid for deficient sets $X = \{x_1, \dots, x_n\}$. More precisely, the worst case error of numerical differentiation formulas on the unit ball of $F_q(K) := F(K, \mathbb{P}_q^d)$,

$$E(u) := \sup_{\substack{f \in F_q(K) \\ \|f\|_{F_q(K)} \leq 1}} \left| Df(x) - \sum_{i=1}^n u_i f(x_i) \right|,$$

can be computed as

$$\begin{aligned}
E^2(u) &= D' D'' K(x, x) - \sum_{i=1}^n u_i (D' K(x, x_i) + D'' K(x_i, x)) \\
&\quad + \sum_{i,j=1}^n u_i u_j K(x_i, x_j).
\end{aligned} \tag{13}$$

The weight vector w has the *optimal recovery* property in the sense that it satisfies

$$E(w) = \min \left\{ E(u) : u \in \mathbb{R}^n, Dp(x) = \sum_{i=1}^n u_i p(x_i) \text{ for all } p \in \mathbb{P}_q^d \right\} \tag{14}$$

as soon as the mixed partial derivatives of K exist at $(x, x) \in \mathbb{R}^d \times \mathbb{R}^d$ up to the order k in each of both d -dimensional variables, and X is such that there exists a vector $u \in \mathbb{R}^n$ with polynomial exactness

$$Dp(x) = \sum_{i=1}^n u_i p(x_i) \text{ for all } p \in \mathbb{P}_q^d.$$

We use here D' and D'' to indicate when $\lambda f = Df(x)$ acts on the first, respectively, the second argument of K .

Note that the equality-constrained quadratic minimization problem (14) provides an alternative way of computing the optimal weight vector on a deficient set. By Theorem 2 we know that its solution w is unique as soon as the feasible region is non-empty.

3 Examples

In this section we illustrate Corollaries 1 and 2 on particular examples where deficient sets X seem useful.

We consider the *polyharmonic kernels* $K_{s,d} : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}$, defined for all real $s > 0$ by $K_{s,d}(x, y) = \varphi_s(\|x - y\|_2)$, where

$$\varphi_s(r) := (-1)^{\lfloor s/2 \rfloor + 1} \begin{cases} r^s \log r, & \text{if } s \text{ is an even integer,} \\ r^s, & \text{otherwise.} \end{cases} \tag{15}$$

The kernel $K_{s,d}$ is conditionally positive definite with respect to \mathbb{P}_q^d for all $q \geq \lfloor s/2 \rfloor + 1$. We cite [6, 9] and references therein for further information on these kernels. If $m = (s + d)/2$ is an integer and q is chosen equal to m , then the native space $F_m(K_{s,d})$ coincides with the Beppo-Levi space $BL_m(\mathbb{R}^d)$, see [9, Theorem 10.43]. For any $q \geq \lfloor s/2 \rfloor + 1$, the space $F_q(K_{s,d})$ can be described with the help of the generalized Fourier transforms as in [9, Theorem 10.21]. By the arguments in

Sect. 2, formulas (13) and (14) apply to $K_{s,d}$ as soon as $s > 2k$, where k is the order of the differential operator D .

3.1 Numerical Differentiation of Laplacian on a Grid

We are looking for numerical differentiation formulas of the type

$$\Delta f(0) \approx \sum_{\alpha \in Z_{d,r}} w_\alpha f(\alpha), \quad Z_{d,r} := \{\alpha \in \mathbb{Z}^d : \|\alpha\|_2 \leq r\}, \quad r > 0, \quad (16)$$

where Δ is the Laplacian $\Delta f = \sum_{i=1}^d \partial^2 f / \partial x_i^2$. The set $Z_{d,r}$ for $0 \leq r < 1$ consists of the origin only and hence is not useful for the approximation of the Laplacian. For $r = 1$ we have $Z_{d,1} = \{0, \pm e_1, \dots, \pm e_d\}$, where e_i is the i -th unit vector in \mathbb{R}^d , and

$$\Delta f(0) \approx -2df(0) + \sum_{i=1}^d f(e_i) + \sum_{i=1}^d f(-e_i) \quad (17)$$

is the classical numerical differentiation formula exact for all cubic polynomials $f = p \in \mathbb{P}_4^d$. Hence (6) is solvable for all $X = Z_{d,r}$, $r \geq 1$, if $P = \mathbb{P}_4^d$.

According to Corollary 1, we have computed the unique weights of the formula (16) satisfying (5)–(6) for the kernel $K_{7,d}$, $P = \mathbb{P}_4^d$ and $X = Z_{d,r}$ for all $d = 2, \dots, 5$ and $r = 1, \sqrt{2}, \sqrt{3}, 2$. As a basis for \mathbb{P}_4^d we choose ordinary monomials. However, the computation is performed using the rescaling of X as X/r according to the suggestion in [5, Section 6.1]

Table 1 presents information about the size $|X|$ of X , dimensions of the null spaces of P_X and P_X^T , the optimal recovery error (14) on $F_4(K_{7,d})$, the stability constant of the weight vector $\|w\|_1 = \sum_{i=1}^n |w_i|$, and the condition number $cond$ of the system (10) we solved in order to compute the weights for $r \neq 1$. A smaller optimal recovery error indicates better approximation quality, whereas $\|w\|_1$ and $cond$ measure the numerical stability of the formulas. Note that $\dim N(P_X^T) = 0$ for $r = 1$, which means that (17) is the only solution of (6) in this case, and hence it provides the optimal recovery on $F_4(K_{7,d})$. For $r = 2$ we have $\dim N(P_X) = 0$ and it follows that $Z_{d,2}$ is a determining set for \mathbb{P}_4^d . For $r = \sqrt{2}, \sqrt{3}$ we obtain examples of optimal recovery weights on deficient sets, with $\dim N(P_X)$ being the dimension of the affine space of weight vectors satisfying the polynomial exactness condition (6). These new weights seem to provide a meaningful choice for the two intermediate sets between the classical polynomial stencil on $Z_{d,1}$, and the standard polyharmonic weights on the determining set $Z_{d,2}$. Indeed, as expected, the optimal recovery error $E(w)$ reduces when $|X|$ increases, whereas the stability constant and condition numbers tend to increase.

Table 1 Numerical differentiation of Laplacian on a grid: $|X|$ is the cardinality of X , $dN = \dim N(P_X)$, $dNt = \dim N(P_X^T)$, $E(w)$ is given by (13), and $cond$ is the condition number of the matrix of (10)

r	$ X $	dN	dNt	$E(w)$	$\ w\ _1$	$cond$
$d = 2, X = Z_{2,r}, \dim \mathbb{P}_4^2 = 10$						
1	5	5	0	13.4	8.0	–
$\sqrt{2}$	9	2	1	10.6	13.5	2.0e+02
$\sqrt{3}$	9	2	1	10.6	13.5	2.0e+02
2	13	0	3	7.4	11.8	3.9e+02
$d = 3, X = Z_{3,r}, \dim \mathbb{P}_4^3 = 20$						
1	7	13	0	17.2	12.0	–
$\sqrt{2}$	19	4	3	12.3	22.7	3.8e+02
$\sqrt{3}$	27	3	10	12.4	24.8	2.5e+03
2	33	0	13	9.0	30.1	5.1e+03
$d = 4, X = Z_{4,r}, \dim \mathbb{P}_4^4 = 35$						
1	9	26	0	20.8	16.0	–
$\sqrt{2}$	33	8	6	14.0	31.8	5.7e+02
$\sqrt{3}$	65	4	34	13.9	39.7	6.9e+03
2	89	0	54	10.4	40.5	3.1e+04
$d = 5, X = Z_{5,r}, \dim \mathbb{P}_4^5 = 56$						
1	11	45	0	24.2	20.0	–
$\sqrt{2}$	51	15	10	15.6	40.9	7.7e+02
$\sqrt{3}$	131	5	80	15.4	56.4	1.3e+04
2	221	0	165	11.7	55.0	9.0e+04

3.2 Interpolation of Data on Ellipse

In this example we compute the kernel interpolant (1) satisfying (3) and $\sigma(x_i) = f(x_i)$, $i = 1, \dots, n$, for the test function $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ given by

$$f(x, y) = \sin(\pi x) \sin(\pi y).$$

We use the polyharmonic kernels $K_{s,2}$ and $P = \mathbb{P}_q^2$ for the pairs

$$(s, q) = (5, 3), (7, 4), (9, 5),$$

and choose sets X with $n = |X| = 5 \cdot 2^i$, $i = 0, 1, \dots, 6$, on the ellipse \mathcal{E} with half-axes $a = 1$ and $b = 0.75$ centered at the origin. The sets are obtained by first choosing parameter values $t_i = ih$, $i = 0, \dots, n - 1$, where $h = 2\pi/n$, then adding to each t_i a random number ϵ_i with uniform distribution in the interval $[-0.3h, 0.3h]$, and selecting $x_i = (a \cos(t_i + \epsilon_i), b \sin(t_i + \epsilon_i))$. The first two sets used in our experiments are shown in Fig. 1.

Since $X \subset \mathcal{E}$ and there exists a nontrivial quadratic polynomial $p \in \mathbb{P}_3^2$ that vanishes on \mathcal{E} ,

$$p(x, y) = x^2/a^2 + y^2/b^2 - 1,$$

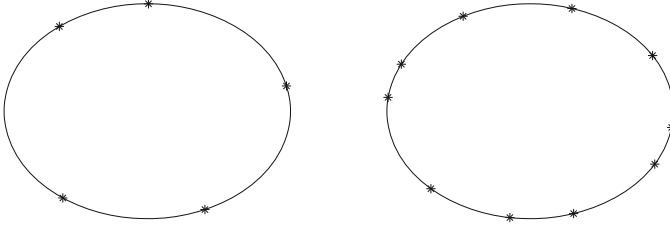


Fig. 1 Interpolation of data on ellipse: The sets X with $|X| = 5$ (left) and 10 (right)

all sets X are deficient for \mathbb{P}_q^2 , $q \geq 3$. Nevertheless, according to Corollary 2, the coefficients c_j of the interpolant σ in (1) are uniquely determined and can be computed by solving the system (10). A polynomial \tilde{p} of (1), $\tilde{p} = v_1 p_1 + \dots + v_m p_m$, can be computed by solving (11). We will use the pseudoinverse as in (12), but in fact the polynomial \tilde{p} is uniquely determined on the ellipse \mathcal{E} as soon as $n \geq 2q - 1$. Indeed, if both $\tilde{p}_1, \tilde{p}_2 \in \mathbb{P}_q^2$ satisfy (11), and $\tilde{p}_1 - \tilde{p}_2 = u_1 p_1 + \dots + u_m p_m$, then $P_X u = 0$, which implies $(\tilde{p}_1 - \tilde{p}_2)|_X = 0$. Hence x_1, \dots, x_n are intersection points of the ellipse and the zero curve of $\tilde{p}_1 - \tilde{p}_2$, an algebraic curve of order $q - 1$. By Bezout theorem, this curve must contain \mathcal{E} as soon as $n > 2(q - 1)$, which implies $\tilde{p}_1|_{\mathcal{E}} = \tilde{p}_2|_{\mathcal{E}}$.

Thus, $\sigma|_{\mathcal{E}}$ is well defined as soon as $|X| \geq 5$ for $q = 3$, $|X| \geq 7$ for $q = 4$ and $|X| \geq 9$ for $q = 5$. We are using $\sigma(x)$ as an approximation of $f(x)$ for $x \in \mathcal{E}$. Moreover, we also approximate the surface gradient

$$\nabla_{\mathcal{E}} f(x) := \nabla f(x) - \nabla f(x)^T \nu(x) \cdot \nu(x), \quad x \in \mathcal{E},$$

where $\nu(x)$ is the unit outer normal to \mathcal{E} at x . The surface gradient $\nabla_{\mathcal{E}} f(x)$ can either be approximated by $\nabla_{\mathcal{E}} \sigma(x)$, or by using a numerical differentiation formula (4), with the same result. For each X , except of $|X| = 5$ for $q = 4, 5$, we evaluated the maximum error of the function and surface gradient,

$$\begin{aligned} \max &= \max_{x \in \mathcal{E}} |f(x) - \sigma(x)|, \\ \max g &= \max_{x \in \mathcal{E}} \|\nabla_{\mathcal{E}} f(x) - \nabla_{\mathcal{E}} \sigma(x)\|_2, \end{aligned}$$

by sampling the parameter t of the ellipse $(a \cos t, b \sin t)$, $t \in [0, 2\pi)$, equidistantly with the step $h/20$. The results are presented in Table 2, where we also included the condition number *cond* of the system (10). Note that we translate and scale X using its center of gravity z , and perform the computations with $K_{s,2}$ and ordinary monomials on the set $Y = (X - z) / \max\{\|x_i - z\|_2 : i = 1, \dots, n\}$, in order to improve the condition numbers. The results in the table demonstrate a fast convergence of the interpolant σ and its surface gradient to f and $\nabla_{\mathcal{E}} f$. Note that although the condition numbers become high when the set X fills the ellipse

Table 2 Interpolation of a test function on an ellipse using the kernel $K_{s,2}$ and $P = \mathbb{P}_q^2$: $|X|$ is the cardinality of X , max and $maxg$ are the maximum errors of the function or the surface gradient, respectively, and $cond$ is the condition number of the matrix of (10)

$ X $	$s = 5, q = 3$			$s = 7, q = 4$			$s = 9, q = 5$		
	max	$maxg$	$cond$	max	$maxg$	$cond$	max	$maxg$	$cond$
5	8.3e-01	2.5e+00	4.7e+00	-	-	-	-	-	-
10	2.5e-01	9.9e-01	3.0e+02	1.8e-01	7.9e-01	1.9e+02	1.6e-01	6.9e-01	2.3e+02
20	2.9e-03	2.0e-02	2.3e+04	1.6e-03	1.1e-02	9.8e+04	1.6e-03	1.5e-02	2.0e+05
40	4.6e-05	6.9e-04	2.5e+06	2.2e-06	3.6e-05	5.3e+07	6.1e-07	1.4e-05	6.3e+08
80	1.0e-06	3.1e-05	2.4e+08	1.4e-08	3.5e-07	2.3e+10	4.2e-10	1.5e-08	1.2e+12
160	2.2e-08	1.2e-06	1.2e+10	7.5e-11	4.2e-09	4.8e+12	8.1e-12	1.8e-09	4.6e+15
320	2.7e-10	3.1e-08	9.2e+11	1.0e-12	2.2e-10	2.9e+15	9.1e-12	4.9e-10	5.8e+17

more densely, they are moderate in comparison to significantly worse conditioned matrices arising if infinitely smooth kernels such as the Gaussian $K_{G,\varepsilon}(x - y) = \exp(-\varepsilon\|x - y\|_2^2)$, $\varepsilon > 0$, are employed, see also discussions in [7, Section 5.1.5].

References

1. Bayona, V., Flyer, N., Fornberg, B., Barnett, G.A.: On the role of polynomials in RBF-FD approximations: II. Numerical solution of elliptic PDEs. *J. Comput. Phys.* **332**, 257–273 (2017). <https://doi.org/10.1016/j.jcp.2016.12.008>
2. Benzi, M., Golub, G.H., Liesen, J.: Numerical solution of saddle point problems. *Acta Numer.* **14**, 1–137 (2005)
3. Davydov, O.: Selection of sparse sets of influence for meshless finite difference methods (2019). arxiv:1908.01567
4. Davydov, O., Schaback, R.: Error bounds for kernel-based numerical differentiation. *Numer. Math.* **132**(2), 243–269 (2016). DOI <https://doi.org/10.1007/s00211-015-0722-9>
5. Davydov, O., Schaback, R.: Minimal numerical differentiation formulas. *Numer. Math.* **140**(3), 555–592 (2018). <https://doi.org/10.1007/s00211-018-0973-3>
6. Davydov, O., Schaback, R.: Optimal stencils in Sobolev spaces. *IMA J. Numer. Anal.* **39**(1), 398–422 (2019). <https://doi.org/10.1093/imanum/drx076>
7. Fornberg, B., Flyer, N.: *A Primer on Radial Basis Functions with Applications to the Geosciences*. Society for Industrial and Applied Mathematics, Philadelphia (2015)
8. Golub, G.H., Van Loan, C.F.: *Matrix Computations*, 3rd edn. The Johns Hopkins University Press, Baltimore (1996)
9. Wendland, H.: *Scattered Data Approximation*. Cambridge University Press, Cambridge (2005)