Available online at www.sciencedirect.com
ScienceDirect
Journal of Computational and Applied Mathematics 212 (2008) 127–149
JOURNAL OF COMPUTATIONAL AND APPLIED MATHEMATICS
www.elsevier.com/locate/cam

Limit problems for interpolation by analytic radial basis functions

Robert Schaback

Institut für Numerische und Angewandte Mathematik, Lotzestraße 16-18, D-37083 Göttingen, Germany

Received 30 January 2006; received in revised form 20 November 2006

Abstract

Interpolation problems for analytic radial basis functions like the Gaussian and inverse multiquadrics can degenerate in two ways: the radial basis functions can be scaled to become *increasingly flat*, or the data points *coalesce* in the limit while the radial basis functions stay fixed. Both cases call for a careful regularization, which, if carried out explicitly, yields a preconditioning technique for the degenerating linear systems behind these interpolation problems. This paper deals with both cases. For the *increasingly flat* limit, we recover results by Larsson and Fornberg together with Lee, Yoon, and Yoon concerning convergence of interpolants towards polynomials. With slight modifications, the same technique can also handle scenarios with *coalescing* data points for fixed radial basis functions. The results show that the degenerating local Lagrange interpolation problems converge towards certain Hermite–Birkhoff problems. This is an important prerequisite for dealing with approximation by radial basis functions adaptively, using freely varying data sites.

MSC: 41A15; 41A25; 41A30; 41A63; 65D10

Keywords: Radial basis functions; Moment conditions; Preconditioning; Hermite interpolation; Polynomial interpolation

1. Introduction

In the last two decades, *radial basis functions* have attracted more and more attention as tools in various fields of computational and applied mathematics. There are now books of Buhmann [1] and Wendland [19] containing the background theory, and a recent survey [17] in Acta Numerica summarizes many applications.

A central problem with a close connection to optimal recovery and supervised learning [17] is d-variate *interpolation* of real values $f_1, \dots, f_N$ at scattered *centers* $x_1, \dots, x_N \in \mathbb{R}^d$ by translates $\phi(\varepsilon\| \cdot - x_j\|_2)$ of a real-valued scalar function $\phi : [0, \infty) \to \mathbb{R}$ with an additional positive *dilation parameter* ε. This means solving a system

$$\sum_{j=1}^{N} a_j(\varepsilon)\phi(\varepsilon\|x_i - x_j\|_2) = f_i, \quad 1 \leqslant i \leqslant N, \tag{1}$$

E-mail address: schaback@math.uni-goettingen.de.

0377-0427/$ - see front matter
doi:10.1016/j.cam.2006.11.023

Table 1
Positive definite analytic radial basis functions

Name	$\phi(r)$
Gaussian	$\exp(-r^2)$
Inverse quadratic	$(1+r^2)^{-1}$
Inverse multiquadrics	$(1+r^2)^{-\beta}$, $\beta>0$
Bessel	$\frac{J_{d/2-1}(r)}{r^{d/2-1}}$

for real coefficients $a_j(\varepsilon)$, $1\leqslant j\leqslant N$, depending on ε and the data, and writing the interpolant as

$$s(\varepsilon,x):=\sum_{j=1}^{N}a_j(\varepsilon)\phi(\varepsilon\|x-x_j\|_2)\quad\text{for all } x\in\mathbb{R}^d.$$

It is well known that the coefficient matrix $S(\varepsilon)$ with entries $\phi(\varepsilon\|x_k-x_j\|_2)$, $1\leqslant j,k\leqslant N$, called the *kernel matrix* in machine learning is positive definite for all $\varepsilon>0$, any choice of N data points $x_1,\ldots,x_N\in\mathbb{R}^d$ for arbitrary dimension d, if the radial basis function is *positive definite*, and this is true in the first three cases given by Table 1. The fourth case, the *Bessel* radial basis function, was introduced in [14] and rediscovered in [5]. It still leads to positive definite systems, but the dimension d enters into the function, as is the case for compactly supported radial basis functions like those of Wendland [18].

However, the interpolation problem still depends on the dilation parameter ε, and the systems usually have a wildly increasing condition when $\varepsilon\to 0$. Nevertheless, by a surprising observation of Driscoll and Fornberg [4], the interpolants $s(\varepsilon,x)$ often converge towards multivariate polynomials when $\varepsilon\to 0$, and this is called the *increasingly flat* limit. Convergence does not always occur, and it is influenced by the geometry of the data points $x_1,\ldots,x_N$ and the radial basis function ϕ in a way which is not yet fully understood. But interpolants using the Gaussian can be proven [15] to converge towards the polynomial interpolant in [3] in all cases, and the same unconditional convergence was reported experimentally for the Bessel case in [11].

For general analytic radial basis functions and general geometric configurations, Larsson and Fornberg [11] provided interesting sufficient conditions which unfortunately were still dependent on the hypothesized nonsingularity of certain matrices. Later, Lee et al. [12] could prove their nonsingularity and extended the results to *conditionally* positive definite analytic radial basis functions.

This paper applies a completely different technique adapted from [15] written in 2002. It is useful also in other situations, e.g., for understanding preconditioning strategies for the matrices $S(\varepsilon)$. In the *increasingly flat* limit case, we recover the result of the cited papers [11,12] with a different and somewhat more direct proof, but we can also analyze the *coalescence scenario* for the first time. This interpolates with a fixed unscaled radial basis function ϕ, but (in the simplest case) in data points coalescing into the origin for $\varepsilon\to 0$. This leads to the system

$$\sum_{j=1}^{N}a_j(\varepsilon)\phi(\|\varepsilon x_i-\varepsilon x_j\|_2)=f(\varepsilon x_i),\quad 1\leqslant i\leqslant N,\tag{2}$$

similar to (1), where now the data are evaluations $f(\varepsilon x_i)$ of a smooth function f around zero. We shall show that the limit will exist in most cases, and it will be a Hermite interpolant of a specific geometry-dependent form.

The outline of the paper is as follows. Since our analysis will have close connections to multivariate polynomial interpolation, we shall start with the latter and postpone radial basis functions as much as we can. Then we turn to multivariate meshless kernel-based interpolation problems and focus on the *increasingly flat* case first, because it partially solves the *coalescing points* case also, as will turn out in Section 11. Following [15], we shall explicitly precondition the degenerating interpolation problem in such a way that the limit of the preconditioned system can be analyzed and calculated. This will solve both the flat limit and the coalescence scenario, but it also helps to a better general understanding of the ill-conditioning of systems arising in radial basis function interpolation. But note that Fornberg and Wright [6] have devised a Contour–Padé algorithm which can stably calculate multiquadric interpolants in the increasingly flat limit case without any need for preconditioning.

We insert some explicit calculations at several places of the text, instead of placing all examples at the end. These calculations serve as illustrations for our theory and are not intended to solve real-world problems.

2. Polynomial interpolation

For multivariate polynomial interpolation on a set $X := \{x_1, \dots, x_N\}$ in $\mathbb{R}^d$ there are a few important quantities to be defined a priori. To this end, we use multiindices $\alpha \in \mathbb{Z}_0^d$ in the standard way, defining the monomials $x^\alpha \in \mathbb{R}^d$ for $x \in \mathbb{R}^d$, the nonnegative integer $|\alpha| := \|\alpha\|_1$, and the multivariate derivative D^α as usual. We order multiindices $\alpha, \beta \in \mathbb{Z}_0^d$ *polynomially* by defining $\alpha < \beta$ if either $|\alpha| < |\beta|$ or $|\alpha| = |\beta|$ with $\alpha < \beta$ lexicographically. A polynomial

$$p(x) := \sum_{\alpha \in \mathbb{Z}_0^d} a_\alpha x^\alpha$$

with finitely many nonzero coefficients is an *interpolant* to data $f_1, \dots, f_N$ on X, if $p(x_i) = f_i$, $1 \leqslant i \leqslant N$. Furthermore, the space $\mathbb{P}_k^d$ of polynomials of degree at most k in d variables has dimension $\binom{k+d}{d}$. Only in rare cases will the number N of given data be equal to one of these numbers. Anyway, there always is an integer $k_1 = k_1(N, d)$ with

$$\binom{k_1 - 1 + d}{d} < N \leqslant \binom{k_1 + d}{d}. \tag{3}$$

However, even in case $N = \binom{k_1+d}{d}$ it is not at all clear whether the monomial basis $\{x^\alpha : |\alpha| \leqslant k_1\}$ is linearly independent on X. Therefore, one has to look at *monomial* or *Vandermonde* matrices formed by entries x_j^α, where we let the row index be j, $1 \leqslant j \leqslant N$, and the column index be the multiindex α. With our polynomial ordering as defined above, we can define *monomial matrices* by

$$\mathbb{M}_{|.| \leqslant k} = (x_j^\alpha)_{1 \leqslant j \leqslant N, \alpha \in \mathbb{Z}_0^d, |\alpha| \leqslant k},$$
$$\mathbb{M}_{|.| < k} = (x_j^\alpha)_{1 \leqslant j \leqslant N, \alpha \in \mathbb{Z}_0^d, |\alpha| < k}$$

with N rows, where the columns are formed by multiindices in ascending order. Existence of an interpolant of degree k for arbitrary data on X is ensured if the monomial matrix $\mathbb{M}_{|.| \leqslant k}$ has rank N. Consequently, the number $k_1 = k_1(N, d)$ of (3) is the minimum possible degree for uniquely solvable polynomial interpolation on well-chosen sets of N points in $\mathbb{R}^d$. It is attained if the determinant of the Vandermonde matrix formed by the values of the first N monomials (in our ordering) on the given N points of X is nonzero, and this can be satisfied inductively by picking suitable points.

But if the N points of X are not that nicely placed, degrees of interpolating polynomials on X can be as large as

$$k_2 := k_2(X) := \min\{k : \operatorname{rank}(\mathbb{M}_{|.| \leqslant k}) = N\}.$$

We have $k_2 \leqslant N - 1$, because the nonzero Lagrange-type polynomials

$$L_i(x) := \prod_{1 \leqslant j \leqslant N, i \neq j} \frac{(x - x_j)^{\mathrm{T}}(x_i - x_j)}{\|x_i - x_j\|_2^2}$$

have degree at most $N - 1$ and are linearly independent on X. The matrix $\mathbb{M}_{|.| \leqslant k_2}$ thus has N linearly independent columns which cannot all occur already in $\mathbb{M}_{|.| < k_2}$. Note that the bound $k_2 \leqslant N - 1$ is sharp in 1D cases, and clearly $k_1 \leqslant k_2$ holds because of (3).

Uniqueness usually is more complicated and will not hold without further assumptions. Of course, one can always select N multiindices $\alpha^1 < \cdots < \alpha^N$ with $|\alpha^i| \leqslant k_2$ such that the monomial matrix with entries $x_j^{\alpha^k}$, $1 \leqslant j, k \leqslant N$, is nonsingular. Then there is a unique interpolant in the span of x^{α^j}, $1 \leqslant j \leqslant N$, which we shall consider later, but any other choice of multiindices with the above properties will lead to a different interpolant. The parameter describing uniqueness is

$$k_0 := k_0(X) := \max\{k : p \in \mathbb{P}_k^d, \, p(X) = \{0\} \Rightarrow p = 0\}$$

defined as the maximal k such that any polynomial from $\mathbb{P}_k^d$ vanishing on X must be identically zero. Equivalently, k_0 is the maximal polynomial degree for which interpolants, if they exist, are unique. The monomial matrix $\mathbb{M}_{|.|\leqslant k_0}$ must then have rank $\binom{k_0+d}{d} \leqslant N$, and we finally get

$$0 \leqslant k_0 \leqslant k_1 \leqslant k_2 \leqslant N-1 \tag{4}$$

as a fundamental relation between the problem parameters. In Section 4 we shall show how to calculate k_0 and k_2 in general.

3. Examples

We consider the bivariate examples given by Larsson and Fornberg in [11] (Fig. 1) and calculate the constants of (4) for them in Table 2.

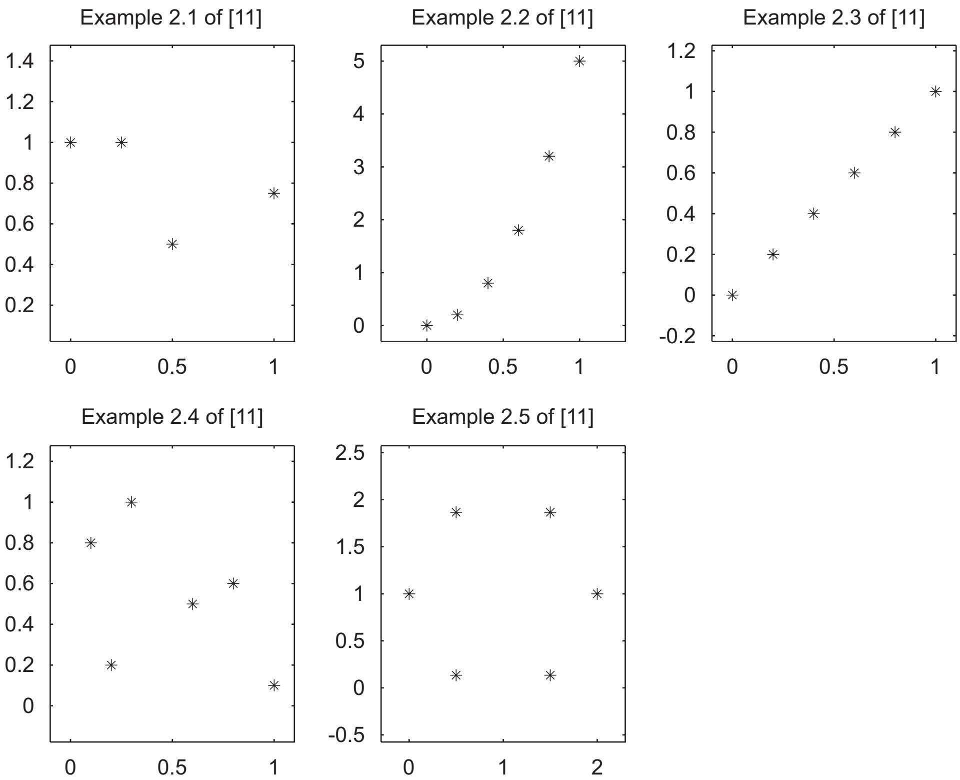

Fig. 1. Examples of [11].

Table 2
Parameters of examples of [11]

Example	k_0	k_1	k_2	N	Data
2.1	1	2	2	4	General
2.2	1	2	3	6	On parabola
2.3	0	2	5	6	On a line
2.4	2	2	2	6	General
2.5	1	2	3	6	On a circle

In order to give the reader an impression of the nontrivial geometric background of multivariate polynomial interpolation, we explain how the table entries arise.

- In 2D, all N between 4 and 6 have $k_1 = 2$ because of (3). Note that k_1 is always independent of the geometry of the given set $X \subset \mathbb{R}^d$ of N points.
- For any N points on a line in $\mathbb{R}^d$, there is a polynomial of degree 1 vanishing on all data points. Thus $k_0 = 0$ for such cases for all N and d.
- Any number of points in 2D lying on nondegenerate conics and, more generally, points in $\mathbb{R}^d$ lying on nondegenerate quadrics have a polynomial of minimal degree 2 vanishing on all data points. Thus $k_0 = 1$ in such cases for all N and d.
- To interpolate arbitrary data values on 4–6 data locations in $\mathbb{R}^2$, one cannot get away with polynomials of degree 1. Thus $k_2 \geqslant 2$ in all cases of Table 2. But degree 2 clearly suffices in cases where there are 4 points not on a line (Example 2.1) or 6 points not on a conic (Example 2.5). Remember that 5 given points in 2D always lie on a suitable conic.
- If N points are all on a line in $\mathbb{R}^d$, we need interpolants of degree $N-1$ along the 1D-parametrized line. Thus $k_2 = N - 1$ holds in such situations for all space dimensions.
- Interpolation on 6 points on a nondegenerate conic in $\mathbb{R}^2$ cannot work for degree 2, because the space of polynomials of degree up to 2 has dimension 6 while 1D is useless because of the nontrivial polynomial vanishing on the conic. Thus necessarily $3 \leqslant k_2 \leqslant 5$ in Examples 2.2 and 2.5, but k_2 depends on the geometry of points. Example 2.2 uses points $x_i = ((i-1)/5, ((i-1)/5)^2)$, $1 \leqslant i \leqslant 6$, while Example 2.5 takes 6 equidistant points on the circle.

Due to the above arguments, N data on a line in $\mathbb{R}^d$ always lead to

$$0 = k_0 \leqslant k_1 \leqslant k_2 = N - 1, \tag{5}$$

the intermediate k_1 being ridiculously dependent on the dimension d of the embedding space. This is why, in contrast to [11,12], we consider k_1 as much less relevant for analysis than the other parameters, and ignore it from now on. Note that the classical geometric situation of data points in *general position* with respect to $\mathbb{R}^d$ is the case of maximal k_0, and this case can be described by $k_0 = k_1 = k_2$ in case $N = \binom{k_1+d}{d}$, while $k_0 = k_1 - 1 = k_2 - 1$ in case $N < \binom{k_1+d}{d}$. This is satisfied in Examples 2.1 and 2.4 of Table 2.

The cited papers [11,12] prove convergence of increasingly flat radial basis function interpolants towards polynomials if, in our simplified notation, the condition

$$0 \leqslant k_2 - k_0 \leqslant 2 \tag{6}$$

holds, the intermediate k_1 being irrelevant. If 2 points are added to the 6 points on a parabola in [11, Example 2.2], we get $k_0 = 1$, $k_2 = 4$, and convergence in the flat limit turns out to fail for certain radial basis functions. Thus inequality (6) is sharp as a sufficient condition for convergence.

The proofs of [11,12] are done by an ingenious recursive analysis of various linear equation systems connecting polynomial coefficients to moments. However, this paper uses techniques of [15] to arrive at the same result and to provide additional information on degeneration caused by coalescing data points.

4. Moment matrices

In contrast to [11,12] we use the concept of a *moment matrix* here. Such matrices arose in [15] as part of the preconditioning technique, and they cared there for the geometry-dependent part of preconditioning, while the scale-dependent part was done by certain positive definite diagonal matrices. As readers will see, moment matrices are closely related to sets of multivariate (nondivided) *differences*. Since they are quite useful, they deserve a detailed introduction, including an algorithm for their calculation and showing their connection to the numbers k_0, k_1, and k_2. Readers should keep in mind that they are connected to polynomial interpolation, not to radial basis functions. This is why we treat them here, well before we deal with radial basis functions.

Given N scattered points in $\mathbb{R}^d$, we take the *monomial matrix* $\mathbb{M}_{|.| \leqslant N-1}$ with entries x_j^α and apply pivoted Gaussian elimination on it, starting from the left and proceeding to the right. Note that we index rows by j from 1 to N, but columns by multiindices $\alpha \in \mathbb{Z}_0^d$ with $|\alpha|_1 \leqslant N-1$ in ascending order. We only allow row permutations for pivoting, which only means reordering of points of X. If no nonzero pivot can be found in a certain elimination step, we proceed to work with the next column to the right, but we never move columns. Since the submatrix $\mathbb{M}_{|.|\leqslant k_2}$ is $N \times k_2$ with $k_2 \geqslant N$ and has full rank N, we do exactly N pivoting steps at N columns whose multiindices we denote by

$$\alpha^1 < \alpha^2 < \cdots < \alpha^N. \tag{7}$$

The degrees of the related monomials x^{α^i} form a sequence

$$0 = t_1 := |\alpha_1| \leqslant t_2 := |\alpha_2| \leqslant \cdots \leqslant t_N := |\alpha_N| = k_2, \tag{8}$$

where we have $t_1 = 0$ because we start with an all-ones column. Furthermore, the definition of k_2 implies that we must reach the degree $k_2 \leqslant N-1$ as soon as we find the last pivot, proving $|\alpha_N| = k_2$. This is a standard way to determine k_2 in general situations.

After suitable reordering of points, this process leads to a matrix factorization $\mathbb{M}_{|.|\leqslant k_2} = L \cdot U$ where L is a standard lower triangular $N \times N$ matrix with 1 on the diagonal, while the $N \times k_2 \geqslant N$ matrix U is not exactly upper triangular, but of full rank with a *staircase* shape of zero elements in the lower left part.

Then we define a nonsingular lower triangular $N \times N$ *moment* matrix $M = (m_{ij})$, $1 \leqslant i, j \leqslant N$, as $M := L^{-1}$ satisfying $M \cdot \mathbb{M}_{|.|\leqslant k_2} = U$. This implies the *moment conditions*

$$\begin{aligned} &\sum_{j=1}^{i} m_{ij} x_j^{\alpha} = 0 \quad \text{for all } \alpha < \alpha^i, \quad 1 \leqslant i \leqslant N, \\ &\sum_{j=1}^{i} m_{ij} x_j^{\alpha^i} \neq 0, \quad 1 \leqslant i \leqslant N, \end{aligned} \tag{9}$$

due to our pivoting process. The connection to nondivided differences is apparent, since (9) shows how the rows of the moment matrix annihilate polynomials.

The moment matrix also teaches us something about the constants k_0, k_1, and k_2. In short, if looking at the staircase-shaped positions of first nonzero elements in rows of U,

- k_0 is the degree after which the staircase leaves the diagonal to move to the right,
- k_1 is the degree necessary for forming the left $N \times N$ submatrix,
- k_2 is the degree at which the staircase hits the bottom row.

This illustrates (4), but we should give an example. Due to our arguments explaining Table 2, we should look at a case where k_2 is nontrivial, and we pick Example 2.2 with 6 points on a parabola. If we take $x_i = ((i-1)/5, (i-1)^2/25)$, $1 \leqslant i \leqslant 6$, as in [11], the left part of the matrix $U * 15,625$ comes out to be

$$\begin{pmatrix} 15,625 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 625 & 3125 & 25 & 125 & 625 & 1 & 5 & 25 & 125 \\ 0 & 0 & -6250 & 300 & 500 & 0 & 60 & 140 & 300 & 500 \\ 0 & 0 & 0 & 900 & 750 & 0 & 540 & 750 & 900 & 750 \\ 0 & 0 & 0 & 0 & -500 & 0 & 1200 & 700 & 0 & -500 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1800 & 600 & 0 & 0 \end{pmatrix} \tag{10}$$

and the above staircase rules give us $k_0 = 1 < k_1 = 2 < k_2 = 3$ and sequence (7) as

$$(0,0), (0,1), (1,0), (0,2), (1,1), (0,3) \tag{11}$$

with

$$0 = t_1 < 1 = t_2 = t_3 < 2 = t_4 = t_5 < 3 = t_6 = k_2. \tag{12}$$

The moment matrix in this case is

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 & 0 & 0 \\ 3 & -4 & 1 & 0 & 0 & 0 \\ -1 & 3 & -3 & 1 & 0 & 0 \\ \frac{5}{3} & -6 & 8 & -\frac{14}{3} & 1 & 0 \\ -1 & 5 & -10 & 10 & -5 & 1 \end{pmatrix}.$$

Up to here, there are no radial basis functions involved, and it is no problem to calculate a polynomial interpolant based on the monomials x^{α^i}, $1 \leqslant i \leqslant N$, for comparison with what comes later. This interpolant was already studied in [15] and shown to be different from other polynomial interpolants like the one in [3]. It is somewhat special in that it uses the minimum possible number of monomials, and it picks those which come first in our ordering when looking for N linearly independent columns of the monomial matrix. Thus we call it the *minimally ordered* polynomial interpolant. In many practical cases, it looks preferable to every other interpolating polynomial, and it was recently investigated and generalized in [2], calling it a *Schaback interpolant* there. We shall see examples later.

Note that Gaussian elimination on a Vandermonde matrix is a notoriously ill-conditioned process. But in view of the role of the moment matrix in preconditioning [15], this has to be expected and cannot be avoided. The computational complexity is of order $\mathcal{O}(N^2 \cdot k_2) \geqslant \mathcal{O}(N^3)$, and thus all of this makes no sense for practical use in case of very large N. But modern large-scale methods like domain decomposition or partition-of-unity techniques [17] will localize large problems anyway, and thus get away with reasonably small values of N.

This finishes our treatment of polynomial interpolation. From now on we consider interpolation by radial basis functions.

5. Function expansions

Following [4,11,15], we assume an analytic radial basis function

$$\phi(r) = \varphi(r^2) = \sum_{n=0}^{\infty} \varphi_{2n}(-1)^n r^{2n}, \quad 0 \leqslant r < R \leqslant \infty,$$

with strictly positive φ_{2n}, $n \geqslant 0$, to be given and scale it into

$$\phi_\varepsilon(r) := \phi(\varepsilon r) = \varphi(\varepsilon^2 r^2) = \sum_{n=0}^{\infty} \varphi_{2n}(-1)^n \varepsilon^{2n} r^{2n}, \quad 0 \leqslant r < R \leqslant \infty.$$

The conditions on the φ_{2n} are motivated from the standard assumption of complete monotonicity [13], but we insist on strictly positive constants here. This includes all standard analytic positive definite cases, e.g., the Gaussian and inverse multiquadrics, but also the Bessel radial basis functions.

If we insert $r := \|x - y\|_2$ into the expansion, we need an expansion of $\|x - y\|_2^{2n}$ into monomials x^β and y^α for vectors $x, y \in \mathbb{R}^d$. To this end, we define two multiindices $\alpha, \beta \in \mathbb{Z}_0^d$ to have *equal parity*, in short $(\alpha, \beta) \in \mathbb{Z}_P^{2d}$ or $\mathrm{EQP}(\alpha, \beta)$ if all components α_j and β_j have equal parity for all j, $1 \leqslant j \leqslant d$. For later use, the reader should be aware that the boolean-valued predicate EQP satisfies rules like

$$\mathrm{EQP}(\alpha, \beta) = \mathrm{EQP}(\alpha, \beta + 2\gamma) = \mathrm{EQP}(\alpha + \gamma, \beta + \gamma)$$

for any choice of multiindices $\alpha, \beta, \gamma \in \mathbb{Z}_0^d$, and likewise for plain integers.

We use Taylor's formula twice and the multinomial formula once to get

$$
\begin{aligned}
(-1)^n \|x-y\|_2^{2n} &= (-1)^n \sum_{\alpha\in\mathbb{Z}_0^d} \frac{(-y)^\alpha}{\alpha!} D^\alpha \|x\|_2^{2n} \\
&= (-1)^n \sum_{\alpha\in\mathbb{Z}_0^d} \sum_{\beta\in\mathbb{Z}_0^d} \frac{x^\beta}{\beta!} \frac{(-y)^\alpha}{\alpha!} D_{|0}^{\alpha+\beta} \|x\|_2^{2n} \\
&= \sum_{\substack{\alpha,\beta\in\mathbb{Z}_P^{2d}\\ |\alpha+\beta|=2n}} \frac{x^\beta}{\beta!} \frac{y^\alpha}{\alpha!} (-1)^{n-|\alpha|} \frac{n!(\alpha+\beta)!}{((\alpha+\beta)/2)!} \\
&= \sum_{\substack{(\alpha,\beta)\in\mathbb{Z}_P^{2d}\\ |\alpha+\beta|=2n}} c(\alpha,\beta) x^\beta y^\alpha \\
&= \sum_{\substack{\alpha,\beta\in\mathbb{Z}_0^{d}\\ |\alpha+\beta|=2n}} c(\alpha,\beta) x^\beta y^\alpha,
\end{aligned}
$$

with the symmetric kernel $c(\alpha, \beta)$ on $\mathbb{Z}_0^d \times \mathbb{Z}_0^d$ defined by

$$
(-1)^{(|\beta|-|\alpha|)/2} c(\alpha,\beta) := \frac{|(\alpha+\beta)/2|!(\alpha+\beta)!}{((\alpha+\beta)/2)!\alpha!\beta!} = \frac{D_{|0}^{\alpha+\beta}\|x\|_2^{|\alpha+\beta|}}{\alpha!\beta!} \tag{13}
$$

in case of $(\alpha, \beta) \in \mathbb{Z}_P^{2d}$ and zero else. By methods of [12], it will turn out in the following section that c is a positive definite kernel on the set $\mathbb{Z}_0^d$. Furthermore, it makes many kernels of the form $\varphi_{|\alpha+\beta|} c(\alpha, \beta)$ positive definite.

For illustration we display the $c(\alpha, \beta)$ values for 2D and $|\alpha|, |\beta| \leqslant 4$. The matrix is positive definite in spite of certain negative entries. Note the sparsity pattern due to the EQP predicate:

$$
\begin{pmatrix}
1 & 0 & 0 & -1 & 0 & -1 & 0 & 0 & 0 & 0 & 1 & 0 & 2 & 0 & 1 \\
0 & 2 & 0 & 0 & 0 & 0 & -4 & 0 & -4 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 2 & 0 & 0 & 0 & 0 & -4 & 0 & -4 & 0 & 0 & 0 & 0 & 0 \\
-1 & 0 & 0 & 6 & 0 & 2 & 0 & 0 & 0 & 0 & -15 & 0 & -18 & 0 & -3 \\
0 & 0 & 0 & 0 & 8 & 0 & 0 & 0 & 0 & 0 & 0 & -24 & 0 & -24 & 0 \\
-1 & 0 & 0 & 2 & 0 & 6 & 0 & 0 & 0 & 0 & -3 & 0 & -18 & 0 & -15 \\
0 & -4 & 0 & 0 & 0 & 0 & 20 & 0 & 12 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & -4 & 0 & 0 & 0 & 0 & 36 & 0 & 12 & 0 & 0 & 0 & 0 & 0 \\
0 & -4 & 0 & 0 & 0 & 0 & 12 & 0 & 36 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & -4 & 0 & 0 & 0 & 0 & 12 & 0 & 20 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & -15 & 0 & -3 & 0 & 0 & 0 & 0 & 70 & 0 & 60 & 0 & 6 \\
0 & 0 & 0 & 0 & -24 & 0 & 0 & 0 & 0 & 0 & 0 & 160 & 0 & 96 & 0 \\
2 & 0 & 0 & -18 & 0 & -18 & 0 & 0 & 0 & 0 & 60 & 0 & 216 & 0 & 60 \\
0 & 0 & 0 & 0 & -24 & 0 & 0 & 0 & 0 & 0 & 0 & 96 & 0 & 160 & 0 \\
1 & 0 & 0 & -3 & 0 & -15 & 0 & 0 & 0 & 0 & 6 & 0 & 60 & 0 & 70
\end{pmatrix}.
$$

Power series with these coefficients have nice convergence properties, since Neumann's series yields

$$\frac{1}{1+\|x-y\|_2^2} = \sum_{\alpha,\beta\in\mathbb{Z}_0^d} c(\alpha,\beta)x^\beta y^\alpha. \tag{14}$$

This is why we do not have to worry about local convergence of series expansions occurring below. Inserting (13) into our expansion, we get

$$\begin{aligned}
\phi_\varepsilon(\|x-y\|_2) &= \sum_{n=0}^{\infty} \varphi_{2n}\varepsilon^{2n}(-1)^n\|x-y\|_2^{2n}\\
&= \sum_{n=0}^{\infty} \varphi_{2n}\varepsilon^{2n} \sum_{\substack{(\alpha,\beta)\in\mathbb{Z}_P^{2d}\\ |\alpha+\beta|=2n}} c(\alpha,\beta)x^\beta y^\alpha\\
&= \sum_{(\alpha,\beta)\in\mathbb{Z}_P^{2d}} \varphi_{|\alpha+\beta|}\varepsilon^{|\alpha+\beta|}c(\alpha,\beta)x^\beta y^\alpha\\
&= \sum_{\alpha,\beta\in\mathbb{Z}_0^d} \varphi_{|\alpha+\beta|}\varepsilon^{|\alpha+\beta|}c(\alpha,\beta)x^\beta y^\alpha,
\end{aligned}$$

where we define φ_n to be zero for n odd.

6. Expansion kernels

We now consider symmetric matrices having elements $\varphi_{|\alpha+\beta|}c(\alpha,\beta)$ for $\alpha,\beta\in I$ from any index set $I\subset\mathbb{Z}_0^d$. Fortunately, following [12], such matrices are nonsingular under mild assumptions.

Lemma 1. *Let $\Phi() := \phi(\|.\|_2)$ be a positive definite radial kernel which is inverse Fourier transformable on $\mathbb{R}^d$ from a generalized Fourier transform which is nonnegative everywhere and positive on a set of positive measure in $\mathbb{R}^d$. Then the kernel $C(\alpha,\beta) := (-1)^{(|\beta|-|\alpha|)/2}c(\alpha,\beta)$ of* (13) *is positive definite, and symmetric matrices formed by elements of the form $\varphi_{|\alpha+\beta|}c(\alpha,\beta)$ are nonsingular.*

Proof. We start with

$$\begin{aligned}
\Phi(x-y) &= \sum_{n=0}^{\infty} \varphi_{2n}(-1)^n\|x-y\|_2^{2n}\\
&= \sum_{n=0}^{\infty} \varphi_{2n} \sum_{\substack{(\alpha,\beta)\in\mathbb{Z}_p^{2d}\\ |\alpha+\beta|=2n}} \varphi_{|\alpha+\beta|}c(\alpha,\beta)x^\beta y^\alpha\\
&= \sum_{(\alpha,\beta)\in\mathbb{Z}_p^{2d}} \varphi_{|\alpha+\beta|}c(\alpha,\beta)x^\beta y^\alpha\\
&= \sum_{\alpha,\beta\in\mathbb{Z}_0^d} (-1)^{|\alpha|}\frac{D^{\alpha+\beta}\Phi(0)}{\alpha!\beta!}x^\beta y^\alpha
\end{aligned}$$

for $x, y \in \mathbb{R}^d$, where the last equality is Taylor's formula. Since Φ is positive definite and inverse Fourier transformable, we look at a specific quadratic form with coefficients b_α for all $\alpha \in I \subset \mathbb{Z}_0^d$ and get

$$\begin{aligned}
0 &\leqslant \int_{\mathbb{R}^d} \hat{\Phi}(\omega) \left| \sum_{\alpha\in I} b_\alpha \omega^\alpha \right|^2 \mathrm{d}\omega \\
&= \sum_{\alpha\in I} \sum_{\beta\in I} b_\alpha b_\beta \int_{\mathbb{R}^d} \hat{\Phi}(\omega)\omega^{\alpha+\beta}\,\mathrm{d}\omega \\
&= \sum_{\alpha\in I} \sum_{\beta\in I} b_\alpha b_\beta (-\mathrm{i}D)^{\alpha+\beta}\Phi(0) \\
&= \sum_{\substack{\alpha,\beta\in I \\ (\alpha,\beta)\in\mathbb{Z}_P^{2d}}} b_\alpha b_\beta (-1)^{|\alpha+\beta|/2} D^{\alpha+\beta}\Phi(0) \\
&= \sum_{\alpha\in I} \sum_{\beta\in I} \alpha!\, b_\alpha \beta! b_\beta \varphi_{|\alpha+\beta|}(-1)^{(|\alpha|-|\beta|)/2} c(\alpha, \beta),
\end{aligned}$$

where we have used that $c(\alpha, \beta)$ vanishes if α, β are not of equal parity. Therefore, all matrices with entries $\varphi_{|\alpha+\beta|}C(\alpha, \beta)$ based on arbitrary index sets I are positive semidefinite. But if the above expression is zero, and if we use our special assumption (which rules out the Bessel kernel), the polynomial in the first integrand must vanish on an open set, thus all coefficients must be zero. This proves positive definiteness. As a byproduct, we get positive definiteness of the C kernel itself, if we use the inverse quadratic (14) with $\varphi_{2n} = 1$ for all n. Furthermore, all symmetric matrices formed with elements $\varphi_{|\alpha+\beta|}\ c(\alpha, \beta)$ will be nonsingular. $\square$

Repeating this argument with $g_\alpha := b_\alpha \alpha!(-1)^{|\alpha|/2}$, the sum above runs over $g_\alpha \overline{g_\beta} \varphi_{|\alpha+\beta|} c(\alpha, \beta)$. This proves:

Lemma 2. *Under the assumptions of Lemma* 1, *all matrices formed by elements* $\varphi_{|\alpha+\beta|}c(\alpha, \beta)$ *are positive definite as quadratic forms over* $\mathbb{C}$.

7. Expansions of interpolants

If we solve an interpolation problem on $X := \{x_1, \ldots, x_N\}$ using ϕ_ε and data $f_1, \ldots, f_N$, the system

$$\sum_{j=1}^{N} a_j(\varepsilon)\phi_\varepsilon(\|x_j - x_\ell\|_2) = f_\ell, \quad 1 \leqslant \ell \leqslant N, \tag{15}$$

has a unique solution for all $\varepsilon > 0$ which can be written as a quotient of determinants by Cramer's rule. The coefficients $a_j(\varepsilon)$ come out as rational functions of ε with a leading term of the form ε^{-2k} for some nonnegative k. This was first observed by Driscoll and Fornberg in [4], whereas [15] looked at similar expansions for the Lagrange basis functions, as studied in the next section.

We start by connecting this k to relevant quantities for polynomial interpolation.

Theorem 1. *Under the assumptions of Lemma* (1), *the coefficients* $a_j(\varepsilon)$ *have expansions starting with* ε^{-2k_2}.

Proof. We proceed very similar to the final section of [15] on preconditioning. From now on, multiindices α, β will always vary in $\mathbb{Z}_0^d$, and we only state additional conditions. Let $S(\varepsilon)$ be the matrix arising in (15) and use the *moment matrix* M from (9) to form the matrix $MS(\varepsilon)M^{\mathrm{T}}$ with the (r, s) entry

$$\begin{aligned}
&\sum_{j=1}^{r} m_{rj} \sum_{\ell=1}^{s} m_{s\ell}\phi_\varepsilon(\|x_j - x_\ell\|_2)\\
&\quad = \sum_{\alpha,\beta\in\mathbb{Z}_0^d} \varphi_{|\alpha+\beta|}\varepsilon^{|\alpha+\beta|}c(\alpha,\beta)\sum_{j=1}^{r} m_{rj}x_j^\beta \sum_{\ell=1}^{s} m_{s\ell}x_\ell^\alpha\\
&\quad = \sum_{\substack{\beta\geqslant\alpha^r\\ \alpha\geqslant\alpha^s\\ \mathrm{EQP}(\alpha,\beta)}} \varphi_{|\alpha+\beta|}\varepsilon^{|\alpha+\beta|}c(\alpha,\beta)v(r,\beta)v(s,\alpha),
\end{aligned}$$

with *moments*

$$v(r,\beta) := \sum_{j=1}^{r} m_{rj}x_j^\beta, \quad 1\leqslant r\leqslant N, \quad \beta\in\mathbb{Z}_0^d, \tag{16}$$

having the properties

$$\begin{aligned}
&v(r,\beta) = 0 \quad \text{for all } \beta<\alpha^r,\ 1\leqslant r\leqslant N,\ \text{in particular},\\
&v(r,\alpha^s) = 0 \quad \text{for all } 1\leqslant s<r\leqslant N,\\
&v(r,\alpha^r) \neq 0, \quad 1\leqslant r\leqslant N,
\end{aligned} \tag{17}$$

due to (9). Note that the matrix of the v values is identical to the matrix U of the LU decomposition we performed for getting the moment matrix. We can collect the terms as

$$\begin{aligned}
&\sum_{j=1}^{r} m_{rj} \sum_{\ell=1}^{s} m_{s\ell}\phi_\varepsilon(\|x_j - x_\ell\|_2)\\
&\quad = \varepsilon^{t_r+t_s} \sum_{\substack{\beta\geqslant\alpha^r\\ \alpha\geqslant\alpha^s\\ \mathrm{EQP}(\alpha,\beta)}} \varphi_{|\alpha+\beta|}\varepsilon^{|\alpha+\beta|-t_r-t_s}c(\alpha,\beta)v(r,\beta)v(s,\alpha)\\
&\quad =: \varepsilon^{t_r+t_s}B_{r,s}(\varepsilon)
\end{aligned}$$

to define a symmetric positive definite $N\times N$ matrix $B(\varepsilon)$ which converges for $\varepsilon\to 0$ to $B(0)$ with entries

$$B_{r,s}(0) = \sum_{\substack{|\beta|=t_r\\ |\alpha|=t_s\\ \mathrm{EQP}(\alpha,\beta)}} \varphi_{t_r+t_s}c(\alpha,\beta)v(r,\beta)v(s,\alpha)$$

for $1\leqslant r, s\leqslant N$ with equal parity of t_r and t_s, and zero else.

In our running example, we do not need to repeat the v values because they are already displayed by (10). Here, they generate the matrix $390,625\cdot B(0)$ with the sparsity pattern determined by (12) as

$$\begin{pmatrix}
390,625 & 0 & 0 & -22,500 & 0 & 0\\
0 & 32,500 & -62,500 & 0 & 0 & -19,200\\
0 & -62,500 & 125,000 & 0 & 0 & 24,000\\
-22,500 & 0 & 0 & 14,976 & -4800 & 0\\
0 & 0 & 0 & -4800 & 3200 & 0\\
0 & -19,200 & 24,000 & 0 & 0 & 124,416
\end{pmatrix}.$$

Lemma 3. *The matrix* $B(0)$ *is nonsingular.*

Proof. We take an arbitrary $u \in \mathbb{R}^N$, define the set

$$I := \{\alpha \in \mathbb{Z}_0^d : |\alpha| = t_r \text{ for some } r,\ 1 \leqslant r \leqslant N\}$$

and a function R which associates to each $\beta \in I$ the set

$$R(\beta) := \{j : |\beta| = t_j,\ 1 \leqslant j \leqslant N\}.$$

Then we evaluate the quadratic form

$$\begin{aligned}
&\sum_{r=1}^{N}\sum_{s=1}^{N} u_r u_s B_{r,s}(0)(-1)^{(t_r-t_s)/2} \\
&\quad = \sum_{r=1}^{N}\sum_{s=1}^{N} u_r u_s (-1)^{(t_r-t_s)/2} \sum_{\substack{|\beta|=t_r \\ |\alpha|=t_s \\ \alpha,\beta\in I \\ \mathrm{EQP}(\alpha,\beta)}} \varphi_{|\alpha+\beta|} c(\alpha,\beta) v(r,\beta) v(s,\alpha) \\
&\quad = \sum_{\substack{\alpha,\beta\in I \\ \mathrm{EQP}(\alpha,\beta)}} \varphi_{|\alpha+\beta|} c(\alpha,\beta)(-1)^{(|\alpha|-|\beta|)/2} \sum_{r\in R(\beta)} u_r v(r,\beta) \sum_{s\in R(\alpha)} u_s v(s,\alpha)
\end{aligned}$$

which clearly is positive semidefinite due to Lemma 1 and because it is the limit of positive definite quadratic forms. It is positive definite, because from

$$\sum_{r\in R(\beta)} u_r v(r,\beta) = 0 \quad \text{for all } \beta \in I$$

we can conclude $u = 0$ by inserting $\beta = \alpha^1, \ldots, \alpha^N$ one after another, applying (17). This finishes the proof of the lemma. □

With an $N \times N$ diagonal matrix $D(\varepsilon)$ with entries ε^{-t_k}, $1 \leqslant k \leqslant N$, system (15) is rewritten as

$$\begin{aligned}
y &= S(\varepsilon)a(\varepsilon), \\
D(\varepsilon)My &= \underbrace{D(\varepsilon)MS(\varepsilon)M^{\mathrm{T}}D(\varepsilon)}_{=:B(\varepsilon)} D^{-1}(\varepsilon)(M^{\mathrm{T}})^{-1}a(\varepsilon) \\
&= B(\varepsilon)D^{-1}(\varepsilon)(M^{\mathrm{T}})^{-1}a(\varepsilon)
\end{aligned}$$

to get the solution as a rational vector-valued function

$$a(\varepsilon) = M^{\mathrm{T}}D(\varepsilon)B^{-1}(\varepsilon)D(\varepsilon)My$$

for all positive ε with an asymptotic behavior which has at most $\varepsilon^{2t_N} = \varepsilon^{2k_2}$ in the denominator. □

We shall not use Theorem 1 directly, because it concerns the coefficients of interpolants in terms of the degenerating basis $\phi(\varepsilon\|x - x_j\|_2)$, $1 \leqslant j \leqslant N$. Naturally, these coefficients are much less stable than coefficients $u_j(x,\varepsilon)$ of a Lagrange basis. This observation motivates the next section.

8. Expansions of Lagrange bases

We write the standard linear system for Lagrange interpolating functions $u_j(x,\varepsilon)$ satisfying $u_j(x_k,\varepsilon) = \delta_{jk}$, $1 \leqslant j, k \leqslant N$, as $S(\varepsilon)u(x,\varepsilon) = \Phi_X(x,\varepsilon)$ with

$$\Phi_X(x,\varepsilon) := (\phi(\varepsilon\|x - x_1\|_2), \ldots, \phi(\varepsilon\|x - x_N\|_2))^{\mathrm{T}}$$

and transform it into

$$\underbrace{D(\varepsilon)MS(\varepsilon)M^{\mathrm{T}}D(\varepsilon)}_{=B(\varepsilon)}\underbrace{D^{-1}(\varepsilon)(M^{-1})^{\mathrm{T}}u(x,\varepsilon)}_{=:v(x,\varepsilon)}=\underbrace{D(\varepsilon)M\Phi_X(x,\varepsilon)}_{=:w(x,\varepsilon)} \tag{18}$$

to make it stably solvable, as we shall see, led by the last section of [15]. Note that proving a stable limit for (18) shows how to precondition the matrix $S(\varepsilon)$ successfully. This is the background link of our technique to preconditioning. Scaling is handled by the diagonal matrices $D(\varepsilon)$, while geometry is cared for by the moment matrix M which is independent of scaling. The diagonal matrices $D(\varepsilon)$ contain the ε-dependent denominators needed to turn $D(\varepsilon)M$ into a matrix of *divided* differences.

We expand the elements of the $B(\varepsilon)$ matrix as follows:

$$\begin{aligned}
B_{r,s}(\varepsilon) &= \sum_{\substack{|\beta|\geqslant t_r\\ |\alpha|\geqslant t_s}} \varphi_{|\alpha+\beta|}\varepsilon^{|\alpha+\beta|-t_r-t_s}c(\alpha,\beta)v(r,\beta)v(s,\alpha)\\
&= \sum_{n=0}^{\infty}\varepsilon^n \sum_{\substack{|\beta|\geqslant t_r\\ |\alpha|\geqslant t_s\\ |\alpha+\beta|=n+t_r+t_s}} \varphi_{|\alpha+\beta|}c(\alpha,\beta)v(r,\beta)v(s,\alpha)\\
&=: \sum_{n=0}^{\infty}\varepsilon^n B_{r,s,n}
\end{aligned} \tag{19}$$

with coefficients

$$B_{r,s,n} = \sum_{\substack{|\beta|\geqslant t_r\\ |\alpha|\geqslant t_s\\ |\alpha+\beta|=n+t_r+t_s}} \varphi_{|\alpha+\beta|}c(\alpha,\beta)v(r,\beta)v(s,\alpha)$$

which are zero unless $\mathrm{EQP}(n, t_r+t_s)$ holds. The components of the right-hand side of (18) are

$$\begin{aligned}
w_j(x,\varepsilon) &:= \varepsilon^{-t_j}\sum_{k=1}^{j} m_{jk}\phi(\varepsilon\|x-x_k\|_2)\\
&= \varepsilon^{-t_j}\sum\nolimits_{k=1}^{j} m_{jk}\sum_{\alpha,\beta}\varphi_{|\alpha+\beta|}\varepsilon^{|\alpha+\beta|}c(\alpha,\beta)x^\beta x_k^\alpha\\
&= \sum_{\substack{\alpha,\beta\\ |\alpha|\geqslant t_j}} \varphi_{|\alpha+\beta|}c(\alpha,\beta)\varepsilon^{|\alpha|-t_j}(\varepsilon x)^\beta v(j,\alpha)\\
&= \sum_{n=0}^{\infty}\varepsilon^n \sum_{\substack{\alpha,\beta\\ |\alpha|=n+t_j}} \varphi_{|\alpha+\beta|}c(\alpha,\beta)(\varepsilon x)^\beta v(j,\alpha)\\
&= \sum_{n=0}^{\infty}\varepsilon^n w_{j,n}(\varepsilon x),
\end{aligned} \tag{20}$$

where we defined

$$\begin{aligned}
w_{j,n}(y) &:= \sum_{\substack{\alpha,\beta\\ |\alpha|=n+t_j}} \varphi_{|\alpha+\beta|}c(\alpha,\beta)y^\beta v(j,\alpha)\\
&=: \sum_{\beta} w_{j,n,\beta}y^\beta
\end{aligned}$$

having coefficients

$$w_{j,n,\beta} = \sum_{\substack{\alpha \\ |\alpha|=n+t_j}} \varphi_{|\alpha+\beta|} c(\alpha,\beta) v(j,\alpha) \tag{21}$$

which can be nonzero only if $\mathrm{EQP}(|\beta|, n+t_j)$ holds. The special representation (20) of the right-hand side leads us to postulate a similar representation

$$v_j(x,\varepsilon) = \sum_{n=0}^{\infty} \varepsilon^n v_{j,n}(x\varepsilon) \tag{22}$$

for the solution. If we plug this into the full system, we get

$$\begin{aligned} \sum_{n=0}^{\infty} \varepsilon^n w_{r,n}(x\varepsilon) &= \sum_{n=0}^{\infty} \varepsilon^n \sum_{s=1}^{N} B_{r,s,n} \sum_{m=0}^{\infty} \varepsilon^m v_{s,m}(x\varepsilon) \\ &= \sum_{k=0}^{\infty} \varepsilon^k \sum_{m=0}^{k} \sum_{s=1}^{N} v_{s,m}(x\varepsilon) B_{r,s,k-m} \end{aligned}$$

and this is satisfied, because the nonsingularity of $B(0)$ proven in Lemma 3 allows to solve the recursive linear system

$$\begin{aligned} w_{r,n}(y) &= \sum_{m=0}^{n} \sum_{s=1}^{N} v_{s,m}(y) B_{r,s,n-m} \\ &= \sum_{s=1}^{N} v_{s,n}(y) B_{r,s,0} + \sum_{m=0}^{n-1} \sum_{s=1}^{N} v_{s,m}(y) B_{r,s,n-m} \end{aligned}$$

for all $n \geqslant 0$, $1 \leqslant r \leqslant N$. This justifies (22) and allows a recursive component wise calculation in the form

$$\begin{aligned} \sum_{\beta \in \mathbb{Z}_0^d} w_{r,n,\beta} y^\beta &= \sum_{m=0}^{n} \sum_{s=1}^{N} \sum_{\beta \in \mathbb{Z}_0^d} v_{s,m,\beta} y^\beta B_{r,s,n-m} \\ &= \sum_{\beta \in \mathbb{Z}_0^d} y^\beta \left(\sum_{s=1}^{N} v_{s,n,\beta} B_{r,s,0} + \sum_{m=0}^{n-1} \sum_{s=1}^{N} v_{s,m,\beta} B_{r,s,n-m} \right), \end{aligned}$$

$$w_{r,n,\beta} = \sum_{s=1}^{N} v_{s,n,\beta} B_{r,s,0} + \sum_{m=0}^{n-1} \sum_{s=1}^{N} v_{s,m,\beta} B_{r,s,n-m} \tag{23}$$

of the representation

$$v_{s,m}(y) =: \sum_{\beta \in \mathbb{Z}_0^d} v_{s,m,\beta} y^\beta.$$

Here, nonzero coefficients can only occur if $\mathrm{EQP}(|\beta|, m+t_s)$ holds, as was the case for the $w_{r,n}$ expansion coefficients. To see this, we look inductively at (23) in case $\mathrm{EQP}(|\beta|, n+t_r)$ fails. Then the left-hand side is zero, and so is the double sum, because it contains only terms with $\mathrm{EQP}(|\beta|, m+t_s)$ and $\mathrm{EQP}(n-m, t_r+t_s)$ which imply $\mathrm{EQP}(|\beta|, n+t_r)$. Thus the solution is zero.

We now exploit $u(x,\varepsilon)=M^{\mathrm{T}}D(\varepsilon)v(x,\varepsilon)$ componentwise with (22) to get

$$\begin{aligned}
u_j(x,\varepsilon) &= \sum_{m=0}^{\infty}\varepsilon^m\sum_{k=j}^{N} m_{k,j}\varepsilon^{-t_k}v_{k,m}(x\varepsilon)\\
&= \sum_{m=0}^{\infty}\varepsilon^m\sum_{k=j}^{N} m_{k,j}\varepsilon^{-t_k}\sum_{\alpha\in\mathbb{Z}_0^d} v_{k,m,\alpha}x^\alpha\varepsilon^{|\alpha|}\\
&= \varepsilon^{-t_N}\sum_{n=0}^{\infty}\varepsilon^n\sum_{k=j}^{N} m_{k,j}\sum_{\substack{\alpha\in\mathbb{Z}_0^d\\ |\alpha|\leqslant n+t_k-t_N}} v_{k,n+t_k-t_N-|\alpha|,\alpha}x^\alpha\\
&= \varepsilon^{-t_N}\sum_{n=0}^{\infty}\varepsilon^n P_{j,n}(x)
\end{aligned} \tag{24}$$

with polynomials and coefficients

$$\begin{aligned}
P_{j,n}(x) &= \sum_{\substack{\alpha\in\mathbb{Z}_0^d\\ |\alpha|\leqslant n}} P_{j,n,\alpha}x^\alpha,\\
P_{j,n,\alpha} &= \sum_{\substack{k=j\\ |\alpha|\leqslant n+t_k-t_N\leqslant n}}^{N} v_{k,n+t_k-t_N-|\alpha|,\alpha}m_{k,j}.
\end{aligned} \tag{25}$$

Note that the degeneration of Lagrange basis functions is only like $\varepsilon^{-k_2}=\varepsilon^{-t_N}$, while the solution of (15) degenerates like ε^{-2k_2}.

9. Convergence conditions

Now it is time to draw conclusions from the above expansions.

Lemma 4. *All polynomials $P_{j,n}$ are zero unless* EQP(n,k_2) *holds.*

Proof. In fact, the equation for $P_{j,n,\alpha}$ contains only terms with

$$\mathrm{EQP}(|\alpha|, n+t_k-t_N-|\alpha|+t_k)=\mathrm{EQP}(0,n-t_N)=EQP(0,n-k_2). \qquad \square$$

As in the cited papers, expansion (24) implies

$$\begin{aligned}
&P_{j,n}(x_k)=0, \quad 1\leqslant j,k\leqslant N,\quad n\geqslant 0,\quad n\neq k_2=t_N,\\
&P_{j,k_2}(x_k)=\delta_{jk}, \quad 1\leqslant j,k\leqslant N.
\end{aligned} \tag{26}$$

Theorem 2. *For analytic positive definite radial basis functions with positive Fourier transforms on a set of positive measure, increasingly flat interpolants will converge to a polynomial if and only if all polynomials $P_{j,n}$, $1\leqslant j\leqslant N$, $0\leqslant n<k_2$, calculated in the previous section are zero. If convergence takes place, the polynomials P_{j,k_2}, $1\leqslant j\leqslant N$, are the* flat limit *Lagrange basis on the given N points. Convergence is guaranteed, if $k_2\leqslant k_0+2$ holds.*

Proof. Assume nonconvergence. Then there are j,n with $1\leqslant j\leqslant N$, $0\leqslant n<k_2$ and EQP(n,k_2) such that $P_{j,n}$ does not vanish. This polynomial then must have a degree larger than k_0, because it vanishes on X and is nonzero. This implies

$$k_0<\deg\ P_{j,n}\leqslant n\leqslant k_2-2<k_2 \quad \text{with EQP}(n,k_2). \qquad \square$$

Note that by [12] the final assertion of Theorem 2 holds also in the conditionally positive definite case, but our proof technique does not currently cover this.

10. Computations in the flat limit case

In our running example, we can explicitly calculate the quantities $B_{r,s,n}$, $w_{j,n,\beta}$, $v_{k,m,\alpha}$, $P_{j,n,\alpha}$ along the lines of the previous section, but we do not display the matrices here. The polynomial $P_{1,3}$ arising as a *flat limit* will be the Lagrange interpolant to data (1, 0, 0, 0, 0, 0) at the 6 points given, and via the coefficients $P_{1,3,\alpha}$ we get

$$P_{1,3}(x,y) = \tfrac{1}{816}(816 + 26{,}034y - 9100x + 46{,}500y^2 - 61{,}500xy + 9750x^2 - 5625y^3 - 4375xy^2 - 1875x^2y - 625x^3) \tag{27}$$

in case of the inverse quadratic. This agrees with the findings of Larsson and Fornberg [11].

But we should explain our calculations in more detail, because we have to determine how far to calculate the expansions. Looking back at (25), it suffices to calculate the $P_{j,n,\alpha}$ for $1 \leqslant j \leqslant N$, $0 \leqslant n \leqslant k_2$ and $|\alpha| \leqslant k_2$. This requires $v_{j,n,\alpha}$ for the same range. From (23) we see that also the $w_{j,n,\alpha}$ share this range, and we need the $B_{r,s,k}$ for $1 \leqslant r, s \leqslant N$, $0 \leqslant k \leqslant k_2$. However, Eqs. (21) and (19) imply that we need the c and v values for multiindices $|\alpha| \leqslant 2k_2$ to calculate those values. Altogether, this fixes finite data to work with, and it is quite straightforward to program all necessary linear algebra calculations. Degeneracy occurs if one of the polynomials $P_{j,n}$ with $n < k_2$ and EQP(n, k_2) comes out to be nonzero, or if one of their coefficients $P_{j,n,\alpha}$ does not vanish. Thus our technique allows a complete analysis of the flat limit scenario, even in the degenerated cases.

We calculate two other results for illustration. Using interpolation by the span of the 6 monomials with exponents (11) and ignoring radial basis functions at all, we get the minimally ordered Lagrange interpolant

$$P(x,y) = \tfrac{1}{72}(72 - 625y^3 + 3500y^2 - 798x + 3101y - 5250xy)$$

which looks simpler than (27) but does the same job.

Finally, we should consider the degenerate Example 2.3 of Larsson and Fornberg [11]. The points are $x_i = ((i-1)/5, (i-1)/5)$, $1 \leqslant i \leqslant 6$, such that everything is annihilated by the polynomial $x - y$. Pivoting must increase the degree at each step, and thus we get $t_i = i - 1$, $1 \leqslant i \leqslant 6$, and the multiindices from (7) as

$$(0,0), (0,1), (0,2), (0,3), (0,4), (0,5)$$

because they are lexicographically minimal. Without looking at radial basis functions, the minimally ordered interpolant generated by these monomials to data (1, 0, 0, 0, 0, 0) is

$$\begin{aligned} P(x,y) &= \frac{1}{24}(24 - 625y^5 - 274y + 1125y^2 - 2125y^3 + 1875y^4) \\ &= \frac{625}{24}\prod_{i=1}^{5}\left(y - \frac{i}{5}\right). \end{aligned}$$

When interpolating with radial basis functions, the Gaussian case must converge in the flat limit due to [15], and our method reproduces the result of [11] in this case, which is identical to the de Boor and Ron [3] interpolant by [15]. For the inverse quadratic, degeneration occurs by nonzero polynomials $P_{j,3}$ for $k_2 = 5$. For example, we get

$$P_{1,3}(x,y) = -\frac{625}{10{,}464}(x + y - 2)(x - y)^2$$

satisfying (26). Though it does not arise as a *flat limit*, the polynomial $P_{1,5}$ is well defined and must also be an interpolant to Lagrange data due to (26). It comes out to be as terrible as

$$\begin{aligned} P_{1,5}(x,y) = \frac{1}{50,643,855,552}&(50,643,855,552 - 289,092,008,776y - 289,092,008,776x\\ &+ 644,704,707,000y^2 + 1,084,521,315,000xy + 644,704,707,000x^2\\ &- 669,165,369,750y^3 - 1,572,880,318,750xy^2 - 1,572,880,318,750x^2y\\ &- 669,165,369,750x^3 + 329,367,480,000y^4 + 996,994,387,500xy^3\\ &+ 1,303,827,480,000x^2y^2 + 996,994,387,500x^3y + 329,367,480,000x^4\\ &- 62,010,168,125y^5 - 231,403,798,125xy^4 - 366,011,236,250x^2y^3\\ &- 366,011,236,250x^3y^2 - 231,403,798,125x^4y - 62,010,168,125x^5). \end{aligned}$$

11. Radial coalescence

We now leave the *increasingly flat* scenario and go over to the *coalescence* scenario as described in (2). Given a set $X := \{x_1, \dots, x_N\}$ of N scattered points in $\mathbb{R}^d$, we want to study what happens if the points are very close to each other, and finally coalesce into a single point. Since everything we do with radial basis functions is translation-invariant, we can model this by considering points $\varepsilon \cdot x_i$, $1 \leqslant i \leqslant N$, for $\varepsilon \to 0$, leading to coalescence into the origin, but we keep the radial basis function ϕ fixed. We call this scenario *radial coalescence*. Consequently, the relative geometry of the points is the same for all positive ε, but the limit of the interpolation process will depend on this geometry. We shall relax this assumption to some extent in Section 14.

Clearly, if we look at smooth interpolants to smooth functions using smooth radial basis functions, we expect the interpolants to converge somehow for $\varepsilon \to 0$, and from univariate polynomial interpolation we can expect that the limit is a nonpolynomial Hermite interpolant and has nothing to do with the flat limit polynomials we studied up to now. Since Wu [20] provided the basics of Hermite interpolation by radial basis functions, we have no problems dealing with the limit problems.

Below, we shall prove that Lagrange interpolation in coalescing points indeed converges to Hermite interpolation, but currently we do not know how to proceed from N coalescing Lagrange data close to the origin to N hopefully linear independent Hermite data taken *in* the origin. Furthermore, we would not like to end up with Hermite data requiring *all* derivatives of order up to $N-1$ in the origin. Unfortunately, exactly this order is necessary if the coalescing Lagrange data are on a line, but in this case we only need *directional* derivatives of order up to $N-1$ in a *single* direction. This leads us to expect that we can always get away with some directional derivatives of maximal order k_2 in the origin, directions and orders being dependent on the local geometry of coalescing points. Thus we want to explicitly know N linear independent data functionals arising in the limit, making up the limiting Hermite problem, and these functionals should be linear combinations of directional derivatives in the origin of order at most k_2.

We now tackle this problem, and we need the whole machinery we developed up to here. From (20) we see that

$$w_j(x,\varepsilon) = \sum_{n=0}^{\infty} \varepsilon^n w_{j,n}(y) = \varepsilon^{-t_j} \sum_{k=j}^{N} m_{jk}\phi(\|y - \varepsilon x_k\|_2), \quad y = \varepsilon x,$$

is the right-hand side of a Lagrange-type system of equations where ϕ is not scaled, but where the data points εx_k coalesce radially into zero for $\varepsilon \to 0$. We define linear functionals

$$\lambda_{j,\varepsilon}(f) := \varepsilon^{-t_j} \sum_{k=1}^{j} m_{jk} f(\varepsilon x_k), \quad 1 \leqslant j \leqslant N, \tag{28}$$

and use them for interpolation with the basis

$$w_j(x,\varepsilon) = \lambda^y_{j,\varepsilon}\phi(\|x-y\|_2)$$

generated by $\lambda_{j,\varepsilon}$ acting with respect to y on $\phi(\|x-y\|_2)$. This leads to system (18) having the interpolation matrix $B(\varepsilon)$ with entries we rewrite as

$$\lambda^x_{r,\varepsilon}\lambda^y_{s,\varepsilon}\phi(\|x-y\|_2)=B_{r,s}(\varepsilon) \tag{29}$$

using our functionals (28) now. This is exactly the setting of generalized interpolation started by Wu in [20]. Our scaling is such that in the dual of the native space for ϕ (see [19] for details of native spaces) we have

$$\|\lambda_{r,\varepsilon}\|^2_\phi=B_{r,r}(\varepsilon)\to B_{r,r}(0)>0$$

for $\varepsilon\to 0$. The finite-dimensional interpolation space which arises in the limit will now not consist of polynomials, but rather be spanned by the functions

$$w_{j,0}(y)=\sum_{\substack{\alpha,\beta\\|\alpha|=t_j}}\varphi_{|\alpha+\beta|}c(\alpha,\beta)y^\beta v(j,\alpha). \tag{30}$$

They span the same space as the functions $v_{j,0}$ we get when taking the limit of (18), because the $v_{j,0}$ are generated from the $w_{j,0}$ by application of $B(0)^{-1}$. The functions above are of the form $w_{j,0}(x,\varepsilon)=\lambda^y_{j,0}\phi(\|x-y\|_2)$ for limit functionals

$$\lambda_{j,0}(x^\alpha):=\begin{Bmatrix} v(j,\alpha) & |\alpha|=t_j,\\ 0 & \text{else}\end{Bmatrix}$$

which act like t_j-fold derivatives at zero. In fact, they can be rewritten as

$$\lambda_{j,0}(f)=\sum_{|\beta|=t_j}v(j,\beta)\frac{(D^\beta f)(0)}{\beta!} \tag{31}$$

when acting on analytic functions f. They still are linearly independent because of

$$(\lambda_{r,0},\lambda_{s,0})_\phi=B_{r,s}(0),\quad 1\leqslant r,s\leqslant N,$$

and they can be read off the rows of the U matrix we obtained when calculating the moment matrix. We summarize:

Theorem 3. *Under the hypotheses of Theorem* (2), *radially coalescent Lagrange interpolation problems with analytic radial basis functions converge towards Hermite interpolation problems with a maximal differentiation order* k_2 *of limit functionals. The limit functionals for the Hermite problems can be explicitly calculated via* (31), *and they are independent of the radial basis function used.*

12. Calculations for the coalescence scenario

If we go back to the matrix in (10), we can read off from the rows and (31) that the 6 limiting Hermite data functionals in this special case are

$$f,\quad \frac{\partial f}{\partial y}+5\frac{\partial f}{\partial x},\quad \frac{\partial f}{\partial x},\quad 3\frac{\partial^2 f}{\partial y^2}+5\frac{\partial^2 f}{\partial x\partial y},\quad \frac{\partial^2 f}{\partial x\partial y},\quad \frac{\partial^3 f}{\partial y^3}+\frac{\partial^3 f}{\partial x\partial y^2}$$

up to multiplicative constants and evaluated at the origin. Since it is rather special to let 6 points of a parabola coalesce radially, we also look at Example 2.5 of Larsson and Fornberg [11]. There, 6 points of a regular hexagon coalesce radially into its center. Unfortunately, the corresponding U matrix is too large to be printed here. The numbers t_i are as in (12), but instead of (11) we have $\alpha^6=(1,2)$. Then, after close inspection of the U matrix, the limiting Hermite functionals come out to be

$$f,\quad \sqrt{3}\frac{\partial f}{\partial y}-\frac{\partial f}{\partial x},\quad \frac{\partial f}{\partial x},\quad \frac{\partial^2 f}{\partial y^2}+\sqrt{3}\frac{\partial^2 f}{\partial x\partial y}-\frac{\partial^2 f}{\partial x^2},\quad \frac{\partial^2 f}{\partial x\partial y},\quad \frac{\partial^3 f}{\partial x\partial^2 y}-\frac{\partial^3 f}{\partial^2 x\partial y}.$$

But this is only an illustration of the expansions we developed here, and not intended for practical use. For small local subsystems, e.g., when using partitions of unity, one can replace the standard system by the stabilized version

$$\sum_{j=1}^{N} \lambda^x_{i,\varepsilon}\lambda^y_{j,\varepsilon}\phi(\|x-y\|_2)b_j(\varepsilon)=\lambda_{i,\varepsilon}(f), \quad 1\leqslant i\leqslant N,$$

which converges to the Hermite limiting case for $\varepsilon\to 0$.

But in practice this approach should be put upside down. A set of close points y_k, $1\leqslant k\leqslant N$, should be rewritten as $y_k=y_0+\varepsilon\cdot x_k$ with a small ε, i.e., the points y_k are *pushed out* towards x_k from a central point y_0 to be chosen. This defines a small ε, and then the above preconditioned system using the x_k and ε is set up and solved. This interprets the given points y_k-y_0 as *nearly coalesced* from the points x_k, and the preconditioned system will not show the usual condition problems, at least for moderate values of ε. However, as in every other standard preconditioning method, bad condition is at least partly moved into the preconditioning transformation.

We should look at the condition behavior somewhat more closely. The preconditioning transformation $D(\varepsilon)M$ can be described as the action of the functionals $\lambda_{j,\varepsilon}$ of (28) on discrete data. The M part is only dependent on geometry, not on scaling. It does not affect the behavior for $\varepsilon\to 0$ at all, neither in the *flat limit* nor in the *coalescence* case. It has increasing condition as a function of N, but this is much less dramatic than the one of the whole process for $\varepsilon\to 0$ which is hidden in the $D(\varepsilon)$ part. In fact, the lowest diagonal entry of $D(\varepsilon)$, which corresponds to the denominator of $\lambda_{N,\varepsilon}$ of (28), is of the form ε^{-k_2}. This makes the direct pointwise evaluation of $\lambda_{N,\varepsilon}$ numerically impossible already for moderate values of N and ε. For instance, if about 50 points in 2D are used, there will be a k_2 of about 9, and an ε of about 0.1 will already be hard to handle. But there is a way out, namely to evaluate the divided difference functionals $\lambda_{j,\varepsilon}$ for small ε via an application to a power series, e.g., like evaluation of the first divided difference in 1D via

$$\frac{f(y)-f(x)}{y-x}=f'(x)+\frac{1}{2}(y-x)f''(x)+\cdots$$

which is perfectly stable for $y\to x$ and can for analytic functions f be done to any accuracy, provided that all derivatives are explicitly available. But this would need quite some coding effort in the multivariate case.

13. Newton interpolation

The foregoing sections contained some rather heavy machinery, but they followed a strategy which is well known from univariate polynomial interpolation. In fact, the transition from Lagrange to Hermite interpolation via the Newton interpolation formula is precisely what happened above. To see this more clearly, we start without the parameter ε in this section.

First, in univariate situations, we make the transition from function values $f(x_1),\dots,f(x_N)$ to the N divided differences

$$\lambda_j(f):=f[x_1,\dots,x_j], \quad 1\leqslant j\leqslant N,$$

in standard terminology, but written as N linear functionals which have form (28) with a moment matrix that does not appear explicitly. The connection between (28) and divided differences is based on the property

$$\lambda_j(x^\alpha)=0, \quad |\alpha|<t_j=|\alpha^j|=j-1,$$

in 1D, as assured by the moment matrix via (9). Then the Newton basis

$$v_j(x):=\prod_{i=1}^{j-1}(x-x_i), \quad 1\leqslant j\leqslant N,$$

generates the Newton interpolant

$$p(x):=\sum_{j=1}^{N}\lambda_j(f)v_j(x), \tag{32}$$

and the Lagrange interpolation process nicely converges for coalescing points into a Hermite interpolation problem. Note that this transition is not apparent as long as the Lagrange interpolation is written via Lagrange basis functions. If both problems are rewritten in Newton form, the limit behavior is obvious.

While our functionals of (28) correspond nicely to divided differences, we still have to see how our new basis of functions $v_{j,\varepsilon}$ of (22) corresponds to the Newton basis. The crucial fact is that the 1D polynomial Newton basis satisfies

$$\lambda_j(v_k) = \delta_{jk}, \quad 1 \leqslant j, k \leqslant N, \tag{33}$$

as it follows from (32). Note that the case $j < k$ relies on the fact that v_k vanishes on $x_1, \ldots, x_{k-1}$, while the case $j > k$ is standard for divided differences, because they annihilate lower-order polynomials.

In our technique, system (18) can be written as

$$\sum_{j=1}^{N} (\lambda_j, \lambda_k)_\phi v_j(x) = w_k(x)$$

using (29), both for positive ε and in the limit $\varepsilon \to 0$. But the definition of the w_k implies $\lambda_j(w_k) = (\lambda_j, \lambda_k)_\phi$, and thus we get (33). If we use the fact that the M matrix is lower triangular by construction, we immediately get something similar to the 1D case:

Lemma 5. *The functions $v_{j,\varepsilon}$ of the transformed interpolation process satisfy*

$$v_{j,\varepsilon}(x_k) = 0, \quad 1 \leqslant k < j \leqslant N, \ \varepsilon > 0.$$

14. General coalescence

We now turn to the harder problem of N more or less freely coalescing points at zero. To this end, we assume that our given coalescing data points $y_k(h)$ move along smooth curves for $h \to 0$ into $0 = y_k(0)$. For simplicity, we assume $\|y_k(h)\|_2 \leqslant h$ throughout. The geometry now is h-dependent, and the characteristic multiindices $\alpha^j(h)$ of (7) and the $t_j(h) := |\alpha^j(h)|$ of (8) will vary with h. But we shall focus on sequences $h_k \to 0$ where these discrete quantities do not vary any more. Thus we ignore their dependence on h again.

If we define points $x_k(h)$ by $y_k(h) = h\, x_k(h)$ such that the $x_k(h)$ still vary smoothly, the geometric quantities derived for the $x_k(h)$ are the same as those for $y_k(h)$, because the columns of the monomial matrices just get different scalar factors. We assume that higher-order monomials of the $x_k(h)$ can be stably calculated via

$$x_k^\alpha(h) = \sum_{j=1}^{N} d(j, h, \alpha) x_k^{\alpha^j}(h), \quad 1 \leqslant k \leqslant N, \ \ |\alpha| > k_2, \tag{34}$$

from lower-order monomials, with uniformly bounded coefficients $d(j, h, \alpha)$. From the definition of k_2 this is clear if the x_k are constant, but we allow them to vary here, allowing a more general but still somewhat regular coalescence of the $y_k(h)$.

The above identity describes how the column with multiindex α of the monomial matrix can be reconstructed from the N linear independent columns corresponding to the α^j, $1 \leqslant j \leqslant N$. In our coalescence scenario, the above identity, when rewritten in terms of the $y_k^\alpha(h)$, turns into

$$y_k^\alpha(h) = \sum_{j=1}^{N} d(j, h, \alpha) h^{|\alpha| - |\alpha^j|} y_k^{\alpha^j}(h), \quad 1 \leqslant k \leqslant N, \ \ |\alpha| > k_2, \tag{35}$$

and describes in a natural way how the larger powers of the y_k vanish faster than the lower ones for $h \to 0$. This provides a good reason why (34) should be assumed.

We can always find h-dependent $N \times N$ moment matrices $M(h) = (m_{jk}(h))$ such that the linear functionals

$$\lambda_{j,h}(f) := \sum_{k=1}^{N} m_{jk}(h) f(y_k(h)), \quad 1 \leqslant j \leqslant N, \tag{36}$$

are orthonormal in the native space of ϕ. This can be done theoretically by orthogonalizing in the span of the functionals $\delta_{y_k(h)}$, $1 \leqslant k \leqslant N$. Due to their normalization, these functionals must be weak-$*$-convergent, and thus there are limit functionals $\lambda_{j,0}$ with norm one in the dual of the native space such that

$$\lambda_{j,0}(f) = \lim_{h\to 0} \lambda_{j,h}(f)$$

for suitable subsequences and all f in the native space of ϕ. The whole problem works in the span of the right-hand sides

$$w_{j,h}(y) := \lambda^x_{j,h}\phi(\|x-y\|_2)$$

which nicely converge in the native space towards

$$w_{j,0}(y) := \lambda^x_{j,0}\phi(\|x-y\|_2), \quad 1 \leqslant j \leqslant N,$$

whatever these functions actually are, and the orthogonalization of our functionals imply the Lagrange property

$$\lambda^y_{k,h} w_{j,h}(y) = \lambda^x_{j,h}\lambda^x_{k,h}\phi(\|x-y\|_2) = \delta_{jk}, \quad 1 \leqslant j,k \leqslant N,$$

for all positive h. Clearly, the limit functionals must be supported in zero only, but we want to figure out that they are necessarily derivatives at zero of order up to k_2. From (35) we get uniform convergence

$$\lambda_{m,h}(x^\alpha) = \sum_{j=1}^{N} d(j,h,\alpha) h^{|\alpha|-|\alpha^j|} \lambda_{m,h}(x^{\alpha^j}) \to 0$$

for $h \to 0$ and all $|\alpha| > k_2$. This proves

$$\lambda_{j,0}(x^\alpha) = 0 \quad \text{for all } |\alpha| > k_2, \quad 1 \leqslant j \leqslant N.$$

But for general functions f the functionals act like

$$\begin{aligned} \lambda_{j,0}(f) &= \sum_{\alpha\in\mathbb{Z}_0^d} \frac{D^\alpha f(0)}{\alpha!}\lambda_{j,0}(x^\alpha) \\ &= \sum_{|\alpha|\leqslant k_2} \frac{D^\alpha f(0)}{\alpha!}\lambda_{j,0}(x^\alpha) \end{aligned}$$

proving that they are derivatives of order at most k_2 as required. Now we can also check the limit of the orthogonality. Since convergence is not strong, we cannot directly conclude

$$\delta_{jk} = \lim_{h\to 0}(\lambda_{j,h}, \lambda_{k,h})_\phi \stackrel{?}{=} (\lambda_{j,0}, \lambda_{k,0})_\phi,$$

but we can consider the limit of

$$\begin{aligned} \delta_{jk} &= (\lambda_{j,h}, \lambda_{k,h})_\phi \\ &= \sum_{\alpha,\beta\in\mathbb{Z}_0^d} c(\alpha,\beta)\varphi_{|\alpha+\beta|}\lambda_{j,h}(x^\alpha)\lambda_{k,h}(y^\beta) \\ &\to \sum_{\substack{\alpha,\beta\in\mathbb{Z}_0^d \\ |\alpha|,|\beta|\leqslant k_2}} c(\alpha,\beta)\varphi_{|\alpha+\beta|}\lambda_{j,0}(x^\alpha)\lambda_{k,0}(y^\beta) \\ &= (\lambda_{j,0}, \lambda_{k,0})_\phi, \quad 1 \leqslant j,k \leqslant N. \end{aligned}$$

Theorem 4. *Under the hypotheses of Theorem* 2, *generally coalescent Lagrange interpolation problems satisfying* (34) *converge to Hermite problems whose functions and functionals are defined by limit functionals being certain* *derivatives of order at most k_2 at the coalescence point. The limit functionals for the Hermite problems can be explicitly calculated, and they are independent of the radial basis function used.*

15. Open problems

For the *increasingly flat limit* case, the conditions given by Theorem 2 are sufficient to guarantee convergence, and they are sharp as far as conditions are formulated using k_0 and k_2 only. However, convergence is equivalent to certain guaranteeing $P_{jn} = 0$ for all $k_0 < n < k_2$, $1 \leqslant j \leqslant N$ with EQP(n, k_2), and these come up as complicated rational expressions involving the data set X and the expansion coefficients $\varphi_{2\ell}$ of the radial basis function ϕ. Thus there may be special cases of X and ϕ where there is convergence outside the validity of Theorem 2. A particular case, conjectured first in [7] and proven in [15], surprisingly states that the Gaussian lets these conditions be satisfied in all cases, no matter what the geometry of X is. In other words the Gaussian overcomes all possible geometric degenerations. By recent numerical experiments and conjectures of Fornberg et al. [5], the same property holds for a new class of oscillatory radial kernels, including the Bessel radial basis function $J_0(r)$ in 2D. But this one fails to satisfy the assumptions of Lemma 1 and leads to singular matrices [12]. The special role of the Gauss and Bessel kernels is still a mystery which deserves investigation. Our MAPLE© procedures, available on request, allow some explicit experimentation along these lines, but there are no theoretical results known so far.

For the *coalescence* scenario, our methods indicate how to cope with data points that come too close and thus spoil the condition of the linear system. Such situations will automatically occur, if adaptive methods calculate approximations of functions that are derivatives of the kernel at a fixed point. Investigations of such methods are under way, since they proved to be rather efficient [10,16] in practice, even for solving partial differential equations by collocation [8,9].

Another future application concerns the investigation of approximations with N free centers. In such nonlinear situations, one has to calculate the closure of the manifold of trial functions under a weak norm like the L_∞ norm. This requires to investigate coalescence carefully, and the current paper is a first step into this direction.

Acknowledgment

Special thanks go to Bengt Fornberg and in particular Elisabeth Larsson for various valuable comments and corrections before submission, and to the referees after submission.

References

[1] M.D. Buhmann, Radial Basis Functions, Theory and Implementations, Cambridge University Press, Cambridge, 2003.
[2] C. de Boor, On interpolation by radial polynomials, Adv. Comput. Math. 24 (2006) 143–153.
[3] C. de Boor, A. Ron, On multivariate polynomial interpolation, Constr. Approx. 6 (1990) 287–302.
[4] T. Driscoll, B. Fornberg, Interpolation in the limit of increasingly flat radial basis functions, Comput. Math. Appl. 43 (2002) 413–422.
[5] B. Fornberg, E. Larsson, G. Wright, A new class of oscillatory radial basis functions, Comput. Math. Appl. 51 (2006) 1209–1222.
[6] B. Fornberg, G. Wright, Stable computation of multiquadric interpolants for all values of the shape parameter, Comput. Math. Appl. 48 (2004) 853–867.
[7] B. Fornberg, G. Wright, E. Larsson, Some observations regarding interpolants in the limit of flat radial basis functions, Comput. Math. Appl. 47 (2004) 37–55.
[8] C. Franke, R. Schaback, Solving partial differential equations by collocation using radial basis functions, Appl. Math. Comput. 93 (1998) 73–82.
[9] C. Franke, R. Schaback, Convergence order estimates of meshless collocation methods using radial basis functions, Adv. Comput. Math. 8 (1998) 381–399.
[10] Y. Hon, R. Schaback, X. Zhou, An adaptive greedy algorithm for solving large RBF collocation problems, Numer. Algorithms 32 (2003) 13–25.
[11] E. Larsson, B. Fornberg, Theoretical and computational aspects of multivariate interpolation with increasingly flat radial basis functions, Comput. Math. Appl. 49 (1) (2005) 103–130.
[12] Y.J. Lee, G.J. Yoon, J. Yoon, Convergence of increasingly flat radial basis interpolants to polynomial interpolants, SIAM J. Math. Anal. 39 (2007) 537–553.
[13] C. Micchelli, Interpolation of scattered data: distance matrices and conditionally positive definite functions, Constr. Approx. 2 (1986) 11–22.

[14] D.St.P. Richards, Positive definite symmetric functions on finite-dimensional spaces II, Statist. Probab. Lett. 3 (1985) 325–329.
[15] R. Schaback, Multivariate interpolation by polynomials and radial basis functions, Constr. Approx. 21 (2005) 293–317.
[16] R. Schaback, H. Wendland, Adaptive greedy techniques for approximate solution of large RBF systems, Numer. Algorithms 24 (2000) 239–254.
[17] R. Schaback, H. Wendland, Kernel techniques: from machine learning to meshless methods, Acta Numer. 15 (2006) 543–639.
[18] H. Wendland, Piecewise polynomial, positive definite and compactly supported radial functions of minimal degree, Adv. Comput. Math. 4 (1995) 389–396.
[19] H. Wendland, Scattered Data Approximation, Cambridge University Press, Cambridge, 2005.
[20] Z. Wu, Hermite–Birkhoff interpolation of scattered data by radial basis functions, Approx. Theory Appl. 8 (1992) 1–10.