Journal of Computational Physics 365 (2018) 149–172

Contents lists available at ScienceDirect

Journal of Computational Physics

www.elsevier.com/locate/jcp

Minimum Sobolev norm interpolation of scattered derivative data

S. Chandrasekaran [a,1], C.H. Gorman [b,*], H.N. Mhaskar [c,2]

[a] Department of Electrical and Computer Engineering, University of California, Santa Barbara, CA 93106, United States
[b] Department of Mathematics, University of California, Santa Barbara, CA 93106, United States
[c] Institute of Mathematical Sciences, Claremont Graduate University, Claremont, CA 91711, United States

ARTICLE INFO

Article history:
Received 21 October 2017
Received in revised form 28 February 2018
Accepted 7 March 2018
Available online 21 March 2018

Keywords:
Minimum Sobolev norm
Birkhoff interpolation
Data-defined manifolds
Diffusion polynomials

ABSTRACT

We study the problem of reconstructing a function on a manifold satisfying some mild conditions, given data of the values and some derivatives of the function at arbitrary points on the manifold. While the problem of finding a polynomial of two variables with total degree $\le n$ given the values of the polynomial and some of its derivatives at exactly the same number of points as the dimension of the polynomial space is sometimes impossible, we show that such a problem always has a solution in a very general situation if the degree of the polynomials is sufficiently large. We give estimates on how large the degree should be, and give explicit constructions for such a polynomial even in a far more general case. As the number of sampling points at which the data is available increases, our polynomials converge to the target function on the set where the sampling points are dense. Numerical examples in single and double precision show that this method is stable, efficient, and of high-order.

1. Introduction

The subject of Lagrange interpolation of univariate functions is a very old one. Thus, one starts with an (infinite) *interpolation matrix* X whose n-th column consists of n real numbers $x_{k,n}$, $k=1,\cdots,n$. It is well known that for any matrix Z whose n-th column consists of n real numbers $z_{k,n}$, $k=1,\cdots,n$, there exists a sequence of polynomials $L_n(X,Z)$ of degree $\le n-1$ such that $L_n(X,Z;x_{k,n})=z_{k,n}$, $k=1,\cdots,n$. In the case when each $z_{k,n}=f(x_{k,n})$ for some continuous function f, then it is customary to denote $L_n(X,Z)$ by $L_n(X,f)$. It is also well known that for any $X\subset[-1,1]$, there exists a continuous function $f:[-1,1]\to\mathbb{R}$ such that the sequence $L_n(X,f)$ does not converge in the uniform norm [1]. This situation is similar to the theory of trigonometric Fourier series, where the Fourier projections of a function do not always converge to the function in the uniform norm, but one can construct summability operators to obtain convergence [2].

Several interpolatory analogues of such summability operators are studied in the literature. For example, if $x_{k,n}=\cos((2k-1)\pi/(2n))$, $k=1,\cdots,n$, $f:[-1,1]\to\mathbb{R}$ is continuous, and one constructs a sequence of polynomials $F_n(f)$ of degree $\le 2n-1$ such that $F_n(x_{k,n})=f(x_{k,n})$, and $F_n'(x_{k,n})=0$, then the sequence $F_n(f)\to f$ uniformly on $[-1,1]$ [1].

* Corresponding author.
E-mail address: gorman@math.ucsb.edu (C.H. Gorman).

1 The research of this author was supported, in part, by grants CCF-0515320 and CCF-0830604 from the NSF.
2 The research of this author is supported in part by ARO Grant W911NF-15-1-0385.

https://doi.org/10.1016/j.jcp.2018.03.014
0021-9991/

In 1906, Birkhoff initiated a study of interpolation in a more general setting, known now as Birkhoff interpolation [3]. For each column of the matrix X, one considers an *incidence matrix* E whose entries are in $\{0,1\}$. If the number of 1's in E is N, one seeks a polynomial B_N of degree $\le N-1$ such that $B_N^{(j)}(x_{k,n}) = y_{j;k,n}$ if the (k,j)-th entry in E is 1. Clearly, such a polynomial may or may not exist. The conditions under which it exists and is unique is the topic of a great deal of research [4]. As expected, the problems are much harder in the multivariate setting [5,6]. For example, it is not always possible to find a bivariate polynomial P of total degree 2 (with 6 parameters) such that P and its derivatives up to order 2 take given values (also 6 conditions).

A great deal of research on this subject is focused on forming a "square" system and determining whether or not the system is solvable for most point distributions. Since there is not always an obvious choice of interpolation space, the space can be constructed in multiple ways. For example, monomial bases [7–9], Newton-type bases [10,7], and cardinal bases [11] have been studied by various authors. Obtaining a unique solution, though, does not ensure convergence to the underlying function when the data become dense. Although some of the previously cited articles have numerical examples, not all of them computed the maximum difference between the test function and its approximation in a suitable domain, and we do not know of any general provably convergent technique for the Birkhoff problem on scattered data points in arbitrary domains.

If we do not require that the dimension of the polynomial space match exactly the number of 1's in the incidence matrix, then it is possible to guarantee not just existence, but also give explicit algorithms and prove the convergence of the resulting polynomials. In the case of interpolation based on the values of the polynomials alone, this has been observed in a series of papers [12–14]. The purpose of this paper is to generalize these results for Birkhoff-like interpolation for the so called diffusion polynomials.

One of our motivations is to study numerical solutions of a system of linear partial differential equations. Given differential operators L_k on a manifold $\mathbb{X}$ (and its boundary), the collocation method involves finding a "polynomial" (i.e., an element of a suitably chosen finite dimensional space) P for which the values of $L_k(P)$ are known at some grid points. We view this question as a generalized Birkhoff interpolation problem, except that we do not require the dimension of the space to be exactly equal to the conditions. On general manifolds, one does not always have standard grids such as equidistant grid on a Euclidean space. Therefore mesh-free methods require a solution of such interpolation problems on scattered data; i.e., when one cannot prescribe the location of the grid points in advance. There are many efforts to reconstruct a function from scattered data; see [15] for one example. The papers [16–18] present work relating to function approximation on the sphere. We show the feasibility of the solution of such interpolation problems provided the dimension of the space is sufficiently high, inversely proportional to the minimal separation among the grid points. We will prove that such solutions can be constructed as minimizers of an optimization problem, and prove that these solutions will converge to the target function at limit points of the grid points. An application of these ideas is already demonstrated in [19].

The main results are presented in Section 2, while the overall assumptions are discussed in Section 3. Preparatory results are developed in Sections 4–6 and are used to prove the main theorems in Section 7. Finally, numerical simulations are given in Section 8. We show that the standard divergence phenomenon is overcome in both one and two dimensions. We also numerically demonstrate the high-order convergence of our method. Furthermore we present results using both *single* and double precision to show that the high ill-conditioning associated with high order methods can be successfully overcome using our efficient numerical techniques.

2. Main results

We wish to study the problem of interpolation of derivative information in a somewhat abstract manner, to accommodate several examples. The most elementary among these is interpolation by multivariate trigonometric polynomials. Other examples include the problem of interpolation by spherical polynomials, or by linear combinations of the eigenfunctions of some elliptic differential operator on a smooth manifold.

Let $\mathbb{X}$ be a metric measure space with a probability measure μ^* and metric ρ. In this paper, a measure will mean a complete, sigma-finite, Borel measure, signed or positive. For a measure ν, $|\nu|$ denotes its total variation measure. If ν is a measure, and $f:\mathbb{X}\to\mathbb{R}$ is ν-measurable, we define

$$\|f\|_{\nu;p} = \begin{cases} \left(\displaystyle\int_{\mathbb{X}} |f(x)|^p d|\nu|(x)\right)^{1/p}, & \text{if } 1\le p<\infty, \\ |\nu| - \displaystyle\operatorname*{ess\,sup}_{x\in\mathbb{X}} |f(x)|, & \text{if } p=\infty. \end{cases}$$

The symbol $L^p(\nu)$ will denote the space of all f such that $\|f\|_{\nu;p}<\infty$, where two functions are considered equal if they are equal $|\nu|$-almost everywhere. If $\nu=\mu^*$, we will often omit the mention of the measure from the notation, if we feel that this should not cause any confusion; e.g., $\|f\|_p = \|f\|_{\mu^*;p}$, $L^p = L^p(\mu^*)$. The space of uniformly continuous and bounded functions on $\mathbb{X}$ will be denoted by UBC. For $1\le p\le\infty$, we define the dual exponent p' by $1/p+1/p'=1$ as usual.

Let $\{\lambda_k\}$ be increasing sequence of nonnegative numbers with $\lambda_0=0$, and $\{\phi_k\}$ be an orthonormal set (of real valued functions) in $L^2\cap UBC$. We define the space

$$\Pi_n = \text{span}\,\{\phi_k : \lambda_k < n\}, \qquad n > 0,\ \Pi_n = \{0\}, \qquad n \le 0.$$

We will write

$$\Pi_\infty = \bigcup_{n\ge 0} \Pi_n.$$

The elements of Π_∞ have been referred to as diffusion polynomials in [20], and we will use the same terminology. If $P \in \Pi_n$, we will refer to n as the degree of P, or more precisely that P is of degree $< n$. From the definition, we do *not* require λ_k to be integers. The L^p closure of the Π_∞ will be denoted by X^p.

Example 2.1 *(The trigonometric case).* Let $q \ge 1$ be an integer. The space $\mathbb{X}$ is the torus (quotient group) $\mathbb{T}^q = \mathbb{R}^q/(2\pi\mathbb{Z}^q)$, ρ is the arc-length, and μ^* is the normalized Lebesgue measure. We take $\{\lambda_k\}$ to be an enumeration of multi-integers in $\mathbb{Z}^q_+$ such that $\mathbf{m}$ comes before $\mathbf{j}$ if either $|\mathbf{m}|_2 \le |\mathbf{j}|_2$ and if $|\mathbf{m}|_2 = |\mathbf{j}|_2$, then $\mathbf{m}$ comes before $\mathbf{j}$ in the alphabetical ordering. If the multi-integer corresponding to λ_k is $\mathbf{k}$, the corresponding ϕ_k's are $\cos(\mathbf{k}\cdot\circ), \sin(\mathbf{k}\cdot\circ)$. In this case, $X^p = L^p$ if $1 \le p < \infty$, and X^∞ is the set of all continuous functions on $\mathbb{T}^q$.

Example 2.2 *(The manifold case).* Let $\mathbb{X}$ be a compact Riemannian manifold, ρ be the geodesic distance, μ^* be the Riemann measure. We take λ_k to be the square roots of eigenvalues of a self-adjoint, second order, regular elliptic differential operator on $\mathbb{X}$, and ϕ_k to be the corresponding eigenfunction. The precise nature of the space X^p will depend upon the manifold and the elliptic operator.

Example 2.3 *(The 2-sphere).* If $\mathbb{X}$ is the 2 dimensional Euclidean sphere $\mathbb{S}^2$, we may choose each ϕ_k to be one of the orthonormalized spherical harmonics and $\lambda_k = k$, the degree of this polynomial. This emphasizes in particular, that λ_k's do not have to be the eigenvalues of any predetermined differential/integral operator. Indeed, in our abstract setting of metric measure spaces, there is no notion of a differentiability structure.

In the absence of any concrete structure, we will need to make several assumptions on the system $\Xi = (\mathbb{X}, \rho, \mu^*, \{\lambda_k\}, \{\phi_k\})$. These are formulated precisely in Section 3. For the clarity of exposition, we will now assume that these are all satisfied. In particular, the symbol q used in the following discussion is defined in (3.4).

Next, we define the smoothness classes needed in order to state our theorems. In the trigonometric case, the Sobolev space $W_{p,\beta}$ is defined as the space of all functions $f : \mathbb{T}^q \to \mathbb{C}$ for which $(|\mathbf{k}|_2^2+1)^{\beta/2}\hat{f}(\mathbf{k}) = \widehat{f^{(\beta)}}(\mathbf{k})$ for some $f^{(\beta)} \in L^p(\mathbb{T}^q)$. In our abstract case, the role of $|\mathbf{k}|_2$ is played by λ_k. However, we would like to introduce a greater flexibility in the definition of the Sobolev class.

If $f \in L^1$, we define

$$\hat{f}(k) := \int_{\mathbb{X}} f(y)\phi_k(y)d\mu^*(y), \qquad k = 0, 1, \cdots. \tag{2.1}$$

To include such multipliers as $(\lambda_k^2+1)^{\beta/2}$ or $(\lambda_k+1)^\beta$, we use a mask of type β, defined below in Definition 2.1. In the sequel, S will be a fixed integer.

Definition 2.1. Let $\beta \in \mathbb{R}$. A function $b : \mathbb{R} \to \mathbb{R}$ will be called a mask of type β if b is an even, $S+1$ times continuously differentiable function (for some integer $S > q$) such that for $t > 0$, $b(t) = (1+t)^{-\beta}F_b(\log t)$ for some $F_b : \mathbb{R} \to \mathbb{R}$ such that $|{F_b}^{(k)}(t)| \le c(b)$, $t \in \mathbb{R}$, $k = 0, 1, \cdots, S+1$, and $F_b(t) \ge c_1(b)$, $t \in \mathbb{R}$.

If $\beta \in \mathbb{R}$, b is a mask of type β, and $f \in L^p$, we say that a function $f \in W_{p,b}$ if there exists $f^{(b)} \in L^p$ such that

$$b(\lambda_k)\widehat{f^{(b)}}(k) = \hat{f}(k), \qquad k = 0, 1, \cdots. \tag{2.2}$$

We will write

$$\|f\|_{W_{p,b}} = \|f^{(b)}\|_p. \tag{2.3}$$

In [21], we have shown that if $\beta > q/p$, and b is a mask of type β, then for every $y \in \mathbb{X}$, there exists $\psi_y := G(b; \circ, y) \in X^{p'}$ such that $\langle \psi_y, \phi_k \rangle = b(\lambda_k)\phi_k(y)$, $k = 0, 1, \cdots$. Moreover,

$$\sup_{y\in\mathbb{X}} \|G(b; \circ, y)\|_{p'} \le c. \tag{2.4}$$

It is then easy to check by comparison of coefficients that if $f \in W_{p,b}$ then for almost all $x \in \mathbb{X}$:

$$f(x) = \int_{\mathbb{X}} G(b; x, y) f^{(b)}(y) d\mu^*(y). \tag{2.5}$$

As a prelude to our main theorem, we formulate first a Golomb–Weinberger-type theorem in our general setting. The theorem gives an explicit expression for a solution of the interpolation problem:

Given linear operators L_k, $k = 1, \cdots, R$, each defined on $W_{2,b}$, such that point evaluations are well-defined on the range of each L_k, points $y_\ell \in \mathbb{X}$, $\ell = 1, \cdots, M$, and data $f_{k,\ell} \in \mathbb{C}$, find a function $g \in W_{2,b}$ such that $L_k(g)(y_\ell) = f_{k,\ell}$, $k = 1, \cdots, R$, $\ell = 1, \cdots, M$.

The Golomb–Weinberger-type theorem gives a solution as a linear combination of the kernel defined in (2.6) below. Let b be a mask of type $\beta > q/2$. Then b^2 is a mask of type $2\beta > q$. We define

$$\mathbb{G}(x, y) = \sum_{j=0}^{\infty} b(\lambda_j)^2 \phi_j(x) \phi_j(y). \tag{2.6}$$

Clearly, $\mathbb{G}$, treated either as a function of x or y, is a function in $W_{2,b}$.

Theorem 2.1. *Let $R, M \geq 1$ be integers, each L_k, $k = 1, \cdots, R$, be a linear operator defined on $W_{2,b}$ such that point evaluations are well defined on the range of L_k. Let $y_\ell \in \mathbb{X}$, $\ell = 1, \cdots, M$, and $\{f_{k,\ell}\}_{k=1,\cdots,R,\ \ell=1,\cdots,M} \subset \mathbb{C}$. We assume that there exist $\phi_{k,j} \in W_{2,b}$, $k = 1, \cdots, R$, $j = 1, \cdots, M$, such that the following condition holds: For $i = 1, \cdots, R$, $\ell = 1, \cdots, M$,*

$$L_i(\phi_{k,j})(y_\ell) = \begin{cases} 1, & \text{if } i = k,\ j = \ell, \\ 0, & \text{otherwise.} \end{cases} \tag{2.7}$$

Then the problem

$$\text{minimize } \|g^{(b)}\|_2 \text{ subject to } L_k(g)(y_\ell) = f_{k,\ell}, \qquad k = 1, \cdots, R,\ \ell = 1, \cdots, M, \tag{2.8}$$

has a solution of the form

$$P(x) = \sum_{k=1}^{R} \sum_{j=1}^{M} a_{k,j} L_{k,1} \mathbb{G}(y_j, x), \qquad x \in \mathbb{X}, \tag{2.9}$$

where, the expression $L_{k,1}\mathbb{G}(y_j, x)$ means that the operator L_k is applied to the 1-st variable in $\mathbb{G}$, and the resulting function is evaluated at (y_j, x).

In the trigonometric case, when $b(t) = (t^2 + 1)^{-\beta/2}$, the kernel $\mathbb{G}$ takes the form

$$\sum_{\mathbf{k} \in \mathbb{Z}} (|\mathbf{k}|_2^2 + 1)^{-\beta} \exp(i\mathbf{k} \cdot \mathbf{x}).$$

A straightforward computation of this series may be slow for small values of β, and in any case, introduces a truncation error. Therefore, we are interested in solving the interpolation problem directly using the diffusion polynomials (trigonometric polynomials in the trigonometric case). Toward this goal, we first introduce some further terminology.

Constant convention: *In the sequel, $c, c_1, \cdots$ will denote generic positive constants independent of any obvious variables such as the degree n, or the target function, etc. Their values may be different at different occurrences, even within the same formula. The symbol $A \sim B$ means that $c_1 A \leq B \leq c_2 A$.*

We will consider an *interpolation matrix*, by which we mean a sequence $\{Y_n\}_{n=1}^{\infty}$ of subsets

$$Y_n = \{y_{j,n}\}_{j=1}^{M_n}, \qquad n = 1, 2, \cdots.$$

This is not a matrix in the usual sense of the word, but the terminology captures the spirit of a similar notion in the theory of classical polynomial interpolation. For any subset $\mathcal{C} \subseteq \mathbb{X}$, we define its minimal separation by

$$\eta(\mathcal{C}) = \min_{x, y \in \mathcal{C},\ x \neq y} \rho(x, y). \tag{2.10}$$

We will write

$$\eta_n = \eta(Y_n), \qquad n = 1, \cdots. \tag{2.11}$$

Our main theorem is the following.

Theorem 2.2. *We assume each of the conditions listed in Section 3, where some of the notation used here is explained. Let* $1 \le p \le \infty$, $\beta > \max_{1\le k\le R} q_k + q/p$, *$b$ be a mask of type* β, $f \in W_{p,b}$. *Then there exists an integer* N^* *with* $N^* \sim \eta_n^{-1}$ *and a mapping* $\mathbf{P}^* = \mathbf{P}_n^* : W_{p,b} \to \Pi_{N^*}$ *such that for every* $f \in W_{p,b}$,

$$L_k(\mathbf{P}^*(f))(y_{j,n}) = L_k(f)(y_{j,n}), \qquad j = 1, \cdots, M_n,\ k = 1, \cdots, R, \tag{2.12}$$

and

$$\|f - \mathbf{P}^*(f)\|_{W_{p,b}} \le c \inf\{\|f - T\|_{W_{p,b}} \ :\ T \in \Pi_{N^*}\}. \tag{2.13}$$

In particular, there exists $\mathbf{P} = \mathbf{P}_n : W_{p,b} \to \Pi_{N^*}$ *such that*

$$\|\mathbf{P}(f)\|_{W_{p.b}} = \min\{\|P\|_{W_{p,b}} : L_k(P)(y_{\ell,n}) = L_k(f)(y_{\ell,n}),\ \ell = 1, \cdots, M_n,\ k = 1, \cdots, R\}. \tag{2.14}$$

We observe that the sets Y_n do not necessarily become dense in $\mathbb{X}$ as $n \to \infty$. In this case, it is clearly impossible to give a construction of the operators P_n^* based entirely on Y_n that satisfies the global convergence implied by (2.14). The second part of Theorem 2.2 helps us to find an interpolatory diffusion polynomial whose $W_{p,b}$ norm is under control on $\mathbb{X}$. Therefore, we do not expect that the sequence $\mathbf{P}_n(f)$ to converge to f on $\mathbb{X}$. However, the sequence converges at limit points of Y_n's.

Theorem 2.3. *With the set up as in Theorem 2.2, if* $x_0 \in \mathbb{X}$ *is a limit point of the family* $\{Y_n\}$, *then* $f(x_0)$ *is a limit point of* $\{\mathbf{P}_n(f)(y) : y \in Y_n,\ n = 1, 2, \cdots\}$.

Theorem 2.3 follows from a much general principle "feasibility implies convergence", which is formulated more precisely in Theorem 7.1 below.

3. Assumptions

3.1. The space

Let $\mathbb{X}$ be a non-empty set, ρ be a metric defined on $\mathbb{X}$, and μ^* be a complete, positive, Borel measure with $\mu^*(\mathbb{X}) = 1$. We fix a non-decreasing sequence $\{\lambda_k\}_{k=0}^\infty$ of nonnegative numbers such that $\lambda_0 = 0$, and $\lambda_k \uparrow \infty$ as $k \to \infty$. Also, we fix a system of continuous, bounded, and integrable functions $\{\phi_k\}_{k=0}^\infty$, orthonormal with respect to μ^*; namely, for all nonnegative integers j, k,

$$\int_{\mathbb{X}} \phi_k(x)\phi_j(x)d\mu^*(x) = \begin{cases} 1, & \text{if } j = k, \\ 0, & \text{otherwise.} \end{cases} \tag{3.1}$$

We will assume that $\phi_0(x) = 1$ for all $x \in \mathbb{X}$.

In our context, the role of polynomials will be played by diffusion polynomials, which are finite linear combinations of $\{\phi_j\}$. In particular, an element of

$$\Pi_n := \mathsf{span}\{\phi_j : \lambda_j < n\}$$

will be called a diffusion polynomial of degree $< n$.

We will formulate our assumptions in terms of a formal heat kernel. The *heat kernel* on $\mathbb{X}$ is defined formally by

$$K_t(x, y) = \sum_{k=0}^{\infty} \exp(-\lambda_k^2 t)\phi_k(x)\phi_k(y), \qquad x, y \in \mathbb{X},\ t > 0. \tag{3.2}$$

Although K_t satisfies the semigroup property, and in light of the fact that $\lambda_0 = 0$, $\phi_0(x) \equiv 1$, we have formally

$$\int_{\mathbb{X}} K_t(x, y)d\mu^*(y) = 1, \qquad x \in \mathbb{X}, \tag{3.3}$$

yet K_t may not be the heat kernel in the classical sense. In particular, we need not assume K_t to be nonnegative.

Definition 3.1. The system $\Xi = (\mathbb{X}, \rho, \mu^*, \{\lambda_k\}_{k=0}^\infty, \{\phi_k\}_{k=0}^\infty)$ is called a **data-defined space** if each of the following conditions are satisfied.

1. For each $x \in \mathbb{X}$ and $r > 0$, the ball $\mathbb{B}(x, r)$ is compact.
2. There exist $q > 0$ and $\kappa_2 > 0$ such that the following power growth bound condition holds:

$$\mu^*(\mathbb{B}(x,r)) = \mu^*\left(\{y \in \mathbb{X} : \rho(x,y) < r\}\right) \le \kappa_2 r^q, \qquad x \in \mathbb{X},\ r > 0. \tag{3.4}$$

3. The series defining $K_t(x, y)$ converges for every $t \in (0, 1$ and $x, y \in \mathbb{X}$. Further, with q as above, there exist $\kappa_3, \kappa_4 > 0$ such that the following Gaussian upper bound holds:

$$|K_t(x,y)| \le \kappa_3 t^{-q/2} \exp\left(-\kappa_4 \frac{\rho(x,y)^2}{t}\right), \qquad x, y \in \mathbb{X},\ 0 < t \le 1. \tag{3.5}$$

There is a great deal of discussion in the literature on the validity of the conditions in the above definition and their relationship with many other objects related to the quasi-metric space in question, (cf. for example, [22–25]). In particular, it is shown in [22, Section 5.5] that all the conditions defining a data-defined space are satisfied in the case of any complete, connected Riemannian manifold with non-negative Ricci curvature. It is shown in [26] that our assumption on the heat kernel is valid in the case when $\mathbb{X}$ is a complete Riemannian manifold with bounded geometry, and $\{-\lambda_j^2\}$, respectively $\{\phi_j\}$, are eigenvalues, respectively eigenfunctions, for a uniformly elliptic second order differential operator satisfying certain technical conditions.

The bounds on the heat kernel are closely connected with the measures of the balls $\mathbb{B}(x, r)$. For example, using (3.5), Proposition 4.2 below, and the fact that

$$\int_{\mathbb{X}} |K_t(x,y)| d\mu^*(y) \ge \int_{\mathbb{X}} K_t(x,y) d\mu^*(y) = 1, \qquad x \in \mathbb{X},$$

it is not difficult to deduce as in [25] that

$$\mu^*(\mathbb{B}(x,r)) \ge c r^q, \qquad 0 < r \le 1. \tag{3.6}$$

In many of the examples cited above, the kernel K_t also satisfies a lower bound to match the upper bound in (3.5). In this case, Grigoryán [25] has also shown that (3.4) is satisfied for $0 < r < 1$.

We remark that the estimates (3.4) and (3.6) together imply that μ^* satisfies the homogeneity condition

$$\mu^*(\mathbb{B}(x,R)) \le c_1 (R/r)^q \mu^*(\mathbb{B}(x,r)), \qquad x \in \mathbb{X},\ r \in (0,1],\ R > 0, \tag{3.7}$$

where $c_1 > 0$ is a suitable constant.

3.2. The operators L_k

In this sub-section, we state our assumptions on the linear operators L_k. For a bivariate function $F : \mathbb{X} \times \mathbb{X} \to \mathbb{R}$, we will denote

$$L_{k,1}F(x,y) := L_k F(x,y) := (L_k(F(\circ, y))(x), \qquad L_{k,2}F(x,y) := (L_k(F(x, \circ))(y).$$

We fix n, letting $\eta = \eta_n$ be the minimum separation between the $M = M_n$ points in Y_n.

1. We assume that each L_k is *closed* linear operator; i.e., if $f_m \to f$ in L^p, and $L_k(f_m) \to g$ in L^p, then f is in the domain of L_k and $L_k(f) = g$.
2. We assume that each L_k is *local*; i.e., if $f(x) = 0$ for almost all x in an open subset U of $\mathbb{X}$, then $L_k(f)(x) = 0$ for almost all $x \in U$.
3. There exists $q_k \ge 0$ such that

$$|L_{k,1} K_t(x,y)| \le c t^{-(q+q_k)/2} \exp\left(-c_1 \frac{\rho(x,y)^2}{t}\right), \qquad x, y \in \mathbb{X},\ 0 < t \le 1. \tag{3.8}$$

4. For each $k = 1, \cdots, R$, $j = 1, \cdots, M$, there exists $\phi_{k,j} \in W_{p,b}$ such that each of the following conditions holds:
   4.1. For $i = 1, \cdots,$

$$L_i(\phi_{k,j}) = 0, \qquad i \ne k,\ i = 1, \cdots, R. \tag{3.9}$$

   4.2. For $i = 1, \cdots,$ $\ell = 1, \cdots, M$,

$$L_i(\phi_{k,j})(y_\ell) = \begin{cases} 1, & \text{if } i = k,\ j = \ell, \\ 0, & \text{otherwise.} \end{cases} \tag{3.10}$$

4.3. $\phi_{k,j}$, $\phi_{k,j}^{(b)}$ are both supported on a neighborhood of y_j with diameter $\le \eta/3$.
4.4. We have

$$\|\phi_{k,j}^{(b)}\|_\infty \le c\eta^{q_k-\beta}. \tag{3.11}$$

We remark that for any $x \in \mathbb{X}$, there is at most one j such that $\rho(x, y_j) \le \eta/3$, and hence,

$$\phi_{k,\ell}(x) = 0, \qquad \ell \ne j,\ j = 1, \cdots, M. \tag{3.12}$$

In the case when $\mathbb{X}$ is the torus, and each L_k is a mixed partial derivative $D^{\mathbf{j}}$ for some multi-integer $\mathbf{j}$, the construction of functions $\phi_{k,j}$ is given in [27]. The bounds (3.11) are established there in the case when β is an even integer. If $\mathbb{X}$ is a manifold (sphere in particular), and η is less than its inradius, (λ_k^2, ϕ_k) are the eigenfunctions of the Laplace–Beltrami operator, and β is an even integer, then the same construction works via exponential coordinates. A slightly more general scenario involving other second order partial differential operators holds also in view of our results in [28]. When the operators L_k are given by

$$L_k(f, x) = \sum_{|\mathbf{j}| \le q_k} a_{k,\mathbf{j}}(x) D^{\mathbf{j}} f(x), \tag{3.13}$$

then one needs to make some assumptions on the matrices $(a_{k,\mathbf{j}}(y_\ell))_{\ell=1,\cdots,M, |\mathbf{j}|\le q_k}$ so that the constructions given in [27] can be used to construct the desired $\phi_{k,j}$.

4. Preparatory results

The proof of Theorem 2.2 is much more involved than those of the other theorems. The goal of this section is to prove a number of auxiliary results which will lead to the proof of Theorem 2.2.

4.1. Regular measures

We start by introducing the concept of what we have called d-regular measures, and some of their properties.

Definition 4.1. Let ν be any measure on $\mathbb{X}$ with $|\nu|(\mathbb{X}) < \infty$, $d > 0$. We say that ν is d-regular if

$$\nu(\mathbb{B}(x, d)) \le cd^q, \qquad x \in \mathbb{X}. \tag{4.1}$$

The infimum of all constants c which work in (4.1) will be denoted by $|||\nu|||_{R,d}$.

For example, (3.4) implies that μ^* is d-regular for every $d > 0$, and $|||\mu^*|||_{R,d} \le c$ for all $d > 0$ with c independent of d. In the sequel, when thinking of μ^* as a regular measure, we will pass to a limit, and use $d = 0$. In the following lemma, we give another example.

Lemma 4.1. *Let $\mathcal{C} = \{y_1, \cdots, y_M\} \subset \mathbb{X}$, $0 < \eta < 2$ be the minimal separation amongst the y_j's (cf. (2.10)), τ be the measure that associates the mass η^q with each y_j. Then ν is η-regular, and $|||\tau|||_{R,\eta} \le c$, where c is a constant independent of η.*

Proof. Let $x_0 \in \mathbb{X}$, and by relabeling if necessary, let $\mathcal{C} \cap \mathbb{B}(x_0, \eta) = \{y_1, \cdots, y_J\}$. Then the caps $\mathbb{B}(y_j, \eta/2)$ are mutually disjoint, and their union is contained in $\mathbb{B}(x_0, 3\eta/2)$. We recall from (3.6) that $\eta^q \le c_1\mu^*(\mathbb{B}(y_j, \eta/2))$ and from (3.4) that $\mu^*(\mathbb{B}(x_0, 3\eta/2)) \le c_2\eta^q$. Therefore, we deduce that

$$\tau(\mathbb{B}(x_0, \eta)) = J\eta^q \le c_1 \sum_{j=1}^{J} \mu^*(\mathbb{B}(y_j, \eta/2)) = c_1\mu^*\left(\cup_{j=1}^{J} \mathbb{B}(y_j, \eta/2)\right) \le c_1\mu^*(\mathbb{B}(x_0, 3\eta/2)) \le c_1c_2\eta^q. \quad \square$$

The following proposition [29, Theorem 5.5(a), Proposition 5.6] lists some equivalent conditions for a measure to be a regular measure.

Proposition 4.1. *Let $N, d > 0$, ν be a signed or positive finite Borel measure.*

(a) *If ν is d-regular, then for each $r > 0$ and $x \in \mathbb{X}$,*

$$|\nu|(\mathbb{B}(x, r)) \le c|||\nu|||_{R,d}\, \mu^*(\mathbb{B}(x, r + d)) \le c_1|||\nu|||_{R,d}(r + d)^q. \tag{4.2}$$

Conversely, if for some $A > 0$, $|\nu|(\mathbb{B}(x, r)) \le A(r + d)^q$ for each $r > 0$ and $x \in \mathbb{X}$, then ν is d-regular, and $|||\nu|||_{R,d} \le 2^q A$.

(b) *For each* $\gamma > 1$,

$$|||\nu|||_{R,\gamma d} \le c_1(\gamma+1)^q |||\nu|||_{R,d} \le c_1(\gamma+1)^q \gamma^q |||\nu|||_{R,\gamma d}, \tag{4.3}$$

where c_1 *is the constant appearing in (4.2).*

(c) *Let* $N \ge 1$. *If* ν *is* $1/N$*-regular, then* $\|P\|_{\nu;p} \le c_1 |||\nu|||_{R,1/N}^{1/p} \|P\|_{\mu^*;p}$ *for all* $P \in \Pi_N$ *and* $1 \le p < \infty$. *Conversely, if for some* $A > 0$ *and* $1 \le p < \infty$, $\|P\|_{\nu;p} \le A^{1/p} \|P\|_{\mu^*;p}$ *for all* $P \in \Pi_N$, *then* ν *is* $1/N$*-regular, and* $|||\nu|||_{R,1/N} \le c_2 A$.

Next, we recall a very general proposition [29, Proposition 6.5] helping us to estimate integrals of quantities such as the right hand side of (4.28).

Proposition 4.2. *Let* $d > 0$, *and* ν *be a* d*-regular measure. If* $g_1 : [0,\infty) \to [0,\infty)$ *is a nonincreasing function, then for any* $N > 0$, $r > 0$, $x \in \mathbb{X}$,

$$N^q \int\limits_{\Delta(x,r)} g_1(N\rho(x,y))d|\nu|(y) \le \frac{2^q(\kappa_1 + (d/r)^q)q}{1-2^{-q}} |||\nu|||_{R,d} \int\limits_{rN/2}^{\infty} g_1(u)u^{q-1}du. \tag{4.4}$$

In particular, if $S > q$,

$$N^q \int\limits_{\Delta(x,r)} \frac{d|\nu|(y)}{\max(1,(N\rho(x,y))^S)} \le \frac{c_1(c+(d/r)^q)q}{\max(1,(rN/2)^{S-q})} |||\nu|||_{R,d}, \tag{4.5}$$

and

$$N^q \int\limits_{\mathbb{X}} \frac{d|\nu|(y)}{\max(1,(N\rho(x,y))^S)} \le c_1(c+(Nd)^q) |||\nu|||_{R,d}. \tag{4.6}$$

4.2. Localized kernels

Localized kernels form the main ingredient in our proofs. To obtain these kernels, we will use the following Tauberian theorem ([30, Theorem 4.3]) with different choices of the function H.

Theorem 4.1. *Let* μ^* *be an extended complex valued measure on* $[0,\infty)$, *and* $\mu^*(\{0\}) = 0$. *We assume that there exist* $Q, r > 0$, *such that each of the following conditions are satisfied.*

1. $$|||\mu^*|||_Q := \sup_{u\in[0,\infty)} \frac{|\mu^*|([0,u))}{(u+2)^Q} < \infty. \tag{4.7}$$

2. *There are constants* $c, C > 0$, *such that*

$$\left| \int\limits_{\mathbb{R}} \exp(-u^2 t) d\mu^*(u) \right| \le c_1 t^{-C} \exp(-r^2/t) |||\mu^*|||_Q, \qquad 0 < t \le 1. \tag{4.8}$$

Let $H : [0,\infty) \to \mathbb{R}$, $S > Q+1$ *be an integer, and suppose that there exists a measure* $H^{[S]}$ *such that*

$$H(u) = \int\limits_0^\infty (v^2 - u^2)_+^S dH^{[S]}(v), \qquad u \in \mathbb{R}, \tag{4.9}$$

and

$$V_{Q,S}(H) = \max\left(\int\limits_0^\infty (v+2)^Q v^{2S} d|H^{[S]}|(v), \int\limits_0^\infty (v+2)^Q v^S d|H^{[S]}|(v) \right) < \infty. \tag{4.10}$$

Then for $n \ge 1$,

$$\left| \int\limits_0^\infty H(u/n) d\mu^*(u) \right| \le c \frac{n^Q}{\max(1,(nr)^S)} V_{Q,S}(H) |||\mu^*|||_Q. \tag{4.11}$$

We observe that if H is compactly supported, has $S+1$ continuous derivatives, and is constant in a neighborhood of 0, then the Taylor formula used with $u \to H(\sqrt{u})$ shows that the representation (4.9) holds with $H^{[S]}$ being the S-th derivative of $u \to H(\sqrt{u})$, and we have

$$V_{Q,S}(H) \le c \max_{0\le k\le S+1} \max_{u\in\mathbb{R}} |H^{(k)}(u)| = c|||H|||_S. \tag{4.12}$$

The following proposition summarizes some general results which we will use in our proofs later. If $\{\psi_j\}$, $\{\tilde{\psi}_j\}$ are sequences of bounded functions on $\mathbb{X}$, we define formally

$$\Phi_n(\{\psi_j\}, \{\tilde{\psi}_j\}, H; x, y) = \sum_{j=0}^{\infty} H\left(\frac{\lambda_j}{n}\right) \psi_j(x)\tilde{\psi}_j(y). \tag{4.13}$$

It is convenient to set

$$\Phi_0(\{\psi_j\}, \{\tilde{\psi}_j\}, H; x, y) = \sum_{j:\lambda_j=0} \psi_j(x)\tilde{\psi}_j(y).$$

Proposition 4.3. *Let $\{\psi_j\}$, $\{\tilde{\psi}_j\}$ be sequences of bounded functions on $\mathbb{X}$, H be as in Theorem 4.1, and for $x, y \in \mathbb{X}$,*

$$\sum_{j:\lambda_j<u} |\psi_j(x)\tilde{\psi}_j(y)| \le cu^Q, \qquad u \ge 1, \tag{4.14}$$

$$\sum_{j=0}^{\infty} \exp(-\lambda_j^2 t)\psi_j(x)\tilde{\psi}_j(y) \le c_1 t^{-C} \exp(-\rho(x, y)^2/t), \qquad 0 < t \le 1. \tag{4.15}$$

Then

$$\left|\Phi_n(\{\psi_j\}, \{\tilde{\psi}_j\}, H; x, y)\right| \le c \frac{n^Q}{\max(1, (n\rho(x, y))^S)} V_{Q,S}(H), \qquad n \ge 1. \tag{4.16}$$

Further, if $d > 0$ and ν is a d-regular measure, then for $x \in \mathbb{X}$, $r > 0$, $n \ge 1$ and $1 \le p < \infty$,

$$\int_{\Delta(x,r)} |\Phi_n(\{\psi_j\}, \{\tilde{\psi}_j\}, H; x, y)| d|\nu|(y) \le n^{Q-q} \frac{c_1(c + (d/r)^q)q}{\max(1, (rN/2)^{S-q})} |||\nu|||_{R,d} V_{Q,S}(H), \tag{4.17}$$

$$\int_{\mathbb{X}} |\Phi_n(\{\psi_j\}, \{\tilde{\psi}_j\}, H; x, y)|^p d|\nu|(y) \le c\left(c_1 + (nd)^q\right) n^{Qp-q} |||\nu|||_{R,d} V_{Q,S}(H)^p. \tag{4.18}$$

Proof. For each $x, y \in \mathbb{X}$, we apply Theorem 4.1 with the measure

$$\mu^*_{x,y}([0, u)) = \sum_{j:\lambda_j<u} \psi_j(x)\tilde{\psi}_j(y).$$

The estimate (4.14) (respectively, (4.15), (4.16)) is equivalent to (4.7) (respectively, (4.8), (4.11)). The estimates (4.17) and (4.18) follow from (4.16) and a straightforward application of (4.5) and (4.6) respectively. □

Corresponding to the formal kernel in (4.13), we have the formal operator:

$$\sigma_n(\nu; \{\psi_j\}, \{\tilde{\psi}_j\}, H; f, x) := \int_{\mathbb{X}} \Phi_n(\{\psi_j\}, \{\tilde{\psi}_j\}, H; x, y) f(y) d\nu(y), \qquad f \in L^1,\ n > 0,\ x \in \mathbb{X}. \tag{4.19}$$

It is convenient to define

$$\sigma_0(\nu; \{\psi_j\}, \{\tilde{\psi}_j\}, H; f, x) := \int_{\mathbb{X}} \Phi_0(\{\psi_j\}, \{\tilde{\psi}_j\}, H; x, y) f(y) d\nu(y), \qquad f \in L^1,\ x \in \mathbb{X}. \tag{4.20}$$

The following proposition lists some norm estimates for these operators.

Proposition 4.4. *We assume the set up as in Proposition 4.3. Let ν_1 be a d_1-regular measure, ν_2 be a d_1-regular measure, and $1 \le p \le r \le \infty$. Then for $f \in L^p(\nu_1)$*

$$\|\sigma_n(\nu_1; \{\psi_j\}, \{\tilde{\psi}_j\}, H; f)\|_{\nu_2;r} \le c\left((c_1 + (nd_1)^q)|||\nu_1|||_{R,d_1}\right)^{1/p'} \times \left((c_1 + (nd_2)^q)|||\nu_2|||_{R,d_2}\right)^{1/r} N^{Q-q+q(1/p-1/r)}\|f\|_{\nu_1;p}. \tag{4.21}$$

Proof. In this proof, we will abbreviate $\Phi_n(\{\psi_j\}, \{\tilde{\psi}_j\}, H; x, y)$ by $\Phi_n(x, y)$ and $\sigma_N(\nu_1; \{\psi_j\}, \{\tilde{\psi}_j\}, H; f)$ by $\sigma_n(f)$. Without loss of generality, we may assume also that $V_{S,Q}(H) = 1$, and ν_j are positive measures. Using Hölder inequality and (4.18), we deduce that for $x \in \mathbb{X}$,

$$|\sigma_n(f, x)| \le \int_{\mathbb{X}} |\Phi_n(x, y)||f(y)|d\nu_1(y) \le \|\Phi_n(x, \circ)\|_{\nu_1;p'}\|f\|_{\nu_1;p}$$
$$\le c\left(c_1 + (nd_1)^q)|||\nu_1|||_{R,d_1}\right)^{1/p'} n^{Q-q/p'}\|f\|_{\nu_1;p};$$

i.e.,

$$\|\sigma_n(f)\|_{\nu_2;\infty} \le c\left((c_1 + (nd_1)^q)|||\nu_1|||_{R,d_1}\right)^{1/p'} n^{Q-q/p'}\|f\|_{\nu_1;p}. \tag{4.22}$$

This proves (4.21) when $r = \infty$.

In particular, with $p = \infty$,

$$\|\sigma_n(f)\|_{\nu_2;\infty} \le c(c_1 + (nd_1)^q)|||\nu_1|||_{R,d_1} n^{Q-q}\|f\|_{\nu_1;\infty}. \tag{4.23}$$

Switching the roles of $\{\psi_j\}$ and $\{\tilde{\psi}_j\}$ in (4.18) (used with ν_2 in place of ν, $r = 1$), the estimate becomes

$$\int_{\mathbb{X}} |\Phi_n(x, y)|d\nu_2(x) \le c(c_1 + (nd_2)^q)|||\nu_2|||_{R,d_2} n^{Q-q}.$$

Therefore,

$$\|\sigma_n(f)\|_{\nu_2;1} \le \int_{\mathbb{X}}\int_{\mathbb{X}} |\Phi_n(x, y)||f(y)|d\nu_1(y)d\nu_2(x) = \int_{\mathbb{X}}\left\{\int_{\mathbb{X}} |\Phi_n(x, y)|d\nu_2(x)\right\}|f(y)|d\nu_1(y)$$
$$\le c(c_1 + (nd_2)^q)|||\nu_2|||_{R,d_2} n^{Q-q}\|f\|_{\nu_1;1}. \tag{4.24}$$

In view of the Riesz–Thorin interpolation theorem [31, Theorem 1.1.1], (4.23) and (4.24) lead to

$$\|\sigma_n(f)\|_{\nu_2;p} \le c\left((c_1 + (nd_1)^q)|||\nu_1|||_{R,d_1}\right)^{1/p'}\left((c_1 + (nd_2)^q)|||\nu_2|||_{R,d_2}\right)^{1/p} n^{Q-q}\|f\|_{\nu_1;p}, \qquad 1 \le p \le \infty. \tag{4.25}$$

Hence, using (4.22), we obtain for $1 \le r < r < \infty$,

$$\int_{\mathbb{X}} |\sigma_n(f, x)|^r d\nu_2(x) = \int_{\mathbb{X}} |\sigma_n(f, x)|^{r-p}|\sigma_n(f, x)|^p d\nu_2(x) \le \|\sigma_n(f)\|_{\nu_2;\infty}^{r-p}\|\sigma_n(f)\|_{\nu_2;p}^p$$
$$\le c\left((c_1 + (nd_1)^q)|||\nu_1|||_{R,d_1}\right)^{(r-p)/p'}\left((c_1 + (nd_1)^q)|||\nu_1|||_{R,d_1}\right)^{p/p'}$$
$$\left((c_1 + (nd_2)^q)|||\nu_2|||_{R,d_2}\right) n^{(Q-q/p')(r-p)+(Q-q)p}\|f\|_{\nu_1;p}^r.$$

This leads to (4.21) when $r < \infty$, and completes the proof. □

In the sequel, we fix an infinitely differentiable, even function $h : \mathbb{R} \to \mathbb{R}$, nonincreasing on $[0, \infty)$, such that $h(t) = 1$ if $|t| \le 1/2$, and $h(t) = 0$ if $|t| \ge 1$. We will write $g(t) = h(t) - h(2t)$. The following identities will be used often without reference: For integers $n, k \ge 0$,

$$h\left(\frac{k}{2^n}\right) = h(k) + \sum_{j=1}^{n} g\left(\frac{k}{2^j}\right), \qquad h\left(\frac{k}{2^n}\right) + \sum_{j=n+1}^{\infty} g\left(\frac{k}{2^j}\right) = 1. \tag{4.26}$$

We summarize the localization properties of three kernels which we will use in our proofs. Let b be a mask of type $\beta \in \mathbb{R}$. In the sequel, if $n > 0$, we will write $b_n(t) = b(nt)$.

Proposition 4.5. *Let $x, y \in \mathbb{X}$, $n, N \geq 1$, $k = 1, \cdots, R$, $d > 0$, ν be a d-regular measure, $1 \leq p \leq r \leq \infty$. We have*

$$|\Phi_n(h; x, y)| \leq c \frac{n^q}{\max(1, (n\rho(x, y))^S)}, \quad \|\Phi_n(h; x, \circ)\|_{\nu;p} \leq c \left((c_1 + (nd)^q)|||\nu|||_{R,d}\right)^{1/p} n^{q/p'}. \tag{4.27}$$

$$|L_{k,1}\Phi_n(h; x, y)| \leq c \frac{n^{q+q_k}}{\max(1, (n\rho(x, y))^S)}, \quad \|L_{k,1}\Phi_n(h; x, \circ)\|_{\nu;p} \leq c \left((c_1 + (nd)^q)|||\nu|||_{R,d}\right)^{1/p} n^{q_k - q/p'}, \tag{4.28}$$

and

$$|L_{k,1}\Phi_n(gb_N; x, y)| \leq cN^{-\beta} \frac{n^{q+q_k}}{\max(1, (n\rho(x, y))^S)},$$
$$\|L_{k,1}\Phi_n(gb_N; x, \circ)\|_{\nu;p} \leq c \left((c_1 + (nd)^q)|||\nu|||_{R,d}\right)^{1/p} n^{q_k - q/p'} N^{-\beta}. \tag{4.29}$$

Further, with the set up as in Proposition 4.4, and writing $C_1 = (c_1 + (nd_1)^q)|||\nu_1|||_{R,d_1}$ and $C_2 = (c_1 + (nd_2)^q)|||\nu_2|||_{R,d_2}$, we have

$$\|\sigma_n(\nu_1; h; f)\|_{\nu_2;r} \leq cC_1^{1/p'} C_2^{1/r} n^{q(1/p - 1/r)} \|f\|_{\nu_1;p}, \tag{4.30}$$

$$\|L_k\sigma_n(\nu_1; h; f)\|_{\nu_2;r} \leq cC_1^{1/p'} C_2^{1/r} n^{q_k + q(1/p - 1/r)} \|f\|_{\nu_1;p}, \tag{4.31}$$

and

$$\|L_k\sigma_n(\nu_1; gb_N; f)\|_{\nu_2;r} \leq cC_1^{1/p'} C_2^{1/r} n^{q_k + q(1/p - 1/r)} N^{-\beta} \|f\|_{\nu_1;p}. \tag{4.32}$$

Proof. In view of Proposition 4.3, the first estimate in (4.27) (respectively, (4.28)) follows from (4.12) and (3.5) (respectively, (3.8)). The second estimate in (4.27) follows from (4.18) with the choices $\tilde{\psi}_j = \psi_j = \phi_j$, $H = h$, by observing from (3.5) that $Q = q$, and from (4.12) that $V_{S,q}(H) \leq c$. Proposition 4.4 yields (4.30) with the same choices.

The second estimate in (4.28) follows similarly, except with the choice $\tilde{\psi}_j = L_k\phi_j$ and the observation that (3.8) implies that $Q = q + q_k$. Proposition 4.4 yields (4.31) with the same choices.

It is easy to verify by induction that

$$\left| t^k \frac{d^k}{dt^k}((1+t)^\beta b(t)) \right| = \left| t^k \frac{d^k}{dt^k} F_b(\log t) \right| \leq c(b)c_2, \qquad t > 0,\ k = 0, \cdots, S,$$

and hence, for $N \geq 1$,

$$\left| t^k \frac{d^k}{dt^k}((1/N+t)^\beta b_N(t)) \right| \leq c(b)c_2 N^{-\beta}, \qquad t > 0,\ k = 0, \cdots, S+1. \tag{4.33}$$

Since $b(t)^{-1}$ is a mask of type $-\beta$, we record that

$$\left| t^k \frac{d^k}{dt^k}((1/N+t)^\beta b_N(t))^{-1} \right| \leq c(b)c_2 N^{\beta}, \qquad t > 0,\ k = 0, \cdots, S+1. \tag{4.34}$$

Finally, if $g : \mathbb{R} \to \mathbb{R}$ is any compactly supported, S times continuously differentiable function, such that $g(t) = 0$ on some neighborhood of 0 then (4.33), (4.34) imply

$$|||gb_N|||_S \leq c(b, g)N^{-\beta}, \qquad |||g/b_N|||_S \leq c(b, g)N^{\beta}, \qquad N \geq 1. \tag{4.35}$$

In particular, (4.12), and (4.35) imply that $V_{S,Q}(gb_N) \leq cN^{-\beta}$. The first estimate in (4.29) follows from Proposition 4.3. The second estimate follows similarly to the second estimate in (4.28). The estimate (4.32) follows from Proposition 4.4 as with (4.31). □

5. Integral representation

The objective of this section is to prove

Theorem 5.1. *Let $1 \leq p \leq \infty$, $\beta > \max_{1 \leq k \leq m} q_k + q/p$, and $f \in W_{p,b}$. Then for every $x \in \mathbb{X}$,*

$$L_k(f, x) = \int_{\mathbb{X}} L_{k,1}G(x, y) f^{(b)}(y) d\mu^*(y). \tag{5.1}$$

Moreover, for $k = 1, \cdots, R$,

$$\|f - \sigma_n(h; f)\|_\infty \leq cn^{-q(\beta - 1/p)} \|f^{(b)}\|_p, \quad \|L_k f - L_k\sigma_n(h; f)\|_\infty \leq cn^{-q(\beta - q_k - 1/p)} \|f^{(b)}\|_p. \tag{5.2}$$

Our proof of Theorem 5.1 requires two inequalities: the Bernstein inequality and the Nikolskii inequality. The Bernstein inequality is the following.

Lemma 5.1. *Let* $1 \le k \le R$. *Then*

$$\|L_k(P)\|_p \le cn^{q_k}\|P\|_p, \qquad P \in \Pi_n,\ n \ge 1. \tag{5.3}$$

Proof. If $P \in \Pi_n$, then a straightforward calculation using the orthogonality of $\{\phi_k\}$ and the facts that $h(t) = 1$ if $|t| \le 1/2$ and $\hat{P}(k) = 0$ if $k > n$ shows that for $x \in \mathbb{X}$,

$$P(x) = \int_{\mathbb{X}} \Phi_{2n}(h; x, y)P(y)d\mu^*(y) = \sigma_{2n}(\mu^*; h, P), \qquad x \in \mathbb{X}. \tag{5.4}$$

Therefore, $L_k P = L_k\sigma_{2n}(\mu^*; h; P)$. We now apply (4.31) with $2n$ in place of n, $\nu_2 = \nu_1 = \mu^*$ (so that $d_1 = d_2 = 0$, $|||\nu_j|||_{d_j} = 1$), and $r = p$. This leads to (5.3). □

The Nikolskii inequality is the following.

Lemma 5.2. *If* $n > 0$, $P \in \Pi_n$, $1 \le p < r \le \infty$, *then*

$$\|P\|_r \le cn^{q(1/p-1/r)}\|P\|_p. \tag{5.5}$$

Proof. We apply (4.30) with $2n$ in place of n, $\nu_2 = \nu_1 = \mu^*$ (so that $d_1 = d_2 = 0$, $|||\nu_j|||_{d_j} = 1$), and P in place of f to obtain

$$\|\sigma_{2n}(\mu^*; h; P)\|_r \le cn^{q(1/p-1/r)}\|P\|_p.$$

The estimate (5.5) follows from (5.4). □

We are now ready to prove Theorem 5.1.

Proof of Theorem 5.1. In this proof only, let L denote any of the operators L_k, and A be the corresponding q_k as defined in (5.3). We observe using (2.2) that

$$\sigma_{2^j}(gb_{2^j}; f^{(b)}) = \sigma_{2^j}(g; f).$$

In view of (4.35), we deduce that for integers $j \ge 1$,

$$\|\sigma_{2^j}(g; f)\|_p = \|\sigma_{2^j}(gb_{2^j}; f^{(b)})\|_p \le c2^{-j\beta}\|f^{(b)}\|_p. \tag{5.6}$$

Since $\sigma_{2^j}(g; f) \in \Pi_{2^j}$, (5.5) implies that

$$\|\sigma_{2^j}(g; f)\|_\infty \le c2^{-j(\beta-1/p)}\|f^{(b)}\|_p. \tag{5.7}$$

Hence, (5.3) shows that

$$\|L\sigma_{2^j}(g; f)\|_\infty \le c2^{-j(\beta-A-1/p)}\|f^{(b)}\|_p. \tag{5.8}$$

Using the second identity in (4.26), we deduce (5.2) easily from (5.7) and (5.8). In particular, since L is a closed operator, this shows that for $x \in \mathbb{X}$,

$$L(f, x) = \sum_{j=0}^{\infty} L\sigma_{2^j}(g; f, x) = \sum_{j=0}^{\infty} L\sigma_{2^j}(gb_{2^j}; f^{(b)}, x),$$

with the convergence being uniform. This leads to (5.1). □

6. An approximation result

The purpose of this section is to prove the following result, analogous to [12, Theorem 6.3].

Theorem 6.1. *Let* $1 \le p \le \infty$, $\beta > \max_{1\le k\le R} q_k + q/p'$, *and*

$$\Psi(x) = \sum_{k=1}^{R}\sum_{j=1}^{M} a_{k,j} L_{k,1} G(y_j, x), \qquad x \in \mathbb{X}. \tag{6.1}$$

There exists an integer $N^* \sim \eta^{-1}$ *such that for* $n \ge N^*$,

$$\|\Psi - \sigma_n(h;\Psi)\|_p \le \frac{1}{2}\|\Psi\|_p. \tag{6.2}$$

As in [12], the main problem is to relate $\|\Psi\|_p$ and the coefficients $c_{k,j}$, via $\sigma_{2^m}(g;\Psi)$. The main technical problem is the following. The mapping $y \mapsto \sigma_{2^m}(g; L_{k,1}G(y,x))$ is not in Π_{2^m}. So, we cannot use an analogue of [12, Proposition 6.3] to obtain a full estimate of the form [12, Theorem 6.2(b)]. The following Proposition 6.1 (which we will call the coefficient inequalities) serves as a substitute.

In order to state the coefficient inequalities, we will use the following notations: $\mathbf{a}_k = (a_{k,1}, \cdots, a_{k,j})$, and

$$\Psi_k(x) = \sum_{j=1}^{M} a_{k,j} L_{k,1} G(y_j, x), \qquad x \in \mathbb{X}, \tag{6.3}$$

so that $\Psi = \sum_{k=1}^{R} \Psi_k$. For any sequence $\mathbf{d}$,

$$\|\mathbf{d}\|_p := \begin{cases} \left\{\sum_{j=1}^{\infty} |d_j|^p\right\}^{1/p}, & \text{if } 1 \le p < \infty, \\ \sup_{1\le j\le\infty} |d_j|, & \text{if } p = \infty, \end{cases}$$

with a Euclidean vector extended to a sequence by padding with 0's. The coefficient inequalities are given in the following proposition.

Proposition 6.1. *Let* $1 \le p \le \infty$, *and* $1 \le k \le R$.

(a) *For integer m with* $2^m\eta \ge 1$, *we have*

$$\|\sigma_{2^m}(g;\Psi_k)\|_p \le c2^{m(q_k-\beta+q/p')}\|\mathbf{a}_k\|_p. \tag{6.4}$$

(b) *We have*

$$\|\mathbf{a}_k\|_p \le c\eta^{q_k-\beta+q/p'}\|\Psi\|_p. \tag{6.5}$$

(c) *For integer m with* $2^m\eta \ge 1$, *we have*

$$\|\sigma_{2^m}(g;\Psi)\|_p \le c\sum_{k=1}^{R}(2^m\eta)^{q_k-\beta+q/p'}\|\Psi\|_p, \tag{6.6}$$

and

$$\sum_{k=1}^{R}\eta^{-q_k}\|\mathbf{a}_k\|_p \le c\eta^{q/p'-\beta}\|\Psi\|_p. \tag{6.7}$$

Proof. The proof of part (a) mimics that of [12, Theorem 6.2(a)]. We observe that for $x \in \mathbb{X}$,

$$\sigma_{2^m}(g; L_{k,1}G(y_j,\circ), x) = \sum_i g(\lambda_i/2^m) b(\lambda_i) L_k(\phi_i)(y_j)\phi_i(x) = L_{k,1}\Phi_{2^m}(gb_{2^m}; y_j, x).$$

Using (4.29) with $p = 1$, $\nu = \mu^*$ (so that $d = 0$, $|||\nu|||_{R,d} = 1$), we obtain

$$\|\sigma_{2^m}(g; L_{k,1}G(y_j,\circ)\|_1 \le c2^{m(q_k-\beta)}. \tag{6.8}$$

Therefore,

$$\|\sigma_{2^m}(g;\Psi_k)\|_1 \le \sum_{j=1}^{M} |a_{k,j}| \|\sigma_{2^m}(g; L_{k,1}G(y_j,\circ))\|_1 \le c2^{m(q_k-\beta)}\|\mathbf{a}_k\|_1. \tag{6.9}$$

Next, we use (4.29) again with $2^m \ge \eta^{-1}$ in place of n, $p=1$ and ν to be the measure τ as defined in Lemma 4.1 (so that $d=\eta$, $|||\tau|||_{R,\eta} \le c$) to deduce that

$$\begin{aligned}\left\|\sum_{j=1}^{M} |\sigma_{2^m}(g; L_{k,1}G(y_j,\circ))|\right\|_\infty &= \left|\eta^{-q}\int_{\mathbb{X}} |\sigma_{2^m}(g; L_{k,1}G(\circ, y))| d\tau(y)\right\|_\infty \\ &\le c2^{m(q+q_k-\beta)}(1+(2^m\eta)^q)(2^m\eta)^{-q} \le c2^{m(q+q_k-\beta)}.\end{aligned} \tag{6.10}$$

Therefore,

$$\begin{aligned}\|\sigma_{2^m}(g;\Psi_k)\|_\infty &\le \sum_{j=1}^{M} |a_{k,j}| \|\sigma_{2^m}(g; L_{k,1}G(y_j,\circ)\|_\infty \\ &\le \|\mathbf{a}_k\|_\infty \left\|\sum_{j=1}^{M} |\sigma_{2^m}(g; L_{k,1}G(y_j,\circ))|\right\|_\infty \le c2^{m(q+q_k-\beta)}\|\mathbf{a}_k\|_\infty.\end{aligned} \tag{6.11}$$

The estimate (6.4) follows from (6.9) and (6.11) by an application of the Riesz–Thorin interpolation theorem [31, Theorem 1.1.1] to the operator $\mathbf{a}_k \mapsto \Psi_k$.

To prove part (b), we fix k, and find $\mathbf{d} \in \mathbb{R}^M$ such that

$$\langle \mathbf{a}_k, \mathbf{d}\rangle = \|\mathbf{a}_k\|_p, \qquad \|\mathbf{d}\|_{p'} = 1. \tag{6.12}$$

We then recall the functions $\phi_{k,j}$ defined in the assumptions on L_k, and set

$$F(x) = \sum_{j=1}^{M} d_j \phi_{k,j}^{(b)}(x), \qquad x \in \mathbb{X}.$$

We estimate $\|F\|_{p'}$. We consider the case when $1 \le p' < \infty$, the case when $p'=\infty$ is only simpler. Since each $\phi_{k,j}^{(b)}$ is supported on a neighborhood of y_j with diameter $\le \eta/3$, for any given $x \in \mathbb{X}$, there is at most one j' such that $\phi_{k,j'}^{(b)}(x) \ne 0$. Then

$$|F(x)|^{p'} = |d_{j'}|^{p'} |\phi_{k,j'}^{(b)}(x)|^{p'},$$

and thus, for any $x \in \mathbb{X}$,

$$|F(x)|^{p'} = \sum_{j=1}^{M} |d_j|^{p'} |\phi_{k,j}^{(b)}(x)|^{p'}. \tag{6.13}$$

Since each $\phi_{k,j}^{(b)}$ is supported on a neighborhood of y_j with diameter $\le \eta/3$, (3.11) and (3.4) shows that

$$\int_{\mathbb{X}} |\phi_{k,j}^{(b)}(x)|^{p'} d\mu^*(x) \le \int_{\mathbb{B}(y_j,\eta/3)} |\phi_{k,j}^{(b)}(x)|^{p'} d\mu^*(x) \le c\eta^{(q_k-\beta)p'+q}.$$

Consequently, (6.13) implies that

$$\|F\|_{p'} \le c\eta^{q_k-\beta+q/p'}. \tag{6.14}$$

Next, using Theorem 5.1 and (2.7) from the assumptions on L_k, we obtain that for any $k' = 1, \cdots, R$,

$$\begin{aligned}\int_{\mathbb{X}} L_{k',1}G(y_{j'}, x)F(x)d\mu^*(x) &= \sum_{j=1}^{M} d_j \int_{\mathbb{X}} L_{k',1}G(y_{j'}, x)\phi_{k,j}^{(b)}(x)d\mu^*(x) \\ &= \sum_{j=1}^{M} d_j L_{k'}(\phi_{k,j})(y_{j'}) = \begin{cases} d_{j'}, & \text{if } k'=k, \\ 0, & \text{otherwise.}\end{cases}\end{aligned}$$

Consequently, the definition (6.12) of $\mathbf{d}$ shows that

$$\int_{\mathbb{X}} \Psi(x)F(x)d\mu^*(x) = \sum_{k'=1}^{R}\sum_{j'=1}^{M} a_{k',j'} \int_{\mathbb{X}} L_{k',1}G(y_{j'},x)F(x)d\mu^*(x) = \sum_{j'=1}^{M} a_{k,j'}d_{j'} = \|\mathbf{a}_k\|_p.$$

Together with (6.14), this shows that

$$\|\mathbf{a}_k\|_p = \int_{\mathbb{X}} \Psi(x)F(x)d\mu^*(x) \le \|\Psi\|_p\|F\|_{p'} \le c\eta^{q_k-\beta+q/p'}\|\Psi\|_p.$$

This proves part (b).

Part (c) is a straightforward consequence of parts (a) and (b) and the fact that $\Psi = \sum_{k=1}^{R}\Psi_k$. $\square$

Proof of Theorem 6.1. In this proof, let $A = \max_{1\le k\le R} q_k$. In light of (4.26) and (6.6) we obtain for $N \ge \eta^{-1}$, $\beta > A + q/p'$ that

$$\|\Psi - \sigma_{2^N}(h;\Psi)\|_p \le \sum_{m=N+1} \|\sigma_{2^m}(g;\Psi)\|_p \le c\sum_{k=1}^{R}\sum_{m=N+1}(2^m\eta)^{q_k-\beta+q/p'}\|\Psi\|_p \le c_1(2^N\eta)^{A-\beta+q/p'}\|\Psi\|_p.$$

If $n \ge 2^{N+1}$, then

$$\|\Psi - \sigma_n(h;\Psi)\|_p \le c \inf_{P\in\Pi_{n/2}} \|\Psi - P\|_p \le c \inf_{P\in\Pi_{2^N}} \|\Psi - P\|_p \le c\|\Psi - \sigma_{2^N}(h;\Psi)\|_p \le c_2(2^N\eta)^{A-\beta+q/p'}\|\Psi\|_p.$$

We may choose N so that $2^N \sim \eta^{-1}$ and the rightmost expression above is $\le (1/2)\|\Psi\|_p$. Then we have proved (6.2) with $N^* = 2^{N+1}$. $\square$

7. Proofs of the main results

Proof of Theorem 2.1. We define the inner product

$$\langle g_1, g_2\rangle = \sum_{k=0}^{\infty} \frac{\hat{g_1}(k)\hat{g_2}(k)}{b(\lambda_k)^2}.$$

We wish to show first that there exist the coefficients $a_{k,j}$ such that

$$L_iP(y_\ell) = f_{i,\ell}, \qquad \ell = 1,\cdots,M_n,\ i = 1,\cdots,R, \tag{7.1}$$

and with this choice, P is the solution of the minimization problem (2.8).

We observe that

$$\sum_{k,j}\sum_{i,\ell} d_{k,j}d_{i,\ell}L_{i,2}L_{k,1}\mathbb{G}(y_j,y_\ell) = \sum_{\nu=0}^{\infty} b(\lambda_\nu)^2 \left\{\sum_{k,j} d_{k,j}L_k(\phi_\nu)(y_j)\right\}^2 \ge 0.$$

If the infinite sum is equal to 0, then

$$\sum_{k,j} d_{k,j}L_k(\phi_\nu)(y_j) = 0, \qquad \nu = 0,1,\cdots.$$

Consequently, for every $x \in \mathbb{X}$,

$$0 = \sum_{\nu=0}^{\infty} b(\lambda_\nu)^2 \sum_{k,j} d_{k,j}L_k(\phi_\nu)(y_j)\phi_\nu(x) = \sum_{k,j} d_{k,j}L_{k,1}\mathbb{G}(y_j,x).$$

Next, we observe by a straightforward calculation that for any f in the domain of the L_k's,

$$\langle L_{k,1}\mathbb{G}(y,\circ), f\rangle = L_k(f)(y). \tag{7.2}$$

So, for each i and ℓ,

$$d_{i,\ell} = L_i(\phi_{i,\ell})(y_\ell) = \left\langle \sum_{k,j} d_{k,j}L_{k,1}\mathbb{G}(y_j,\circ), \phi_{i,\ell}\right\rangle = 0.$$

Thus, the matrix of the system of equations (7.1) is positive definite, and hence, (2.9) has a unique solution given by P.

Let g be another candidate for the minimization problem. Then

$$\begin{aligned}\langle P, g-P\rangle &= \sum_{k,j} a_{k,j}\langle L_{k,1}\mathbb{G}(y_j,\circ), g-P\rangle = \sum_{k,j} a_{k,j}(L_k(g)(y_j) - L_k(P)(y_j)) \\ &= \sum_{k,j} a_{k,j}(L_k(f)(y_j) - L_k(f)(y_j)) = 0.\end{aligned}$$

Hence,

$$\|g^{(b)}\|_2^2 = \langle g, g\rangle = \langle g-P, g-P\rangle + \langle P, P\rangle \geq \|P^{(b)}\|_2^2.$$

This proves that P is the solution of the minimization problem. □

Proof of Theorem 2.2. This proof is almost verbatim the same as the proof of [12, Theorem 5.1]. We will point out only the differences. We will omit the notation n in our proof as in the other paper. Let $\mathbf{a} \in \mathbb{R}^{RM}$ be a row-major ordering of the matrix $(a_{k,j})_{k=1,\cdots,R,\ j=1,\cdots,M}$. We define

$$|||\mathbf{a}|||_{RM}^* := \left\| \sum_{k=1}^{R}\sum_{j=1}^{M} a_{k,j} L_{k,1} G(y_j,\circ)\right\|_{p'}. \tag{7.3}$$

In view of (6.7) this is a norm on $\mathbb{R}^{RM}$. Let N^* be chosen so that Theorem 6.1 holds with p' in place of p, and D be the dimension of Π_{N^*}. For $\mathbf{d} \in \mathbb{R}^D$, we define

$$|||\mathbf{d}|||_D := \left\| \sum_{\lambda_j < N^*} b(\lambda_j)^{-1} d_j \phi_j \right\|_p. \tag{7.4}$$

In place of $F^{(-s)}$ in the proof in [12], we need $F^{(1/b)}$. In place of the matrix $\mathbf{A}$ in [12], we take the matrix appropriate for the interpolation problem which we are dealing with; i.e., a matrix indexed by (k, j) and m so that the $((k, j), m)$-th entry is $L_k(\phi_m)(y_j)$. The rest of the proof is the same as in [12] with obvious minor changes; e.g., $\mathbb{X}$ in place of $[-\pi,\pi]^q$, bivariate kernels in place of convolutions, etc. □

We will prove Theorem 2.3 in much greater generality for future reference in the form of Theorem 7.1 below.

Let $\mathbb{X}$ be a separable Banach space with norm $\|\circ\|$, $\mathbb{X}^*$ be its dual space with dual norm $\|\circ\|^*$. Let $R \geq 1$ be an integer, and for each integer $n \geq 1$, $1 \leq k \leq R$, $x_{k,n}^* \in \mathbb{X}^*$, with $\|x_{k,n}^*\|^* \leq 1$. We assume that for each k, $x_{k,n}^* \to x_k^*$ in the weak-* topology. Necessarily, each $x_k^* \in \mathbb{X}^*$ and $\|x_k^*\|^* \leq 1$.

Let $\mathbb{Y}$ be another normed linear space with norm $|\circ|_{\mathbb{Y}}$, continuously embedded, and hence, identified with a subspace of $\mathbb{X}$. We assume that the unit ball

$$\mathcal{B} = \{g \in \mathbb{Y} : |g|_{\mathbb{Y}} \leq 1\}$$

is compact in $\mathbb{X}$. Let $V_1 \subseteq \cdots \subseteq V_n \subseteq V_{n+1} \subseteq \cdots$ be a sequence of subsets of $\mathbb{Y}$.

Theorem 7.1. *Let $f \in \mathbb{Y}$, and we assume that there exists $v_n \in V_n$ such that*

$$|v_n|_{\mathbb{Y}} = \min\{|g|_{\mathbb{Y}} : g \in V_n,\ x_{k,n}^*(g) = x_k^*(f),\ k = 1,\cdots,R\} \leq c|f|_{\mathbb{Y}}. \tag{7.5}$$

Then $x_k^*(v_n) \to x_k^*(f)$ as $n \to \infty$.

This theorem easily follows from the following lemma. If $\delta > 0$, and $K \subset \mathbb{X}$ is a compact set, we say that a subset A of K is δ separated if

$$\min_{\substack{f_1,f_2\in A\\ f_1\neq f_2}} \|f_1 - f_2\| \geq \delta.$$

Since K is compact, such a set is necessarily finite.

Lemma 7.1. *Let $\mathbb{X}$ be a separable Banach space, $\{y_m^*\}_{m=0}^{\infty}$ be a sequence in the unit ball of $\mathbb{X}^*$ converging to $y^* \in \mathbb{X}^*$ in the weak-* topology. If $K \subset \mathbb{X}$ is a compact set, then*

$$\lim_{m\to\infty} \sup_{f\in K} |y_m^*(f) - y^*(f)| = 0.$$

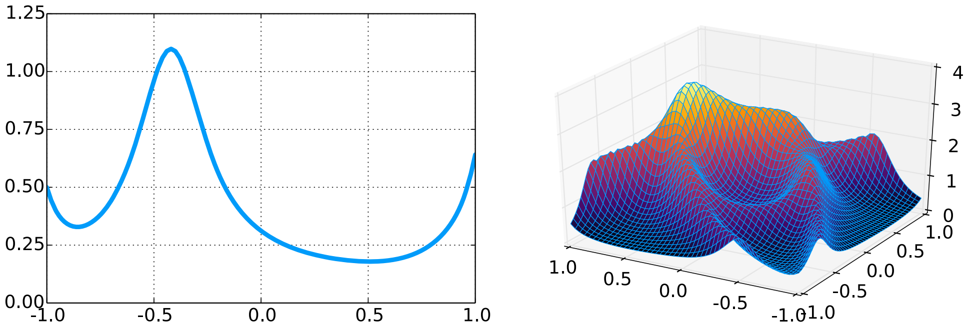

Fig. 1. Test functions $f_{25}(x, -0.96)$ and $f_{25}(x, y)$ from Eq. (8.1).

Proof. Let $\epsilon > 0$. We consider the set A to be the maximal $\epsilon/3$ separated subset of K. If

$$f_0 \in K \setminus \bigcup_{g\in A} \{f \in \mathbb{X} : \|g - f\| \le \epsilon/3\},$$

then we may add f_0 to A to obtain a larger $\epsilon/3$ separated subset of K. Therefore,

$$K \subseteq \bigcup_{g\in A} \{f \in \mathbb{X} : \|g - f\| \le \epsilon/3\}. \tag{7.6}$$

In view of the weak-* convergence, we obtain an integer $N \ge 1$ such that

$$\sup_{m\ge N,\ g\in A} |y^*_m(g) - y^*(g)| \le \epsilon/3.$$

Let $f \in K$. In view of (4.22), we obtain $g \in A$ such that

$$\|f - g\| \le \epsilon/3.$$

Using the fact that $\|y^*_m\|^* \le 1$, $\|y^*\|^* \le 1$, we obtain for $m \ge N$,

$$|y^*_m(f) - y^*(f)| \le |y^*_m(f) - y^*_m(g)| + |y^*_m(g) - y^*(g)| + |y^*(g) - y^*(f)| \le \|f - g\| + \epsilon/3 + \|f - g\| \le \epsilon.$$

This completes the proof. □

Proof of Theorem 2.3. We apply Theorem 7.1 with the following choices. The space $\mathbb{X}$ is defined as follows. We consider the space of all functions $f \in L^p$ for which $L_k(f)$ is well defined, with the norm

$$\|f\| = \sum_{k=1}^{R} \|L_k(f)\|_p.$$

Then $\mathbb{X}$ is the closure of the space of all diffusion polynomials in the sense of this norm. The functionals are defined by

$$x^*_{k,n}(f) = d_k L_k(f)(x_n), \quad x^*_k(f) = c_k L_k(f)(x_0),$$

where the constants are chosen to bring the linear functionals into the unit ball of $\mathbb{X}^*$. These will depend only on L_k and $\mathbb{X}$, not on the individual points x_n. □

8. Numerical simulations

In this section we present numerical simulations of Birkhoff interpolation. By noting the one-to-one correspondence between even trigonometric polynomials and algebraic polynomials of the same degree, we focus on interpolation on $[-1, 1]$ and $[-1, 1]^2$. To show that this method works for a general basis, we also include examples of interpolation on the sphere $\mathbb{S}^2$ using real spherical harmonics. Examples of both 1D and 2D interpolation are included, but we emphasize 2D interpolation due to its challenging nature. In two dimensions, we restrict ourselves to interpolating on subsets of $[-1, 1]^2$ and $\mathbb{S}^2$. In particular, we look at interpolating the function (see Fig. 1)

$$\begin{aligned} f_R(x, y) = {} & \frac{1}{1 + R(x^2 + y - 0.3)^2} + \frac{1}{1 + R(x + y - 0.4)^2} + \frac{1}{1 + R(x + y^2 - 0.5)^2} \\ & + \frac{1}{1 + R(x^2 + y^2 - 0.25)^2}, \end{aligned} \tag{8.1}$$

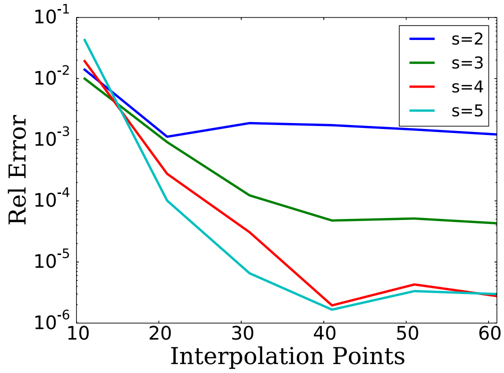

Fig. 2. Interpolation error of $f_{25}(x, -0.96)$ using MSN 1D Birkhoff interpolation for various s values. Function and derivative values are given at n equally spaced points. Error has been normalized by maximum function value. **Single precision**. (For interpretation of the colors in the figures, the reader is referred to the web version of this article.)

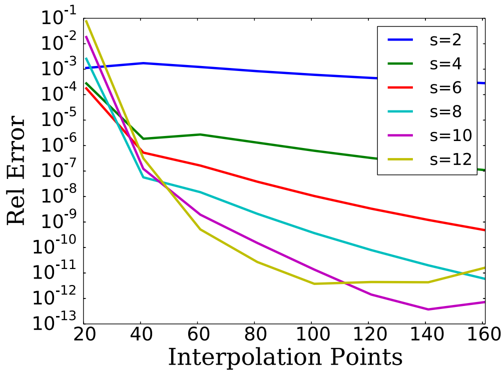

Fig. 3. Interpolation error of $f_{25}(x, -0.96)$ using MSN 1D Birkhoff interpolation for various s values. Function and derivative values are given at n equally spaced points. Error has been normalized by maximum function value.

with Runge functions on two parabolas, one line, and one circle in $[-1, 1]^2$. Here, the R parameter controls the difficulty of the interpolation, with larger R values corresponding to more difficult interpolation problems. Most of the tests use $R = 25$, although some examples (Figs. 17 and 18) use $R = 9$ in order to show we can obtain small errors in 2D interpolation: error on the order of 10^{-8} using double precision. In one dimension, we interpolate $f_{25}(x, -0.96)$. On the sphere, we look at interpolating the function

$$g(x, y, z) = \frac{1}{1 + (\cos 7x + \cos 7y + \cos 7z)^2}, \tag{8.2}$$

where we restrict g to $\mathbb{S}^2$. Unless stated otherwise, the examples are computed using double precision and the derivative information consists of directional derivatives along the coordinate axes. There does not appear to be any standard software packages for Birkhoff interpolation when the point distribution is arbitrary, so we are not able to compare our method with others.

For $p = 2$, MSN Birkhoff interpolation reduces to solving

$$\min_{Va=f} ||D_s a||_2, \tag{8.3}$$

where D_s is a diagonal positive-definite matrix with condition number $O(n^s)$, V is a Chebyshev–Vandermonde matrix, a is the vector containing the Chebyshev interpolation coefficients, and f is the vector containing the function and derivative values. Here, s takes the role of β in the previous discussion. The D_s matrix allows us to control the sth derivative of our approximation. Solving this linear system takes great care and we present details below.

In the one dimensional case, we interpolate $f_{25}(x, -0.96)$ on $[-1, 1]$ using both single precision and double precision on equally-spaced points. Results for function and derivative information at equally spaced points are presented in Figs. 2 and 3. We obtain similar results in Figs. 4 and 5 when we interlace the function and derivative information. In the 2D interpolation below, the point $(-0.96, -0.96)$ was the point with the largest error on the 21×21 tensor grid for $s = 8$, and this is why we chose to the 1D function to be $f_{25}(x, -0.96)$. In both single and double precision the interpolation error

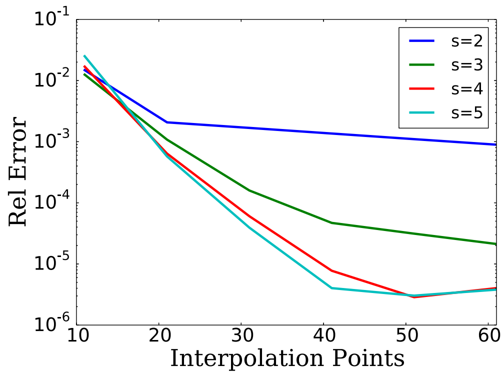

Fig. 4. Interpolation error of $f_{25}(x, -0.96)$ using MSN 1D Birkhoff interpolation for various s values. Function values are given at n equally spaced points while derivative values are given at $n-1$ points in between. Error has been normalized by maximum function value. **Single precision**.

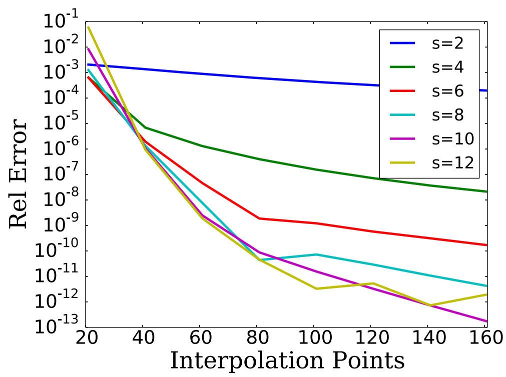

Fig. 5. Interpolation error of $f_{25}(x, -0.96)$ using MSN 1D Birkhoff interpolation for various s values. Function values are given at n equally spaced points while derivative values are given at $n-1$ points in between. Error has been normalized by maximum function value.

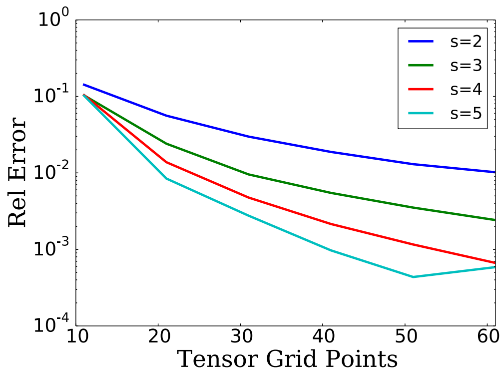

Fig. 6. Interpolation error of $f_{25}(x, y)$ using MSN 2D Birkhoff interpolation for various s values. Function and derivative values are given on $n \times n$ tensor grid. Error has been normalized by the maximum function value. **Single precision**.

approaches machine epsilon with an appropriately chosen s-value and increasing points. The error is computed by taking the difference between the function and the approximation on $10n$ equally spaced points, normalized by the maximum function value. The mesh norm is

$$m = \left\lceil \frac{2\pi}{\min_{i \neq j} \left|\cos^{-1}(x_i) - \cos^{-1}(x_j)\right|} \right\rceil, \tag{8.4}$$

and the interpolation order is taken to be m.

We showcase the true power of MSN Birkhoff interpolation by using a variety of point distributions in two dimensions. In particular, the interpolation of $f_{25}(x, y)$ is performed on a tensor grid (Figs. 6 and 7), on a tensor grid intersected with

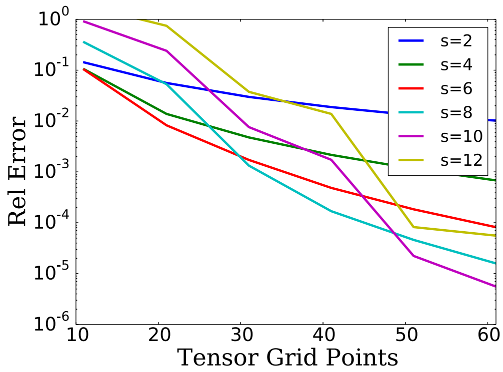

Fig. 7. Interpolation error of $f_{25}(x, y)$ using MSN 2D Birkhoff interpolation for various s values. Function and derivative values are given on $n \times n$ tensor grid. Error has been normalized by the maximum function value.

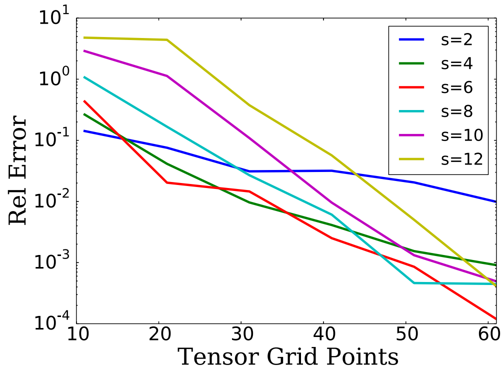

Fig. 8. Interpolation error of $f_{25}(x, y)$ using MSN 2D Birkhoff interpolation for various s values. Function and derivative values are given on $n \times n$ tensor grid that has been intersected with an annulus (inner radius 0.5 and outer radius 1). Error has been normalized by the maximum function value.

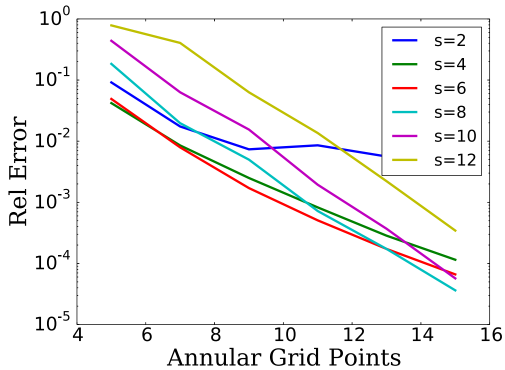

Fig. 9. Interpolation error of $f_{25}(x, y)$ using MSN 2D Birkhoff interpolation for various s values. Function and derivative values are given on an $(n, 8n)$ annular grid: inner radius 0.5, outer radius 1, n equally spaced points in the radial direction, and $8n$ equally spaced points in the angular direction. Error has been normalized by the maximum function value.

an annulus (Fig. 8), and on an annular grid (Fig. 9). The maximum error is computed by taking the maximum difference on a $10n \times 10n$ grid for an $n \times n$ interpolation grid. When we interpolate with function and derivative values interlaced, we also see convergence, whether the directional derivatives are parallel to the coordinate axes (Figs. 10 and 11) or not (Figs. 12 and 13). Furthermore, MSN Birkhoff interpolation can use a variety of bases. Figs. 14 and Fig. 15 show results for interpolation on $\mathbb{S}^2$ using real spherical harmonics. Additionally, we present results in Fig. 16 for when points are clustered near the north and south poles, requiring interpolation points to have polar angle $\theta \in [0, \frac{\pi}{3}] \cup [\frac{2\pi}{3}, \pi]$. We see a general trend of decreasing error in these regions. By interpolating $f_9(x, y)$ on tensor grids on $[-1, 1]^2$ (Figs. 17 and 18), we see that for easier 2D problems we still find that the error decreases as we increase the interpolation points. The interpolation order for two-dimensional problems is computed similar to the one-dimensional case. An archived version of this paper includes

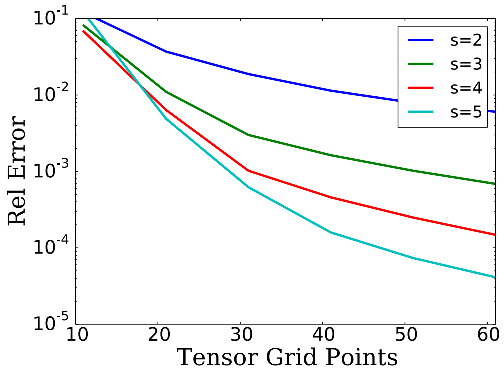

Fig. 10. Interpolation error of $f_{25}(x, y)$ using MSN 2D Birkhoff interpolation for various s values. Function values are given at $n \times n$ tensor grid while derivative values are given at $(n-1) \times (n-1)$ tensor grid in between. Error has been normalized by maximum function value. **Single precision**.

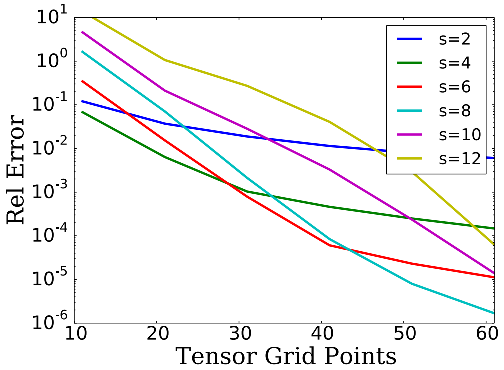

Fig. 11. Interpolation error of $f_{25}(x, y)$ using MSN 2D Birkhoff interpolation for various s values. Function values are given at $n \times n$ tensor grid, while derivative values are given at $(n-1) \times (n-1)$ tensor grid in between. Error has been normalized by maximum function value.

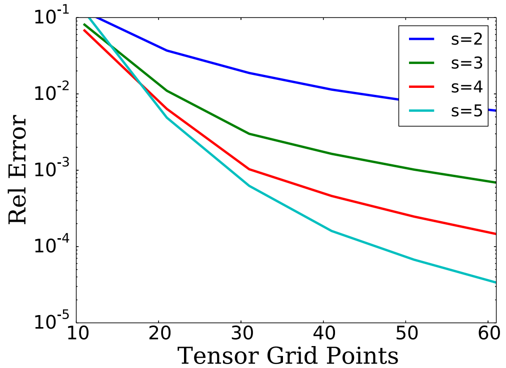

Fig. 12. Interpolation error of $f_{25}(x, y)$ using MSN 2D Birkhoff interpolation for various s values. Function values are given at $n \times n$ tensor grid, while derivative values are given at $(n-1) \times (n-1)$ tensor grid in between. **Directional derivatives are not parallel to the coordinate axes** and are in the directions $\left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)$ and $\left(\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\right)$. Error has been normalized by maximum function value. **Single precision**.

the same data presented here in table form to allow for other people to easily compare their results with ours [32]. The difficulty in obtaining low interpolation error is due to the long run time required for the dense matrix computations. Future work will be devoted to investigating and implementing fast algorithms arising from the structured equations.

Larger s values give greater derivative control and, by looking at the error plots, generally more accurate results; however, large s values lead to high condition numbers, so care must be taken to arrive at an accurate solution. We use a variant of the complete orthogonal decomposition of [33] because the ill-conditioning in our method results primarily from a diagonal matrix. From Eq. (8.3), this is equivalent to solving

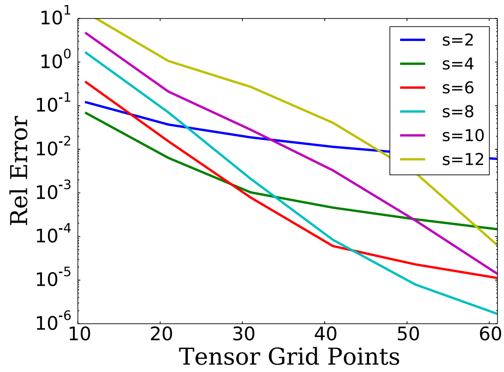

Fig. 13. Interpolation error of $f_{25}(x, y)$ using MSN 2D Birkhoff interpolation for various s values. Function values are given at $n \times n$ tensor grid, while derivative values are given at $(n-1) \times (n-1)$ tensor grid in between. **Directional derivatives are not parallel to the coordinate axes** and are in the directions $\left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)$ and $\left(\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\right)$. Error has been normalized by maximum function value.

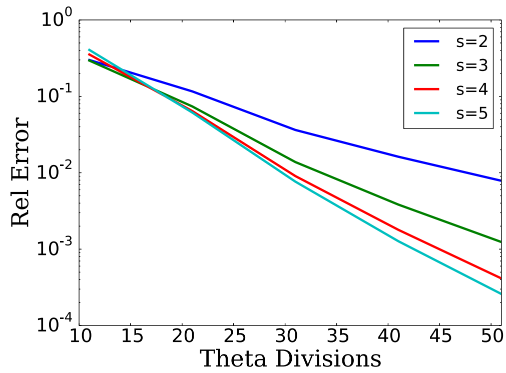

Fig. 14. Interpolation error of $g(x, y, z)$ using MSN 2D Birkhoff interpolation for various s values on $\mathbb{S}^2$. Function and derivative values are given on a scattered grid with minimum separation approximately π/d, where d is the Theta Divisions. Error has been normalized by the maximum function value. **Single precision**.

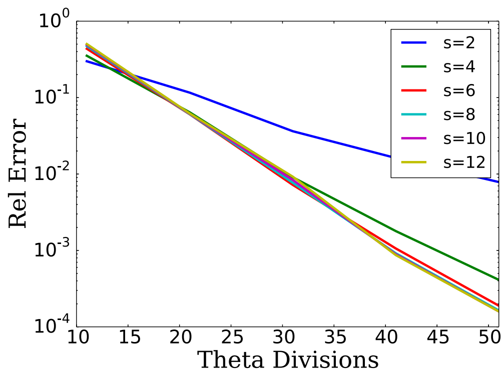

Fig. 15. Interpolation error of $g(x, yz)$ using MSN 2D Birkhoff interpolation for various s values on $\mathbb{S}^2$. Function and derivative values are given on a scattered grid with minimum separation approximately π/d, where d is the Theta Divisions. Error has been normalized by the maximum function value.

$$\min_{V D_s^{-1} x = f} ||x||_2 \tag{8.5}$$

and setting $a = D_s^{-1} x$. Thus, the main computational cost is an LQ factorization of $V D_s^{-1}$. Large s values may render standard pivoted LQ algorithms (especially those based on QR Factorization with Column Pivoting) useless. Thus, the best algorithm may be to implement a Rank-Revealing LQ factorization based on [34]; we implement a poor man's version that is sufficient for our work here.

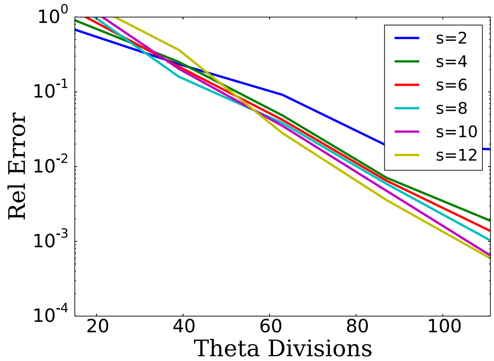

Fig. 16. Interpolation error of $g(x, y, z)$ using MSN 2D Birkhoff interpolation for various s values on $\mathbb{S}^2$. Function and derivative values are given on a scattered grid with minimum separation approximately π/d, where d is the Theta Divisions, with the restriction that the polar angle $\theta \in [0, \frac{\pi}{3}] \cup [\frac{2\pi}{3}, \pi]$. Error has been normalized by the maximum function value.

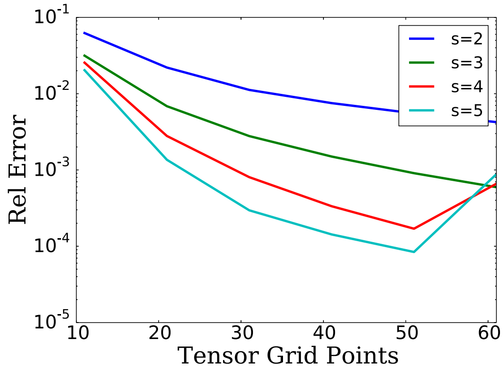

Fig. 17. Interpolation error of $f_9(x, y)$ using MSN 2D Birkhoff interpolation for various s values. Function and derivative values are given on $n \times n$ tensor grid. Error has been normalized by the maximum function value. **Single precision**.

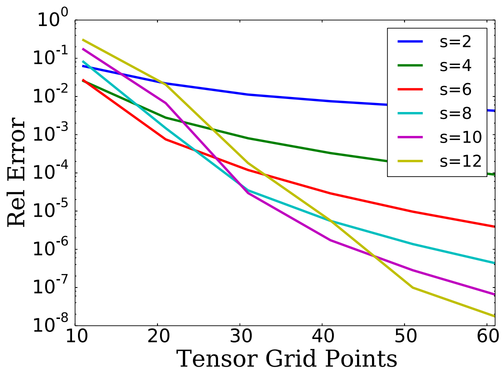

Fig. 18. Interpolation error of $f_9(x, y)$ using MSN 2D Birkhoff interpolation for various s values. Function and derivative values are given on $n \times n$ tensor grid. Error has been normalized by the maximum function value.

Specifically, we use the following algorithm:

1. Compute $V = P_1 L_1 Q_1$ using an LQ factorization based on QRCP.
2. Determine permutation Π such that $Q_1 D_s^{-1} \Pi$ has decreasing column norms.
3. Compute the singular value decomposition: $Q_1 D_s^{-1} \Pi = U \Sigma V^*$; only U is used.
4. Compute $U^* Q_1 D_s^{-1} \Pi = P_2 L_2 Q_2$.
5. Solve $L_1 z = P_1^* f$.
6. Solve $L_2 y = P_2^* U^* z$.
7. Set $a = D_s^{-1} \Pi Q_2^* y$.

The main cost of the algorithm is 2 pivoted LQ factorizations and 1 SVD, and this algorithm is used for both 1D and 2D problems. We now look at the important feature: in Step 3, we compute the SVD of $Q_1 D_s^{-1} \Pi$, so that $U^* Q_1 D_s^{-1} \Pi \approx \Sigma V^*$ to machine precision. This ensures that our LQ factorization in Step 4 is accurate. In essence, U is a preconditioner for numerical stability. This algorithm appears to give results independent of s, so our results depend only on the condition number in V, not $V D_s^{-1}$, and is similar to what was used in [12,19]. Now, the underlying condition number of V does appear to be the limiting factor in the accuracy of the approximations; the 1D results (Figs. 2–5) seem to show this as well as Figs. 6 and 17 in the 2D case, both of which use single precision and which appear to have the "U"-error shape that is expected when rounding errors start to dominate computations.

We have not shown this method to be numerically stable; however, all of the numerical examples are solved using the same algorithm in both single and double precision. The one-dimensional data indicate that we approach machine epsilon in both single and double precision, while the two-dimensional data show that increasing data points generally leads to decreased error.

9. Conclusions

We proved a general Birkhoff interpolation result: with minimal restrictions it is possible to interpolate function and derivative information with diffusion polynomials of degree $N^* \sim \eta^{-1}$, where η is the minimum separation distance between points. This extends previous work related to function interpolation [12] and is needed for numerical approximations arising in the solution of partial differential equations [19]. One and two dimensional numerical examples were presented to demonstrate the abilities of this method with indifference to interpolation location.

References

[1] I. Natanson, Constructive Function Theory, vol. 3, Ungar, New York, 1965, Google Scholar, 1982.
[2] A. Zygmund, Trigonometric Series, vol. 1, Cambridge University Press, 2002.
[3] G.D. Birkhoff, General mean value and remainder theorems with applications to mechanical differentiation and quadrature, Trans. Am. Math. Soc. 7 (1906) 107–136.
[4] G.G. Lorentz, K. Jetter, S.D. Riemenschneider, Birkhoff Interpolation, Addison–Wesley, Reading, Massachusetts, 1983.
[5] R.A. Lorentz, Multivariate Birkhoff Interpolation, Springer, 1992.
[6] R.A. Lorentz, Multivariate Hermite interpolation by algebraic polynomials: a survey, J. Comput. Appl. Math. 122 (2000).
[7] J. Chai, N. Lei, Y. Li, P. Xia, The proper interpolation space for multivariate Birkhoff interpolation, J. Comput. Appl. Math. 235 (2011) 3207–3214.
[8] N. Lei, J. Chai, P. Xia, Y. Li, A fast algorithm for the multivariate Birkhoff interpolation problem, J. Comput. Appl. Math. 236 (2011) 1656–1666.
[9] K. Cui, N. Lei, Stable monomial basis for multivariate Birkhoff interpolation problems, J. Comput. Appl. Math. 277 (2015) 162–170.
[10] X. Wang, S. Zhang, T. Dong, Newton basis for multivariate Birkhoff interpolation, J. Comput. Appl. Math. 228 (2009) 466–479.
[11] G. Allasia, C. Bracco, Multivariate Hermite–Birkhoff interpolation by a class of cardinal basis functions, Appl. Math. Comput. 218 (2012) 9248–9260.
[12] S. Chandrasekaran, K.R. Jayaraman, H.N. Mhaskar, Minimum Sobolev norm interpolation with trigonometric polynomials on the torus, J. Comput. Phys. 249 (2013) 96–112.
[13] J. Szabados, P. Vértesi, Interpolation of Functions, World Scientific, 1990.
[14] H.N. Mhaskar, F.J. Narcowich, N. Sivakumar, J.D. Ward, Approximation with interpolatory constraints, Proc. Am. Math. Soc. 130 (2002) 1355–1364.
[15] H. Wendland, Scattered Data Approximation, Cambridge Monographs on Applied and Computational Mathematics, Cambridge University Press, 2004, https://books.google.com/books?id=qy4cbWUmSyYC.
[16] Q.T. Le Gia, H.N. Mhaskar, Polynomial operators and local approximation of solutions of pseudo-differential equations on the sphere, Numer. Math. 103 (2006) 299–322.
[17] Q.T. Le Gia, H.N. Mhaskar, Localized linear polynomial operators and quadrature formulas on the sphere, SIAM J. Numer. Anal. (2008) 440–466.
[18] A. Townsend, H. Wilber, G.B. Wright, Computing with functions in spherical and polar geometries I. The sphere, SIAM J. Sci. Comput. 38 (2016) C403–C425.
[19] S. Chandrasekaran, H.N. Mhaskar, A minimum Sobolev norm technique for the numerical discretization of PDEs, J. Comput. Phys. 299 (2015) 649–666.
[20] M. Maggioni, H.N. Mhaskar, Diffusion polynomial frames on metric measure spaces, Appl. Comput. Harmon. Anal. 24 (2008) 329–353.
[21] H.N. Mhaskar, Eignets for function approximation on manifolds, Appl. Comput. Harmon. Anal. 29 (2010) 63–87.
[22] E.B. Davies, L^p spectral theory of higher-order elliptic differential operators, Bull. Lond. Math. Soc. 29 (1997) 513–546.
[23] A. Grigor'yan, Estimates of Heat Kernels on Riemannian Manifolds, Lond. Math. Soc. Lect. Note Ser., vol. 273, 1999, pp. 140–225.
[24] A. Grigor'yan, Heat kernels on weighted manifolds and applications, Contemp. Math. 398 (2006) 93–191.
[25] A. Grigorlyan, Heat kernels on metric measure spaces with regular volume growth, in: Handbook of Geometric Analysis, vol. 2, 2010.
[26] Y.A. Kordyukov, L^p-theory of elliptic differential operators on manifolds of bounded geometry, Acta Appl. Math. 23 (1991) 223–260.
[27] H.N. Mhaskar, On bounded interpolatory and quasi-interpolatory polynomial operators, Front. Interpol. Approx. (2006) 345.
[28] F. Filbir, H.N. Mhaskar, A quadrature formula for diffusion polynomials corresponding to a generalized heat kernel, J. Fourier Anal. Appl. 16 (2010) 629–657.
[29] F. Filbir, H.N. Mhaskar, Marcinkiewicz–Zygmund measures on manifolds, J. Complex. 27 (2011) 568–596.
[30] H.N. Mhaskar, A unified framework for harmonic analysis of functions on directed graphs and changing data, Appl. Comput. Harmon. Anal. 44 (2018) 611–644.
[31] J. Bergh, J. Löfström, Interpolation Spaces: An Introduction, vol. 223, Springer, Berlin, 1976.
[32] S. Chandrasekaran, C.H. Gorman, H.N. Mhaskar, Minimum Sobolev norm interpolation of derivative data, arXiv e-prints, 2017.
[33] P. Hough, S.A. Vavasis, Complete orthogonal decomposition for weighted least squares, SIAM J. Matrix Anal. Appl. 18 (1995) 369–392.
[34] M. Gu, S.C. Eisenstat, Efficient algorithms for computing a strong rank-revealing qr factorization, SIAM J. Sci. Comput. 17 (1996) 848–869.