



Regular families of kernels for nonlinear approximation

Keaton Hamm ^{a,*}, Jeff Ledford ^b



^a Department of Mathematics, University of Arizona, Tucson, AZ, 85721, United States of America
^b Department of Mathematics and Computer Science, Longwood University, Farmville, VA, 23901, United States of America

ARTICLE INFO

Article history:

Received 26 April 2018
 Available online 12 March 2019
 Submitted by S. Tikhonov

Keywords:

Nonlinear approximation
 General multiquadratics
 Radial basis functions
 Triebel–Lizorkin space
 Sobolev space

ABSTRACT

This article studies sufficient conditions on families of approximating kernels which provide N -term approximation errors from an associated nonlinear approximation space which match the best known orders of N -term wavelet expansion. These conditions provide a framework which encompasses some notable approximation kernels including splines, cardinal functions, and many radial basis functions such as the Gaussians and general multiquadratics. Examples of such kernels are given to justify the criteria. Additionally, the techniques involved allow for some new results on N -term Greedy interpolation of Sobolev functions via radial basis functions.

© 2019 Elsevier Inc. All rights reserved.

1. Introduction

Among the many notions of smoothness spaces of functions on \mathbb{R}^d are spaces of fractional smoothness which may be defined by, or intimately related to, a wavelet expansion. The primary examples to be considered here are the *Besov* and *Triebel–Lizorkin* spaces. For an overview of the basic facts about these spaces, the reader is invited to consult [30] and the many references therein. DeVore, Jawerth, and Popov [16] studied approximation orders for best N -term approximation of functions from these spaces with respect to a given wavelet system; in particular, it was shown that if f is in the Triebel–Lizorkin space $F_{\tau,q}^s$ (where $s \in \mathbb{R}_+$ is the smoothness) then the error of the best N -term approximation in L_p (for $p = (1/\tau - s/d)^{-1}$) via many wavelet systems is $O(N^{-s/d})$.

This article studies N -term approximation of functions in these smoothness spaces from nonlinear spaces associated with different kernels, which are typically related to a spline system or a *radial basis function*, i.e. one for which $\phi(x) = \psi(|x|)$ for a univariate function ψ .

More specifically, suppose we have a family of continuous kernels depending upon some parameter, $(\phi_\alpha)_{\alpha \in A}$ for some unbounded $A \subset (0, \infty)$. Considering the following nonlinear approximation space:

* Corresponding author.

E-mail addresses: hamm@math.arizona.edu (K. Hamm), ledfordjp@longwood.edu (J. Ledford).

$$\Phi_N := \left\{ \sum_{j=1}^N a_j \phi_{\alpha_j}(\cdot - x_j) : (a_j) \subset \mathbb{C}, (\alpha_j) \subset A, (x_j) \subset \mathbb{R}^d \right\}, \quad (1)$$

we seek to answer the following problem.

Problem 1.1. Find sufficient conditions on the family (ϕ_α) such that for $f \in F_{\tau,q}^s$, there exists an $S_{f,N} \in \Phi_N$ such that

$$\|f - S_{f,N}\|_{L_p} \leq C_f N^{-s/d}.$$

Similar problems have appeared in various contexts throughout the literature. In the shift invariant setting, de Boor, DeVore, and Ron investigated similar problems, [6,7]. The methodology used in those papers exploits features of the periodization of $|\hat{\phi}|^2$ to provide approximation results analogous the one above. In [31], Kyriazis and Petrushev develop sufficient conditions for unconditional bases in Triebel–Lizorkin spaces involving moment conditions from which an answer to their version of the above problem arose. Interestingly, although their techniques are different, there is overlap in the examples that they produce. In [17], DeVore and Ron study approximation from both linear and nonlinear spaces using kernels that arise as the solution of certain elliptic differential operators. This technique encompasses some growing kernels, but does not allow the use of some notable examples, such as multiquadratics. Hangelbroek and Ron [26] analyzed best N -term approximations via Gaussians. The analyticity and rapid decay of the Gaussian kernel allows for very fast approximation in theory; however, their analysis does not directly extend to allow kernels of finite smoothness, and certainly not growing kernels. Additionally, Buhmann and Dai [12] carried out a related analysis for quasi-interpolation schemes based around polynomial reproduction. Our approximation scheme (detailed in Section 3) is similar to theirs in that it is essentially one of quasi-interpolation, but the kernel they use is formed indirectly from the initial kernel to guarantee polynomial reproduction, and so is of a different form. Our main results give sufficient conditions on finite-smoothness kernels (both decaying and growing ones) which allow for the desired approximation orders using properties of the kernel. Some notable implications of the current analysis are that other growing kernels such as the *multiquadratics* of all orders may be used, and also our methods allow for the use of *cardinal functions*, which gives rise to some new nonlinear methods involving interpolation. The use of cardinal functions in turn allows for error estimates for Sobolev spaces, which cannot be achieved directly from the Besov and Triebel–Lizorkin space estimates.

The primary concern of this work is to demonstrate that the approximation orders of the nonlinear spaces Φ_N associated with a large variety of kernels match the known rates for nonlinear wavelet systems, which are determined by the smoothness of the target functions. However, in the course of the proof, we exhibit a concrete approximant to a given function f which attains the error bound of $O(N^{-s/d})$. As an intermediate step in the procedure, we use the fact that such $f \in F_{\tau,q}^s$ admits a wavelet expansion of a certain type, though the choice of the wavelet is not overly important, and we stress that the use of wavelets here is merely an intermediate step in the analysis.

Specifically, we employ a two-stage approximation process as follows:

Step 1: Form a linear approximation space which provides extremely good error bounds for approximating mother wavelets which are bandlimited Schwartz functions. Then truncate the approximant. If ψ is the wavelet, let $T_N \psi$ be its truncated approximant from the linear space, which is also an N -term approximant in Φ_N .

Step 2: Given a wavelet expansion $f = \sum_I f_I \psi_I$, where I are dyadic cubes, and ψ is as in Step 1, approximate $S_{f,N} = \sum_I f_I T_{N_I} \psi_I$, for some cost distribution (N_I) with $\sum_I N_I \leq N$.

For a particular cost distribution (i.e. choice of (N_I) based on the wavelet coefficients (f_I)), this $S_{f,N}$ obtains the desired error bounds for the given f . In general terms, we now state the main theorems contained below (Theorems 4.5 and 7.6):

Theorem. Suppose that $(\phi_\alpha)_{\alpha \in A}$ is a family of kernels satisfying conditions (A1)–(A6) or (B1)–(B4) below. Then for the appropriate choices of parameters s, τ, q , and p , every $f \in F_{\tau,q}^s$ has an N -term approximant $S_{f,N} \in \Phi_N$ such that

$$\|f - S_{f,N}\|_{L_p} \leq CN^{-s/d} |f|_{F_{\tau,q}^s}.$$

As a final note, let us point out that the approximation spaces of the form Φ_N considered here are both nonlinear as mentioned above, but also nonuniform in the sense that no structure is required *a priori* on the shifts (x_j) – although the particular method illustrated in the sequel will use regular shifts at different levels. This is in contrast to the shift-invariant space literature, in which the approximation spaces are of the form $\overline{\text{span}}^{L_p}\{\phi(\cdot - j) : j \in \mathbb{Z}^d\}$, though the resulting approximation bounds are similar. For approximation and structural results of these spaces, consult [6, 7, 9, 27–29]. Typical results therein apply to approximating (or interpolating) Sobolev classes, and provide approximation rates of order h^k , where k is the specified smoothness of the Sobolev space. Approximation spaces which are nonuniform but linear, i.e. of the form $\overline{\text{span}}^{L_p}\{\phi(\cdot - x_j) : j \in \mathbb{Z}^d\}$, have been studied for quite some time as well. In recent works, these have been called quasi shift-invariant spaces, e.g. [3, 15, 20, 24]. For a more extensive history, Wendland's text on scattered-data interpolation is an excellent reference [45]. For approximation orders for interpolation of Sobolev functions similar to the theorem above, consult [21, 32].

The rest of the paper is laid out as follows: Section 2 begins with some basic definitions, followed by the conditions on decaying kernels for rapid approximation of bandlimited Schwartz functions in Section 3. Approximation orders using nonlinear approximation spaces associated with these kernels are given in Sections 4 and 5, and examples of kernels satisfying the regularity conditions are given in Section 6. Section 7 provides sufficient conditions on growing kernels to allow for similar approximation rates, and examples of such families of kernels are given in Section 8. In Section 9, we move the discussion to the related topic of cardinal interpolation, which allows us to give approximation orders for broader classes of smooth functions, namely Sobolev spaces. This analysis also develops some Greedy interpolation schemes related to Gaussians and multiquadratics. To give a better idea of the limitations, a special case of the cost distribution for the approximation scheme is discussed in more detail in Section 10.

2. Preliminaries

Let \mathcal{S} be the space of Schwartz functions on \mathbb{R}^d , that is the collection of infinitely differentiable functions ϕ such that for all multi-indices α and β , $\sup_{x \in \mathbb{R}^d} |x^\alpha D^\beta \phi(x)| < \infty$. The Fourier transform and inverse Fourier transform of a Schwartz function ϕ are given, respectively, by

$$\widehat{\phi}(\xi) := \int_{\mathbb{R}^d} \phi(x) e^{-i\langle \xi, x \rangle} dx, \quad \xi \in \mathbb{R}^d; \quad \phi^\vee(x) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \phi(\xi) e^{i\langle x, \xi \rangle} d\xi, \quad x \in \mathbb{R}^d,$$

where $\langle \cdot, \cdot \rangle$ is the usual scalar product on \mathbb{R}^d . Let $\Omega \subset \mathbb{R}^d$ be an open set. Then let $L_p(\Omega)$, $1 \leq p \leq \infty$, be the usual space of p -integrable functions on Ω with its usual norm. If no set is specified, we mean $L_p(\mathbb{R}^d)$. Additionally, we will use \mathbb{N}_0 to denote the set of natural numbers including 0, and \mathbb{R}_+ to denote the positive real numbers, i.e. $(0, \infty)$.

We denote by $F_{\tau,q}^s$ the Triebel–Lizorkin space (defined in Section 4, where $s \in \mathbb{R}_+$ essentially determines the smoothness of the functions in the space). For $1 \leq p \leq \infty$ and $k \in \mathbb{N}$, let $W_p^k(\Omega)$ be the Sobolev space of functions in $L_p(\Omega)$ whose weak derivatives of order up to k are also in $L_p(\Omega)$. The norm and seminorm on the Sobolev space may be defined, respectively, via

$$\|f\|_{W_p^k(\Omega)} := \|f\|_{L_p(\Omega)} + |f|_{W_p^k(\Omega)}, \quad \text{and} \quad |f|_{W_p^k(\Omega)} = \max_{|\beta|=k} \|D^\beta f\|_{L_p(\Omega)}.$$

Let PW_S be the Paley–Wiener space of bandlimited L_2 functions whose Fourier transform is supported on S . If $B(0, R)$ is the Euclidean ball of radius R centered about the origin, define the space $\mathcal{S}_B := \mathcal{S} \cap PW_{B(0, R)}$; in the sequel, R will be determined from the context. For a set S , let χ_S be the function which takes value 1 on S and 0 elsewhere, and let $|S|$ denote the Lebesgue measure of the set S . We use C to denote constants, and where appropriate add subscripts to emphasize the parameters they may depend on.

3. Regularity criteria and approximation in \mathcal{S}_B

We begin with Step 1 described above, which is to give criteria on kernels which allows for rapid approximation of bandlimited Schwartz functions from a linear space involving the kernels. For now, we resolve our problem for decaying kernels, and turn to growing ones in later sections. Let A be an infinite subset of $(0, \infty)$, and $(\phi_\alpha)_{\alpha \in A}$ be a family of continuous kernels depending on the parameter α , and let Φ_N be as defined in (1).

Letting $R > 0$ be fixed, but arbitrary, and $B := B(0, R)$, we introduce the following regularity conditions and assume that they hold from here on unless otherwise noted.

- (A1) Φ_N is closed under translation and dilation.
- (A2) Each ϕ_α is continuous, and $\sup_{\alpha \in A} \|\phi_\alpha\|_{L_\infty} \leq C$.
- (A3) Let $A_{\alpha, h, j} := \frac{\widehat{\phi_\alpha}(\cdot + 2\pi j/h)}{\widehat{\phi_\alpha}}$. Suppose that $\sum_{j \neq 0} \|A_{\alpha, h, j}\|_{L_\infty(B)} \leq g_\alpha(h)$ for some function $g_\alpha(h)$, and that for every $k \in \mathbb{N}_0$, there exists an $\alpha_k \in A$ such that for every $\alpha \geq \alpha_k$, $g_\alpha(h) \leq Ch^k$ for some absolute constant C .
- (A4) For every $k \in \mathbb{N}$, there exists an $\alpha'_k \in A$ such that for every $\alpha \geq \alpha'_k$, $D^\gamma(1/\widehat{\phi_\alpha}) \in L_2(B)$ for every $|\gamma| \leq k$.
- (A5) For every $k \in \mathbb{N}$, there exists an $\alpha''_k \in A$ such that for every $\alpha \geq \alpha''_k$, $|\phi_\alpha(x)| = O(|x|^{-2k})$, $|x| \rightarrow \infty$.
- (A6) For every $\alpha \in A$, $|\phi_\alpha(x)| + |\widehat{\phi_\alpha}(x)| \leq C(1 + |x|)^{-d-\varepsilon}$ for some $C, \varepsilon > 0$.

The analysis in [26] is aided by the fact that the Gaussian has rapid decay – a requirement we drop in order to consider a family of general inverse multiquadratics which serves as the typical example of a collection of kernels which satisfies (A1)–(A6). These conditions allow us to provide norm and pointwise estimates for functions in \mathcal{S}_B . The requirement (A1) is to ensure the ease of approximation of the wavelet basis, while (A6) allows one to use the Poisson summation and Fourier inversion formulas (of course (A6) may be replaced by any other condition which allows these to hold). The utility of the remaining criteria will reveal itself more clearly in the sequel, but essentially (A2)–(A5) guarantee that after some value of the parameter α , the kernels are sufficiently localized, smooth, and decaying. Since the Triebel–Lizorkin spaces are defined by wavelet systems, these criteria guarantee that for any resolution level of the wavelet, the parameter may be chosen so that the kernel is sufficiently localized. There are other feasible ways to determine such criteria; for instance, various sets of sufficient conditions can be found in [12, 17, 26, 31], but our aim here is to give readily checked conditions based on the kernels themselves. Presently, we define an approximation suited to our purposes.

For $f \in \mathcal{S}_B$ and $\phi \in (\phi_\alpha)$, we first define f_ϕ to be the function such that $\widehat{f_\phi} = \widehat{f}/\widehat{\phi}$. The main obstruction is that f_ϕ need not be smooth. Nevertheless, we define

$$T_h^\sharp f(x) := \sum_{j \in \mathbb{Z}^d} f_\phi(hj) \phi(x - hj).$$

Proposition 3.1. *If $\alpha \geq \alpha_0$ and $h < \pi/R$, then for every $f \in \mathcal{S}_B$,*

$$\|f - h^d T_h^\sharp f\|_{L_\infty} \leq C \|\widehat{f}\|_{L_1} g_\alpha(h),$$

where $C > 0$ is independent of f and h .

Proof. By the inversion formula and the Poisson summation formula (both are justified by (A6)),

$$\begin{aligned} h^d T_h^\sharp f(x) &= \frac{h^d}{(2\pi)^d} \int_{\mathbb{R}^d} \widehat{f}(\xi) \left[\frac{1}{\widehat{\phi}(\xi)} \sum_{j \in \mathbb{Z}^d} \phi(x - hj) e^{i\langle \xi, hj \rangle} \right] d\xi \\ &= \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \widehat{f}(\xi) \left[\frac{e^{i\langle \xi, x \rangle}}{\widehat{\phi}(\xi)} \sum_{j \in \mathbb{Z}^d} \widehat{\phi}(\xi + 2\pi j/h) e^{i\langle x, 2\pi j/h \rangle} \right] d\xi. \end{aligned}$$

The $j = 0$ term is nothing but $f(x)$, hence

$$f(x) - h^d T_h^\sharp f(x) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \widehat{f}(\xi) \left[\frac{e^{i\langle \xi, x \rangle}}{\widehat{\phi}(\xi)} \sum_{j \neq 0} \widehat{\phi}(\xi + 2\pi j/h) e^{i\langle x, 2\pi j/h \rangle} \right] d\xi.$$

Recalling that $\text{supp}(\widehat{f}) \subset B$, we have the straightforward estimate

$$\|f - h^d T_h^\sharp f\|_{L_\infty} \leq \frac{1}{(2\pi)^d} \left[\sum_{j \neq 0} \|A_{\alpha, h, j}\|_{L_\infty(B)} \right] \|\widehat{f}\|_{L_1},$$

which is at most $C \|\widehat{f}\|_{L_1} g_\alpha(h)$ by (A3). \square

Note that by the restriction $\alpha \geq \alpha_k$ above, we mean that for any $\phi \in (\phi_\alpha)_{\alpha \geq \alpha_k}$, the conclusion of the proposition holds.

We move now to truncate our approximant to put it in the space Φ_N . Define $T_h^\flat f$ via

$$T_h^\flat f(x) := \sum_{j \in \mathbb{Z}^d \cap B_h} f_\phi(hj) \phi(x - hj),$$

where B_h is the ball of radius h^{-2} centered at the origin. This is a finite sum of $N \sim h^{-2d}$ terms, hence $T_h^\flat f \in \Phi_N$. The next pair of propositions show that this truncated approximant performs quite well both in L_∞ and pointwise.

Proposition 3.2. *Let $k \in \mathbb{N}$ and $f \in \mathcal{S}_B$. For $h < \pi/R$, there is a constant $C_{f,k} > 0$, independent of h , so that for all $\alpha \geq \max\{\alpha_k, \alpha'_{k+d}\}$,*

$$\|f - h^d T_h^\flat f\|_{L_\infty} \leq C_{f,k} h^k.$$

Proof. Note that the norm in question is at most $C \|\widehat{f}\|_{L_1} g_\alpha(h) + h^d \|T_h^\sharp f - T_h^\flat f\|_{L_\infty}$ by the triangle inequality and Proposition 3.1, but the first term is at most $C_f h^k$ by (A3) and the assumption $\alpha \geq \alpha_k$. Thus, it remains to show that $\|T_h^\sharp f - T_h^\flat f\|_{L_\infty} \leq Ch^{k-d}$. By (A2), we have

$$\|T_h^\sharp f - T_h^\flat f\|_{L_\infty} \leq C \sum_{|j| \geq h^{-2}} |f_\phi(hj)|. \quad (2)$$

Now if $\alpha \geq \alpha'_{k+d}$, then for all multi-indices $|\gamma| \leq k+d$, $D^\gamma(1/\hat{\phi}) \in L_2(B)$. Consequently, by standard Fourier transform techniques,

$$|f_\phi(x)| \leq \left(\sum_{|\gamma| \leq k+d} \|D^\gamma(\hat{f}/\hat{\phi})\|_{L_1} \right) (1+|x|)^{-k-d},$$

which is, by Leibniz's rule, the Cauchy–Schwarz inequality, the fact that $f \in \mathcal{S}$, and (A4), majorized by

$$\sum_{|\gamma| \leq k+d} \sum_{\beta \leq \gamma} \|D^\beta \hat{f}\|_{L_2} \left\| D^{\gamma-\beta} \left(\frac{1}{\hat{\phi}} \right) \right\|_{L_2} (1+|x|)^{-k-d} \leq C_{f,k} (1+|x|)^{-k-d}.$$

Thus the series on the right side of (2) is at most

$$C_{f,k} h^{-k-d} \sum_{|j| > h^{-2}} \frac{1}{|j|^{k+d}} \leq C_{f,k} h^{-k-d} \int_{h^{-2}}^{\infty} r^{d-1} r^{-k-d} dr = C_{f,k} h^{k-d},$$

concluding the proof. \square

Proposition 3.3. *Let $k \in \mathbb{N}$ and $f \in \mathcal{S}_B$. Then for all $\alpha \geq \max\{\alpha_{2k}, \alpha'_{2k+d}, \alpha''_k\}$ and sufficiently small h , there is a constant C , independent of h , such that*

$$|f(x) - h^d T_h^\flat f(x)| \leq C h^k (1+|x|)^{-k}.$$

Proof. First notice that if $|x| \leq 2/h$, then $(1+|x|)^{-k} \geq (h/3)^k$, and so the desired inequality arises from Proposition 3.2 (note this requires $\alpha \geq \max\{\alpha_{2k}, \alpha'_{2k+d}\}$).

If $|x| > 2/h$, $h^k (1+|x|)^{-k} \geq C|x/2|^{-2k}$, and so it suffices to show that $|f(x) - h^d T_h^\flat f(x)| \leq C(1+|x|)^{-2k}$ in this range. Note that $|f(x)| \leq C(1+|x|)^{-2k}$ since it is in \mathcal{S} . On the other hand,

$$h^d |T_h^\flat f(x)| \leq h^d \sum_{j \in \mathbb{Z}^d} |f_\phi(hj)| \sup_{|j| \leq h^{-2}} |\phi(x - hj)|.$$

Using the same bound as in the proof of Proposition 3.2, we see that the series above is bounded as long as $\alpha \geq \alpha'_{d+1}$. Moreover, in this case, the series is bounded by a constant times h to a positive power, in which case, we may say $h^d \sum_{j \in \mathbb{Z}^d} |f_\phi(hj)| \leq C$ for some constant independent of h . Next, notice that $|x - hj| \geq |x|/2$, and so if $\alpha \geq \alpha''_k$, (A5) implies that $|\phi(x - hj)| = O(|x - hj|^{-2k}) = O(|x|^{-2k})$ for sufficiently small h . Therefore, $|h^d T_h^\flat f(x)| \leq C(1+|x|)^{-2k}$ as required. \square

Now to make the dependence on N more explicit, we can replace each occurrence of h above with $N^{-1/(2d)}$ and rewrite $T_h^\flat f$ as

$$T_N f(x) := N^{-1/2} T_{N^{-1/(2d)}}^\flat f(x). \quad (3)$$

We denote by N_0 the smallest such N that satisfies the requirements of Proposition 3.3.

4. Approximation in L_p , $1 \leq p < \infty$

As suggested in Step 2 above, we use wavelets to bridge the gap between \mathcal{S}_B and L_p . In order to use the results of Section 3, we need a wavelet system whose generators are in \mathcal{S}_B where $B = B(0, R)$ contains the

support of the Fourier transform of the mother wavelet. Fortunately, such wavelet systems are well known; for example the Meyer wavelet (see [14,19,35]) forms one such system. So that the proofs extend easily to arbitrarily high dimensions (see especially Section 7), we restrict to considering multivariate wavelet systems which are tensor products of univariate ones. Since we use the same methodology as [26, Section 3], we only list the relevant details adapted to our set up. Using the notation we find there, $I = c(I) + [0, \ell(I)]^d$ is a cube with corner $c(I) \in \mathbb{R}^d$ and side length $\ell(I) > 0$. Encumbered upon the wavelet structure, we make use of dyadic cubes, i.e.

$$I \in \mathcal{D} := \{2^m(n + [0, 1]^d) : m \in \mathbb{Z}, n \in \mathbb{Z}^d\}.$$

Let \mathcal{D}_j be the subset of cubes with common edge-length 2^j . We will denote by ψ_I the natural affine change of variables:

$$\psi_I(x) := \psi((x - c(I))/\ell(I)).$$

If $\psi \in \mathcal{S}_B$, then since (A1) is satisfied, we have the following theorem as a direct result of Proposition 3.3.

Theorem 4.1. *Let $k \in \mathbb{N}$, $\psi \in \mathcal{S}_B$, $N > N_0$ and let $I \in \mathcal{D}$. There exists $\alpha_K \in A$ such that for all $\alpha \geq \alpha_K$, there is a constant C , independent of N , such that*

$$|\psi_I(x) - (T_N \psi)_I(x)| \leq CN^{-k/d} \left(1 + \frac{|x - c(I)|}{\ell(I)}\right)^{-2k}.$$

In particular, Theorem 4.1 says that to approximate any element of the wavelet system, it suffices to consider the approximant of the mother wavelet under the same affine change of variables.

We will assume throughout that we have a wavelet system Ψ formed from a mother wavelet ψ as described above. This allows us to use the results of the previous section because we have (for smooth enough f)

$$f = \sum_{I \in \mathcal{D}} f_I \psi_I.$$

To approximate f with N terms, we begin by assigning to each cube I a cost c_I , and subsequently a budget

$$N_I := \begin{cases} \lfloor c_I \rfloor, & \lfloor c_I \rfloor \geq N_0, \\ 0, & \text{otherwise,} \end{cases} \quad (4)$$

where N_0 is the same as in Theorem 4.1 and depends only on the wavelet system. Requiring that $\sum_I N_I \leq N$, we then simply approximate $f_I \psi_I$ with N_I terms, and set

$$S_{f,N} := \sum_{I \in \mathcal{D}} f_I T_{N_I} \psi_I \quad (5)$$

to be our N -term approximant in Φ_N which will exhibit the desired rate of convergence. The cost distribution depends on several auxiliary quantities, which we define below.

Definition 4.2. Given $s, q > 0$, define the maximal function $M_{s,q}f$ via

$$M_{s,q}f(x) := \left(\sum_{I \in \mathcal{D}} |I|^{-sq/d} |f_I|^q \chi_I(x) \right)^{1/q}.$$

For a fixed dyadic cube I , we define a partial maximal function by

$$M_{s,q,I}f(x) := \left(\sum_{I \subset I' \in \mathcal{D}} |I'|^{-sq/d} |f_{I'}|^q \chi_{I'}(x) \right)^{1/q}.$$

Now given $\tau, s, q > 0$, we define the *Triebel–Lizorkin space* $F_{\tau,q}^s$ via

$$F_{\tau,q}^s := \{f : |f|_{F_{\tau,q}^s} := \|M_{s,q}f\|_{L_\tau(\mathbb{R}^d)} < \infty\},$$

where $|f|_{F_{\tau,q}^s}$ is a quasi-seminorm.

For the reader unfamiliar with the Triebel–Lizorkin spaces, it should be mentioned that they may be viewed from many different equivalent vantage points. While the definition above is in terms of mixed sequence and function space norms of a maximal function, they may also be defined via a Littlewood–Paley decomposition, or additionally in terms of a wavelet system; for more details, consult Section 3 of [30]. Returning to the problem at hand, we are now in position to define the cost distribution which will yield the approximant giving optimal approximation orders. Note this cost distribution may also be found in [17,26].

Definition 4.3. Let $s > 0$ and $1 \leq p < \infty$. Define τ and q by $1/\tau := 1/p + s/d$ and $1/q := 1 + s/d$. Let $f \in F_{\tau,q}^s$, with the wavelet expansion $\sum_{I \in \mathcal{D}} f_I \psi_I$. The *cost of a dyadic cube* $I \in \mathcal{D}$ is

$$c_I := |f|_{F_{\tau,q}^s}^{-\tau} m_{s,q,I}^{\tau-q} |f_I|^q |I|^q N,$$

where $m_{s,q,I} := \sup_{x \in \mathbb{R}^d} M_{s,q,I}f(x)$.

As shown in [26], the sum of all costs is bounded by N . Given (N_I) , we can use Theorem 4.1 to see that we are estimating $f_I \psi_I$ with N_I translates of the kernel ϕ_α . Additionally, we can see that for $\alpha \geq \alpha_K$:

$$\begin{aligned} |\psi_I(x)| &\leq C_{k,d} \min\{1, N^{-k/d}\} \left(1 + \frac{\text{dist}(x, I)}{\ell(I)}\right)^{-2k} := R_I(x); \\ |R_I(x)| &\leq C_{k,d} \min\{1, c_I^{-k/d}\} \left(1 + \frac{\text{dist}(x, I)}{\ell(I)}\right)^{-2k}. \end{aligned}$$

The proof of Theorem 4.5 relies on the following lemma, whose proof may be found in [26].

Lemma 4.4. Let $1 \leq p < \infty$ and $\alpha \geq \alpha_K$, then

$$\left\| \sum_{I \in \mathcal{D}} |f_I| R_I \right\|_{L_p} \leq C_{k,d} \left\| \sum_{I \in \mathcal{D}} \min\left\{1, c_I^{-k/d}\right\} |f_I| \chi_I \right\|_{L_p}.$$

Theorem 4.5. Given $s > 0$ and $1 \leq p < \infty$, there exists a constant $C_{s,p,d} > 0$ such that for $f \in F_{\tau,q}^s$, with $1/\tau = 1/p + s/d$ and $1/q = 1 + s/d$, there is a function $S_{f,N} \in \Phi_N$ so that

$$\|f - S_{f,N}\|_{L_p} \leq C_{s,p,d} N^{-s/d} |f|_{F_{\tau,q}^s}.$$

In particular, we may take $S_{f,N}$ as in (5).

Proof. The proof is the same as the one given in [26, Theorem 9], *mutatis mudandis*. \square

While the proof of the main theorem is essentially the same as in [26], it depends on the propositions in Section 3 which, on account of conditions (A1)–(A6), allow for the use of kernels that have only finite smoothness.

5. Approximation in L_∞

It should be noted that the proof of Lemma 4.4 relies upon the Fefferman–Stein inequality, which is not valid in L_∞ . Thus, for L_∞ convergence phenomena, we restrict our attentions to Besov spaces, which are alternative (slightly smaller) smoothness spaces than the Triebel–Lizorkin spaces which do not rely on maximal functions in their definitions.

Here, we list the relevant results suited to our set up. It closely matches that of [26, Section 4], thus the proofs are omitted.

Definition 5.1. For $\tau = d/s \in (0, \infty)$ and $q \in (0, \infty)$, the *Besov space* $B_{\tau,p}^s$ is defined by the finiteness of the quasi-seminorm

$$|f|_{B_{\tau,q}^s} := \left\| k \mapsto \left(\sum_{I \in \mathcal{D}_k} |f_I|^\tau \right)^{1/\tau} \right\|_{\ell_q},$$

that is

$$B_{\tau,q}^s := \left\{ f \in L_\tau : |f|_{B_{\tau,q}^s} < \infty \right\}.$$

Lemma 5.2. Let $k > d$ and $\alpha > \alpha_K$. Suppose that $(a_I)_{I \in \mathcal{D}_j}$ is a finitely supported sequence of coefficients; then

$$\left\| \sum_{I \in \mathcal{D}_j} a_I \psi_I - \sum_{I \in \mathcal{D}_j} a_I [T_{N_I} \psi_I] \right\|_{L_\infty} \leq C_{k,d} \sup_{I \in \mathcal{D}_j} |a_I N_I^{-k/d}|.$$

Theorem 5.3. Given $s > 0$, there exists a constant $C_{s,d} > 0$ such that for $f \in B_{\tau,q}^s$, with $1/\tau = s/d$ and $1/q = 1 + s/d$, there is a function $S_{f,N} \in \Phi_N$ so that

$$\|f - S_{f,N}\|_{L_\infty} \leq C_{s,d} N^{-s/d} |f|_{B_{\tau,q}^s}.$$

For the cost distribution for the Besov space case that produces the function $S_{f,N}$, see [26, Section 4.1].

6. Examples

In this section, we provide numerous examples of kernels which satisfy the conditions prescribed above whose associated N -term approximation spaces Φ_N exhibit optimal rates for nonlinear approximation of smooth functions.

6.1. Gaussians

This example is due to Hangelbroek and Ron and can be found in [26]. It served as the inspiration for our analysis.

Here we will let ϕ_α be the Gaussian:

$$\phi_\alpha(x) := e^{-|x/\alpha|^2},$$

whose Fourier transform is given by

$$\widehat{\phi_\alpha}(\xi) = (\alpha\sqrt{\pi})^d \phi_{2/\alpha}(\xi).$$

We check conditions (A1)–(A6) below, which are straightforward since the Gaussian is nice.

Condition (A1) is clear from the definition of Φ_N and the fact that $\phi_\alpha(x/\beta) = \phi_{\alpha\beta}(x)$. The constant in (A2) can be taken to be $\phi_\alpha(0) = 1$. The quantity in (A3) is calculated in [26] as follows:

$$\sum_{j \neq 0} \left| \frac{\widehat{\phi_\alpha}(\xi + 2\pi j/h)}{\widehat{\phi_\alpha}(\xi)} \right| \leq C \widehat{\phi_\alpha}(2(\pi/h - R)) \leq C_{\alpha,R} e^{c/h^2},$$

where the constant $c > 0$ depends on α and R . We may take our parameter $\alpha_k = 0$.

Condition (A4) is checked easily by noting that $1/\widehat{\phi_\alpha} = e^{\frac{\alpha}{4|\cdot|^2}}$, which is in L_2^{loc} since it is continuous. As with (A3), we may take $\alpha'_k = 0$. Likewise, conditions (A5) and (A6) are both clearly satisfied (with $\alpha''_k = 0$) because of the exponential decay of the Gaussian.

6.2. Inverse multiquadratics

We will now consider the general (inverse) multiquadric,

$$\phi_{\alpha,c}(x) := (|x|^2 + c^2)^{-\alpha},$$

where $|\cdot|$ denotes the Euclidean distance on \mathbb{R}^d ; $c > 0$ is called the *shape parameter*, and we take $\alpha \in A = (d + 1/2, \infty)$. Consider

$$\Phi_N = \left\{ \sum_{j=1}^N a_j \phi_{\alpha_j, c_j}(\cdot - x_j) : (a_j) \subset \mathbb{C}, (\alpha_j) \subset A, (c_j) \subset \mathbb{R}^+, (x_j) \subset \mathbb{R}^d \right\}.$$

Since Φ_N is manufactured to be closed under translations, we check only dilation. If $\delta > 0$, we have $\phi_{\alpha,c}(\delta \cdot) = \delta \phi_{\alpha,c}/\delta$; hence Φ_N is dilation invariant, and (A1) is satisfied.

Condition (A2) is straightforward by noting that $\phi_{\alpha,c}$ is decreasing radially when $|x| \in (0, \infty)$; hence $|\phi_\alpha(x)| \leq c^{-2\alpha}$, which is bounded by c^{-2d+1} for all $\alpha \in A$. Condition (A3) requires the use of the Fourier transform of $\phi_{\alpha,c}$, which is given by (see, for example, [45, Theorem 8.15]):

$$\widehat{\phi_{\alpha,c}}(\xi) = (2\pi)^{\frac{d}{2}} \frac{2^{1-\alpha}}{\Gamma(\alpha)} \left(\frac{c}{|\xi|} \right)^{\frac{d}{2}-\alpha} K_{\frac{d}{2}-\alpha}(c|\xi|), \quad \xi \in \mathbb{R}^d \setminus \{0\}, \quad (6)$$

where

$$K_\nu(r) := \int_0^\infty e^{-r \cosh t} \cosh(\nu t) dt, \quad r > 0, \nu \in \mathbb{R}. \quad (7)$$

The function K_ν is called the modified Bessel function of the second kind (see [1, p. 376]). We note that $\phi_{\alpha,c}$, and consequently its Fourier transform, are radial functions (i.e. $\phi_{\alpha,c}(x) = \phi_{\alpha,c}(|x|)$). It is also clear

from (7) that K is symmetric in its order; that is, $K_{-\nu} = K_\nu$ for any $\nu \in \mathbb{R}$. Additionally, the decay of the Bessel function governs the decay of $\widehat{\phi_{\alpha,c}}$, which is exponential away from the origin.

Using the estimates from [23], we may estimate $A_{\alpha,h,j}$ as follows:

$$\begin{aligned} |A_{\alpha,h,j}(\xi)| &\leq \left(\frac{R}{|\xi + 2\pi j/h|} \right)^{d/2-\alpha} e^{cR-c|\xi+2\pi j/h|} \\ &\leq C_{\alpha,R} e^{2cR} e^{-\pi c|j|/h}, \end{aligned}$$

where we have estimated the polynomial term with an exponential. Thus the series obtained by summing over $j \neq 0$ is bounded by $C_{\alpha,R} e^{2cR} e^{-\pi c/h}$. Consequently, we may take $\alpha_k > d+1/2$ to satisfy the condition.

Estimates from [23] show that $1/\widehat{\phi_{\alpha,c}}$ is only $(2\alpha - d)$ -times differentiable near the origin, hence we must have $\alpha'_k > k/2 + d/2$. Additionally, each of these derivatives are bounded, hence $1/\widehat{\phi_{\alpha,c}} \in L_2^{loc}$ as desired. Condition (A5) is obvious by taking $\alpha''_k > 2k$. Finally, (A6) follows from the decay of $\widehat{\phi_{\alpha,c}}$ when $\alpha \in A$ and the exponential decay of $\widehat{\phi_{\alpha,c}}$.

As a final remark, our use of the multiquadratics here is not entirely standard in that, typically, a fixed multiquadric order α is used in a given approximation scheme; however, our method detailed above forces the use of multiquadratics with differing parameters α to achieve good approximation for the wavelet expansions in the Triebel–Lizorkin spaces. On the other hand, if so desired, one could specify a given α to be fixed in the approximation from Φ_N , but only *after* knowledge of the cost distribution for the target function f .

6.3. Matérn kernels

The Matérn kernels are simply the Fourier transform of the inverse multiquadratics, i.e. $\rho_{\alpha,c}(x) := \widehat{\phi_{\alpha,c}}(x)$, where $\widehat{\phi_{\alpha,c}}$ is given by (6). Consequently, $\widehat{\rho_{\alpha,c}}(\xi) = (|\xi|^2 + c^2)^{-\alpha}$. The approximation space associated with these kernels is closed under dilations since $\rho_{\alpha,c}(\delta x) = \delta^{2\alpha-d} \rho_{\alpha,c\delta}(x)$. It is a simple exercise to verify that condition (A3) holds if $\alpha > d/2$, whilst the other conditions are also readily checked in a fashion similar to the inverse multiquadratics.

6.4. Cardinal functions

Associated with many kernels are so-called *cardinal functions* which exhibit the interpolatory condition on the integer lattice that the cardinal sine function does, i.e. they are functions L such that $L(j) = \delta_{0,j}$, $j \in \mathbb{Z}^d$. Given a kernel ϕ , formally define

$$\widehat{L}_\phi(\xi) := \sum_{j \in \mathbb{Z}^d} \frac{\widehat{\phi}(\xi)}{\widehat{\phi}(\xi - 2\pi j)}, \quad \xi \in \mathbb{R}^d.$$

Then as long as the Fourier inversion formula holds and $\widehat{\phi}$ decays suitably, one may show that L_ϕ given by the inverse Fourier transform is a cardinal function. For sufficient conditions on the kernel ϕ for L_ϕ to be a cardinal function, see [33]. In particular, we assume from here on that $\widehat{\phi} > 0$ on \mathbb{R}^d .

Cardinal functions associated with radial basis functions have been studied rather extensively [4,5,10, 11,22,23,25,33,34,36–39,41]. Specifically, the cardinal functions associated with the Gaussian and general multiquadratics are known to have nice decay (exponential in the former case and polynomial based on the exponent α in the latter).

In the examples given below, we also have $L_\phi(x) = \sum_{j \in \mathbb{Z}^d} c_j \phi(x - j)$, where convergence is at least uniform on compact subsets of \mathbb{R}^d , but in many cases, the convergence is uniform and in $L_2(\mathbb{R}^d)$. We note that the impetus for considering such cardinal functions may be found in the ubiquitous literature on spline

interpolation (particularly cardinal splines), see for example [8,40] and references therein. For some works involving splines used in sampling, see [2,43,44].

Lemma 6.1. *Let $(\phi_\alpha)_{\alpha \in A}$ satisfy (A3), (denote the quantity there by $A_{\alpha,h,j}^\phi$). Then the associated cardinal functions $(L_\alpha)_{\alpha \in A}$ satisfy (A3) (denote the associated functions in the condition by $A_{\alpha,h,j}^L$), and moreover*

$$\sum_{j \neq 0} \|A_{\alpha,h,j}^L\|_{L_\infty(B)} \leq C \sum_{j \neq 0} \|A_{\alpha,h,j}^\phi\|_{L_\infty(B)} \left(1 + \sum_{j \neq 0} \|A_{\alpha,1,j}^\phi\|_{L_\infty(B)} \right).$$

Proof. Note that

$$A_{\alpha,h,j}^L(\xi) = \frac{\widehat{L_\alpha}(\xi + 2\pi j/h)}{\widehat{L_\alpha}(\xi)} = \frac{\widehat{\phi_\alpha}(\xi + 2\pi j/h)}{\widehat{\phi_\alpha}(\xi)} \frac{\sum_{\ell \in \mathbb{Z}^d} \widehat{\phi_\alpha}(\xi - 2\pi \ell)}{\sum_{k \in \mathbb{Z}^d} \widehat{\phi_\alpha}(\xi + 2\pi j/h - 2\pi k)}.$$

The series in the denominator on the right is at least $\widehat{\phi_\alpha}(\xi)$ by the assumption of positivity. Thus

$$A_{\alpha,h,j}^L(\xi) \leq A_{\alpha,h,j}^\phi(\xi) \left(1 + \sum_{\ell \neq 0} A_{\alpha,1,\ell}^\phi(\xi) \right),$$

and the required inequality follows from summing over $j \neq 0$. \square

It follows immediately from this that $g_\alpha^L(h) \leq C g_\alpha^\phi(h)$, where the functions g_α are the rates assumed in (A3).

For a given set of cardinal functions (L_α) , define the N -term approximation space via

$$\mathbb{L}_N := \left\{ \sum_{j=1}^N a_j L_{\alpha_j} \left(\frac{\cdot - x_j}{c_j} \right) : (a_j) \subset \mathbb{C}, (\alpha_j) \subset A, (x_j) \subset \mathbb{R}^d, (c_j) \subset \mathbb{R} \right\}.$$

Note that these spaces are defined to be closed under translation and dilation, so (A1) is automatically satisfied; this is done because the cardinal functions themselves are not generally preserved under these operations. However, defining the approximation space this way is not unnatural since often in such methods, one considers interpolation using cardinal functions for different lattices. Therefore, the use of shifted and dilated cardinal functions will correspond to interpolation of the wavelet at differently scaled lattices to form the approximant. We note also that (A2) is satisfied for all cardinal functions since $|L_\alpha(x)| \leq 1$ for all $x \in \mathbb{R}^d$ for the examples listed in the sequel.

6.4.1. Gaussian cardinal function

The cardinal function associated with the Gaussian kernel ϕ_α described in section 6.1 was studied extensively by Riemenschneider and Sivakumar [36–39,41]. In particular, they showed that the cardinal function decays exponentially away from the origin, as does its Fourier transform, which implies (A5) with $\alpha_k'' = 0$ and (A6). Additionally, for every α , $|L_\alpha(x)| \leq 1$ for every $x \in \mathbb{R}^d$. Finally, (A3) follows from Lemma 6.1, while (A4) is a simple exercise based on the exponential decay of the Gaussian.

6.4.2. Multiquadric cardinal functions

Details on the behavior of the cardinal functions associated with the general multiquadratics may be found in [22,23] for a broad range of exponents α , whereas the particular cases of $\alpha = \pm 1/2, 1, -k + 1/2$, $k \in \mathbb{N}$ were considered previously [4,10,13].

In fact, checking the condition (A3) was done in Section 7 of [22] for the univariate cardinal function, while a different estimate in [23] demonstrates (A3) in higher dimensions. (A4) and (A5) also follow from those estimates (even with the same parameters $\alpha_k, \alpha'_k, \alpha''_k$ as in Section 6.2). For any $\alpha > d/2$, the associated multiquadric cardinal function satisfies (A1)–(A6).

It should also be noted that the growing multiquadratics $\phi_{\beta,c} := (|\cdot|^2 + c^2)^\beta$ for $\beta \geq 1/2$ have well-defined cardinal functions which also satisfy (A1)–(A6) [22,23].

7. Regularity criteria for growing kernels

In this section we investigate the possibilities for growing kernels, which requires a reworking of conditions (A1)–(A6). Moreover, the approximants defined in this section differ from those previously discussed, but nonetheless achieve the same approximation orders. To begin, given $N \geq N_0$, we seek an $M := M_N$ term approximation space

$$\Phi_M := \left\{ \sum_{j=1}^M a_j \phi_{\alpha_j}(\cdot - x_j) : (a_j) \subset \mathbb{C}, (\alpha_j) \subset A, (x_j) \subset \mathbb{R}^d \right\}$$

which is closed under translation and dilation. Relying on the analysis above, we seek functions ϕ_α such that a finite linear combination of their translates decay. For the moment, consider the univariate case, where we could hope to use the divided difference to obtain (A2) and perhaps (A5). In principle and practice, condition (A3) is the hardest to check. Notice that a growing kernel leads to a singularity at the origin in the Fourier domain; however, the function $\sum_{j \in \mathbb{Z}} A_{\alpha,h,j}^\phi$ has no such singularity. With this and the formulation in Proposition 3.1 in mind, we define the bivariate kernel $k_{\phi,h}$ whose Fourier transform in the first variable is given by

$$\widehat{k_{\phi,h}}(\xi, x) := \sum_{j \in \mathbb{Z}} \frac{\widehat{\phi}(\xi + 2\pi j/h) e^{2\pi i j x/h}}{\widehat{\phi}(\xi)} \chi_B(\xi). \quad (8)$$

By (A3), this kernel is well defined for all x , since

$$\|\widehat{k_{\phi,h}}(\cdot, x)\|_{L_1} \leq C \left(1 + \sum_{j \neq 0} \|A_{h,j}^\phi\|_{L_\infty(B)} \right).$$

Thus we make our approximation of $f \in \mathcal{S}_B$ of the form

$$\tilde{T}_h^\sharp f(x) := [f * k_h(\cdot, x)](x) \quad (9)$$

(i.e. the convolution is taken in the first variable) and note that when everything is smooth and decays well enough, the Poisson summation formula holds and we recover our previous work; that is, $\tilde{T}_h^\sharp f = T_h^\sharp f$. The advantage of this technique is that we can consider growing kernels from the outset. For illustration purposes, consider the family of univariate multiquadratics

$$\tau_\alpha(x) := (x^2 + c^2)^{\alpha-1/2}, \quad \alpha \in \mathbb{N}.$$

We know that by taking sufficiently many divided differences, we will be able to obtain any polynomial rate of decay. In fact, $\phi_\alpha(x) := \lfloor \tau_\alpha \rfloor_{2\alpha}(x) = O(|x|^{-2\alpha-1})$, where $\lfloor f \rfloor_{2n} = \frac{1}{(2n)!} \sum_{j=-n}^n (-1)^{j+n} \binom{2n}{j+n} f(\cdot + j)$. Notice that for this choice of kernel, we have

$$A_{\alpha,h,j}^\phi(\xi) = \frac{(1 - \cos(\xi + 2\pi j/h))^\alpha \widehat{\tau_\alpha}(\xi + 2\pi j/h)}{(1 - \cos(\xi))^\alpha \widehat{\tau_\alpha}(\xi)},$$

so that if we choose $h = 1/N$, we get

$$A_{\alpha,1/N,j}^\phi(\xi) = \frac{\widehat{\tau_\alpha}(\xi + 2\pi jN)}{\widehat{\tau_\alpha}(\xi)} = A_{\alpha,1/N,j}^\tau(\xi).$$

The interplay between the divided difference and the choice of $h = 1/N$ is vital and allows us to proceed. To do so, we need to truncate our approximation, so we set $k_{\phi,1/N}^\flat(\cdot, x)$ to be the function whose Fourier transform is given by

$$\widehat{k_{\phi,1/N}^\flat}(\xi, x) := \sum_{|j| \leq N^2} \frac{\widehat{\phi}(\xi + 2\pi jN) e^{2\pi i x j N}}{\widehat{\phi}(\xi)} \chi_B(\xi);$$

then we define

$$\widetilde{T}_{1/N}^\flat f(x) := [f * k_{\phi,1/N}^\flat(\cdot, x)](x). \quad (10)$$

Henceforth we make the following assumptions:

- (B1) Φ_M is closed under translation and dilation.
- (B2) $\phi_\alpha = \lfloor \tau_\alpha \rfloor_{n_\alpha}$ for some $(n_\alpha) \subset \mathbb{N}$; that is, the kernels are built out of divided differences.
- (B3) (A3) holds for each ϕ_α and $h = 1/N$.
- (B4) For every $k \in \mathbb{N}$ and $j \neq 0$, there exists $\tilde{\alpha}_k \in A$ and C independent of N such that for every $\alpha \geq \tilde{\alpha}_k$ and $0 \leq l \leq k$, $D^l A_{\alpha,1/N,j}^\tau \in L_\infty(B)$, and

$$\sum_{j \neq 0} \|D^l A_{\alpha,1/N,j}^\tau\|_{L_\infty(B)} \leq C.$$

We now use these hypotheses to prove the results analogous to those found in Section 3; note that here we have $B = [-R, R]$.

Lemma 7.1. *There exists $\alpha_0 \in A$ such that for all $\alpha \geq \alpha_0$, $N > R/\pi$, and $f \in \mathcal{S}_B$,*

$$|(1/N) \widetilde{T}_{1/N}^\sharp f(x)| \leq \|\widehat{f}\|_{L_1} (1 + g_\alpha(1/N)).$$

Proof. Using (9) and the inversion formula, we have

$$|(1/N) \widetilde{T}_{1/N}^\sharp f(x)| \leq \left\| \widehat{k_{\phi,1/N}^\flat}(\cdot, x) \right\|_{L_1} \leq \|\widehat{f}\|_{L_1} (1 + g_\alpha(1/N)),$$

the last inequality coming from (B3). \square

Proposition 7.2. *There exists $\alpha_0 \in A$ such that for all $\alpha \geq \alpha_0$, $N > R/\pi$, and $f \in \mathcal{S}_B$,*

$$\|f - (1/N) \widetilde{T}_{1/N}^\sharp f\|_{L_\infty} \leq C \|\widehat{f}\|_{L_1} g_\alpha(1/N),$$

where $C > 0$ is independent of f and N .

Proof. We have, from (B3),

$$\left| f(x) - (1/N)\tilde{T}_{1/N}^\sharp f(x) \right| \leq \left\| \widehat{f} \sum_{j \neq 0} A_{\alpha, 1/N, j}^\phi \right\|_{L_1(B)} \leq C \|\widehat{f}\|_{L_1} g_\alpha(1/N). \quad \square$$

Lemma 7.3. Let $k \in \mathbb{N}$ and $f \in \mathcal{S}_B$. There exists $\alpha_k \in A$, such that for sufficiently large N , there is a constant $C_{f,k} > 0$ independent of N , so that for all $\alpha \geq \alpha_k$,

$$\|f - \tilde{T}_{1/N}^\flat f\|_{L_\infty} \leq C_{f,k} N^{-k}.$$

Proof. On account of (B3) and Proposition 7.2, we have

$$\begin{aligned} \left| f(x) - (1/N)\tilde{T}_{1/N}^\flat f(x) \right| &\leq \left| f(x) - (1/N)\tilde{T}_{1/N}^\sharp f(x) \right| + \left| (1/N)\tilde{T}_{1/N}^\sharp f(x) - (1/N)\tilde{T}_{1/N}^\flat f(x) \right| \\ &\leq C \|\widehat{f}\|_{L_1(B)} g_\alpha(1/N) + (1/N) \|\widehat{f}\|_{L_1(B)} \sum_{|j| > N^2} \|A_{\alpha, 1/N, j}^\phi\|_{L_\infty(B)} \\ &\leq C_{f,k} N^{-k}. \quad \square \end{aligned}$$

Proposition 7.4. Let $k \in \mathbb{N}_0$ and $f \in \mathcal{S}_B$. There exists $\alpha'_k \in A$ so that for all $\alpha \geq \alpha'_k$ and sufficiently large N , there is a constant C , independent of N such that

$$|f(x) - (1/N)\tilde{T}_{1/N}^\flat f(x)| \leq C N^{-k} (1 + |x|)^{-k}.$$

Proof. Similar to Proposition 3.3, we need only check that if $|x| > 2/h$, then

$$\left| f(x) - (1/N)\tilde{T}_{1/N}^\flat f(x) \right| \leq C_{k,f} (1 + |x|)^{-2k}.$$

Note that since $f \in \mathcal{S}$, $|f(x)| \leq C(1 + |x|)^{-2k}$. From (B4), when $\alpha \geq 2\tilde{\alpha}_k$, we have

$$\begin{aligned} \left| (1 + |x|)^{2k} (1/N)\tilde{T}_{1/N}^\flat f(x) \right| &\leq C_{f,k} \sum_{|j| > N^2} \sum_{l=0}^{2k} \|D^l A_{\alpha, 1/N, j}^\phi\|_{L_\infty(B)} \\ &\leq \tilde{C}_{f,k}. \quad \square \end{aligned}$$

For a given k , we set $\alpha_K := \max\{\alpha_0, \alpha'_k, \alpha'_k\}$. Thus the conclusions of the results above hold for every $\alpha \geq \alpha_K$.

Theorem 7.5. Let $\psi \in \mathcal{S}_B$ be given. Suppose $k \in \mathbb{N}_0$, $N \geq N_0$, and let I be a cube. Then there exists a constant $C > 0$, independent of N and I , such that for all N large enough and suitably large α

$$\left| \psi_I(x) - \tilde{T}_{1/N}^\flat \psi_I(x) \right| \leq C N^{-k} \left(1 + \frac{|x - c(I)|}{\ell(I)} \right)^{-2k}.$$

We need only double α_K to achieve the stated bound.

In order to obtain results in the multivariate case we use tensor products, which is sufficient due to the fact that the wavelets considered here are tensor products themselves. The univariate estimates above lead to the main results of this section. Note that $\Phi_M^{\otimes d}$ is the space of d -fold tensor products of functions in Φ_M , i.e. $h \in \Phi_M^{\otimes 2}$ has the form $h(x, y) = f(x)g(y)$, for $f, g \in \Phi_M$.

Theorem 7.6. Suppose that $s > 0$, $N \geq N_0$, and $1 \leq p < \infty$. There is a constant $C_{p,s,d} > 0$, $M_{s,N} \in \mathbb{N}$ so that for $f \in F_{\tau,q}^s$, with $1/\tau = 1/p + s/d$ and $1/q = 1 + s/d$, there is $S_f \in \Phi_M^{\otimes d}$ so that

$$\|f - S_f\|_{L_p} \leq C_{p,s,d} N^{-s/d} |f|_{F_{\tau,q}^s}.$$

Proof. We begin with the wavelet expansion of f , given by

$$f = \sum f_I \psi_I,$$

and define S_f in terms of this expansion

$$S_f := \sum f_I \tilde{T}_{1/N_I}^\flat \psi_I.$$

Recall that $\tilde{T}_{1/N_I}^\flat \psi_I = \sum_{j=1}^{N_I} a_{I,j} \phi((\cdot - c(I)/\ell(I)))$, so choose $\phi = \phi_\alpha$ with $\alpha = \alpha_s$ large enough so that Theorem 7.5 holds. Notice also that we have used no more than $\sum M_{s,N_I} \leq M_{s,N}$ total centers. If everything is smooth, we have $M_{s,N} = N$, while if we use divided differences, then we have $M_{s,N} = (2\alpha_s + 1)^d (N^2 + 1)^d$ terms involving the original kernel.

Since the bound in Theorem 7.5 justifies Lemma 4.4 with no further restriction, we use this to obtain

$$\|f - S_f\|_{L_p} \leq C_{k,d} \left\| \sum_I \min\{1, c_I^{-k/d}\} |f_I| \chi_I \right\|_{L_p}.$$

This estimate also makes use of the tensor product structure of the wavelet and approximant S_f . To complete the proof, we must have $k > s$, which provides the restriction on α_s . The rest of the proof follows the same reasoning given in the proof of Theorem 4.5. \square

The conclusion of Theorem 7.6 does not completely resolve the problem of approximation from the space Φ_M as the construction therein relies on taking tensor products for approximation in dimension larger than 1. It remains of interest to determine an approximation method from the space Φ_M in 1 dimension that does not require taking divided differences of the growing kernel and which can be extended in a natural way to higher dimensions.

It should be noted that the techniques of [17] give convergence rates as in Theorem 7.6 for certain growing kernels such as surface splines, but the proofs rely on the kernels inverting an elliptic differential operator of a certain type. The approximants \tilde{T}_h^\sharp and \tilde{T}_h^\flat used here are defined in different ways than those in [17,26] and allow more flexibility in the choice of kernel (indeed they need not be associated to any differential operator), and yet achieve a similar goal.

8. Examples of growing kernels

8.1. Multiquadratics

We will show that the family of divided differences of multiquadratics mentioned above satisfies properties (B1)–(B4). To wit, consider

$$\phi_\alpha = \lfloor \tau_\alpha \rfloor_{2\alpha},$$

where $\tau_\alpha(x) = (x^2 + \alpha^2)^{\alpha-1/2}$, $\alpha \in \mathbb{N}$, $x \in \mathbb{R}$.

Property (B1) is satisfied by using the same reasoning as that for (A1) in Section 6.2. Property (B2) is obvious, and note the fact that we want the divided difference of order 2α of τ_α . To see that (B3) holds, note that

$$A_{\alpha,1/N,j}^\phi(\xi) = \frac{\widehat{\tau_\alpha}(\xi + 2\pi j N)}{\widehat{\tau_\alpha}(\xi)} = A_{\alpha,1/N,j}^\tau(\xi),$$

so using estimates from Lemma 1 in [22], we have

$$\|A_{\alpha,1/N,j}^\tau(\xi)\|_{L_\infty(B)} \leq e^{2cR} e^{-2\pi c|j|N} (2|j| - 1)^{-\alpha},$$

hence (A3) is satisfied. Estimates from [23] show that $A_{\alpha,1/N,j}^\tau$ is $(2\alpha - 1)$ -times differentiable and Theorem 5.1 there may be adapted to our needs, showing that the sum in (B4) can be bounded independent of h .

8.2. Power kernels

Our next example is $\phi_\alpha = \lfloor \tau_\alpha \rfloor_{2\alpha}$, where $\tau_\alpha(x) = |x|^\alpha$ for $\alpha \in \mathbb{R}^+ \setminus 2\mathbb{N}$. We note that

$$\phi_\alpha(x) = O(|x|^{-\alpha}), \quad |x| \rightarrow \infty.$$

We find the Fourier transform in Section 8.3 of [45]:

$$\widehat{\tau_\alpha}(\xi) = (2\pi)^{\frac{d}{2}} \frac{2^{\alpha+\frac{1}{2}} \Gamma(\frac{\alpha+1}{2})}{\Gamma(-\frac{\alpha}{2})} |\xi|^{-\alpha-1},$$

which allows us to move forward with our computations. Conditions (B1) and (B2) are evident, so we begin with (B3):

$$A_{\alpha,1/N,j}^\phi(\xi) = \frac{|\xi|^{\alpha+1}}{|\xi + 2\pi j N|^{\alpha+1}},$$

thus

$$\|A_{\alpha,1/N,j}^\phi\|_{L_\infty(B)} \leq \frac{R^{\alpha+1}}{N^{\alpha+1} \pi^{\alpha+1}} (2|j| - 1)^{-\alpha-1},$$

hence (B3) is satisfied. To see (B4), we apply the quotient rule repeatedly and find that

$$|D^l A_{\alpha,1/N,j}^\phi(\xi)| \leq C_{\alpha,l} N^{-l} (2|j| - 1)^{-\alpha},$$

hence as long as $\alpha > 1$, (B4) is satisfied.

9. Sobolev interpolation using cardinal functions

Let us now look more closely at the example of using cardinal functions as the family of kernels. Due to some existing theory, the use of cardinal functions allows us to obtain error estimates for the classical Sobolev smoothness spaces which are not included in the framework above due to the restrictions on the smoothness s of the Triebel–Lizorkin space. The method here is one of interpolation, whereby more details may be found in the cardinal interpolation literature mentioned previously.

9.1. Decaying functions

To begin, formally define an alternate approximant of f via

$$I_h^\sharp f(x) := \sum_{j \in \mathbb{Z}^d} f(hj) L_{\tau(h)} \left(\frac{x}{h} - j \right),$$

where $L_{\tau(h)}$ is one of the cardinal functions discussed in Section 6.4, and $\tau(h) = h^2$ in the case of the Gaussian, and h^{-1} for the multiquadratics. Note that this is different from $T_h^\sharp f$ due to the fact that we use the samples of f at the lattice $h\mathbb{Z}^d$ in the approximant rather than the values of $f_{L_{\tau(h)}}$ as defined previously. This object has been studied in various instances before [10, 23, 25], and is actually an interpolant of f . By definition of the cardinal functions, it is easy to see that $I_h^\sharp f(hk) = f(hk)$, $k \in \mathbb{Z}^d$. Note that for the general multiquadric cardinal functions, there is an additional parameter α governing the power of the multiquadric $(|x|^2 + \tau(h)^2)^\alpha$. We suppress this subscript, and simply consider α to be fixed but arbitrary in the range $(-\infty, -d - 1/2) \cup [1/2, \infty) \setminus \mathbb{N}$.

It is known that for certain classes of Sobolev functions, these interpolants exhibit nice approximation rates. For example, the following holds.

Theorem 9.1 ([25], Theorem 2.1 and [23], Theorem 3.1). *Let $L_{\tau(h)}$ be the cardinal function associated with the Gaussian or the general multiquadratics. If $1 < p < \infty$, and $k > d/p$, then there exists a constant C , independent of h , such that for every $f \in W_p^k(\mathbb{R}^d)$,*

$$\|I_h^\sharp f - f\|_{L_p} \leq Ch^k \|f\|_{W_p^k}.$$

If $p = 1, \infty$, then the bound changes to $C(1 + |\ln h|)^d h^k \|f\|_{W_p^k}$.

With Theorem 9.1 in mind, one would naturally desire some estimate for a related interpolant which makes use of only finitely many samples of the function. While there are many feasible ways to do this, we focus here on a method similar to the preceding analysis; to wit, let

$$I_h^\flat f(x) := \sum_{j \in \mathbb{Z}^d \cap B_h} f(hj) L_{\tau(h)} \left(\frac{x}{h} - j \right),$$

where B_h is the ball of radius h^{-2} centered about the origin. Similarly, let $I_N f$ be as in (3) where $N \sim h^{-2d}$.

For arbitrary Sobolev functions, it is difficult to ascertain the behavior of the truncated interpolant; however, if one assumes a certain asymptotic decay rate on the function itself, then something may be said. Define \mathcal{S}_κ to be the class of functions on \mathbb{R}^d with decay $O(|x|^{-\kappa})$, $|x| \rightarrow \infty$. Then the following holds.

Theorem 9.2. *With the notation and parameters as in Theorem 9.1, there exists a constant C , independent of h , such that for every $f \in W_p^k(\mathbb{R}^d) \cap \mathcal{S}_\kappa$,*

$$\|I_h^\flat f - f\|_{L_p} \leq Ch^\rho \|f\|_{W_p^k},$$

where $\rho = \min\{k, \kappa - 2d + d/p\}$ for multiquadric interpolation, and $\rho = \min\{k, \kappa - d\}$ for Gaussian interpolation.

Before supplying the complete proof, we collect some useful lemmas.

Lemma 9.3. *If $f \in \mathcal{S}_\kappa$, then*

$$\sum_{|j|>h^{-2}} |f(hj)| \leq Ch^{\kappa-2d}.$$

Proof. By the assumption on the decay of f and the same estimate as in the proof of Lemma 3.2, we find that

$$\sum_{|j|>h^{-2}} |f(hj)| \leq C \sum_{|j|>h^{-2}} \frac{1}{|j|^\kappa h^\kappa} \leq C_d h^{-\kappa} \int_{h^{-2}}^{\infty} r^{d-1-\kappa} dr = C_{d,\kappa} h^{-\kappa} h^{-2d+2\kappa},$$

which is at most $Ch^{\kappa-2d}$ as required. \square

Lemma 9.4. *Let $1 < p < \infty$. If $L_{\frac{1}{h}}$ is the cardinal function associated with the general multiquadric, then*

$$\|L_{\frac{1}{h}}(\cdot/h)\|_{L_p} \leq Ch^{d/p},$$

whereas if L_{h^2} is that associated with the Gaussian, then

$$\|L_{h^2}(\cdot/h)\|_{L_p} \leq Ch^d.$$

Proof. Let us begin with the multiquadric case. From [23, Corollary 5.3], we see that $|L_{\frac{1}{h}}(x/h)| \leq C_d \min\{1, h^d |x|^{-d}\}$ (we also note that the estimate may be obtained in a straightforward manner from Section 4 of [22]). Therefore, the p -th power of the L_p norm in question is bounded by

$$\int_{B(0,h)} dx + \int_{\mathbb{R}^d \setminus B(0,h)} \frac{h^{dp}}{|x|^{dp}} dx \leq Ch^d + Ch^{pd} \int_h^{\infty} r^{-dp} r^{d-1} dr,$$

which is at most $C_{\alpha,p,d} h^d$.

For the Gaussian case, one need only notice that the multivariate Gaussian cardinal function is nothing but the d -fold tensor product of the univariate version. Consequently, we may use the bound of [25, Eq. 4.4], which says that for the univariate Gaussian cardinal function, $L_{h^2}(x/h) \leq C \min\{h, h|x|^{-1}\}$. Consequently,

$$\|L_{h^2}(\cdot/h)\|_{L_p(\mathbb{R})}^p \leq Ch^p \int_{-1}^1 dx + Ch^p \int_1^{\infty} \frac{1}{|x|^p} dx \leq Ch^p.$$

Thus the multivariate estimate is Ch^{dp} , whence taking p -th roots gives the desired inequality. \square

With these ingredients in hand, we are now ready to supply the proof of the theorem.

Proof of Theorem 9.2. First note that $\|I_h^\flat f - f\|_{L_p} \leq \|I_h^\sharp f - f\|_{L_p} + \|I_h^\sharp f - I_h^\flat f\|_{L_p}$, and the first term is majorized by Ch^k by Theorem 9.1. Now to estimate the second term, it follows from Minkowski's integral inequality [18, Theorem 6.19, p. 194] that

$$\begin{aligned}
\|I_h^\sharp f - I_h^\flat f\|_{L_p} &= \left(\int_{\mathbb{R}^d} \left| \sum_{|j|>h^{-2}} f(hj) L_{\tau(h)} \left(\frac{x}{h} - j \right) \right|^p dx \right)^{\frac{1}{p}} \\
&\leq \sum_{|j|>h^{-2}} \left(\int_{\mathbb{R}^d} |f(hj)|^p \left| L_{\tau(h)} \left(\frac{x}{h} - j \right) \right|^p dx \right)^{\frac{1}{p}} \\
&= \sum_{|j|>h^{-2}} |f(hj)| \|L_{\tau(h)}(\cdot/h)\|_{L_p}.
\end{aligned}$$

From Lemmas 9.3 and 9.4, we see that for the multiquadric, $\|I_h^\sharp f - I_h^\flat f\|_{L_p} \leq Ch^{\kappa-2d+d/p}$, while the bound for the Gaussian is $Ch^{\kappa-d}$, thus completing the proof. \square

Corollary 9.5. *With the parameters as in Theorem 9.2, if $N \sim h^{-2d}$, then for every $f \in W_p^k(\mathbb{R}^d) \cap \mathcal{S}_\kappa$,*

$$\|I_N f - f\|_{L_p} \leq CN^{-\rho/(2d)} \|f\|_{W_p^k}.$$

Evidently, the N -term interpolants considered here work well for Sobolev functions which decay away from the origin. Of course, if a given function peaks far away from the origin but still decays away from the peak, one should translate the peak to the origin and then interpolate, in which case the same estimate as in Corollary 9.5 holds for $I_N(f_I) - f_I$.

It follows easily from the lemmas above that one can still estimate the difference of the full and truncated interpolants in L_1 and L_∞ and get analogous bounds in terms of h . However, the method of proof of Theorem 9.2 is not sufficient to estimate approximation orders of $I_h^\flat f - f$ in these spaces due to the logarithmic term found in Theorem 9.1. Moreover, it has proven elusive to estimate the truncated interpolant via other means.

9.2. Greedy interpolation

Of interest, especially in light of the recent advances in Greedy approximations (e.g. [42]), is the *greedy interpolant* of f , which for a given lattice would be formed by keeping only the N largest (in absolute value) samples $f(hj)$. That is, $G_N f(x) = \sum_{j \in \Lambda} f(hj) L_{\tau(h)}(x/h - j)$, where $|f(hj)| \geq |f(hk)|$, $j \in \Lambda$, $k \notin \Lambda$, and $|\Lambda| = N$. Inspired by the estimates in [12] for greedy quasi-interpolants, we may provide similar estimates here for the greedy Gaussian and multiquadric interpolants.

The rest of the argument mimics that of [12], although the proofs of the intermediate steps do not follow directly from their argument, but rather come from other known estimates for the Gaussian and multiquadric cardinal functions. Before stating the theorem, we need an auxiliary lemma on the Lebesgue constants for the given cardinal functions.

Lemma 9.6. *Let $\alpha \in (-\infty, -d - 1/2) \cup [1/2, \infty) \setminus \mathbb{N}$ be a fixed parameter for the general multiquadric, and let $L_{\tau(h)}$ be either the cardinal function associated to the multiquadric or the Gaussian. Then for $h > 0$, we have*

$$\sup_{y \in \mathbb{R}^d} \sum_{j \in \mathbb{Z}^d} |L_{\tau(h)}(y - j)| \leq C |\ln h|^d,$$

for some C independent of h .

Proof. The estimate is well-known for the Gaussian, and may be found in [37, Theorem 5.2], while for the multiquadratics, the estimate in dimension 1 for general α is given by [22, Proposition 7] (though for $\alpha = 1/2$,

the estimate is in [36]), and the proof of the bound in higher dimensions follows the same line of reasoning upon using the bounds provided in [23, Corollary 4.7]. \square

Theorem 9.7. *Let the parameters p, k and α be as in Theorem 9.1, and let G_N be the greedy Gaussian or multiquadric interpolant with parameter α . If $h \sim N^{-\frac{1}{kp+d}}$, then*

$$\|f - G_N f\|_{L_\infty} \leq C N^{-\frac{k}{kp+d}} (1 + |\ln N|)^d \|f\|_{W_p^k}, \quad f \in W_p^k.$$

Proof. First of all, by Theorem 9.1, we have that

$$\|f - I_h^\sharp f\|_{L_\infty} \leq C N^{-\frac{k}{kp+d}} (1 + |\ln N|)^d \|f\|_{W_p^k},$$

whereby it suffices to estimate $\|I_h^\sharp f - G_N f\|_{L_\infty}$. The quantity in question is, by definition,

$$\left\| \sum_{j \notin \Lambda} f(hj) L_{\tau(h)} \left(\frac{\cdot}{h} - j \right) \right\|_{L_\infty}.$$

Since $f \in W_p^k$, it follows from the Sobolev embedding Theorem that $\sum_{j \in \mathbb{Z}^d} |f(hj)|^p \leq C h^{-d} \|f\|_{W_p^k}$, whereby (as noted in [12], there can be at most n elements of \mathbb{Z}^d for which $|f(hj)| \geq C h^{-\frac{d}{p}} N^{-\frac{1}{p}} \|f\|_{W_p^k}$. Consequently, since $|\Lambda| = N$, the reverse inequality is true for $|f(hj)|$ for all $j \notin \Lambda$. Finally, we need only appeal to Lemma 9.6 to see that

$$\|I_h^\sharp f - G_N f\|_{L_\infty} \leq C |\ln N|^d h^{-\frac{d}{p}} N^{-\frac{1}{p}} \|f\|_{W_p^k}.$$

The result follows upon combining this estimate with that from Theorem 9.1, with the requirement on h . \square

Of particular future interest would be to give estimates on $\|f - G_N f\|_{L_p}$ in the above setting; however, this remains an elusive task at present.

10. Cost distribution

10.1. Examples

In this section, we analyze what the cost distribution looks like for some particular types of functions. Suppose that ψ is the mother wavelet, and that ψ_I is as defined previously for a dyadic cube I . Then suppose that $\{I_j\}_{j=1}^M$ are disjointly supported dyadic cubes, and we consider the cost distribution for the function

$$f = \sum_{j=1}^M a_j \psi_{I_j}.$$

We recall the relation of the parameters: for a fixed p and s , we have $1/\tau = 1/p + s/d$, and $1/q = 1 + s/d$. Then

$$M_{s,q} f(x) = \left(\sum_{j=1}^M |I_j|^{-\frac{sq}{d}} |a_j|^q \chi_{I_j}(x) \right)^{\frac{1}{q}}.$$

Table 1
Cost distribution for f with sparse wavelet representation.

N	Cost (c_1, \dots, c_7)
100	(10, 16, 16, 9, 28, 20, 2)
200	(21, 31, 31, 17, 56, 39, 5)
300	(31, 47, 47, 26, 83, 59, 7)
400	(42, 62, 62, 34, 111, 79, 10)
500	(52, 78, 78, 43, 139, 98, 12)

Therefore, since the I_j are disjoint,

$$\begin{aligned}
 |f|_{F_{\tau,q}^s} &= \left(\int_{\mathbb{R}} \left(\sum_{j=1}^M |I_j|^{-\frac{sq}{d}} |a_j|^q \chi_{I_j}(x) \right)^{\frac{\tau}{q}} dx \right)^{\frac{1}{\tau}} \\
 &= \left(\sum_{j=1}^M \int_{I_j} |I_j|^{-\frac{s\tau}{d}} |a_j|^\tau dx \right)^{\frac{1}{\tau}} \\
 &= \left(\sum_{j=1}^M |a_j|^\tau |I_j|^{\frac{\tau}{p}} \right)^{\frac{1}{\tau}},
 \end{aligned}$$

the final step coming from the observation that $1 - s\tau/d = \tau/p$.

Next, recall that $m_{s,q,I_j} := M_{s,q,I_j} = |I_j|^{-\frac{s}{d}} |a_j|$, and by definition $f_{I_j} = a_j$. Then the cost of the cube I_j is

$$\begin{aligned}
 c_{I_j} &= \frac{1}{\sum_{m=1}^M |a_m|^\tau |I_m|^{\frac{\tau}{p}}} |I_j|^{-\frac{s}{d}(\tau-q)} |a_j|^{\tau-q} |a_j|^q |I_j|^q N \\
 &= \frac{|a_j|^\tau |I_j|^{\frac{\tau}{p}}}{\sum_{m=1}^M |a_m|^\tau |I_m|^{\frac{\tau}{p}}} N.
 \end{aligned} \tag{11}$$

Recalling that $N_{I_j} = \lfloor c_{I_j} \rfloor$ if the right-hand side is at least N_0 , and $N_{I_j} = 0$ otherwise, we have that

$$S_f = \sum_{j=1}^M a_j T_{N_j} \psi_{I_j}.$$

As an example of the above, we illustrate the cost distribution for a function having a disjoint wavelet expansion. Consider a fixed 7-term wavelet, $f = a_1\psi(2^{j_1}(\cdot+3)) + a_2\psi(2^{j_2}(\cdot+2)) + \dots + a_7\psi(2^{j_7}(\cdot-3))$ where the coefficients a_1, \dots, a_7 and the dilations were randomly generated to yield coefficients (9, 5, 5, 3, 8, 4, 1) and dilations of (3, 1, 1, 2, 0, 0, 4). The number of terms is of course arbitrary, but the cost distribution is displayed in Table 1 for increasing budget for the parameters $s = 1, d = 1, p = 1, \tau = \frac{1}{2}$.

It can be seen from this example that dilations giving larger supports are more favored in the cost distribution, as the terms with dilation 1 and 0 have the largest associated cost, while those with large dilation (hence smaller measure of $|I_j|$) have smaller costs. Additionally, the coefficients a_j scale the cost according to the power of τ , for example $c_5 \approx \sqrt{2}c_6$, where $a_5 = 2a_6$ (recall that $\tau = \frac{1}{2}$).

10.2. Limitations

Here, we present an inherent limit to this scheme. It should not be surprising based on the discussion above, but the cost distribution scheme can fail to capture enough information on a signal whose wavelet expansion is very spread out. Let us suppose that N is fixed, and arbitrarily large. We will exhibit a signal whose N -term approximant $S_{f,N}$ is identically 0 because the cost of each dyadic cube is 0.

Let $M > N^\tau$, and let $(I_j)_{j=1}^M$ be disjointly supported cubes of unit volume, and define $f = \sum_{j=1}^M \psi_{I_j}$. By (11), we find that for each $1 \leq j \leq M$,

$$c_{I_j} = M^{-\frac{1}{\tau}} N < 1,$$

whereby, $N_{I_j} = 0$ for every j .

Finally, it should be noted that the theoretical cost distribution above is difficult to implement in practice. In particular, it assumes knowledge of all the wavelet coefficients of the target function, f . Moreover, computing the approximants is not typically fast in this case. For the future, it would be interesting to consider other approximations from the spaces Φ_N above which may obtain the optimal recovery rates and which are more readily implemented numerically. However, the main purpose of this work was to demonstrate criteria on general approximation spaces which yield best N -term approximation orders analogous to the wavelet decomposition of the space itself.

Acknowledgments

Work for this article was completed while the first author was an Assistant Professor at Vanderbilt University. In the latter stages, the first author was partially supported by the NSF TRIPODS program under grant CCF-1423411. The second author would like to thank the Mathematics Department at Vanderbilt for hosting him during his stay, and supporting collaboration on this article.

References

- [1] M. Abramowitz, I.A. Stegun (Eds.), *Handbook of Mathematical Functions: With Formulas, Graphs, and Mathematical Tables*, vol. 55, Courier Dover Publications, 1972.
- [2] A. Aldroubi, M. Unser, Sampling procedures in function spaces and asymptotic equivalence with Shannon's sampling theory, *Numer. Funct. Anal. Optim.* 15 (1–2) (1994) 1–21.
- [3] N.D. Atreas, On a class of non-uniform average sampling expansions and partial reconstruction in subspaces of $L_2(\mathbb{R})$, *Adv. Comput. Math.* 36 (1) (2012) 21–38.
- [4] B.J.C. Baxter, The asymptotic cardinal function of the multiquadratic $\phi(r) = (r^2 + c^2)^{\frac{1}{2}}$ as $c \rightarrow \infty$, *Comput. Math. Appl.* 24 (12) (1992) 1–6.
- [5] B.J.C. Baxter, N. Sivakumar, On shifted cardinal interpolation by Gaussians and multiquadratics, *J. Approx. Theory* 87 (1) (1996) 36–59.
- [6] C. de Boor, R. DeVore, A. Ron, Approximation from shift-invariant subspaces of $L_2(\mathbb{R}^d)$, *Trans. Amer. Math. Soc.* 341 (2) (1994) 787–806.
- [7] C. de Boor, R. DeVore, A. Ron, The structure of finitely generated shift-invariant spaces in $L_2(\mathbb{R}^d)$, *J. Funct. Anal.* 119 (1994) 37–78.
- [8] C. De Boor, K. Höllig, S.D. Riemenschneider, *Box Splines*, vol. 98, Springer Science & Business Media, 2013.
- [9] M. Bownik, The structure of shift-invariant subspaces of $L_2(\mathbb{R}^n)$, *J. Funct. Anal.* 177 (2) (2000) 282–309.
- [10] M.D. Buhmann, Multivariate cardinal interpolation with radial-basis functions, *Constr. Approx.* 6 (3) (1990) 225–255.
- [11] M.D. Buhmann, *Radial Basis Functions: Theory and Implementations*, vol. 12, Cambridge University Press, 2003.
- [12] M.D. Buhmann, F. Dai, Compression using quasi-interpolation, *Jaen J. Approx.* 7 (2) (2015) 203–230.
- [13] M.D. Buhmann, C.A. Micchelli, Multiquadric interpolation improved, *Comput. Math. Appl.* 24 (12) (1992) 21–25.
- [14] I. Daubechies, *Ten Lectures on Wavelets*, vol. 61, Society for Industrial and Applied Mathematics, Philadelphia, 1992.
- [15] L. de Carli, P. Vellucci, p -Riesz bases in quasi shift invariant spaces, in: Y. Kim, K.N. Sivaram, G. Picioroaga, E.S. Weber (Eds.), *Frames and Harmonic Analysis*, vol. 706, American Mathematical Society, 2018.
- [16] R. DeVore, B. Jawerth, V. Popov, Compression of wavelet decompositions, *Amer. J. Math.* 114 (1992) 737–785.
- [17] R. DeVore, A. Ron, Approximation using scattered shifts of a multivariate function, *Trans. Amer. Math. Soc.* 362 (12) (2010) 6205–6229.

- [18] G.B. Folland, Real Analysis: Modern Techniques and Their Applications, second edition, John Wiley & Sons, 1999.
- [19] L. Grafakos, Classical and Modern Fourier Analysis, Pearson Education, Inc., Upper Saddle River, NJ, 2004.
- [20] K. Gröchenig, J. Stöckler, Gabor frames and totally positive functions, Duke Math. J. 162 (6) (2013) 1003–1031.
- [21] K. Hamm, Approximation rates for interpolation of Sobolev functions via Gaussians and allied functions, J. Approx. Theory 189 (2015) 101–122.
- [22] K. Hamm, J. Ledford, Cardinal interpolation with general multiquadratics, Adv. Comput. Math. 42 (5) (2016) 1149–1186.
- [23] K. Hamm, J. Ledford, Cardinal interpolation with general multiquadratics: convergence rates, Adv. Comput. Math. 44 (4) (2018) 1205–1233.
- [24] K. Hamm, J. Ledford, On the structure and interpolation properties of quasi shift-invariant spaces, J. Funct. Anal. 274 (7) (2018) 1959–1992.
- [25] T. Hangelbroek, W. Madych, F. Narcowich, J.D. Ward, Cardinal interpolation with Gaussian kernels, J. Fourier Anal. Appl. 18 (2012) 67–86.
- [26] T. Hangelbroek, A. Ron, Nonlinear approximation using Gaussian kernels, J. Funct. Anal. 259 (2010) 203–219.
- [27] R.Q. Jia, Shift-invariant spaces on the real line, Proc. Amer. Math. Soc. 125 (3) (1997) 785–793.
- [28] M.J. Johnson, On the approximation order of principal shift-invariant subspaces of $L_p(\mathbb{R}^d)$, J. Approx. Theory 91 (1997) 279–319.
- [29] M.J. Johnson, Scattered data interpolation from principal shift-invariant spaces, J. Approx. Theory 113 (2) (2001) 172–188.
- [30] N.J. Kalton, S. Mayboroda, M. Mitrea, Interpolation of Hardy-Sobolev-Besov-Triebel-Lizorkin spaces and applications to problems in partial differential equations, Contemp. Math. 445 (2007) 121–178.
- [31] G. Kyriazis, P. Petrushev, New bases for Triebel-Lizorkin and Besov spaces, Trans. Amer. Math. Soc. 354 (2) (2002) 749–776.
- [32] J. Ledford, Recovery of Paley-Wiener functions using scattered translates of regular interpolators, J. Approx. Theory 173 (2013) 1–13.
- [33] J. Ledford, On the convergence of regular families of cardinal interpolators, Adv. Comput. Math. 41 (2015) 357–371.
- [34] J. Ledford, Convergence properties of spline-like cardinal interpolation operators acting on ℓ^p data, J. Fourier Anal. Appl. 23 (1) (2017) 229–244.
- [35] Y. Meyer, Wavelets and Operators, Cambridge Stud. Adv. Math., vol. 37, Cambridge University Press, 1992, translated from the 1990 French original by D.H. Salinger.
- [36] S.D. Riemenschneider, N. Sivakumar, On the cardinal-interpolation operator associated with the one-dimensional multiquadric, East J. Approx. 7 (4) (1999) 485–514.
- [37] S.D. Riemenschneider, N. Sivakumar, On cardinal interpolation by Gaussian radial-basis functions: properties of fundamental functions and estimates for Lebesgue constants, J. Anal. Math. 79 (1999) 33–61.
- [38] S.D. Riemenschneider, N. Sivakumar, Gaussian radial basis functions: cardinal interpolation of ℓ_p and power growth data, Adv. Comput. Math. 11 (1999) 229–251.
- [39] S.D. Riemenschneider, N. Sivakumar, Cardinal interpolation by Gaussian functions: a survey, J. Anal. 8 (2000) 157–178.
- [40] I.J. Schoenberg, Cardinal Spline Interpolation, vol. 12, Society for Industrial and Applied Mathematics, Philadelphia, 1973.
- [41] N. Sivakumar, A note on the Gaussian cardinal-interpolation operator, Proc. Edinb. Math. Soc. (2) 40 (1997) 137–150.
- [42] V. Temlyakov, Greedy Approximation, vol. 20, Cambridge University Press, 2011.
- [43] M. Unser, Sampling—50 years after Shannon, Proc. IEEE 88 (4) (2000) 569–587.
- [44] M. Unser, T. Blu, Cardinal exponential splines: part I—theory and filtering algorithms, IEEE Trans. Signal Process. 53 (4) (2005) 1425–1438.
- [45] H. Wendland, Scattered Data Approximation, vol. 17, Cambridge University Press, 2005.