Approximation using scattered shifts of a multivariate function *

Ronald DeVore and Amos Ron

February 17, 2008

Abstract

The approximation of a general d-variate function f by the shifts $\phi(\cdot - \xi)$, $\xi \in \Xi \subset \mathbb{R}^d$, of a fixed function ϕ occurs in many applications such as data fitting, neural networks, and learning theory. When $\Xi = h\mathbb{Z}^d$ is a dilate of the integer lattice, there is a rather complete understanding of the approximation problem [6, 18] using Fourier techniques. However, in most applications the *center* set Ξ is either given, or can be chosen with complete freedom. In both of these cases, the shift-invariant setting is too restrictive. This paper studies the approximation problem in the case Ξ is arbitrary. It establishes approximation theorems whose error bounds reflect the local density of the points in Ξ. Two different settings are analyzed. The first is when the set Ξ is prescribed in advance. In this case, the theorems of this paper show that, in analogy with the classical univariate spline approximation, improved approximation occurs in regions where the density is high. The second setting corresponds to the problem of non-linear approximation. In that setting the set Ξ can be chosen using information about the target function f. We discuss how to 'best' make these choices and give estimates for the approximation error.

AMS subject classification: 42C40, 46B70, 26B35, 42B25

Key Words: image/signal processing, computation, nonlinear approximation, optimal approximation, radial basis functions, scattered data, thin-plate splines, surface splines, approximation order

*This work has been supported by the Office of Naval Research Contracts ONR-N00014-03-1-0051, ONR/DEPSCoR N00014-03-1-0675, ONR/DEPSCoR N00014-05-1-0715; the Army Research Office Contracts DAAD 19-02-1-0028, W911NF-05-1-0227, and W911NF-07-1-0185; the National Institute of General Medical Sciences under Grant NIH-1-R01-GM072000-01; the National Science Foundation under Grants DMS-0221642, DMS-9872890, DMS-354707, DBI-9983114, ANI-0085984 and DMS-0602837

1 Introduction

The mathematical problem of data fitting in the d-variate Euclidean space $\mathbb{R}^d$ has vast applications in science and engineering. Many algorithms address this problem by approximating the data by a linear combination $F = \sum_{\xi\in\Xi} c(\xi)\phi(\cdot - \xi)$, with $\Xi \subset \mathbb{R}^d$, and ϕ a carefully-selected, often radial, function defined on $\mathbb{R}^d$. One of the primary motivations for this approach is that if the data themselves are defined on Ξ and ϕ is chosen properly, then there is a unique function related to the above that interpolates the data [23, 29, 30]. For example, if ϕ is the so-called surface spline (a fundamental solution of the m-fold iterated Laplacian) and Ξ is a given finite set of points in $\mathbb{R}^d$, then for given data (ξ, y_ξ), $\xi \in \Xi$, there is a unique interpolant to the data from the span $S_\Xi(\phi)$ of the $\phi(\cdot - \xi)$, $\xi \in \Xi$.[1] We refer to the book [32] for more details on radial basis functions in general, and their use in interpolation, in particular.

The problem of estimating the interpolation error in the above setting was studied extensively in the literature. We refer the reader to [21, 33], where the interpolated function f is assumed to come from the so-called "native space" (this approach originated in the work of Duchon, [13, 14]) and to the more recent [28, 35, 36, 24], where error estimates are established for general smooth functions in Sobolev spaces. It should be noted that the interpolation problem is usually analyzed for functions defined on bounded subdomains of $\mathbb{R}^d$, and it is well-understood that the interpolation error for this setup suffers significantly from the so-called "boundary effect". Typically, the rates of decay of the error for smooth functions are about half the corresponding decay rates that are valid in the boundary-free shift-invariant case, [19, 17].

Interpolation is not necessarily the best approach to the data fitting problem for various reasons including possible noise in the data, the computational overhead, and possible lack of stability in the algorithms. If the data are given by a function f defined on $\mathbb{R}^d$, i.e. $y_\xi = f(\xi)$ (or $y_\xi \approx f(\xi)$ in the noisy version of the problem), the primary question is how well can f be approximated in a given metric (typically L_p-norms) from the given information. This is governed, and in part determined, by the related question of how well f can be approximated from the span of the translates $\phi(\cdot - \xi)$, $\xi \in \Xi$, in the given metric.

In this paper, we shall be solely concerned with the latter *approximation* problem. We start with a countable set Ξ of points in $\mathbb{R}^d$ and define $S_\Xi(\phi)$ to be the set of all functions which are finite linear combinations of the shifts $\phi(\cdot - \xi)$, $\xi \in \Xi$. We are interested in how well a given function $f \in L_p(\mathbb{R}^d)$ can be approximated (in the L_p-norm) by the elements of $S_\Xi(\phi)$ (more precisely by elements in the closure of this space in the given metric).

[1] The interpolant is actually selected from a space of the form $S'_\Xi(\phi) \oplus P$, with $S'_\Xi(\phi)$ a certain subspace of $S_\Xi(\phi)$, and P a finite dimensional space of polynomials, that depends only on ϕ.

Such approximation problems have been well studied, especially in the case that Ξ is a dilate of the integer lattice ($\Xi = h\mathbb{Z}^d$ with $h > 0$), [7, 3, 8, 6, 27, 18]. The case where the centers Ξ are scattered was studied in [1, 15, 2, 20, 34]. The error bounds in all these references are given in terms of a global mesh density parameter. In contrast, error bounds that depend on the local density of the scattered centers (i.e., provide improved error bounds on subdomains that contain dense clusters of centers) are less studied and less understood, even though it is often the natural setting in applications. The most notable exception, is, of course, spline approximation in one variable. The fact that the error bounds in linear approximation by splines reflects the local mesh ratio, [5], is a key property of spline approximation. Furthermore, the development and analysis of non-linear approximation schemes for univariate splines [25, 10, 11] presented the first challenge for the development of the substantial theory of non-linear approximation. In more than one variable, however, far less is known. We refer to [26], where low-rate strongly local error estimates are established, and to the approximation scheme based on the "power function approach" in [33].

We shall consider two types of problems for scattered center approximation. In the first, we assume that the set Ξ is fixed and we derive results that show improved approximation in regions where the density is high. These results are described in §3 and §5. The second setting that we consider allows the centers to be chosen dependent on the function f. The basic goal is to establish error bounds that depend on the cardinality N of the chosen center set Ξ. This is a form of nonlinear approximation known as N-term approximation which has been well studied in other settings, primarily for wavelet bases. Our result here is similar to the results on nonlinear wavelet approximation. We show that a function can be approximated in $L_p(\mathbb{R}^d)$ with error $O(N^{-s/d})$ once it lies in the Triebel-Lizorkin space $F^s_{\tau,q}(\mathbb{R}^d))$ where s, p, and τ are related (as in the Sobolev embedding theorem) by $\frac{1}{\tau} - \frac{1}{p} = \frac{s}{d}$ and $q = (1 + \frac{s}{d})^{-1}$. From this result and standard embeddings for Triebel-Lizorkin spaces, we derive corresponding theorems for N-term approximation in terms of the Besov classes. While our actual results in this direction are close in nature to the wavelet results, the non-linear approximation algorithm that leads to the above error bounds differs from its wavelet counterpart: the thresholding algorithm that is employed in the wavelet case is sub-optimal in the present case; as such, we introduce and analyse a more sophisticated algorithm. Details of this result are given in §6.

We begin in the following section by describing the assumptions we make about ϕ and Ξ. We then give our first results for the linear approximation problem in §3. In §4 we recall some results about wavelet decompositions and the use of such decompositions in the characterization of smoothness spaces (Triebel-Lizorkin spaces; they include the more standard Sobolev spaces). We stress that our paper is not concerned with wavelets: we merely use wavelets as a tool for defining our approximation schemes. In §5, we complete our study of the linear approximation problem. Finally, we prove in §6 our results on nonlinear approximation.

We shall treat approximation on all of $\mathbb{R}^d$. In most applications, one would be interested in the case of approximation on a compact domain Ω of $\mathbb{R}^d$. Results on domains can be derived easily from our results (if the approximand f is defined on a domain that includes Ω in its interior, and if one agrees to allow centers outside Ω) but we do not pursue this here.

2 The setting

We describe in this section the setting that will be analyzed in the first part of the paper. There are two main ingredients in our setting. The first is the set Ξ of centers which can be allowed. We do not make direct assumptions on the geometry of the set Ξ: almost any set Ξ will do. Once the set Ξ is given, we associate it with a density function

$$h : \mathbb{R}^d \to \mathbb{R}_+.$$

The value $h(t)$ of the density function depends strongly on the local density of Ξ around t: roughly speaking, there should be L centers of Ξ is a ball of radius $L'h(t)$ centered at t, with L, L' dependent only on ϕ and some parameters that we choose for our approximation scheme (and that we fix throughout). The density $h(t)$ depends also on the geometry of the centers around t. This dependence is generally mild; also, our assumptions never spell out this dependence explicitly. It is embedded implicitly in other assumptions.

We assume that our set Ξ of centers is a countable set in $\mathbb{R}^d$ and

A1: (i) *Any finite ball in $\mathbb{R}^d$ contains a finite number of points from Ξ.*
(ii) *For each integer n, there is an $R = R(n)$ such that each ball of radius R contains at least n points from Ξ.*

Property (i) prevents the occurrence of accumulation points in the set Ξ while property (ii) prohibits the existence of arbitrarily large regions on which there are no centers. Neither of the two conditions in **A1** is essential, and the entire assumption is adopted merely in order to simplify the presentation and the analysis.

The first *essential* ingredient in our setting is the identification of the class of basis functions ϕ that our analysis applies to. We will make assumptions on ϕ that are convenient and make the ensuing analysis as transparent as possible and yet general enough to be valid for certain (but not all) types of multivariate splines and radial basis functions.

We assume that ϕ is a locally integrable function which, when viewed on all of $\mathbb{R}^d$, is a tempered distribution. We denote by

$$S_\Xi(\phi)$$

the finite linear span of the functions $\phi(\cdot - \xi)$, $\xi \in \Xi$. We put forward three assumptions: one on ϕ, one on Ξ, and one that connects between ϕ and Ξ. At the end of this section, we analyse these assumptions for specific choices of ϕ.

Let $C_0^\infty := C_0^\infty(\mathbb{R}^d)$ denote the set of all functions that are infinitely differentiable with compact support (test functions) and C_0^k the analogous spaces of k-times continuously differential functions with compact support ($C_0 := C_0^0$). Our first assumption about ϕ is:

A2: *There is a positive integer $\kappa > 0$ and a linear operator T mapping C_0^κ into C_0, such that for all $f \in C_0^\kappa$,*

$$f(x) = \int_{\mathbb{R}^d} T(f)(t)\phi(x-t)\,dt, \quad x \in \mathbb{R}^d. \tag{2.1}$$

Note that the integration is well-defined: since ϕ is locally in L_1, and $T(f) \in C_0$, we have that $T(f)\phi(x - \cdot) \in L_1$, for every $x \in \mathbb{R}^d$.

The typical example of T is a homogeneous elliptic differential operator of order κ with constant coefficients. In this case ϕ is its fundamental solution on $\mathbb{R}^d$. Note that $\hat{\phi}$ is, in this case, a smooth function on $\mathbb{R}^d\backslash 0$. It will be sometimes convenient to add this assumption to **A2**:

$$\hat{\phi} \text{ is smooth on } \mathbb{R}^d\backslash 0.$$

On the other hand, assumption **A2** may appear too strong for certain applications, since it excludes some interesting examples corresponding to fractional differentiation. Those who may wish to extend our theory in such directions may allow $T(f)$ to have global support; it will still need to decay suitably at infinity. All in all, there is some flexibility in the formulation of **A2**.

That said, **A2**, either in its current variant or in some related one, is fundamental: it determines the maximal decay rate of the error that our approximation scheme can yield (this is the parameter κ), and determines the space $W(L_p(\mathbb{R}^d), \phi)$ of smooth functions that can be approximated at this rate. Let us discuss now this latter issue.

The function space $W(L_p(\mathbb{R}^d), \phi)$ corresponds to the Sobolev space $W^\kappa(L_p(\mathbb{R}^d))$ in the case T is an elliptic differential operator of order κ (with constant coefficients). For more general T, the definition is more abstract: Fixing $1 \le p \le \infty$, we define the semi-norm

$$|f|_{W(L_p(\mathbb{R}^d),\phi)} := \|Tf\|_{L_p(\mathbb{R}^d)}, \quad f \in C_0^\infty, \tag{2.2}$$

and the norm:

$$\|f\|_{W(L_p(\mathbb{R}^d),\phi)} := \|f\|_{L_p(\mathbb{R}^d)} + |f|_{W(L_p(\mathbb{R}^d),\phi)}. \tag{2.3}$$

Using the above, we define $W(L_p(\mathbb{R}^d),\phi)$ to be the completion of C_0^∞ in this topology. Since the norm in (2.3) is stronger than the L_p-norm, $W(L_p,\phi)$ is precisely the space of all functions $f \in L_p(\mathbb{R}^d)$ for which there is a sequence $(f_n)_{n\geq 1} \subset C_0^\infty$ such that f_n converges to f and $(T(f_n))$ converges to a function $g \in L_p(\mathbb{R}^d)$ both in the sense of $L_p(\mathbb{R}^d)$. By making suitable assumptions on ϕ we can conclude that g depends on f, but not on the specific sequence (f_n). The operator T, which was initially defined only for test functions, now extends naturally to a linear operator on $W(L_p,\phi)$ by defining $Tf := g$.

We are guided in the above setup by the following example, [13, 14]: if T is the m-fold Laplacian, then $\kappa = 2m$ and the function ϕ is then the fundamental solution of T:

$$\phi = c_m \begin{cases} |\cdot|^{2m-d}, & d \text{ odd}, \\ |\cdot|^{2m-d}\log|\cdot|, & d \text{ even}. \end{cases} \tag{2.4}$$

(Here, $|\cdot|$ stands for the Euclidean norm in $\mathbb{R}^d$.) The function ϕ is also called a *surface spline*. In this case, $W(L_p(\mathbb{R}^d),\phi)$ is simply the Sobolev space $W^{2m}(L_p(\mathbb{R}^d))$ equipped with its usual semi-norm and norm.

Our remaining assumption about ϕ concerns how well its translates can be approximated from $S_\Xi(\phi)$. Consider the translate $\phi(\cdot - t)$ where $t \in \mathbb{R}^d$ is fixed for the moment. We look for a local approximation of the form

$$K(\cdot,t) = \sum_{\xi\in\Xi(t)} A(t,\xi)\phi(\cdot - \xi), \tag{2.5}$$

for suitable t-dependent coefficients $A(t,\xi)$ and a finite set $\Xi(t) \subset \Xi$. A key component in the success of our approach is the availability of kernels $K(\cdot,\cdot)$ that are local and bounded on the one hand, and approximate well the convolution kernel $(x,t) \mapsto \phi(x-t)$. We break this assumption into two: **A3** deals with basic qualitative properties of the scheme that is used to define K, viz., the coefficient functions $A(\cdot,\xi)$. The companion property, **A4**, deals with the way K approximates the convolution kernel.

A3: *There is an integer $n' > 0$ and a real number M_0 such that for any $t \in \mathbb{R}^d$ the set $\Xi(t)$ consists of at most n' points all lying in the ball of radius M_0 centered at t and the coefficients of the approximation kernel* (2.5) *satisfy $A(\cdot,\xi) \in L_1(\mathbb{R}^d)$ for all $\xi \in \Xi$.*

Similar to condition **A1**, Condition **A3** is secondary, and is formulated and adopted in order to exclude pathological kernels K. This brings us to our last assumption. As said, the last assumption is concerned with the way $K(\cdot,t)$ approximates the translate $\phi(\cdot - t)$. This assumption must be dealt with care: the error $E(\cdot,t) := \phi(\cdot - t) - K(\cdot,t)$ should reflect not only the basic properties of ϕ, but also the local distribution of the center set Ξ around t. To this end, we define for each $t \in \mathbb{R}^d$

$$h(t) := \inf\{\rho : A(t,\xi) \equiv 0,\ |\xi - t| \geq \rho\}. \tag{2.6}$$

In other words, for each $t \in \mathbb{R}^d$, the only centers from Ξ used in the approximation kernel $K(\cdot,t)$ lie in a ball $B_t(h(t))$ of radius $h(t)$ centered at t. In our approach, $h(t)$ measures the "effective" local density of the set Ξ around the point t. We will discuss this issue in the sequel. Right now, let us complete our basic assumptions. We define the error kernel

$$E(x,t) := \phi(x-t) - K(x,t), \quad x,t \in \mathbb{R}^d. \tag{2.7}$$

Notice that for each $t \in \mathbb{R}^d$, E is a finite linear combination of translates of ϕ using centers from $t \cup \Xi$. We shall assume that

A4: *There is a positive real number $\nu > d$ and a constant $C > 0$ depending only on ϕ such that*

$$|E(x,t)| \leq C h(t)^{\kappa-d}\Big(1 + \frac{|x-t|}{h(t)}\Big)^{-\nu}, \quad x,t \in \mathbb{R}^d, \tag{2.8}$$

where κ is the integer in **A2**.

As we will see in the examples that follow, the local density $h(t)$ must be chosen to satisfy two properties: first, the ball $B_t(h(t)) := \{x \in \mathbb{R}^d :\ |x-t| \leq h(t)\}$ must contain a minimal number of centers from Ξ. This number is determined by the parameter ν. However, choosing $h(t)$ at this minimal value leads to error kernel $E(\cdot,t)$ that are too large, forcing us to select a large constant C. In such a case, it is usually preferable to increase $h(t)$, so that more centers are captured in the enlarged ball $B_t(h(t))$. By playing this game correctly at all points t, we can control the global constant C. Needless to say, this comes at a price, since the density function h will enter our error bounds as well.

The remainder of this section will discuss two examples where the assumptions **A1-4** are satisfied. These examples will provide a better understanding of the assumptions as well as of the nature of the smoothness spaces W and the density function h.

Example 1: Univariate splines.
We consider the truncated power $\phi(t) := t_+^{\kappa-1}$ defined on $\mathbb{R}$. We have the elementary and well-known representation

$$f(x) = \frac{1}{(\kappa-1)!}\int_{-\infty}^{\infty} f^{(\kappa)}(t)\phi(x-t)\,dt, \tag{2.9}$$

which holds for all functions in $C_0^\kappa(\mathbb{R})$. This means that **A2** holds for $T := D^\kappa/(\kappa-1)!$. Now given $t \in \mathbb{R}$, let $\Xi(t) = \{\xi_1(t), \ldots, \xi_\kappa(t)\} \subset \Xi$ be the set of the κ points in Ξ that are closest to t. The divided difference

$$[t, \xi_1(t), \ldots, \xi_\kappa(t)]\phi(x - \cdot) =: a(t)\phi(x - t) - \sum_{\xi \in \Xi(t)} A_0(t, \xi)\phi(x - \xi) \tag{2.10}$$

is the B-spline $x \mapsto M(x)$ associated with the knots $\{t, \xi_1(t), \ldots, \xi_\kappa(t)\}$. Thus, we can take $A(t, \xi) = A_0(t, \xi)/a(t)$ for $\xi \in \Xi(t)$ and $A(t, \xi) = 0$ otherwise. It follows that $h(t) = \max_{j=1,\ldots,\kappa} |t - \xi_j(t)|$. A well-known property of divided differences (see [12], p. 121) gives

$$|a(t)| = \prod_{j=1}^{\kappa} |t - \xi_j(t)|^{-1} \geq h(t)^{-\kappa}. \tag{2.11}$$

Since $|E(x, t)| = M(x)/|a(t)| \leq h(t)^\kappa M(x)$, assumption **A4** follows from the facts that $M(x) \leq 1/h(t)$ and that $M(x)$ vanishes for $x \notin (t - h(t), t + h(t))$. As to the constant C, it can be chosen as $C := 4$ (for $\nu := 2$). The verification of **A3** is straightforward (via **A1**).

It is worth stressing the fact that our ability to provide tight error estimates for univariate splines is not only due to the banded structure of the error kernel E. It was also due to the fact that univariate spline theory tells us that $M(x) \leq 1/h(t)$: thus, while the actual coefficient $A(\cdot, \xi_i)$ of the truncated power $(\cdot - \xi_i)_+^{\kappa-1}$ in the representation of $M(x)$ can be arbitrarily large, the size of this coefficient does not affect $\|M\|_\infty$. Unfortunately, we are not aware of a multivariate counterpart of this result. ∎

Example 2: Surface splines.

The multivariate analog of the truncated power is the surface spline (see (2.4); it is also known as the polyharmonic spline.) The best known surface spline is the bivariate *thin-plate spline*

$$\phi = |\cdot|^2 \log |\cdot|,$$

which is, up to a constant, the fundamental solution for Δ^2 when $d = 2$.

We have noted earlier that property **A2** holds for surface splines ϕ of any dimension. We will analyse in detail **A4**, and will briefly discuss **A3**.

Let $\nu > d$ be the number that appears in **A4**, and define

$$n := \kappa - d + \nu.$$

Let P be the space of all polynomials of degree $< n$ in d variables. Given a finite set $Z \subset \mathbb{R}^d$, we denote by Λ_Z the span of the functionals

$$\delta_z \in P', \; z \in Z, \quad \delta_z(p) := p(z), \; p \in P.$$

Every $\lambda := \sum_{z\in Z} a(z)\delta_z \in \Lambda_Z$ extends to $C(\mathbb{R}^d)'$, with the norm of the extension being

$$\|\lambda\| = \sum_{z\in Z} |a(z)|.$$

Now, for every $t \in \mathbb{R}^d$, we select a finite subset $\Xi(t) \subset \Xi$ that satisfies the following properties:

(i) $h(t) := \max\{|t-\xi| : \xi \in \Xi(t)\}$ is "as small as possible".

(ii) There exists $\lambda_t \in \Lambda_{\Xi(t)}$ such that λ_t agrees with δ_t on P.

(iii) $\|\lambda_t\| \le C$, for some t-independent constant C that we choose in advance.

Let us first remark that there are always sets $\Xi(t)$ satisfying (ii) and (iii).[2] We have said nothing about the size of the constant C in (iii). We do not provide specific algorithms for choosing optimally C. The general rule of thumb is that by choosing $\Xi(t)$ to contain the $r + \dim P$ points in Ξ that are closest to t, with r a small positive integer, we should be able to find (with high probability, for a generic distribution of centers) λ_t that satisfies the above.

Once we have chosen the functional $\lambda_t =: \sum_{\xi\in\Xi(t)} A(t,\xi)\delta_\xi$, we define

$$A(t,\xi) = 0, \quad \text{on } \Xi\backslash\Xi(t),$$

and define the kernels

$$K(\cdot,t) := \sum_{\xi\in\Xi} A(t,\xi)\phi(\cdot-\xi), \qquad E(\cdot,t) := \phi(\cdot-t) - K(\cdot,t).$$

Next, recall that (up to a constant that depends only on d and κ) $\phi = |\cdot|^{\kappa-d}L$, with $L = \log|\cdot|$ whenever $\kappa - d$ is an even integer, and $L = 1$ otherwise.

We now complete the proof of property **A4**. Let Ξ and t be given. While we must verify **A4** for every t and every set Ξ satisfying (i-iii), we can (by translating both t and Ξ) assume without loss of generality that $t = 0$. Let $h := h(t)$ be as in (i). Suppose first that $|x| > 2h$. If R is any polynomial of degree $< n$, then

$$|\phi(x) - K(x,0)| = |\phi(x) - \lambda_0(\phi(x-\cdot))| \le |\phi(x) - R(x)| + |\lambda_0(\phi(x-\cdot) - R(x-\cdot))|. \quad (2.12)$$

[2] A sketch of the argument is as follows. First, the claim is definitely correct when $\Xi = \mathbb{Z}^d$. Therefore, there exists $\delta > 0$, such that the claim is correct, provided that Ξ has non-zero intersection with any ball $B_a(\delta)$, $a \in \mathbb{Z}^d$. Since $\|\lambda\|$ is invariant under dilation, the claim is thus correct provided that Ξ has non-zero intersection with any ball $B_a(R(1))$, $a \in \frac{R(1)}{\delta}\mathbb{Z}^d$, and with $R(1)$ as in **A1**. Thanks to assumption **A1**, our Ξ satisfies, indeed, this last property. The argument as here is of mostly theoretical value, since it employs a localization process that involves only a small subset of Ξ, and results therefore in a density function that is prohibitively large.

In particular, choosing R as the Taylor polynomial of degree $n-1$ at x of ϕ, the first term on the right side of (2.12) is zero and we obtain

$$|\phi(x) - K(x,0)| \le C\|\phi - R\|_{L_\infty(B)}$$

where C is the constant in (iii) and B is the ball of radius h about x. From the Taylor remainder formula we obtain

$$|\phi(x) - K(x,0)| \le \frac{C|\phi|_{W^n(L_\infty(B))}h^n}{n!} \le CC'|x|^{\kappa-d-n}h^n \le CC''h^{\kappa-d}[1+\frac{|x|}{h}]^{-\nu}, \quad (2.13)$$

where the constants C', C'' depend on ϕ and ν but are independent of Ξ, t, x. Here, we used the fact that $|x|$ and $|x'|$ are comparable for $x' \in B$ because $|x| > 2h$.

If $|x| \le 2h$, then $|x-\xi| \le 3h$ for every $\xi \in \operatorname{supp}\lambda_0$. Assume momentarily that $h = 1$. Then we simply estimate

$$|\phi(x) - K(x,0)| = |\phi(x) - \lambda_0(\phi(x-\cdot))| \le (1 + \|\lambda_0\|)\,\|\phi\|_{L_\infty(B)} \le CC''', \qquad (2.14)$$

where now B is the ball of radius 3 about the origin, C''' depends only on ϕ and n, and C is the constant in(iii).

Now, suppose that $h := h(0) \ne 1$. Then, dilating ϕ, Ξ and λ_0 by h, we note that

$$\phi(x/h) - \lambda_0(\phi((x-\cdot)/h)) = h^{d-\kappa}(\phi(x) - \lambda_0(\phi(x-\cdot))) + (q(x) - \lambda_0(q(x-\cdot))),$$

with q a polynomial of degree $\le \kappa - d$ (viz., $q = 0$ for odd $\kappa - d$, and $q = -|\cdot|^{\kappa-d}h^{d-\kappa}\log h$ otherwise). Since $\delta_0 - \lambda_0$ annihilates all such polynomials, we conclude that

$$\phi(x/h) - \lambda_0(\phi((x-\cdot)/h)) = h^{d-\kappa}(\phi(x) - \lambda_0(\phi(x-\cdot))). \qquad (2.15)$$

Invoking now the analysis of the $(h = 1)$-case, we have that

$$|\phi(x/h) - \lambda_0(\phi((x-\cdot)/h))| \le CC''',$$

and hence, by (2.15),

$$|\phi(x) - \lambda_0(\phi(x-\cdot))| \le CC'''h^{\kappa-d},$$

provided that $|x| \le 2h$. Altogether, for $x \in \mathbb{R}^d$,

$$|\phi(x) - K(x,0)| \le CC_1h^{\kappa-d}[1+\frac{|x|}{h}]^{-\nu}, \qquad (2.16)$$

with C as in (iii), and C_1 a universal constant.

For general $t \in \mathbb{R}^d$, an argument identical to the above leads to

$$|\phi(x-t) - K(x,t)| \le CC_1h(t)^{\kappa-d}(1 + |x-t|/h(t))^{-\nu}.$$

This validates **A4**. It also shows the constant that appears in **A4** is the product of a constant that is independent of Ξ, x, t by the uniform bound for the norms $\|\lambda_t\|$, $t \in \mathbb{R}^d$.

Concerning property **A3**, for each fixed ξ, the function $A(\cdot, \xi)$ has compact support because of assumption **A1** and our remarks above about the choice of $\Xi(t)$. Since this function is also uniformly bounded as we have shown in the discussion of (iii), we see that **A3** is also satisfied.

■

3 Approximation with a prescribed set Ξ of centers

In this section, we assume that the set Ξ of centers is fixed in advance. We work under the assumption that Ξ, ϕ satisfy **A1-A4**. We shall prove a theorem for the approximation of a given function $f \in W(L_p(\mathbb{R}^d), \phi)$ by the elements of $S_\Xi(\phi)$. In §5, we will extend the results of this section to more general functions in $L_p(\mathbb{R}^d)$.

Since our goal is to derive error estimates that reflect the local density of Ξ, it may seem that we can employ our measure of density $t \mapsto h(t)$ in such estimates. However, it turns out that h may change too rapidly to allow effective error analysis (unless one replaces **A4** by the stronger assumption that $E(x,t)$ is supported in the domain $\{(x,t) : |x-t| \le Ch(t)\}$. However, the only interesting example that satisfies this stronger condition is univariate splines). To circumvent this difficulty, we introduce a companion density function H that varies more slowly than the original h.

Given Ξ, ϕ and a local density h that satisfy **A1-A4**, we define

$$H(x) := \sup_{t \in \mathbb{R}^d} h(t) \left(1 + \frac{|x-t|}{h(t)}\right)^{-r}, \quad x \in \mathbb{R}^d, \tag{3.1}$$

where r is any fixed number satisfying

$$0 < r < \frac{\nu - d}{\kappa}, \tag{3.2}$$

and with ν as in assumption (2.8). The larger the value of r, the smaller the density function H and the better the estimates that we shall obtain.[3] Notice also that in the examples in the previous section, the number ν appearing in **A4** can be chosen arbitrary large; however, the constant that appears in **A4** and the density function h depend both on the selection of this ν.

[3]It is therefore natural to try to take $r = \frac{\nu-d}{\kappa}$ in the analysis we give below; but this fails to work (barely). One could introduce logarithmic factors in the definition of H and get slightly improved results but at the expense of notational complications that we want to avoid.

We assume that $f \in W(L_p(\mathbb{R}^d), \phi)$ and then derive an error estimate for approximating f by the elements of $S_\Xi(\phi)$. We first want to enlarge the space $S_\Xi(\phi)$ to include certain infinite sums. Given $1 \le p \le \infty$, we define

$$S_\Xi(\phi)_p$$

to be the closure of $S_\Xi(\phi)$ in the topology of convergence in $L_p(\Omega)$ for each compact $\Omega \subset \mathbb{R}^d$.

To describe the approximation procedure we are going to use, we recall the kernel K (given by (2.5)) which describes how $\phi(\cdot - t)$ is approximated. Also recall the error bound (2.8) for $E := \phi(\cdot - t) - K(\cdot, t)$ which we assume to hold for this approximation.

Given any positive weight function w defined on $\mathbb{R}^d$, we define the norm

$$\|g\|_{L_p(w)} := \|wg\|_{L_p(\mathbb{R}^d)}, \quad 1 \le p \le \infty, \tag{3.3}$$

and the approximation error

$$\mathcal{E}(f, S_\Xi(\phi))_{L_p(w)} := \inf_{S \in S_\Xi(\phi)_p} \|f - S\|_{L_p(w)}, \quad 1 \le p \le \infty. \tag{3.4}$$

Notice that $\|\cdot\|_{L_p(w)}$ differs from the more usual definition of weighted L_p-norms.

Theorem 3.1 *Suppose that Ξ, ϕ satisfy* **A1-4**, *and let $1 \le p \le \infty$. If $f \in W(L_p(\mathbb{R}^d), \phi)$ and $w := H^{-\kappa}$, then we have*

$$\mathcal{E}(f, S_\Xi(\phi)_p)_{L_p(w)} \le C_0 |f|_{W(L_p(\mathbb{R}^d), \phi)} \tag{3.5}$$

with $C_0 = CC'$, C' being dependent only on ϕ, ν and r, and C is the constant that appears in **A4**.

The parameter κ (which was introduced in Assumptions **A2, A4**) determines the rate of decay of the error. Note that κ appears on both sides of (3.5) (in the definition of $L_p(w)$, as well as in the definition of $W(L_p(\mathbb{R}^d), \phi)$).

The theorem provides local error estimates in terms of the density H. Where H is small, i.e. Ξ is dense, the approximation bound is better. The local nature of the error estimates is best captured in the case $p = \infty$:

Corollary 3.2 *In the notations and assumptions of Theorem 3.1, the error bound in the case $p = \infty$ can be restated as follows: for every compact $\Omega \subset \mathbb{R}^d$ and every $\epsilon > 0$, there exists $g \in S_\Xi(\phi)$, such that, for every $x \in \Omega$,*

$$|f(x) - g(x)| \le \epsilon + C_0 H(x)^\kappa |f|_{W(L_p(\mathbb{R}^d), \phi)}. \tag{3.6}$$

The constant C_0 is independent of f, g, ϵ, Ω and x.

Proof of Theorem 3.1. We begin by assuming that $f \in C_0^\kappa(\mathbb{R}^d)$ and later use a completion argument to derive the general case. We shall establish the estimate (3.5) for $p = 1, \infty$ and then derive the general case by interpolation. For any function $g \in L_1(\mathbb{R}^d) + L_\infty(\mathbb{R}^d)$, we define

$$L(g,x) := \int_{\mathbb{R}^d} H(x)^{-\kappa} g(t) E(x,t)\, dt. \tag{3.7}$$

Then L is a linear operator and we shall show that it maps L_p boundedly into itself for $p = 1, \infty$. Once this is established, the Marcinkiewicz interpolation theorem implies that L maps $L_p(\mathbb{R}^d)$ boundedly into itself for all $1 \le p \le \infty$.

First consider the case $p = 1$. We invoke the estimate (2.8) and the definition of H given in (3.1) to find

$$H(x)^{-\kappa}|E(x,t)| \le C H(x)^{-\kappa} h(t)^{\kappa-d} \left(1 + \frac{|x-t|}{h(t)}\right)^{-\nu} \le C h(t)^{-d} \left(1 + \frac{|x-t|}{h(t)}\right)^{-\nu + r\kappa}. \tag{3.8}$$

Thus

$$\begin{aligned} \|L\|_{L_1(\mathbb{R}^d)} &= \sup_{t\in\mathbb{R}^d} \int_{\mathbb{R}^d} H(x)^{-\kappa}|E(x,t)|\, dx \\ &\le \sup_{t\in\mathbb{R}^d} C \int_{\mathbb{R}^d} h(t)^{-d} \left(1 + \tfrac{|x-t|}{h(t)}\right)^{-\nu + r\kappa} dx \\ &= C \int_{\mathbb{R}^d} (1 + |y|)^{-\nu + r\kappa}\, dy \le C', \end{aligned} \tag{3.9}$$

where we used the fact that $-\nu + r\kappa < -d$.

For the case $p = \infty$, we fix $x \in \mathbb{R}^d$ and define for each $j \in \mathbb{Z}$ the set

$$\Omega_j := \{t \in \mathbb{R}^d : 2^{j-1} \le \frac{h(t)}{H(x)} < 2^j\} \tag{3.10}$$

Then,

$$\int_{\mathbb{R}^d} H(x)^{-\kappa}|E(x,t)|\, dt = \sum_{j\in\mathbb{Z}} \int_{\Omega_j} H(x)^{-\kappa}|E(x,t)|\, dt =: \sum_{j\in\mathbb{Z}} I_j. \tag{3.11}$$

We can estimate each of the integrals I_j appearing in the sum by using (2.8) to obtain

$$I_j \le C 2^{j\kappa} \int_{\Omega_j} [2^j H(x)]^{-d} \left(1 + \frac{|x-t|}{2^j H(x)}\right)^{-\nu} dt = C 2^{j\kappa} \int_{\Omega_j'} (1 + |y|)^{-\nu}\, dy, \tag{3.12}$$

where $\Omega_j' = [2^j H(x)]^{-1}(x - \Omega_j)$. Since $\nu > r\kappa + d > d$, it is clear that $\sum_{j\le 1} I_j \le C'$. For $j > 1$, we use the definition of H to find

$$\left(1 + \frac{2|x-t|}{2^j H(x)}\right)^r \ge \left(1 + \frac{|x-t|}{h(t)}\right)^r \ge \frac{h(t)}{H(x)} \ge 2^{j-1}, \quad t \in \Omega_j. \tag{3.13}$$

In other words,

$$\frac{|x-t|}{2^j H(x)} \geq \frac{2^{(j-1)/r}-1}{2} =: a_j, \quad t \in \Omega_j. \tag{3.14}$$

This means that

$$I_j \leq C 2^{j\kappa} \int\limits_{|y|\geq a_j} (1+|y|)^{-\nu}\, dy \leq C 2^{j\kappa} a_j^{-\nu+d} \tag{3.15}$$

Since $\kappa < \frac{\nu-d}{r}$, we have that $\sum_{j>1} 2^{j\kappa} a_j^{-\nu+d}$ is finite. This, together with (2.13), yields

$$\|L\|_{L_\infty(\mathbb{R}^d)} = \sup_{x\in\mathbb{R}^d} \int\limits_{\mathbb{R}^d} H(x)^{-\kappa} |E(x,t)|\, dt \leq C. \tag{3.16}$$

Consequently, we have proved that L boundedly maps L_p into itself for every $1 \leq p \leq \infty$.

Now, if $f \in W(L_p(\mathbb{R}^d), \phi)$, then by the definition of this space, $Tf \in L_p(\mathbb{R}^d)$. Hence, from what we have already proved,

$$\|L(Tf)\|_{L_p(\mathbb{R}^d)} \leq C_0 \|Tf\|_{L_p(\mathbb{R}^d)} = C_0 |f|_{W(L_p(\mathbb{R}^d),\phi)}, \tag{3.17}$$

with $C_0 = CC'$, with C' an absolute constant, and C the constant that appears in **A4**. Assume next that $f \in C_0^\kappa(\mathbb{R}^d)$ and define

$$F := \int\limits_{\mathbb{R}^d} Tf(t) K(\cdot, t)\, dt. \tag{3.18}$$

From **A1-2** and the fact that $\operatorname{supp} Tf$ is compact, we deduce that the sum that defines $K(\cdot,\cdot)$ is finite on $\mathbb{R}^d \times \operatorname{supp} Tf$. It follows then that

$$F = \sum_{\xi\in\Xi} a(\xi)\phi(x-\xi), \quad a(\xi) := \int\limits_{\mathbb{R}^d} A(t,\xi) Tf(t)\, dt. \tag{3.19}$$

As said, the sum that defines F is actually finite. Thus, $F \in S_\Xi(\phi)$.

Continuing under the assumption that $f \in C_0^\kappa(\mathbb{R}^d)$, we have

$$H(x)^{-\kappa}[f(x) - F(x)] = H(x)^{-\kappa} \int\limits_{\mathbb{R}^d} Tf(t) E(x,t)\, dt = L(Tf)(x). \tag{3.20}$$

It follows from what we have already proved that

$$\mathcal{E}(f, S_\Xi(\phi))_{w,p} \leq \|f - F\|_{L_p(w)} \leq C_0 \|Tf\|_{L_p(\mathbb{R}^d)} = C_0 |f|_{W(L_p(\mathbb{R}^d),\phi)}, \quad 1 \leq p \leq \infty. \tag{3.21}$$

We now want next to extend (3.21) to all of $W(L_p(\mathbb{R}^d), \phi)$. Fix $p \in [1, \infty]$, and let $f \in W(L_p(\mathbb{R}^d), \phi)$. By the definition of $W(L_p(\mathbb{R}^d), \phi)$ there is a sequence of compactly supported functions f_n, $n = 1, 2, \ldots$, from C_0^κ such that $f_n \to f$ in the norm of

$W(L_p(\mathbb{R}^d),\phi)$. Let F_n be defined by (3.18) for f_n. We know that each of these F_n is in $S_\Xi(\phi)$. For any compact set Ω, we have $w(x)\ge c_\Omega>0$, $x\in\Omega$, and therefore by writing $F_m-F_n=F_m-f_m-(F_n-f_n)+(f_m-f_n)$ we find

$$\begin{aligned}\|F_m-F_n\|_{L_p(\Omega)} &\le \|f_m-f_n\|_{L_p(\Omega)}+c_\Omega^{-1}\|f_m-f_n-(F_m-F_n)\|_{L_p(w)}\\ &\le \|f_m-f_n\|_{L_p(\mathbb{R}^d)}+C_\Omega\|T(f_m-f_n)\|_{L_p(\mathbb{R}^d)},\end{aligned} \tag{3.22}$$

where the last inequality uses (3.21) for $f=f_m-f_n$. This shows that (F_n) is a Cauchy sequence in the topology of L_p-convergence on compact sets. By definition, its limit G is in $S_\Xi(\phi)_p$. Again, for any compact set Ω in $\mathbb{R}^d$, we have

$$\|w(f-G)\|_{L_p(\Omega)}\le \lim_{n\to\infty}\|f_n-F_n\|_{L_p(w)}\le C_0\lim_{n\to\infty}\|T(f_n)\|_{L_p(\mathbb{R}^d)}=C_0\|T(f)\|_{L_p(\mathbb{R}^d)}, \tag{3.23}$$

with C_0 the constant of (3.21). Since Ω is arbitrary, we find

$$\|w(f-G)\|_{L_p(\mathbb{R}^d)}\le C_0\|T(f)\|_{L_p(\mathbb{R}^d)}. \tag{3.24}$$

Since $G\in S_\Xi(\phi)_p$, we can replace the left side of (3.24) by $\mathcal{E}(f,S_\Xi(\phi)_p)_{L_p(w)}$. This completes the proof of the theorem. ∎

Theorem 3.1 deals with the approximation of functions that are optimally smooth, i.e., in the space $W(L_p(\mathbb{R}^d),\phi)$. In §5, we establish results concerning the approximation of functions that are less smooth. For such functions, the weight $H^{-\kappa}$ is too strong. To this end, we state a counterpart of Theorem 3.1 for mollified versions of the original weight w. We still assume here that f is optimally smooth. This assumption will be dropped in §5.

Suppose that $0<s<\kappa$. We continue to work under the assumptions **A1-4**. If $f\in W(L_p(\mathbb{R}^d),\phi)$ then

$$\|h^{\kappa-s}Tf\|_{L_p(\mathbb{R}^d)} \tag{3.25}$$

is finite because h is bounded (see **A1 (ii)**).

Theorem 3.3 *If $0<s<\kappa$ and $f\in W(L_p(\mathbb{R}^d),\phi)$ (for some $1\le p\le\infty$), then for $w:=H^{-s}$ we have*

$$\mathcal{E}(f,S_\Xi(\phi))_{L_p(w)}\le C_0\|h^{\kappa-s}Tf\|_{L_p(\mathbb{R}^d)}. \tag{3.26}$$

with C_0 as in Theorem 3.1.

Proof: The proof is very similar to the proof of Theorem 3.1. We first remark that the two bounds

$$\sup_{t\in\mathbb{R}^d}\int_{\mathbb{R}^d}H(x)^{-s}|E(x,t)|h(t)^{s-\kappa}\,dx\le C,\quad \sup_{x\in\mathbb{R}^d}\int_{\mathbb{R}^d}H(x)^{-s}|E(x,t)|h(t)^{s-\kappa}\,dt\le C, \tag{3.27}$$

hold with C a constant depending only on d, s. Indeed, thanks to Assumption **A4** it is sufficient to prove the above boundedness with the integrand replaced by

$$H(x)^{-s}h(t)^{\kappa-d}h(t)^{s-\kappa}\left(1+\frac{|x-t|}{h(t)}\right)^{-\nu}=H(x)^{-s}h(t)^{s-d}\left(1+\frac{|x-t|}{h(t)}\right)^{-\nu}.$$

Since we assume $\nu > \kappa r + d \geq sr + d$, the argument as given in the proof of Theorem 3.1 applies here *verbatim* to yield (3.27) .

The bounds in (3.27) now imply that the linear operator

$$L(g) := H(x)^{-s}\int_{\mathbb{R}^d} g(t)E(x,t)h^{s-\kappa}(t)\,dt$$

is bounded on $L_p(\mathbb{R}^d)$ for $p = 1, \infty$. By interpolation, we derive that this operator is bounded on $L_p(\mathbb{R}^d)$ for all $1 \leq p \leq \infty$. Using this for $g = h^{\kappa-s}T(f)$, we derive (3.26) for all $f \in W(\phi, L_p(\mathbb{R}^d))$ in the same way we have proven Theorem 3.1. ∎

4 Wavelet decompositions

In the remaining sections of this paper, we shall be in need of a local multiscale basis on $\mathbb{R}^d$. We shall employ a standard multivariate wavelet basis for this purpose. This basis will be used only as a tool for proving various results. In this section, we recall the form of such a basis and some of its properties which will be important to us. In particular, we shall need its characterization of Triebel-Lizorkin spaces. There are several books that discuss wavelet decompositions and their characterization of these spaces (see e.g. [22]). We also refer to the article of Daubechies [4] for the construction of wavelet bases of the type we want to use.

Let $\mathcal{D}$ denote the set of dyadic cubes in $\mathbb{R}^d$ and $\mathcal{D}_j$ the set of dyadic cubes of side length 2^{-j} (thus $\mathcal{D} = \cup_{j=-\infty}^{\infty}\mathcal{D}_j$). Each $I \in \mathcal{D}_j$ is of the form

$$I = 2^{-j}[k_1, k_1+1]\times\cdots\times 2^{-j}[k_d, k_d+1] = 2^{-j}(k+[0,1]^d), \quad k = (k_1,\ldots,k_d)\in\mathbb{Z}^d. \tag{4.1}$$

For each such I, we denote its side length by $\ell(I)$:

$$\ell(I) := 2^{-j}, \quad \forall I \in \mathcal{D}_j.$$

Finally, let

$$E := \{1,\ldots,2^d-1\}, \quad \mathcal{V} := \mathcal{D}\times E.$$

Given $v = (I_v, e_v)$ ($I_v \in \mathcal{D}$, $e_v \in E$), we denote

$$\ell(v) := \ell(I_v),$$

and by

$$|v| := \ell(v)^d$$

the volume of the cube I_v.

A *wavelet basis* is an orthonormal basis for $L_2(\mathbb{R}^d)$ with particular structure and properties. The wavelets are indexed by the set $\mathcal{V}$:

$$w_v, \quad v \in \mathcal{V}.$$

Each wavelet w_v, with $I_v = 2^j(k + [0,1]^d) \in \mathcal{D}$, is supported in a cube $\bar{I}_v$, with

$$\bar{I}_v = 2^j(k + A_0[0,1]^d),$$

with A_0 some fixed constant that depends only on the specifics of the wavelet system we choose.

We normalized initially the wavelet system in $L_2(\mathbb{R}^d)$. The L_p-norm of the p-normalized wavelets, $1 \le p \le \infty$:

$$\psi_{v,p} := |v|^{\frac{1}{2}-\frac{1}{p}} w_v \tag{4.2}$$

depends only on their *type*, i.e., on the index $e_v \in E$. Each locally integrable function f defined on $\mathbb{R}^d$ has a wavelet decomposition

$$f = \sum_{v \in \mathcal{V}} f_v \psi_{v,p}, \quad f_v := f_{v,p} := \langle f, \psi_{v,p'} \rangle, \quad 1/p + 1/p' = 1. \tag{4.3}$$

Here f_v depends on the p-normalization that has been chosen but the product $f_v\psi_{v,p}$ is independent of p. In this paper, it will be convenient to normalize the wavelets in L_∞:

$$\psi_v := \psi_{v,\infty} = |v|^{\frac{1}{2}} w_v.$$

The series (4.3) converges absolutely to f in the L_p-norm in the case $f \in L_p(\mathbb{R}^d)$ and $1 \le p < \infty$, with $H_1(\mathbb{R}^d)$ replacing $L_1(\mathbb{R}^d)$, and conditionally in the case $p = \infty$ with $L_\infty(\mathbb{R}^d)$ replaced by $C(\mathbb{R}^d)$.

One of the most important properties of wavelet systems (and the one we need in this paper) is the characterization of smoothness spaces in terms of the wavelet decomposition. This means that we can use the wavelet decomposition in order to *define* those smoothness spaces. For the definition of Triebel-Lizorkin spaces $F^s_{p,q}$ in terms of wavelet coefficients, we fix $s > 0$. Then, for $0 < p, q < \infty$,

$$|f|_{F^s_{p,q}} := \|M_q(f)\|_{L_p(\mathbb{R}^d)}, \tag{4.4}$$

where

$$M_s(f)(x) := M_{s,q}(f)(x) := \left(\sum_{v \in \mathcal{V}} \ell(v)^{-qs} |f_{v,\infty}|^q \chi_{\bar{I}_v}(x) \right)^{1/q}. \tag{4.5}$$

The definition does not depend on the wavelet system we choose, provided that the wavelets are m-times differentiable, and have m vanishing moments (viz., their Fourier transform has a m-fold zero at the origin), for sufficiently large m.[4]

We should also make some specific remarks about our definition. In our definition we have defined the maximal function with $\chi_{\bar{I}_v}$ where $\bar{I}_v$ the support cube of the ψ_v. The usual definition of $F^s_{p,q}$ uses χ_{I_v} instead of $\chi_{\bar{I}_v}$. It is easy to see that these two definitions give equivalent norms by using the Fefferman-Stein inequality mentioned in the proof of Lemma 6.1. We also remark that Triebel-Lizorkin spaces are usually defined using Littlewood-Paley decompositions. Our definition agrees with the classical definition if (the wavelet system is chosen appropriately, cf. the above footnote)and the space $F^s_{p,q}$ continuously embeds into L_1.

The definition extends naturally to the $q = \infty$ case, with $M_\infty(f)$ defined by

$$M_\infty(f)(x) := M_{s,\infty}(f)(x) := \sup_{v \in \mathcal{V}} \ell(v)^{-s} |f_{v,\infty}| \chi_{\bar{I}_v}(x).$$

While we followed so far the tradition of of assuming $p < \infty$, we will use the space $F^s_{\infty,\infty}$, whose semi-norm is

$$|f|_{F^s_{\infty,\infty}} := \sup_{v \in \mathcal{V}} \ell(v)^{-s} |f_{v,\infty}| \chi_{\bar{I}_v}(x),$$

and which usually appears in the literature as the Besov space $B^s_{\infty,\infty}$. Notice that $B^s_{\infty,\infty}$ is compactly embedded in $C(\mathbb{R})$.

The quasi-norm in $F^s_{p,q}(\mathbb{R}^d)$ is defined by

$$\|f\|_{F^s_{p,q}(\mathbb{R}^d)} := \|f\|_{L_p(\mathbb{R}^d)} + |f|_{F^s_{p,q}(\mathbb{R}^d)}. \tag{4.6}$$

5 Approximation of functions with lower smoothness

The estimate (3.26) is unsatisfactory because it can be applied only to a small subset of functions in L_p. In this section, we shall remove this deficiency. Our method for doing this is an 'interpolation of operators' type argument which decomposes a general function into a smooth part to which (3.26) can be applied and a second nonsmooth part which is small. We shall restrict our discussion to the case where the operator T is an elliptic differentiable operator of order κ with constant coefficients and ϕ is its fundamental solution (on $\mathbb{R}^d$). For such T, we can choose the wavelet system in the previous section to satisfy

$$T(\psi_v)(x) \le C' \ell(v)^{-\kappa} \chi_{\bar{I}_v}(x), \quad v \in \mathcal{V}. \tag{5.1}$$

[4]The basic requirement is that $m > s$. Additional requirements are imposed in case $p < 1$ or $q < 1$. For us, the only thing that matters is the existence of some wavelet system that can be used to define the Triebel-Lizorkin space $F^s_{p,q}$. In particular, we can, and do, allow the wavelet system to depend on s, p, q.

We fix $1 \le p \le \infty$ and continue to work under assumptions **A1-4**. We recall in the present case $W(L_p(\mathbb{R}^d), \phi))$ is the Sobolev space $W^\kappa(L_p(\mathbb{R}^d))$. Then, each $f \in L_p(\mathbb{R}^d)$ has a series representation

$$f = \sum_{v \in \mathcal{V}} f_v \psi_v,$$

with the sum convergent in $L_p(\mathbb{R}^d)$. We shall work in this section exclusively with L_∞-normalized wavelets ψ_v.

Now, let f be a function in the Triebel-Lizorkin space $F^s_{p,\infty}$, $s < \kappa$, and let h be the density function for Ξ. We decompose $f = f_h^+ + f_h^-$ in the following way:

$$f_h^+ := \sum_{\ell(v) \ge h(v)} \psi_v f_v, \quad f_h^- := f - f_h^+,$$

where

$$h(v) := \|h\|_{L_\infty(\bar{I}_v)}.$$

We shall first estimate how well f_h^+ can be approximated by elements from $\mathcal{S}_\Xi(\phi)$.

Lemma 5.1 *Let $1 \le p \le \infty$ and $0 < s < \kappa$. If $f \in F^s_{p,\infty}$, then for $w := H^{-s}$, we have*

$$\operatorname{dist}(f_h^+, \mathcal{S}_\Xi(\phi))_{L_p(w)} \le C(s,d) \|f\|_{F^s_{p,\infty}}, \tag{5.2}$$

with a constant C as in Theorem 3.1.

Proof: ¿From (5.1), we obtain for any $x \in \mathbb{R}^d$,

$$\begin{aligned}
|h^{\kappa-s}(x) T(f_h^+)(x)| &\le h^{\kappa-s}(x) \sum_{\ell(v) \ge h(v)} |f_v||T(\psi_v)(x)| \\
&\le C' h^{\kappa-s}(x) \sum_{\ell(v) \ge h(v)} \ell(v)^{-\kappa+s} \ell(v)^{-s} |f_v| \chi_{\bar{I}_v}(x) \\
&\le C' M_{s,\infty}(f)(x)\, h(x)^{\kappa-s} \sum_{\ell(v) \ge h(x)} \ell(v)^{-\kappa+s} \chi_{\bar{I}_v}(x) \\
&\le C M_{s,\infty}(f)(x) h(x)^{\kappa-s} h(x)^{-\kappa+s} \le C M_{s,\infty}(f)(x). \qquad (5.3)
\end{aligned}$$

Here we have used the fact that there is an absolute constant C_1 depending only on the support size of the wavelet such that for any dyadic level j, there are at most C_1 $I \in \mathcal{D}_j$ which contain the given point x. This means that the above series can be compared with a geometric series and can be bounded by a fixed multiple of its largest term. Thus, the constant C depends only s and d. Therefore, by Theorem 3.3,

$$\operatorname{dist}(f_h^+, S_\Xi(\phi))_{L_p(H^{-s})} \le C \left\| h^{\kappa-s} T(f_h^+) \right\|_{L_p(\mathbb{R}^d)} \le C \left\| M_{s,\infty}(f) \right\|_{L_p} = C \left\| f \right\|_{F^s_{p,\infty}},$$

which completes the proof of the lemma. ∎

Remark. Formally, the proof given in the lemma does not cover the case $p = 1$. The reason is that the wavelet representation of $f \in L_1(\mathbb{R}^d)$ does not always converge to f. However, the lemma does extend to all $f \in L_1(\mathbb{R}^d)$, provided that we use an inhomogeneous wavelet representation. Such representation takes the form

$$f = \sum_{\ell(v)<2^J} \langle f, \psi_v \rangle \psi_v + \sum_{\ell(v)=2^J} \langle f, \widetilde{\psi}_v \rangle \widetilde{\psi}_v =: f_1 + f_2,$$

with the modified wavelets $\widetilde{\psi}_v$ supported in exactly the same cube $\bar{I}_v$ as their original wavelet counterparts, and satisfy $T(\widetilde{\psi}_v) \le C' 2^{-J\kappa} \chi_{\bar{I}_v}$. The integer J can be chosen at will. This modified expansion converges for every $f \in L_1(\mathbb{R}^d)$. We can use the above inhomogeneous wavelet expansion in the proof of the Lemma, since we know (see the discussion on surface splines in §2) that the density function is *bounded*, which implies that the term f_h^+ in the decomposition of f contains the entire expansion of the above f_2 (for a suitable large J that depends on the bound we have on h, but on nothing else.) The argument in the proof of the lemma can be then repeated *verbatim* for the case $p = 1$. However, the smoothness space that is characterized by the inhomogeneous expansion is the inhomogeneous Triebel-Lizorkin space. For this reason, we have stated the result with respect to the full norm $\|f\|_{F^s_{p,\infty}}$. For $1 < p \le \infty$, the result is also valid with $\|f\|_{F^s_{p,\infty}}$ replaced by $|f|_{F^s_{p,\infty}}$.

We are left with bounding the $L_p(H^{-s})$-norm of f_h^-.

Lemma 5.2 *Fix $1 \le p \le \infty$. If $0 < s < \kappa$ and $w := H^{-s}$, then*

$$\|f_h^-\|_{L_p(w)} \le C(s,d)\|f\|_{F^s_{p,\infty}}, \tag{5.4}$$

where $C(s,d)$ depends only on s and d.

Proof: Since $f_h^- = \sum_{\ell(v)<h(v)} f_v \psi_v(x)$ and $|\psi_v(x)| \le C' \chi_{\bar{I}_v}(x)$ for an absolute constant C', we have

$$|f_h^-(x)| \le C M_{s,\infty}(f)(x) \sum_{\ell(v)<h(v)} \ell(v)^s \chi_{\bar{I}_v}(x) \tag{5.5}$$

Given an $x \in \mathbb{R}^d$ for which $x \in \bar{I}_v$ and $\ell(v) < h(v)$, we know that there is a $t \in \bar{I}_v$ such that $h(v) = h(t)$. For this t, we have

$$H(x) \ge h(v)\left(1 + \frac{|x-t|}{h(v)}\right)^{-r}. \tag{5.6}$$

Since, $|x - t| \le \sqrt{d}\ell(\bar{I}_v) \le \sqrt{d}A_0\ell(v) \le \sqrt{d}A_0 h(v)$, we find that $h(v) \le CH(x)$ with C depending only on r. Thus,

$$\sum_{\ell(v)<h(v)} \ell(v)^s \chi_{\bar{I}_v}(x) \le \sum_{\bar{I}_v \ni x,\, \ell(v) \le CH(x)} \ell(v)^s \le C' H(x)^s, \tag{5.7}$$

where C' depends on s and d. Here, as in the previous lemma we compared the series in (5.7) with a geometric series and bounded it by a fixed multiple of its largest term. ¿From (5.7), we obtain

$$\left\| H^{-s} f_h^- \right\|_{L_p} \leq C \left\| M_{s,\infty}(f) \right\|_{L_p} = C \left\| f \right\|_{F^s_{p,\infty}(\mathbb{R}^d)} ,$$

which proves the lemma. ∎

Combining the two lemmas, we arrive at the following result.

Theorem 5.3 *Assume that ϕ and Ξ satisfy the assumptions* **A1-4** *with respect to a homogeneous differential operator T with constant coefficients (and degree κ). For every $1 \leq p \leq \infty$, for every $0 < s < \kappa$, and for every $f \in F^s_{p,\infty}$ we have*

$$\operatorname{dist}(f, S_\Xi(\phi))_{L_p(H^{-s})} \leq C(s,d) \left\| f \right\|_{F^s_{p,\infty}} , \tag{5.8}$$

where C as in Theorem 3.1, and H is the majorant density function associated with Ξ.

We can also derive results for approximation in Sobolev spaces, as well as in Besov spaces. For the Sobolev space W^s_p, $1 \leq p < \infty$, $0 < s < \kappa$, we have that $W^s_p = F^s_{p,2} \subset F^s_{p,\infty}$, [31], and hence Theorem 5.3 implies that

$$\operatorname{dist}(f, S_\Xi(\phi))_{L_p(H^{-s})} \leq C \left\| f \right\|_{W^s_p(\mathbb{R}^d)} . \tag{5.9}$$

As to Besov spaces, since $B^s_p(L_p(\mathbb{R}^d))$ is continuously embedded into $F^s_{p,\infty}$ (see [31]), we have, by the same theorem, that, for $1 \leq p < \infty$ and $0 < s < \kappa$,

$$\operatorname{dist}(f, S_\Xi(\phi))_{L_p(H^{-s})} \leq C \left\| f \right\|_{B^s_p(L_p(\mathbb{R}^d))} . \tag{5.10}$$

This latter statement extends trivially to $p = \infty$, since $F^s_{\infty,\infty} = B^s_\infty(L_\infty)$.

Note that we have restricted out attention to a differential operator T. The reason is that our analysis depends on two properties of the wavelet system. The first is the characterization of Triebel-Lizorkin spaces in terms of wavelet decompositions, and the second is (5.1). In order to extend the result of this section to more general T, we will need a representation system that will satisfy, first and foremost, a bound analogous to (5.1) with respect to the more general T. We can then define using that system smoothness spaces that are analogous to the Triebel-Lizorkin spaces from §4, and establish an analog of Theorem 5.3.

6 Nonlinear approximation

We shall now turn to a different setting. We assume $f \in L_p(\mathbb{R}^d)$ is a function that we wish to approximate using the shifts of ϕ. In contrast to the problems studied so far where the set of centers is prescribed in advance, we shall now allow the choice of the centers to be made dependent on f. We are interested in how well we can approximate f using at most N such centers.

Let $\Sigma_N := \Sigma_N(\phi)$ be the set of all functions S for which there is a set Ξ of cardinality N such that

$$S = \sum_{\xi \in \Xi} a_\xi \phi(\cdot - \xi). \tag{6.1}$$

We then define the approximation error

$$\sigma_N(f)_p := \inf_{S \in \Sigma_N} \|f - S\|_{L_p(\mathbb{R}^d)}. \tag{6.2}$$

This form of nonlinear approximation is known as N*-term approximation.*

Our setting is different than that considered in previous sections. We begin with a function ϕ that satisfies assumption **A2**. For simplicity, we also suppose that the operator T is a homogeneous differential operator of order κ. The reader can easily abstract the conditions we use about T to get a more general theorem.

We do not assume **A1, A3, A4** since Ξ is not specified in advance. Rather, we shall assume that for any dyadic cube there is a collection of points near this cube that locally satisfy **A3-4**. To make this assumption precise, we fix the order κ of the differential operator T and fix a wavelet system $\{\psi_v\}_{v \in \mathcal{V}}$ of κ-times differentiable compactly supported wavelets of the form described in §4. As in that section, for each v, we denote by $\bar{I}_v$ the smallest cube which contains the support of ψ_v. Recall that we know that $\bar{I}_v$ has size comparable to that of I_v, namely $\ell(\bar{I}_v) \le A_0 \ell(v)$ for a fixed constant A_0 that depends only on κ.

We make the following assumption (about ϕ) in this section.

A5: *Given $\nu > d$, there is an absolute constant C_0 and an integer N_0 such that for any $N \ge N_0$, and any $v \in \mathcal{V}$, there is a set $\Xi_{v,N} \subset \mathbb{R}^d$ consisting of N points with the following property: There is a linear combination*

$$K_{v,N}(\cdot, t) = \sum_{\xi \in \Xi_{v,N}} A_{v,N}(t, \xi)\phi(\cdot - \xi), \quad t \in \bar{I}_v,$$

with $A_{v,N}(\cdot, \xi)$ in $L_\infty(\bar{I}_v)$ such that $\phi(x - t) - K_{v,N}(x, t)$ satisfies

$$|\phi(x - t) - K_{v,N}(x, t)| \le C_0 h_{v,N}^{\kappa - d} \left(1 + \frac{|x - t|}{h_{v,N}}\right)^{-\nu}, \quad t \in \bar{I}_v, \ x \in \mathbb{R}^d, \tag{6.3}$$

with

$$h_{v,n} := \frac{\ell(v)}{N^{1/d}}.$$

The new assumption **A5** can be easily shown to be satisfied for surface splines (Example 2 in §2). Indeed, given N, we choose m such that $m^d \le N$, and choose $\Xi_{v,N}$ to be the vertices of a uniform grid on $\bar{I}_v$ with mesh size $\widetilde{h}_{v,N} := \ell(\bar{I}_v)/(m-1)$. Assuming that $m \ge \kappa - d + \nu$ (see Example 2), it is then easy to follow the dilation argument given in Example 2 to conclude that the linear functionals λ_t, $t \in \bar{I}_v$ are uniformly bounded, independently of $\bar{I}_v$ and m (hence of N). Also, $\widetilde{h}_{v,N}$ as above satisfies $\widetilde{h}_{v,N} \le C h_{v,N}$. Thus, (6.3) follows from the validity of **A4** for surface splines (see Example 2 in §2). The constant C depends only on d and the number A_0, hence is universal for all of our purposes. Thus, N_0 in this case is $(\kappa - d + \nu)^d$.

To prove theorems about N-term approximation of a function f by shifts of the function ϕ, we will first represent f in a wavelet decomposition and approximate the individual terms in this decomposition. We begin by seeing how well we can approximate the wavelet ψ_v (normalized in L_∞), by using a budget of N_v centers. Given v, and any integer $N_v \ge N_0$, we let Ξ_{v,N_v} be the set of points satisfying **A5**. Let us denote

$$S_{v,N_v} := \int_{\mathbb{R}^d} T(\psi_v)(t) K(\cdot, t)\, dt, \quad K(x,t) := \sum_{\xi \in \Xi_{v,N_v}} A(t,\xi)\phi(x-\xi). \tag{6.4}$$

Notice that since $T(\psi_v) = 0$, $t \notin \bar{I}_v$, the integral in (6.4) only goes over $t \in \bar{I}_v$ and therefore by **A5**, this integral is finite.

Now, an important property of the wavelets that we need here is that ψ_v satisfies $\|T\psi_v\|_{L_\infty} \le C_0 \ell(v)^{-\kappa}$. Also, thank to **A2**, $\psi_v = \int_{\mathbb{R}^d} T(\psi_v)(t)\phi(\cdot - t)\, dt$. We use this together with **A5** to derive the following bound

$$\begin{aligned} |\psi_v(x) - S_{v,N_v}(x)| &\le C_0 C_1 \left(\frac{h_{v,N_v}}{\ell(v)}\right)^{\kappa} \int_{\bar{I}_v} \left(1 + \frac{|x-t|}{h_{v,N_v}}\right)^{-\nu} h_{v,N_v}^{-d}\, dt \\ &\le C N_v^{-\kappa/d} \left(1 + \frac{dist(x, \bar{I}_v)}{\ell(v)}\right)^{-\nu+d}. \end{aligned} \tag{6.5}$$

Let us describe now our approximation algorithm and analyze its performance. In the algorithm, we are given a budget N of centers, and invest a nominal amount $c_v \ge 0$ in each wavelet ψ_v, $v \in \mathcal{V}$; we refer to c_v as **cost**. We ensure that the total cost $\sum_v c_v$ does not exceed the given budget N. Since c_v may not be an integer, and since the minimal number of centers that we can use is N_0, the cost c_v allows us to approximate the term $f_v\psi_v$ in the wavelet expansion (4.3) of f by investing

$$N_v := \lfloor c_v \rfloor \tag{6.6}$$

centers, provided $c_v \ge N_0$. In this case, (6.5) gives us the estimate

$$|\psi_v(x) - S_{v,N_v}(x)| \le C N_v^{-\kappa/d} \left(1 + \frac{dist(x, \bar{I}_v)}{\ell(v)}\right)^{-\nu+d} \le C' c_v^{-\kappa/d} \left(1 + \frac{dist(x, \bar{I}_v)}{\ell(v)}\right)^{-\nu+d}. \tag{6.7}$$

If $c_v < N_0$, we do not approximate ψ_v at all, and get then

$$|\psi_v(x) - S_{v,N_v}(x)| = |\psi_v(x)| \le C \chi_{\bar{I}_v}(x), \quad x \in \mathbb{R}^d, \tag{6.8}$$

since the wavelets are uniformly bounded.

Now, suppose that we are given a budget of N centers, determine a cost distribution $(c_v)_v$ (with $\sum_v c_v \le N$), and would like to estimate the L_p-norm of the error when approximating the term $f_v \psi_v$ in the wavelet expansion of f using the designated cost c_v. According to (6.7) and (6.8), our error will be determined, up to a universal constant, by the p-norm of

$$\sum_v |f_v||\psi_v - S_{v,N_v}| =: \sum_v |f_v| R_v. \tag{6.9}$$

Here, (f_v) are the wavelet coefficients of the approximand f.

Lemma 6.1 *Let $1 \le p \le \infty$ and suppose that the constant ν appearing in* **A5** *is $> 2d$. If R_v is defined as in* (6.9)*, then*

$$\left\| \sum_{v \in \mathcal{V}} |f_v| R_v \right\|_{L_p(\mathbb{R}^d)} \le C \left\| \sum_{v \in \mathcal{V}} [\max(1, c_v)]^{-\kappa/d} |f_v| \chi_{I_v} \right\|_{L_p(\mathbb{R}^d)}. \tag{6.10}$$

Proof: This inequality can be derived using the method of proof for Fefferman-Stein inequalities [16]. In fact, for $1 \le p < \infty$, it can be derived directly from these inequalities as follows. Let M_0 be the Hardy-Littlewood maximal operator

$$M_0(g)(x) := \sup_{Q \ni x} \frac{1}{|Q|} \int_Q |g(u)|\, du \tag{6.11}$$

where the supremum is taken over all cubes Q that contain x. Then, for $1 < p < \infty$ and any real numbers (a_v), the Fefferman-Stein inequality says

$$\left\| \sum_{v \in \mathcal{V}} |a_v| M_0(\chi_{I_v}) \right\|_{L_p(\mathbb{R}^d)} \le C \left\| \sum_{v \in \mathcal{V}} |a_v| \chi_{I_v} \right\|_{L_p(\mathbb{R}^d)}, \tag{6.12}$$

where C depends only on p as p gets close to 1 and ∞. Now a direct calculation shows that for a constant C_0 depending only on d we have

$$M_0(\chi_{I_v})(x) \ge C_0 \left(1 + \frac{dist(x, I_v)}{\ell(v)}\right)^{-d} \ge C_0' \left(1 + \frac{dist(x, \bar{I}_v)}{\ell(v)}\right)^{-\nu+d} \tag{6.13}$$

where, in the last inequality, we used our assumption that $\nu > 2d$ and the fact that $1 + \frac{\mathrm{dist}(x,I_v)}{\ell(v)} \le C(1 + \frac{\mathrm{dist}(x,\bar{I}_v)}{\ell(v)})$ for all $x \in \mathbb{R}^d$ with a constant C depending only on the space dimension d. It follows from this and (6.5), (6.8) that

$$R_v(x) \le C[\max(1, c_v)]^{-\kappa/d} M_0(\chi_{I_v})(x), \quad x \in \mathbb{R}^d,$$

with C again depending only on d. Using this with (6.12) we derive the lemma for $1 < p < \infty$.

One can derive the lemma in the case $p = 1$ and also obtain a constant not depending on p as $p \to 1$ by using a modified Hardy-Littlewood maximal function

$$M_0'(g)(x) := \sup_{Q \ni x} \left(\frac{1}{|Q|} \int_Q |g(u)|^\mu \, du \right)^{1/\mu}$$

with $\mu < 1$. The Fefferman-Stein inequality now holds for $p = 1$ if this new maximal function is used in place of M_0. For M_0' one has an analogue of (6.13) where in the first inequality the exponent d is replaced by d/μ. Thus if μ is sufficiently close to one so that $\mu(\nu - d) > d$, we again arrive at the lemma for $p = 1$. When $p = \infty$, one can again derive the lemma using the fact that $\nu > 2d$ by an analogous argument to the proof of (6.12). ∎

The following is the main result of this section.

Theorem 6.2 *Let $1 \le p < \infty$ be given, and let $f \in F^s_{\tau,q}$, with $s \le \kappa$, $\tau = (1/p + s/d)^{-1}$, and $q := (1 + s/d)^{-1}$. Then*

$$\sigma_N(f)_p \le C N^{-s/d} \|f\|_{F^s_{\tau,q}}.$$

The constant C here is independent of f and N.

In the proof of the theorem, we will use the following elementary observation.

Lemma 6.3 *Let $\sum_{j=-\infty}^{\infty} z_j$ be a non-negative series with limit $Z < \infty$, and with partial sum sequence $Z_k := \sum_{j=-\infty}^{k} z_j$. For any $\epsilon > 0$, there is a $C_\epsilon > 0$ depending only on ϵ such that*

$$\sum_{j=-\infty}^{\infty} \frac{z_j}{Z_j^{1-\epsilon}} \le C_\epsilon Z^\epsilon.$$

Proof of the lemma: For each positive integer k, let j_k be the minimal integer for which $Z_{j_k} \ge 2^{-k} Z$, $j_0 := \infty$. Then

$$\sum_{j=j_k}^{j_{k-1}-1} \frac{z_j}{Z_j^{1-\epsilon}} \le 2^{k(1-\epsilon)} Z^{\epsilon-1} 2^{-(k-1)} Z = 2Z^\epsilon 2^{-\epsilon k}.$$

Summing over all positive k, we obtain the stated result. ∎

Proof of the theorem: Fix $f \in F^s_{\tau,q}$ with f not the zero function, and fix a positive integer N. For any given $x \in \mathbb{R}^d$, we consider the set $\mathcal{V}_x$ of all $v \in \mathcal{V}$ such that $\chi_{I_v}(x) \neq 0$. We can order the $v \in \mathcal{V}_x$ as follows. We take any fixed ordering for E and then we say $v > v'$ if either $|v| > |v'|$ or $|v| = |v'|$ and $e_v > e_{v'}$. Given this order, we now define for each $x \in \mathbb{R}^d$ and each $v' \in \mathcal{V}$,

$$M_{q,v'}(x) := \Big(\sum_{v \in \mathcal{V}_x : v \geq v'} |v|^{-qs/d} |f_v|^q \chi_{I_v}(x) \Big)^{1/q}.$$

Notice that $M_{q,v'}(x)$ is actually constant on $I_{v'}$ and so we denote

$$M_{q,v'} := M_{q,v'}(x), \quad x \in I_{v'}. \tag{6.14}$$

Also, we clearly have

$$M_{q,v'} \leq M_q(f)(x), \quad x \in I_{v'}. \tag{6.15}$$

We determine our cost function by the rule

$$c_v = a|v|^q |f_v|^q M_{q,v}^{\tau-q}, \quad v \in \mathcal{V},$$

where a will be specified in a moment. Now $\tau - q \geq 0$ and $q = 1 - qs/d$. Therefore, from (6.15), we see that

$$c_v = \left\| a|v|^{-qs/d} |f_v|^q M_{q,v}^{\tau-q} \chi_{I_v} \right\|_{L_1(\mathbb{R}^d)} \leq \left\| a|v|^{-qs/d} |f_v|^q M_q(f)^{\tau-q} \chi_{I_v} \right\|_{L_1(\mathbb{R}^d)}.$$

Thus,

$$\begin{aligned} \sum_{v \in \mathcal{V}} c_v &\leq a \left\| M_q(f)^{\tau-q} \sum_{v \in \mathcal{V}} |v|^{-qs/d} |f_v|^q \chi_{I_v} \right\|_{L_1(\mathbb{R}^d)} \\ &\leq a \left\| M_q(f)^\tau \right\|_{L_1(\mathbb{R}^d)} = a \left\| f \right\|^\tau_{F^s_{\tau,q}} < \infty, \end{aligned} \tag{6.16}$$

where we have used the fact that $\chi_{I_v} \leq \chi_{\bar{I}_v}$ for all $v \in \mathcal{V}$. Thus, we can choose a so that $a \|f\|^\tau_{F^s_{\tau,q}} = N$, and obtain that $\sum_v c_v \leq N$.

It remains to estimate the L_p-error produced by the scheme. In view of Lemma 6.1, we need to estimate the L_p-norm of

$$\sum_{v \in \mathcal{V}} [\max(1, c_v)]^{-\kappa/d} |f_v| \chi_{I_v} \leq \sum_{v \in \mathcal{V}} [\max(1, c_v)]^{-s/d} |f_v| \chi_{I_v} \leq \sum_{v \in \mathcal{V}} c_v^{-s/d} |f_v| \chi_{I_v} =: \sum_v E_v.$$

Here we have used the fact that $s \leq k$.

For $x \in I_v$, we have

$$E_v(x) = a^{-s/d}|v|^{-qs/d}|f_v|^{1-qs/d}M_{q,v}^{(q-\tau)s/d} = a^{-s/d}|v|^{-qs/d}|f_v|^q M_{q,v}^{\tau/p-q}.$$

Here, we have used the fact that $1 - qs/d = q$. If $p = 1$, then $\tau/p - q = 0$, and, fixing x, we obtain

$$\sum_{v\in\mathcal{V}} E_v(x) \le a^{-s/d}M_q^q(f)(x) = a^{-sp/d}M_q^\tau(f)(x).$$

We can prove a similar estimate when $p > 1$. Namely, we fix $x \in \mathbb{R}^d$ and invoke Lemma 6.3 with $z_v := |v|^{-qs/d}|f_v|^q\chi_{I_v}(x)$ using the ordering on $\mathcal{V}_x$ introduced earlier. Hence, $M_{q,v}^q = Z_v$ and $M_{q,v}^{\tau/p-q} = Z_v^{\epsilon-1}$, with $\epsilon = \tau/(pq) > 0$. Also, $Z \le M_q^q(f)(x)$ again because $\chi_{I_v} \le \chi_{\bar{I}_v}$. By the lemma,

$$\sum_{x\in\bar{I}_v} |v|^{-qs/d}|f_v|^q M_{q,v}^{\tau/p-q} \le C(\tau,p,q)M_q(f)^{\tau/p}.$$

Thus,

$$(\sum_{v\in\mathcal{V}} E_v(x))^p \le Ca^{-sp/d}M_q(f)^\tau(x),$$

and we conclude that, with $A := \|f\|_{F^s_{\tau,q}}$

$$\left\|\sum_v E_v\right\|_{L_p(\mathbb{R}^d)} \le Ca^{-s/d}A^{\tau/p} = C(aA^\tau)^{-s/d}A = CN^{-s/d}\|f\|_{F^s_{\tau,q}}.$$

■

We can derive from the theorem a corresponding result for the Besov space $B_q^s(L_\tau)$ (see any of the standard texts for a definition of these spaces.) This Besov space is continuously embedded in $F^s_{\tau,q}$. Hence we obtain

Corollary 6.4 *Let $1 \le p < \infty$ be given, and let $f \in B_q^s(L_\tau)$, with $s \le \kappa$, $\tau = (1/p + s/d)^{-1}$, and $q := (1 + s/d)^{-1}$. Then*

$$\sigma_N(f)_p \le CN^{-s/d}\|f\|_{B_q^s(L_\tau)}.$$

The constant C here is independent of f and N.

Finally, we compare this theorem with the classical results on N-term wavelet approximation given in [9]. For wavelet approximation one obtains the same bounds with the assumption $f \in B_\tau^s(L_\tau) = F^s_{\tau,\tau}$. Since $q < \tau$, the wavelet assumption is (slightly) weaker than what is assumed in Theorem 6.2. The two assumptions agree when $p = 1$. We do not know if q can be replaced by τ for other values of p. We believe that it cannot, i.e., that the value of q in our theorems is the best possible one.

Acknowledgment

We are indebted to Thomas Hangelbroek for fruitful discussions concerning this article.

References

[1] M. Buhmann, N. Dyn, and D. Levin, *On quasi-interpolation by radial basis functions with scattered centers shift-invariant spaces*, Constructive Approximation, 11 (2) (1995), pp. 239–254.

[2] M. Buhmann and A. Ron, *Radial basis functions: L_p-approximation orders with scattered centers*, in Curves and Surfaces II, P.J. Laurent, A. Le Méhauté and L.L. Schumaker, eds., AK Peters, 1994.

[3] M. D. Buhmann, *Multivariate cardinal interpolation with radial-basis functions*, Constr. Approx., 6 (1990), pp. 225–255.

[4] I. Daubechies, *Ten Lectures on Wavelets*, vol. 61 of CBMS-NSF Lecture Notes, SIAM, 1992.

[5] C. de Boor, *A Practical Guide to Splines, Revised Edition*, vol. 27 of Applied Mathematical Sciences, Springer-Verlag, Berlin, 2001.

[6] C. de Boor, R. A. DeVore, and A. Ron, *Approximation from shift-invariant subspaces of $L_2(\mathbf{R}^d)$*, Trans. Amer. Math. Soc., 341 (1994), pp. 787–806.

[7] C. de Boor and A. Ron, *The exponentials in the span of the integer translates of a compactly supported function shift-invariant spaces*, Journal of the London Mathematical Society, 45 (1992), pp. 519–535.

[8] ——, *Fourier analysis of approximation power of principal shift-invariant spaces*, Constructive Approximation, 8 (1992), pp. 427–462.

[9] R. DeVore, B. Jawerth, and V. Popov, *Compression of wavelet decompositions*, American Journal of Mathematics, 114 (1992), pp. 737–785.

[10] R. DeVore and V. Popov, *Interpolation spaces and non-linear approximation*, in Function Spaces and Applications, eds. M. Cwikel, J. Peetre, Y. Sagher, and H. Wallin, Springer, Berlin, 1988, pp. 191–205. Lecture Notes in Math., Vol. 1302.

[11] R. DeVore and X. M. Yu, *Degree of adaptive approximation*, Math. Comp., 55 (1990), pp. 625–635.

[12] R. A. DeVore and G. G. Lorentz, *Constructive approximation*, vol. 303 of Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], Springer-Verlag, Berlin, 1993.

[13] J. Duchon, *Splines minimizing rotation-invariant semi-norms in Sobolev spaces*, in Constructive theory of functions of several variables (Proc. Conf., Math. Res. Inst., Oberwolfach, 1976), Springer, Berlin, 1977, pp. 85–100. Lecture Notes in Math., Vol. 571.

[14] ——, *Sur l'erreur d'interpolation des fonctions de plusieurs variables par les D^m-splines*, RAIRO Anal. Numér., 12 (1978), pp. 325–334.

[15] N. Dyn and A. Ron, *Radial basis function approximation: from gridded centres to scattered centres*, Proc. London Math. Soc. (3), 71 (1995), pp. 76–108.

[16] C. Fefferman and E. Stein, *Some maximal inequalities*, Amer. J. Math, 93 (1971), pp. 107–115.

[17] T. Hangelbroek, *Error estimates for thin plate spline approximation in the disk*, Constr. Approx., 28 (2008), pp. 27–59.

[18] M. J. Johnson, *On the approximation order of principal shift-invariant subspaces of $L_p(\mathbf{R}^d)$*, J. Approx. Theory, 91 (1997), pp. 279–319.

[19] M. J. Johnson, *A bound on the approximation order of surface splines*, Constr. Approx., 14 (1998), pp. 429–438.

[20] M. J. Johnson, *Approximation in $L_p(R^d)$ from spaces spanned by the perturbed integer translates of a radial function*, J. Approx. Theory, 107 (2000), pp. 163–203.

[21] W. R. Madych and S. A. Nelson, *Multivariate interpolation: a variational theory*, unpublished, (1983).

[22] Y. Meyer, *Ondelettes et opérateurs, I: Ondelettes, II: Opérateurs de Calderón-Zygmund, III: Opérateurs multilinéaires*, Hermann, Paris, 1990.

[23] C. A. Micchelli, *Interpolation of scattered data: distance matrices and conditionally positive definite functions*, Constr. Approx., 2 (1986), pp. 11–22.

[24] F. J. Narcowich and J. D. Ward, *Scattered-data interpolation on $\mathbf{R}^n$: error estimates for radial basis and band-limited functions*, SIAM J. Math. Anal., 36 (2004), pp. 284–300.

[25] P. Petrushev, *Direct and converse theorems for rational and spline approximation and Besov spaces*, in Function Spaces and Applications, eds. M. Cwikel, J. Peetre, Y. Sagher, and H. Wallin, Springer, Berlin, 1988, pp. 363–377. Lecture Notes in Math., Vol. 1302.

[26] M. J. D. Powell, *The uniform convergence of thin plate spline interpolation in two dimensions*, Numer. Math., 68 (1994), pp. 107–128.

[27] A. Ron, *The L_2-approximation orders of principal shift-invariant spaces generated by a radial basis function*, in Numerical Methods of Approximation Theory, Vol. 9, D. Braess & L.L. Schumaker eds., International Series of Numerical Mathematics, 1992, pp. 245–268.

[28] R. Schaback, *Approximation by radial basis functions with finitely many centers*, Constr. Approx., 12 (1996), pp. 331–340.

[29] I. J. Schoenberg, *Metric spaces and completely monotone functions*, Ann. of Math. (2), 39 (1938), pp. 811–841.

[30] ——, *Metric spaces and positive definite functions*, Trans. Amer. Math. Soc., 44 (1938), pp. 522–536.

[31] H. Triebel, *Interpolation Theory, Function Spaces, and Differential Operators*, North-Holland, Amsterdam, 1978.

[32] H. Wendland, *Scattered data approximation*, vol. 17 of Cambridge Monographs on Applied and Computational Mathematics, Cambridge University Press, Cambridge, 2005.

[33] Z. M. Wu and R. Schaback, *Local error estimates for radial basis function interpolation of scattered data*, IMA J. Numer. Anal., 13 (1993), pp. 13–27.

[34] J. Yoon, *Approximation in $L_p(\mathbf{R}^d)$ from a space spanned by the scattered shifts of a radial basis function*, Constr. Approx., 17 (2001), pp. 227–247.

[35] ——, *Interpolation by radial basis functions on Sobolev space*, Journal of Approximation Theory, 112 (2001), pp. 1–15.

[36] ——, *Spectral approximation orders of radial basis function interpolation on the Sobolev space*, Siam Journal on Mathematical Analysis, 33 (2001), pp. 946–958.

Ronald DeVore, Department of Mathematics, University of South Carolina, Columbia, SC 29208, USA, devore@math.sc.edu
Amos Ron, Computer Science Department, University of Wisconsin-Madison, Madison, WI 53706, USA, amos@cs.wisc.edu