Constr. Approx. (2006) 23: 261–304
DOI: 10.1007/s00365-005-0607-8

CONSTRUCTIVE
APPROXIMATION

"Push-the-Error" Algorithm for Nonlinear n-Term Approximation

W. Dahmen and P. Petrushev

Abstract. This paper is concerned with further developing and refining the analysis of a recent algorithmic paradigm for nonlinear approximation, termed the "Push-the-Error" scheme. It is especially designed to deal with L_∞-approximation in a multilevel framework. The original version is extended considerably to cover all commonly used multiresolution frameworks. The main conceptually new result is the proof of the quasi-semi-additivity of the functional $N(\varepsilon)$ counting the number of terms needed to achieve accuracy ε. This allows one to show that the improved scheme captures *all* rates of best n-term approximation.

1. Introduction

The understanding of nonlinear approximation has greatly benefitted from recent multilevel and wavelet concepts. Norm equivalences induced by wavelet bases in a Hilbert space context play a major role in the analysis of best n-term approximation, part of which can even be retained for L_p-norms for $1 < p < \infty$, see, e.g., [18]. Near best n-term approximation is simply obtained by keeping the (properly scaled) n largest coefficients in the wavelet expansion. However, many applications involve more complex geometries for which wavelet bases with the desired properties are hard to construct or are not available at all. In the absence of such bases the realization of best n-term approximation is far less obvious, let alone approximation in L_∞. A significant advance in best n-term approximation, in settings where explicit wavelet bases may not be available, is offered by the approach in [16], [23], [25].

The situation is again quite different when approximating in the *uniform norm* which is the primary concern and guiding issue in this paper. The "piling up" effect of multilevel structures is not well aligned with the L_∞-norm. This principal obstruction concerns *any* sort of multilevel expansion, even those for "ideal" wavelet bases. Nevertheless, an efficient way of realizing optimal L_∞-approximation rates, for approximation spaces induced by best n-term approximation in the above-mentioned flexible settings, is offered by another algorithmic paradigm, called the "Push-the-Error" algorithm. This has been

Date received: March 30, 2004. Date revised: February 27, 2005. Date accepted: June 6, 2005. Communicated by Albert Cohen. Online publication: October 21, 2005.
AMS classification: 41A15, 41A17, 41A25, 41A46, 65D15, 46E35.
Key words and phrases: Uniform norm approximation, Multilevel expansions, Wavelet bases, Best n-term approximation.

developed in [24] for the specific case of nonlinear n-term approximation from Courant elements (piecewise linear finite elements) in the uniform norm and dimension $d = 2$. The essence of this algorithm originates from [18]. In view of its importance as a paradigm that works in the uniform norm (even in the absence of good multilevel bases), it is interesting to explore the scope of applicability of its conceptual foundation.

Our primary goal in this paper is therefore to further refine and extend this algorithm in several directions. The key new steps in this paper are the following:

(i) We generalize the “Push-the-Error” algorithm to nonlinear n-term approximation from the “scaling functions” of a general multiresolution analysis (MRA) on compact domains in $\mathbf{R}^d$.

(ii) We refine the algorithm from [24] and its analysis considerably. In particular, we prove the quasi-semi-additivity of the functional $N(\varepsilon)$ counting the number of terms in the approximation needed to achieve accuracy ε. This enables us to show that the improved algorithm captures *all* rates of the best n-term approximation.

It should be stressed that the “Push-the-Error” paradigm is, in principle, very flexible in that it essentially requires only refinability of single scale basis functions, i.e., it has a potential to work under fairly general circumstances. For instance, complex domain geometries pose much less of an obstruction than for the construction of wavelet bases thresholding concepts in L_p are typically based upon. The main idea is to complement thresholding strategies, i.e., keeping terms with large coefficients, while transferring small terms to higher levels with the aid of refinement equations. This accounts for the fact that small terms may add up over different levels to form eventually a significant contribution in the uniform norm, because even the best multilevel bases are no longer able to properly separate the contributions from different length scales.

In addition, we briefly relate our findings to the somewhat wider context of nonlinear n-term approximation in L_p. As mentioned before, for $1 < p < \infty$, best n-term approximation is provided by thresholding wavelet expansions. We show here first that even for $0 < p < \infty$ the usual thresholding strategy can be utilized for nonlinear n-term approximation in L_p for the more flexible setting of multilevel scaling function representations in general MRAs so as to capture the rate of the best n-term approximation. This thresholding scheme can be shown to emerge from extending “Push-the-Error” to the L_p case for $0 < p < \infty$.

In [24] there is another algorithm (named “Trim & Cut”) developed for nonlinear n-term approximation in L_p, $0 < p \leq \infty$. The idea of this algorithm originates in the proof of the Jackson estimate in [20]. A similar algorithm has also been suggested by Yu. Brudnyi and I. Kozlov (see [2] and the references therein). The execution of the “Trim & Cut” algorithm relies heavily on a coloring procedure used to represent the set of all supports of basis functions as a disjoint union of trees with respect to the inclusion relation. This renders the scheme practically infeasible. Consequently, it is less valuable compared to the “Push-the-Error” algorithm.

Finally, we note that the “Push-the-Error” algorithm is not restricted to only approximation from MRAs consisting of continuous functions. It can successfully be used for nonlinear approximation of continuous functions from discontinuous (isotropic or anisotropic) hierarchical bases in the L_∞-norm. All results from this paper have analogues in such settings under less restrictive conditions. We shall not present the details here.

The paper is organized as follows. In Section 2 we collect some prerequisites. First, in Section 2.1 we describe a general multiresolution setting which is designed to host all commonly used setups. In doing so, we extract the abstract requirements on such multiresolution hierarchies of spaces that make "Push-the-Error" work and collect the tools needed in this context. In Section 2.2 we outline several examples covered by the general framework while we collect, in Section 2.3, some further consequences and prerequisites for later use. In Section 2.4 we introduce a family of local projectors that serve as a tool for forming multilevel decompositions. In Section 3 we introduce a scale of "Besov-like" spaces (B-spaces) associated with the MRA needed to prove optimality of the "Push-the-Error" scheme. In Section 4 we characterize the approximation spaces generated by nonlinear n-term approximation from the scaling functions of an MRA, placing special emphasis on the L_∞-case. In Section 5 we describe the improved "Push-the-Error" algorithm, present its error analysis, and discuss its complexity. In Section 6 we describe and give the error analysis of the "Threshold" algorithm for nonlinear n-term approximation in L_p, $0 < p < \infty$, from the scaling functions of an MRA. In Section 7 we give the proof of the main results concerning the quasi-semi-additivity of the functional counting the number of terms generated by the scheme, and the error estimation theorem. Finally, Section 8 is an appendix where we place the proofs of the Bernstein estimate and the norm equivalence in the B-spaces.

Throughout the paper, we use the following notation: $\mathbf{N} := \{1, 2, \ldots\}$, $\mathbf{N}_0 := \mathbf{N} \cup \{0\}$. For any set $E \subset \mathbf{R}^d$, $\mathbf{1}_E$ denotes the characteristic function of E, and $|E|$ denotes the Lebesgue measure of E while E° means the interior of E. For a finite set E, $\#E$ denotes the cardinality of E. Positive constants are denoted by $c, c_1, c_*, \ldots$ (if not specified, they may vary at every occurrence), $A \approx B$ means $c_1 A \le B \le c_2 B$, and $A := B$ or $B =: A$ stands for "A is by definition equal to B." Whenever an L_p-norm refers to the fixed underlying domain Ω, we write briefly $\|\cdot\|_p$, whereas $\|f\|_{L_p(G)}$ indicates the reference to a particular subdomain $G \subset \Omega$.

2. Preliminaries

2.1. *Multiresolution Analysis* (*MRA*)—*Basic Properties*

We consider the general case of a hierarchy of spaces

$$V_0 \subset V_1 \subset \cdots \tag{2.1}$$

on a compact domain $\Omega \subset \mathbf{R}^d$ ($d \ge 1$) such that $\overline{\bigcup V_m} = C(\Omega)$ (usually Ω is a polyhedral (polygonal if $d = 2$) domain in $\mathbf{R}^d$). We set $\mathcal{M} := \{V_m\}_{m \ge 0}$. In what follows we shall specify our requirements on such hierarchies. These assumptions are designed to accommodate all commonly used setups as well as possible further settings that could be anticipated in the future.

We assume that each V_m is spanned by a basis $\Phi_m = \{\varphi_\theta\}_{\theta \in \Theta_m}$, consisting of compactly supported and continuous basis functions, normalized in L_∞ ($\|\varphi_\theta\|_\infty = 1$), which should be viewed as scaling functions when dealing with the classical wavelet setting. Here Θ_m is an index set and for convenience we use these indices simultaneously to denote sets satisfying $\operatorname{supp} \varphi_\theta \subset \theta$ for $\theta \in \Theta_m$. We denote $\Theta := \bigcup_{m \in \mathbf{N}_0} \Theta_m$ and $\Phi := \bigcup_{m \in \mathbf{N}_0} \Phi_m$.

At times we shall loosely call θ the "support" of φ_θ although $\operatorname{supp}\varphi_\theta$ may actually be strictly contained in θ. However, θ and the true support will always "scale" in the same way which will be made precise later. In particular, Θ_m may contain more than one (although always a uniformly bounded number) copy of a set θ.

More specific properties of the *single scale bases* Φ_m can typically be related to some underlying mesh or, more generally, to some partition of the spatial domain. We shall formalize next our requirements on such partitions that will cover all cases of interest.

Cells (*Cubes*, *Simplexes*). We shall always assume that there is an underlying sequence of partitions of Ω: $\mathcal{P}_0, \mathcal{P}_1, \dots$ with $\mathcal{P} := \bigcup_{m\in\mathbf{N}_0} \mathcal{P}_m$ which satisfy the following conditions:

(a) Every level $\mathcal{P}_m$ is a partition of Ω, consisting of finitely many compact connected sets (cells) with disjoint interiors. Usually these cells are cubes, simplexes (triangles), or polyhedral subdomains of Ω.
(b) The partitions $(\mathcal{P}_m)$ are nested, i.e., $\mathcal{P}_{m+1}$ is a refinement of $\mathcal{P}_m$.
(c) Each cell $I \in \mathcal{P}_m$ has (contains) at least two and at most ν_0 children in $\mathcal{P}_{m+1}$ with $\nu_0 \geq 2$ an absolute constant.
(d) There exist constants $0 < r < \rho < 1$ such that, for each $I \in \mathcal{P}$ and any child I' of I,

$$r|I| \leq |I'| \leq \rho|I|. \tag{2.2}$$

(e) *Local quasi-uniformity*. There exists a constant $\vartheta \geq 1$ such that if $I, J \in \mathcal{P}_m$ $(m \geq 0)$ and $I \cap J \neq \emptyset$, then

$$\vartheta^{-1} \leq |I|/|J| \leq \vartheta. \tag{2.3}$$

Further conditions on the "supports" $\theta \in \Theta$ of the basis functions φ_θ are specified in the following:

(α) Each $\theta \in \Theta_m$ as well as $\operatorname{supp}\varphi_\theta$ is a connected compact set which can be "paved" by cells from $\mathcal{P}_m$, that is, $\theta = \bigcup_{I\in\mathcal{N}_\theta} I$, where $\mathcal{N}_\theta \subset \mathcal{P}_m$ and $\#\mathcal{N}_\theta \leq \nu_1$ with ν_1 an absolute constant.
(β) The interiors of at most ν_2 sets $\theta \in \Theta_m$ $(m \geq 0)$ may intersect at a time, where ν_2 is another absolute constant.

For a given $\theta \in \Theta$, we denote by $l(\theta)$ the *level* of θ, i.e., $l(\theta) = m$ if $\theta \in \Theta_m$, and we similarly denote by $l(I)$ the *level* of $I \in \mathcal{P}$.

For later use it will be convenient to record for direct reference the following consequences of the properties (a)–(e) and (α)–(β):

(γ) If $I \subset \theta$ and $l(I) = l(\theta)$, then

$$|\theta| \leq \beta_0|I|, \qquad \beta_0 = \text{constant}. \tag{2.4}$$

(δ) For each $\theta \in \Theta_m$ $(m \geq 0)$,

$$\#\{\eta \in \Theta_{m+1} : \eta \subset \theta\} \leq \nu_3, \qquad \nu_3 = \text{constant}. \tag{2.5}$$

Remark 2.1. It is an important observation that the above conditions involve essentially only measures of cells but not the shape of cells and, consequently, cover the case of anisotropic partitions of the types considered in [16], [23], [24], [27].

Since Φ_m is a basis for V_m, each $f \in V_m$ has a unique representation

$$f = \sum_{\theta \in \Theta_m} c_\theta(f)\varphi_\theta, \tag{2.6}$$

where $\{c_\theta(f)\}_{\theta \in \Theta_m}$ are the dual functionals, i.e., $c_\theta(\varphi_{\theta'}) = \delta_{\theta,\theta'}$.

Aside from the locality of the φ_θ's, a crucial further requirement on the MRA $\mathcal{M}$ concerns the locality of the dual functionals. We assume that each linear functional $c_\theta(\cdot)$ is supported on θ and satisfies the condition

$$|c_\theta(g)| \le \frac{\beta_1}{|\theta|} \int_\theta |g(x)|\, dx \qquad \text{for} \quad \theta \in \Theta_m \quad \text{and} \quad g \in V_m, \tag{2.7}$$

where $\beta_1 \ge 1$ is a constant. We shall assume that the linear functionals $c_\theta(\cdot)$ are extended to $L_1(\Omega)$ (retaining the same notation) so that

$$|c_\theta(f)| \le \frac{\beta_1}{|\theta|} \int_\theta |f(x)|\, dx \qquad \text{for} \quad f \in L_1(\Omega). \tag{2.8}$$

Due to the Hahn–Banach theorem this is always possible. We pause to record a few important consequences of (2.7).

A first consequence of (2.7) is the *stability* of the *single scale bases* $(\Phi_m)_{m \in \mathbf{N}_0}$. There exists a constant $\beta_2 \ge 1$ such that for each $g \in V_m$ with representation $g = \sum_{\theta \in \Theta_m} c_\theta \varphi_\theta$, we have

$${\beta_2}^{-1}\|g\|_p \le \left(\sum_{\theta \in \Theta_m} \|c_\theta \varphi_\theta\|_p^p \right)^{1/p} \le \beta_2 \|g\|_p, \qquad 1 \le p \le \infty, \tag{2.9}$$

uniformly in m, with the usual modification when $p = \infty$. Moreover, using also (2.3) and property (β) of the Θ_m's, it follows that, for any $0 < q \le \infty$ and $\gamma \in \mathbf{R}$,

$$\left(\sum_{I \in \mathcal{P}_m} (|I|^\gamma \|g\|_{L_p(I)})^q \right)^{1/q} \approx \left(\sum_{\theta \in \Theta_m} (|\theta|^\gamma \|c_\theta \varphi_\theta\|_p)^q \right)^{1/q}. \tag{2.10}$$

Condition (2.7) readily implies that

$$\|\varphi_\theta\|_p \approx |\theta|^{1/p - 1/q} \|\varphi_\theta\|_q, \qquad \theta \in \Theta, \quad 1 \le p, q \le \infty, \tag{2.11}$$

where the constants of equivalence depend only on β_1.

When dealing with nonlinear approximation in $L_p, 0 < p \le 1$, we shall be additionally assuming that, for any $g \in V_m$ and $I \in P_m$ $(m \ge 0)$,

$$\|g\|_{L_q(I)} \approx |I|^{1/q - 1} \|g\|_{L_1(I)}, \qquad 0 < q < 1, \tag{2.12}$$

with constants independent of g and m. Evidently, this condition yields (2.9)–(2.11) when $0 < p < 1$.

From (2.1) we know that each element of Φ_m can be written as a linear combination of elements in Φ_{m+1}. Furthermore, due to the locality of the dual functionals, this expansion is local, i.e., we have

$$\varphi_\theta = \sum_{\eta\in\Theta_m,\, \eta\subset\theta} a_{\theta,\eta}\varphi_\eta, \qquad \theta \in \Theta_{m-1}. \tag{2.13}$$

Moreover, by (2.7) and the L_∞-normalization of φ_θ, we have $|a_{\theta,\eta}| = |c_\eta(\varphi_\theta)| \le \beta_1$.

We conclude our list of basic general assumptions with one which can be viewed as strengthening our assumptions on the dual functionals. We shall assume that there exist constants $0 < \delta < 1$ and $\beta_3 \ge 1$ such that, for each $g \in V_m$, $I \in \mathcal{P}_m$ $(m \ge 0)$, and any set $E \subset I$ with $|E| \le \delta|I|$, we have

$$\|g\|_{L_\infty(I)} \le \beta_3\|g\|_{L_\infty(I\setminus E)}. \tag{2.14}$$

This condition is in essence the local linear independence of the φ_θ's which is known to hold in many cases of interest, see the examples below.

For the purpose of nonlinear approximation in L_p, $0 < p < \infty$, we shall assume that the L_p-analogue of (2.14) is valid:

$$\|g\|_{L_p(I)} \le \beta_3\|g\|_{L_p(I\setminus E)}. \tag{2.15}$$

The only use of (2.14) and (2.15) is in the proof of the corresponding Bernstein estimates (see Theorem 4.2 below).

Depending on the domain Ω in some settings one can even construct wavelet or prewavelet bases. For simplicity, whenever we assume in this paper the existence of wavelets, we assume the existence of a biorthogonal wavelet basis $\Psi = \{\psi_\lambda : \lambda \in \mathcal{L}\}$ on Ω with a dual $\widetilde{\Psi} = \{\widetilde{\psi}_\lambda : \lambda \in \mathcal{L}\}$, where $\mathcal{L} = \bigcup_{m\in\mathbf{N}_0}\mathcal{L}_m$ is the index set of the "true" wavelets. Then each $f \in L_p(\Omega)$ $(1 \le p \le \infty)$ has the representation

$$f = \sum_{\theta\in\Theta_0} c_\theta(f)\varphi_\theta + \sum_{m\in\mathbf{N}_0}\sum_{\lambda\in\mathcal{L}_m} c_\lambda(f)\psi_\lambda, \qquad c_\lambda(f) := \langle f, \widetilde{\psi}_\lambda\rangle, \tag{2.16}$$

which is assumed to be unconditional if $1 < p < \infty$. In addition, we assume that ψ_λ, $\widetilde{\psi}_\lambda$ are compactly supported with $\operatorname{supp}\psi_\lambda$, $\operatorname{supp}\widetilde{\psi}_\lambda \subset \lambda$, and $\lambda = \bigcup_{I\in\mathcal{N}_\lambda} I$, where $\mathcal{N}_\lambda \subset \mathcal{P}_{m+1}$ if $\lambda \in \mathcal{L}_m$, and $\#\mathcal{N}_\lambda \le \nu_w$ with $\nu_w =$ constant. Also, we assume that for $\lambda \in \mathcal{L}_m$, $\psi_\lambda \in V_{m+1}$, i.e., $\psi_\lambda = \sum_{\theta\in\Theta_{m+1}} a_{\lambda,\theta}\varphi_\theta$, and $|a_{\lambda,\theta}| \le \beta_4$ with β_4 a constant. Our last assumption is that ψ_λ are at least continuous, $\|\psi_\lambda\|_\infty = 1$, and $\|\widetilde{\psi}_\lambda\|_\infty < \infty$.

2.2. *Examples of MRAs*

In this section we briefly outline some examples covered by the above framework. This list is by no means meant to be exhaustive.

Shift-Invariant Refinable Functions. The classical approach to constructing wavelets on $\mathbf{R}$ is based on hierarchies of nested shift-invariant spaces spanned by the dilated translates $\varphi(2^m \cdot -k)$, $k \in \mathbf{Z}$, of a single *scaling function* φ or, more generally, of a finite number $\varphi^i(2^m \cdot -k)$, $i = 1, \ldots, r$, $k \in \mathbf{Z}$, of *multiscaling functions*, which are refinable, i.e.,

$$\varphi = \sum_{k \in \mathbf{Z}} a_k \varphi(2 \cdot -k) \qquad \text{or} \qquad \varphi^i = \sum_{j=1}^{r} \sum_{k \in \mathbf{Z}} a_k^j \varphi^j(2 \cdot -k)$$

holds for some mask sequences $(a_k^i)_{k \in \mathbf{Z}}$. These translates are usually required to have some stability properties such as *linearly independent* integer translates, i.e., $\sum_{k \in \mathbf{Z}} c_k \varphi(\cdot - k) = 0$ implies $c_k = 0, k \in \mathbf{Z}$. It is known that this latter fact implies the existence of local dual functionals in the sense of (2.7). For the most prominent examples, such as cardinal B-splines or the family of orthonormal Daubechies scaling functions, one even has that the dual functionals are also refinable scaling functions [14], [7]. In this case even local linear independence of the scaling functions is known to hold [5]. This means that, whenever a linear combination of such scaling functions vanishes on any given open neighborhood, the coefficients of those scaling functions whose support intersects this neighborhood have to be zero. This setting hosts the well-known local orthonormal or biorthogonal bases for $L_2(\mathbf{R})$.

As mentioned above, the local independence implies property (2.14). Moreover, fixing any interval Ω, say, we can take here

$$\Phi_m = \{\varphi(2^m \cdot -k) : k \in \mathbf{Z},\ \operatorname{supp} \varphi(2^m \cdot -k) \subset \Omega \ (\text{or } (\operatorname{supp} \varphi(2^m \cdot -k))^\circ \cap \Omega \neq \emptyset)\}.$$

Here $\mathcal{P}_m$ consists of the dyadic intervals of length 2^{-m} contained in Ω, while the $\theta \in \Theta_m$ are unions of finitely many dyadic intervals.

Of course, taking tensor products provides analogous MRAs on domains Ω which are finite connected unions of integer translates of the unit cube, the cells being dyadic cubes now.

A classical class of nontensor product shift-invariant multivariate MRAs satisfying the above requirements is based on the notion of *box-spline*. In this case, stability, linear independence, and local linear independence are known to be equivalent properties whose validity can be characterized completely in terms of the generating set of directions, see, e.g., [10].

Wavelets on the Interval. The biorthogonal or orthogonal shift-invariant MRAs on $\mathbf{R}$ can be used as a starting point for constructing an MRA on a fixed finite interval $[0, M]$, say, along with corresponding biorthogonal or orthogonal bases, see, e.g., [5], [9], [8]. Instead of taking just basis functions whose supports are contained in a given domain or its restrictions to such a domain, one proceeds as outlined next, first again for the univariate case and a fixed integer interval Ω. The idea is to generate V_0 as the span of all integer translates of a scaling function φ whose supports are fully contained in $(0, M)$ and by finitely many additional basis functions near the endpoints of the interval, which are formed as finite linear combinations of the $\varphi(\cdot - k)$ so as to retain some polynomial exactness and refinability. The V_j, $j > 0$, are obtained by scaling. One still has local biorthogonal bases so that (2.7) and (2.14) remain valid. These boundary adaptations allow one to construct a dual pair of biorthogonal MRAs on Ω which in turn lead to the construction of wavelet bases on Ω.

Parametrically Lifted MRA and Wavelets on Domains. Once boundary adapted MRAs of the above type are available, one can construct MRAs on more complicated domains whose boundary is not necessarily aligned with the coordinate axes. In fact, on can deal with domains of the type

$$\Omega = \bigcup_{\kappa \in \mathcal{K}} \kappa(\square),$$

where $\square$ is again the unit d-cube and the κ are regular parametric mappings. Corresponding parametric liftings of the $\mathrm{MRA}_\square$ on $\square$ can be stitched together to form even a globally smooth MRA on Ω which inherits the relevant properties of $\mathrm{MRA}_\square$. For details the reader is referred, e.g., to [3], [4], [11], [12], [6].

Finite Elements. Suppose that $\mathcal{P}_0$ is a locally quasi-uniform, shape regular triangulation of the polyhedral domain Ω and that each $\mathcal{P}_m$ for $m > 0$ arises from $\mathcal{P}_0$ through m successive regular subdivisions. Examples for $d = 2$ are based on decomposing each triangle into four congruent children or into two triangles by splitting the longest edge. Similar procedures are known for $d = 3$. In this case the cells are triangles or, more generally, simplices. Finite element spaces of degree k on such partitions are usually defined as linear spans of nodal basis functions which are (globally continuous, sometimes even C^1) piecewise polynomials on these partitions which are dual to suitable collections of nodal values (point values or derivatives) at the vertices or midpoints of edges. The simplest examples are continuous Lagrange finite elements of degree k in the plane where the nodal values are associated with a regular "k-mesh" which is the refined triangulation obtained by subdividing each triangle in $\mathcal{P}_m$ into k^2 congruent subtriangles. Since on each cell the same number of basis functions overlap, namely the dimension of the generated polynomial space, local linear independence and hence property (2.14) holds. Moreover, the construction of a local dual basis, consisting of (discontinuous) piecewise polynomials of the same degree, is straightforward, so that all the above assumptions can be verified in this case as well, see, e.g., [13] for wavelet bases in the finite element context.

Anisotropic Spline Bases over Multilevel Nested Triangulations. For a given bounded polygonal domain $\Omega \subset \mathbf{R}^2$, consider a sequence of triangulations $(\mathcal{P}_m)_{m\in\mathbf{N}_0}$ such that each level $\mathcal{P}_m$ is a partition of Ω into triangles and a refinement of the previous level $\mathcal{P}_{m-1}$. Write $\mathcal{P} := \bigcup_{m\in\mathbf{N}_0} \mathcal{P}_m$. Each such sequence of triangulations generates an MRA of spaces $\mathcal{S}_0 \subset \mathcal{S}_1 \subset \cdots$ consisting of piecewise linear functions, where $\mathcal{S}_m$ ($m \geq 0$) is spanned by all Courant elements φ_θ supported on cells θ at the mth level $\mathcal{P}_m$. Natural mild conditions should be imposed on the triangulations in order that this MRA satisfies our conditions from Section 2.1 (see [23], [24] for the exact conditions; $\mathcal{P}$ is then called a *locally regular triangulation*). These conditions essentially do not allow the areas of the triangles to change uncontrollably when moving away from a fixed triangle in $\mathcal{P}$ with regard to scale and spatial location. On the other hand, the conditions still allow the triangles in $\mathcal{P}$ to change in size, shape, and orientation quickly when moving around at a given level or across the levels. In particular, triangles with arbitrarily sharp angles are permitted in any location and at any level. The above-described hierarchy of linear splines provides a simple example of an MRA which may have a very anisotropic structure.

To give an example of more general anisotropic MRAs, consider now the hierarchy $\mathcal{S}_0 \subset \mathcal{S}_1 \subset \cdots$, where $\mathcal{S}_m := \mathcal{S}^{k,r}(\mathcal{P}_m)$ is the space of all r-times differentiable piecewise polynomials over the triangles of $\mathcal{P}_m$ of degree $< k$ ($k \geq 1$). In [16], a construction of spline basis Φ_m in $\mathcal{S}^{k,r}(\mathcal{P}_m)$ is given, whenever $r \geq 1$ and $k > 4r + 1$, in the case of $\Omega = \mathbf{R}^2$. It is shown that under some reasonable conditions on the triangulations $(\mathcal{P}_m)_{m\in\mathbf{Z}}$ of $\mathbf{R}^2$ the bases $(\Phi_m)_{m\in\mathbf{Z}}$ satisfy our conditions on MRAs from Section 2.1. In particular, these conditions admit arbitrarily sharp angles and offer considerable flexibility. The triangulations satisfying these conditions are called *strong locally regular triangulations*. If one considers a sequence of triangulations $(\mathcal{P}_m)_{m\in\mathbf{N}_0}$ on a compact domain $\Omega \subset \mathbf{R}$, then the usual modifications (see [15]) of the basis functions corresponding to boundary edges or vertices lead again to bases satisfying our conditions. The construction in [16] can be extended to the spaces $\mathcal{S}^{k,r}(\mathcal{P}_m)$, $k > r2^d + 1$, in dimensions $d > 2$.

MRAs Consisting of Discontinuous Functions. MRAs consisting of (discontinuous) piecewise polynomials are completely legitimate as well. Such hierarchies can be defined over regular (uniform) or irregular simplicial or other partitions of a compact domain in $\mathbf{R}^d$. See [23], [26] for more details in the anisotropic case. Due to the more enhanced locality of corresponding basis functions (e.g. supports and cells agree in this case) the analysis becomes simpler in many ways. In this paper we therefore focus our attention on MRAs consisting of continuous or even more regular functions.

2.3. *Geometric Properties and Further Prerequisites*

Refined properties of the above examples involve, in one way or another, the geometry of the supports of the basis functions. In spite of the difference of respective geometric settings the relevant properties turn out to be governed by the same abstract mechanism. The goal of this section is to extract and bring out the essential mechanism in order to allow us to provide a unified treatment of the above and many other cases.

In order to deal with neighborhood relations in such partitions under possibly general circumstances it is convenient to employ the notion of the mth *level star* of a set. For a given set $E \subset \Omega$ and level $m \geq 0$, we define

$$\mathrm{Star}^{(m)}(E) := \mathrm{Star}_1^{(m)}(E) := \bigcup\{I \in \mathcal{P}_m : I \cap E \neq \emptyset\}$$

and, inductively,

$$\mathrm{Star}_j^{(m)}(E) := \mathrm{Star}_1^{(m)}(\mathrm{Star}_{j-1}^{(m)}(E)), \qquad j > 1.$$

One can easily show that

$$\mathrm{Star}_{j_1+j_2}^{(m)}(E) = \mathrm{Star}_{j_1}^{(m)}(\mathrm{Star}_{j_2}^{(m)}(E)), \qquad j_1, j_2 \geq 1.$$

We shall drop the reference to m whenever the level is clear from the context which is, for instance, the case when the set E has a specific level such as the indices $\theta \in \Theta_m$ or the cells $I \in \mathcal{P}_m$. When E consists of a single point x we write, with a slight abuse of notation, briefly, $\mathrm{Star}_j^{(m)}(x)$ instead of $\mathrm{Star}_j^{(m)}(\{x\})$.

The extent to which the supports θ overlap can be conveniently expressed in terms of stars as well. We record for later use the following consequence of (α):

(ε) For each $\theta \in \Theta_m$,

$$\theta \subset \mathrm{Star}_{\nu_*}^{(m)}(x) \qquad \text{for} \quad x \in \theta, \tag{2.17}$$

where $\nu_* \le \nu_1$ is also a constant, see property (α) in Section 2.1.

We now state one more condition on the cells which guarantees that they are properly refined, i.e., as in all our examples, all "sides" of the cells are subdivided (in a weakly isotropic fashion).

(f) There exists a constant $\tilde{\nu} \ge 1$ such that

$$\mathrm{Star}_2^{(m+\tilde{\nu})}(E) \subset \mathrm{Star}_1^{(m)}(E), \qquad E \subset \Omega. \tag{2.18}$$

The fact that the supports θ overlap causes some "spatial pollution" across different levels. The following notion helps us to quantify this effect.

Connecting by n-Stars. For $\theta, \eta \in \Theta$ with $l(\eta) \ge l(\theta)$, we say that θ is connected with η by n-stars ($n \ge 1$) if there exist cells I_j, $j = 1, \dots, r$, such that:

(i) $l(I_1) \ge l(\theta) + 1, l(I_{j+1}) \ge l(I_j) + 1, j = 1, \dots, r-1, l(I_r) \le l(\eta)$;
(ii) $I_1 \subset \mathrm{Star}_n^{(l_0)}(\theta)$, $I_2 \subset \mathrm{Star}_n^{(l_1)}(I_1)$, $\dots$, $I_r \subset \mathrm{Star}_n^{(l_{r-1})}(I_{r-1})$, $\eta \subset \mathrm{Star}_n^{(l_r)}(I_r)$, where $l_0 := l(\theta), l_j := l(I_j)$.

Lemma 2.2. *If $\theta, \eta \in \Theta$ with $l(\eta) \ge l(\theta)$, and θ is connected with η by n-stars ($n \ge 1$), then independently of the number of the connecting cells*

$$\eta \subset \mathrm{Star}_{2\tilde{\nu}n}^{(m)}(\theta), \qquad m := l(\theta), \tag{2.19}$$

where $\tilde{\nu}$ is from (2.18).

Proof. Using the monotonicity of the nth stars with respect to the levels ($\mathrm{Star}_n^{(l+1)}(E) \subset \mathrm{Star}_n^{(l)}(E)$) it is easy to see that for the proof of the lemma it suffices to show that, for any $k \ge 1$,

$$\mathrm{Star}_n^{(m+k)}(\mathrm{Star}_n^{(m+k-1)}(\cdots \mathrm{Star}_n^{(m)}(\theta) \cdots)) \subset \mathrm{Star}_{2\tilde{\nu}n}^{(m)}(\theta). \tag{2.20}$$

Let $k =: \tilde{\nu} j + i$, where $j \ge 0$ and $0 \le i < \tilde{\nu}$. Again, by the monotonicity of the stars, it is clear that (2.20) will follow if we prove that

$$\mathrm{Star}_{\tilde{\nu}n}^{(m+\tilde{\nu}j)}(\mathrm{Star}_{\tilde{\nu}n}^{(m+\tilde{\nu}(j-1))}(\cdots \mathrm{Star}_{\tilde{\nu}n}^{(m)}(\theta) \cdots)) \subset \mathrm{Star}_{2\tilde{\nu}n}^{(m)}(\theta). \tag{2.21}$$

One proves this by applying recursively the inclusion

$$\mathrm{Star}_{2\tilde{\nu}n}^{(l+\tilde{\nu})}(\mathrm{Star}_{\tilde{\nu}n}^{(l)}(E)) \subset \mathrm{Star}_{2\tilde{\nu}n}^{(l)}(E), \qquad l \ge 0, \quad E \subset \Omega,$$

which follows by applying (2.18) $\tilde{\nu}n$ times. ■

The possibly significant overlap of the supports θ is a severe obstruction to localizing estimates. In order still to be able to manage such pollution effects, we require an auxiliary multilevel system of overlapping cells that are, on one hand, simple enough to be disentangled while, on the other hand, they essentially scale like the actual supports.

Extended Cells. We assume the existence of a collection of overlapping *extended cells*

$$\mathcal{O} = \bigcup_{m \in \mathbf{N}_0} \mathcal{O}_m$$

with the following properties:

(i) Every level $\mathcal{O}_m$ is a cover of Ω, i.e., $\Omega = \bigcup_{\omega \in \mathcal{O}_m} \omega$.
(ii) Each extended cell $\omega \in \mathcal{O}_m$ can be "paved" by cells from the same level $\mathcal{P}_m$, i.e., $\omega = \bigcup_{I \in \mathcal{N}_\omega} I$ with $\mathcal{N}_\omega \subset \mathcal{P}_m$.
(iii) If $\omega \in \mathcal{O}_m$, then

$$\omega \subset \mathrm{Star}_{\nu_4}^{(m)}(x) \qquad \text{for} \quad x \in \omega, \tag{2.22}$$

where ν_4 is a constant satisfying $1 \le \nu_4 \le \nu_*$.
(iv) For every $\omega \subset \mathcal{O}_m$, $m \ge 1$, there exists $\omega' \in \mathcal{O}_{m-1}$ such that $\omega \subset \omega'$.
(v) For every $\omega_1 \subset \mathcal{O}$ there exists $\omega_2 \in \mathcal{O}$ such that

$$\mathrm{Star}_1(\omega_1) \subset \omega_2 \qquad \text{and} \qquad l(\omega_2) \ge l(\omega_1) - \nu_5, \tag{2.23}$$

whenever $l(\omega_1) \ge \nu_5$, where $\nu_5 \ge 1$ is a constant and $l(\omega)$ denotes the level of ω.
(vi) *Coloring property*. The set $\mathcal{O}$ can be represented as a finite disjoint union of subsets $\{\mathcal{O}^j\}_{j=1}^J$ such that each set $\mathcal{O}^j$ is a tree with respect to the inclusion relation, that is, if $\omega', \omega'' \in \mathcal{O}^j$ and $(\omega')^\circ \cap (\omega'')^\circ \ne \emptyset$, then either $\omega' \subset \omega''$ or $\omega'' \subset \omega'$.

The existence of the extended cells $\omega \in \mathcal{O}$ is, in general, not a consequence of the conditions on the cells $I \in \mathcal{P}$ and supports $\theta \in \Theta$. One should think of extended cells $\omega \in \mathcal{O}_m$ as simple regions of type $\omega = \mathrm{Star}_1^{(m)}(v)$ with v a point in Ω (which in the case of Courant elements agrees with the supports θ) or $\omega = \mathrm{Star}_1^{(m)}(I)$ with $I \in \mathcal{P}_m$. This is the case in all examples mentioned in Section 2.2. The supports $\theta \in \Theta$, however, can be much larger than the extended cells $\omega \in \mathcal{O}$.

The coloring property (vi) of the extended cells is the reason for introducing them here. It is not clear whether this holds directly for the supports θ. In the case when $\mathcal{P}$ consists of dyadic cubes in $\mathbf{R}^d$, the coloring property is established in [20], and in the case when $\mathcal{P}$ consists of triangles (in $\mathbf{R}^2$), such a result is proved in [24]. The proof of the coloring lemma from [24] can be carried over to spatial dimensions $d \ge 3$.

Our final assumption on the supports θ, which is also satisfied in the examples listed in Section 2.2, couples the system of extended cells with the supports θ.

(ζ) For each $\omega \in \mathcal{O}_m$ ($m \ge 0$) there exists $\theta \in \Theta_m$ such that $\omega \subset \theta$.

Lemma 2.3. *Suppose $m \ge jK$, where $j \ge 1$ and $K := \nu_* \nu_5$. For any $\theta \in \Theta_m$ there exists $\omega \in \mathcal{O}$ such that*

$$\mathrm{Star}_j^{(m)}(\theta) \subset \omega \qquad \textit{and} \qquad l(\omega) = m - jK. \tag{2.24}$$

Moreover,

$$\mathrm{Star}_j^{(m)}(\theta) \subset \mathrm{Star}_{\nu_4}^{(m-jK)}(x) \qquad \textit{for} \quad x \in \theta. \tag{2.25}$$

Proof. In view of (2.23), it suffices to prove the lemma only in the case $j = 1$. Choose $I \in \mathcal{P}$ and $\omega \in \mathcal{O}$ so that $l(I) = l(\omega) = l(\theta)$, $I \subset \theta$, and $I \subset \omega$. Then, by (2.17),

$$\mathrm{Star}_1(\theta) \subset \mathrm{Star}_{\nu_*}(I) \subset \mathrm{Star}_{\nu_*}(\omega).$$

Using (2.23), there exist extended cells $\omega_0 := \omega, \omega_1, \dots, \omega_{\nu_*}$ such that

$$\mathrm{Star}_1(\omega_i) \subset \omega_{i+1} \qquad \text{and} \qquad l(\omega_{i+1}) \geq l(\omega_i) - \nu_5.$$

and hence

$$\mathrm{Star}_{\nu_*}(\theta) \subset \omega_{\nu_*} \qquad \text{and} \qquad l(\omega_{\nu_*}) \geq l(\theta) - \nu_*\nu_5,$$

which implies (2.24).

Fix $x \in \theta$. By (2.24) and (2.22), we obtain $\mathrm{Star}_j^{(m)}(\theta) \subset \omega \subset \mathrm{Star}_{\nu_4}^{(m-jK)}(x)$. ■

In the following all constants will depend on $r, \rho, \vartheta, \delta, \beta_0, \dots, \beta_4, \nu_0, \dots, \nu_5, \tilde{\nu}, \nu_*, \nu_w$, and $\#\Theta_0$ (or at least some of them), which are not completely independent. We shall refer to them as parameters of the MRA which is being currently used.

2.4. *Local Approximation from V_m and Projectors*

As in [16], [23], [24] a scale of B-spaces induced by the multiresolution hierarchy will play an essential role in the subsequent analysis. The local approximation from the spaces V_m will be an important element in the definition of these B-spaces. We first define, for a given cell $I \in \mathcal{P}_m$ ($m \geq 0$), the extension $\widehat{I}$ by

$$\widehat{I} := \bigcup_{\theta \in \Theta_m :\, I \subset \theta} \theta. \tag{2.26}$$

Clearly, $|\widehat{I}| \leq c|I|$ with c depending only on the parameters of the MRA.

For a given function $f \in L_q(\Omega)$ and $I \in \mathcal{P}_m$ ($m \geq 0$), the error of L_q-approximation to f on $\widehat{I}$ from V_m is defined by

$$\mathcal{E}(f, \widehat{I})_q := \inf_{g \in V_m} \|f - g\|_{L_q(\widehat{I})}. \tag{2.27}$$

We define

$$Q_m(f) := \sum_{\theta \in \Theta_m} c_\theta(f)\varphi_\theta, \qquad f \in L_1(\Omega), \tag{2.28}$$

where $c_\theta(f)$ are extensions of the linear functionals from (2.6) which satisfy (2.8). Clearly, $Q_m : L_1(\Omega) \to V_m$ is a linear projector onto V_m.

Lemma 2.4. *If $f \in L_q(\Omega)$, $1 \leq q \leq \infty$, and $I \in \mathcal{P}_m$, $m \geq 0$, then*

$$\|Q_m(f)\|_{L_q(I)} \leq c^\flat \|f\|_{L_q(\widehat{I})} \tag{2.29}$$

and

$$\|f - Q_m(f)\|_{L_q(I)} \leq c^\flat \mathcal{E}(f, \widehat{I})_q, \tag{2.30}$$

where $c^\flat$ depends only on q and the parameters of the MRA.

Proof. Estimates (2.29)–(2.30) readily follow by property (2.8) of the linear functionals $c_\theta(f)$ (see also [16], [23]). ■

We use the projectors Q_m for decomposing a given function into multilevel components. We denote by

$$q_m := Q_m - Q_{m-1} \qquad \text{where} \quad Q_{-1} := 0, \tag{2.31}$$

the "detail" of f between the levels m and $m-1$. Whenever a wavelet basis is available, q_m is understood to arise from the associated canonical projectors, i.e.,

$$q_m(f) = \sum_{\lambda \in \mathcal{L}_{m-1}} c_\lambda(f)\psi_\lambda, \qquad c_\lambda(f) := \langle f, \tilde{\psi}_\lambda \rangle.$$

In general, for a given function $f \in L_1(\Omega)$, one has $q_m(f) \in V_m$ and hence

$$q_m(f) =: \sum_{\theta \in \Theta_m} b_\theta(f)\varphi_\theta. \tag{2.32}$$

From the approximation properties of the spaces V_m, we therefore know that, for $f \in L_q(\Omega)$, $1 \le q \le \infty$, the expansion

$$f = \sum_{m \in \mathbf{N}_0} q_m(f) = \sum_{m \in \mathbf{N}_0} \sum_{\theta \in \Theta_m} b_\theta(f)\varphi_\theta \tag{2.33}$$

converges in L_q.

For the purposes of nonlinear approximation in L_p, $0 < p \le 1$, we modify the above construction in a standard way as described in the following. Denote by V_m' the linear space of all piecewise V_m-functions over $\mathcal{P}_m$, i.e., $g \in V_m'$ if $g = \sum_{J \in \mathcal{P}_m} g_J \cdot \mathbf{1}_J$, where $g_J \in V_m$. For a given $I \in \mathcal{P}_m$, let $P_{I,q} : L_q(I) \to V_m|_I$ be a (nonlinear) projector such that

$$\|f - P_{I,q}(f)\|_{L_q(I)} \le c\mathcal{E}(f, I)_q \qquad \text{with} \qquad \mathcal{E}(f, I)_q := \inf_{g \in V_m} \|f - g\|_{L_q(I)}.$$

We now define the operator (projector) $p_{m,q} : L_q(\Omega) \to V_m'$ by

$$p_{m,q}(f) := \sum_{J \in \mathcal{P}_m} P_{I,q}(f) \cdot \mathbf{1}_J.$$

Finally, we consider the operator $Q_{m,q} : L_q(\Omega) \to V_m$ defined by $Q_{m,q}(f) := Q_m(p_{m,q}(f))$. It is easy to see that $Q_m := Q_{m,q}$ satisfies (2.29)–(2.30) if $0 < q \le \infty$. In going further, we set $q_m := q_{m,q} := Q_{m,q} - Q_{m-1,q}$ with $Q_{-1,q} := 0$, and define $\{b_{\theta,q}(f)\}_{\theta \in \Theta_m}$ similarly as in (2.32). Now, we have the following representation of any $f \in L_q(\Omega)$, $0 < q \le \infty$,

$$f = \sum_{m \in \mathbf{N}_0} q_{m,q}(f) = \sum_{m \in \mathbf{N}_0} \sum_{\theta \in \Theta_m} b_{\theta,q}(f)\varphi_\theta \quad \text{in } L_q. \tag{2.34}$$

See [16] for more details of the above in the spline case.

3. B-Spaces and Besov spaces

We first introduce the B-spaces, which will be an important vehicle in showing that the "Push-the-Error" algorithm captures the rates of the best nonlinear n-term approximation. As elsewhere, we assume that $0 < p \le \infty$, and $\alpha \ge 1$ if $p = \infty$ and $\alpha > 0$ if $p < \infty$. In both cases, we set $1/\tau := \alpha + 1/p$. Here we impose the restriction $\alpha \ge 1$ when $p = \infty$ since otherwise the important embedding of the space $\mathcal{B}_\tau^\alpha(\mathcal{M})$ (defined below) in $L_\infty(\Omega) := C(\Omega)$ is not true (see Remark 3.3).

The Case $1 < p \le \infty$. Given an MRA $\mathcal{M}$ with a set of basis (scaling) functions $\Phi = \bigcup_{m\in\mathbf{N}_0} \Phi_m$, we define the B-space $\mathcal{B}_\tau^\alpha = \mathcal{B}_\tau^\alpha(\mathcal{M})$ as the set of all $f \in L_1(\Omega)$ such that

$$|f|_{\mathcal{B}_\tau^\alpha(\mathcal{M})} := \left(\sum_{I\in\mathcal{P}} (|I|^{-\alpha-1+1/\tau}\mathcal{E}(f,\widehat{I})_1)^\tau\right)^{1/\tau} < \infty, \tag{3.1}$$

where $\mathcal{E}(f,\widehat{I})_1$ denotes the error of L_1-approximation to f on $\widehat{I}$ from V_m if $I \in \mathcal{P}_m$ (see (2.27)). Clearly, $|\cdot|_{\mathcal{B}_\tau^\alpha(\mathcal{M})}$ is a semi-norm if $\tau \ge 1$ and semi-quasi-norm if $\tau < 1$. For different purposes it will be convenient to employ different equivalent norms. We shall next introduce these variants.

The local approximation in L_1 above can be replaced by approximation in L_q with an arbitrary $q < p$ (but not with $q \ge p$). Namely, for $1 \le q < p$, we define

$$|f|^{\mathcal{E}_q}_{\mathcal{B}_\tau^\alpha(\mathcal{M})} := \left(\sum_{I\in\mathcal{P}} (|I|^{-\alpha-1/q+1/\tau}\mathcal{E}(f,\widehat{I})_q)^\tau\right)^{1/\tau} \approx |f|_{\mathcal{B}_\tau^\alpha(\mathcal{M})}. \tag{3.2}$$

For the proof of the above equivalence, see Theorem 3.4 below.

By (3.10) below, it follows that $\mathcal{B}_\tau^\alpha$ is embedded in L_p and hence it is natural to define a (quasi-)norm in $\mathcal{B}_\tau^\alpha$ by

$$\|f\|_{\mathcal{B}_\tau^\alpha} := \|f\|_p + |f|_{\mathcal{B}_\tau^\alpha}. \tag{3.3}$$

We also set

$$\|f\|^{\mathcal{E}_q}_{\mathcal{B}_\tau^\alpha} := \|f\|_p + |f|^{\mathcal{E}_q}_{\mathcal{B}_\tau^\alpha}. \tag{3.4}$$

The space $\mathcal{B}_\tau^\alpha$ has an atomic decomposition. We define

$$\|f\|^{A}_{\mathcal{B}_\tau^\alpha(\mathcal{M})} := \inf_{f=\sum_{\theta\in\Theta} a_\theta\varphi_\theta} \left(\sum_{\theta\in\Theta} (|\theta|^{-\alpha-1+1/\tau}\|a_\theta\varphi_\theta\|_1)^\tau\right)^{1/\tau}, \tag{3.5}$$

where the infimum is over all representations of f in $L_1(\Omega)$. By (2.11), we have

$$\|f\|^{A}_{\mathcal{B}_\tau^\alpha(\mathcal{M})} \approx \inf_{f=\sum_{\theta\in\Theta} a_\theta\varphi_\theta} \left(\sum_{\theta\in\Theta} \|a_\theta\varphi_\theta\|_p^\tau\right)^{1/\tau}. \tag{3.6}$$

Another important fact is that the norm in $\mathcal{B}_\tau^\alpha$ can be realized by decompositions using simple projectors. Let $f = \sum_{\theta\in\Theta} b_\theta(f)\varphi_\theta$ be the decomposition of f from (2.33). We define

$$\|f\|^{Q}_{\mathcal{B}^\alpha_\tau(\mathcal{M})} := \left(\sum_{\theta\in\Theta}(|\theta|^{-\alpha-1+1/\tau}\|b_\theta(f)\varphi_\theta\|_1)^\tau\right)^{1/\tau}. \tag{3.7}$$

The norm equivalence (2.11) yields

$$\|f\|^{Q}_{\mathcal{B}^\alpha_\tau(\mathcal{M})} \approx \left(\sum_{\theta\in\Theta}\|b_\theta(f)\varphi_\theta\|_p^\tau\right)^{1/\tau}. \tag{3.8}$$

Finally, the B-spaces have equivalent norms through wavelets or prewavelets, whenever the latter are available. Suppose a wavelet basis exists and satisfies the conditions from Section 2.1. Let $f \in L_1(\Omega)$ and

$$f = \sum_{\theta\in\Theta_0} c_\theta\varphi_\theta + \sum_{\lambda\in\mathcal{L}} c_\lambda\psi_\lambda.$$

We define

$$\|f\|^{W}_{\mathcal{B}^\alpha_\tau(\mathcal{M})} := \left(\sum_{\theta\in\Theta_0}\|c_\theta\varphi_\theta\|_p^\tau + \sum_{\lambda\in\mathcal{L}}\|c_\lambda\psi_\lambda\|_p^\tau\right)^{1/\tau}. \tag{3.9}$$

The Case $0 < p \le 1$. We recall our standing assumptions: $\alpha > 0$ and $1/\tau := \alpha + 1/p$. In this case we define $|f|^{\mathcal{E}_q}_{\mathcal{B}^\alpha_\tau(\mathcal{M})}$, $0 < q < p$, as in (3.2) and set $|f|_{\mathcal{B}^\alpha_\tau(\mathcal{M})} := |f|^{\mathcal{E}_\tau}_{\mathcal{B}^\alpha_\tau(\mathcal{M})}$. We also define the quasi-norms $\|f\|_{\mathcal{B}^\alpha_\tau(\mathcal{M})}$ and $\|f\|^{\mathcal{E}_q}_{\mathcal{B}^\alpha_\tau(\mathcal{M})}$ as in (3.3)–(3.4). Further, we introduce the atomic quasi-norm $\|f\|^{A}_{\mathcal{B}^\alpha_\tau(\mathcal{M})}$ by the quantity on the right-hand side in (3.6) and define the quasi-norm $\|f\|^{Q}_{\mathcal{B}^\alpha_\tau(\mathcal{M})}$ by the right-hand side quantity in (3.8) with $b_\theta(f)$ replaced by $b_{\theta,q}(f)$ from (2.34) for some $0 < q < p$. The only substantial difference in the definition of the B-spaces, when $0 < p \le 1$ compared with the case $1 < p \le \infty$, is that in the definition of $\|f\|^{Q}_{\mathcal{B}^\alpha_\tau(\mathcal{M})}$ the projectors Q_m defined in (2.28) are replaced by the projectors $Q_{m,q}$ (see the end of Section 2.4) and also wavelets are no longer usable.

Remark 3.1. In the above definition of $\|f\|^{\mathcal{E}_q}_{\mathcal{B}^\alpha_\tau(\mathcal{M})}$, $\|f\|_{\mathcal{B}^\alpha_\tau(\mathcal{M})} := \|f\|^{\mathcal{E}_1}_{\mathcal{B}^\alpha_\tau(\mathcal{M})}$ $(q = 1)$, and $\|f\|^{Q}_{\mathcal{B}^\alpha_\tau(\mathcal{M})}$ via $\{b_{\theta,q}(f)\}$ or $\{b_\theta(f)\}$ $(q = 1)$ it is imperative to have $q < p$. Therefore, it is important that (Q_m) satisfy (2.29)–(2.30) for some $q < p$, which essentially follows by condition (2.8) on the duals $\{c_\theta(\cdot)\}$. In turn, condition (2.8) can be relaxed somewhat; it can be replaced by $|c_\theta(\cdot)| \le c|\theta|^{-1/q}\|f\|_q$ with $1 < q < p$.

The following embedding result, proved in [23], [26], will play an important role.

Theorem 3.2. *For any collection of real numbers* $\{c_\theta\}_{\theta\in\Theta}$, *and* $0 < \tau < p < \infty$ *or* $p = \infty$ *and* $0 < \tau \le 1$, *we have*

$$\left\|\sum_{\theta\in\Theta}|c_\theta|\,\varphi_\theta\right\|_p \le c\left(\sum_{\theta\in\Theta}\|c_\theta\varphi_\theta\|_p^\tau\right)^{1/\tau}, \tag{3.10}$$

where c depends only on τ, p, *and the parameters of the MRA.*

Remark 3.3. It is easy to see that estimate (3.10) is not true if $p = \infty$ and $\tau < 1$, and consequently the space $\mathcal{B}^\alpha_\tau(\mathcal{M})$ is not embedded in $C(\Omega)$, which is critical for our further development. This is the reason for imposing the restriction $\alpha \geq 1$ when $p = \infty$.

The announced equivalence result now reads as follows:

Theorem 3.4. *For a given MRA the norms* $\|\cdot\|_{\mathcal{B}^\alpha_\tau}$, $\|\cdot\|^{\mathcal{E}_q}_{\mathcal{B}^\alpha_\tau}$, $\|\cdot\|^{A}_{\mathcal{B}^\alpha_\tau}$, $\|\cdot\|^{Q}_{\mathcal{B}^\alpha_\tau}$, *and* $\|\cdot\|^{W}_{\mathcal{B}^\alpha_\tau}$, *if (pre)wavelets exist and* $p > 1$, *are equivalent with constants of equivalence depending only on* p, α, *and the parameters of the MRA.*

The proof of this theorem is quite similar (but not identical) to the proofs of the corresponding results in [16], [23]. For completeness, we give this proof in the Appendix (Section 8).

The following Sobolev-type embedding result follows immediately by (3.6) or (3.8): If $0 < \alpha_0 < \alpha_1$ and $\tau_j := (\alpha_j + 1/p)^{-1}$, $j = 0, 1$, then $\mathcal{B}^{\alpha_1}_{\tau_1}(\mathcal{M}) \subset \mathcal{B}^{\alpha_0}_{\tau_0}(\mathcal{M})$, i.e., if $f \in \mathcal{B}^{\alpha_1}_{\tau_1}(\mathcal{M})$, then $f \in \mathcal{B}^{\alpha_0}_{\tau_0}(\mathcal{M})$ and $\|f\|_{\mathcal{B}^{\alpha_0}_{\tau_0}(\mathcal{M})} \leq c\|f\|_{\mathcal{B}^{\alpha_1}_{\tau_1}(\mathcal{M})}$.

Since the B-spaces are essentially sequence spaces (retracts of sequence spaces [1]) they are easy to interpolate. In particular, the analogue of Theorem 2.12 from [16] holds with a similar proof. We skip the details.

For a given MRA $\mathcal{M}$, more general B-spaces $B^\alpha_{pq}(\mathcal{M})$, $0 < p, q \leq \infty$, $\alpha > 0$, can be defined similarly as in [23] and then $B^\alpha_\tau(\mathcal{M}) = B^\alpha_{\tau\tau}(\mathcal{M})$. The B-spaces should be viewed as nonclassical smoothness spaces which are specifically designed for the needs of nonlinear n-term approximation. A crucial property of the B-spaces is that the basis functions $\{\varphi_\theta\}_{\theta\in\Theta}$ of an MRA $\mathcal{M}$ are infinitely smooth with respect to the scale of the B-spaces $B^\alpha_\tau(\mathcal{M})$. This is reflected by the estimate $\|\varphi_\theta\|_{B^\alpha_\tau(\mathcal{M})} \leq c\|\varphi_\theta\|_p$ for $0 < \alpha < \infty$ (see Theorem 4.2 below). As a consequence, our direct, inverse, and characterization theorems as well as our algorithms impose no restriction on the rates of approximation.

In regular settings the scale of Besov spaces $B^s_\tau(L_\tau(\Omega))$, $1/\tau = s/d + 1/p$, usually arises in nonlinear approximation in $L_p(\Omega)$ (see, e.g., [17]). Note that the smoothness parameters of the Besov spaces and B-spaces are normalized differently. Thus the Besov space $B^{d\alpha}_\tau(L_\tau(\Omega))$ corresponds to the B-space $B^\alpha_\tau(\mathcal{M})$. The Besov regularity of the basis functions $\{\varphi_\theta\}$ determines the smoothness range where the Besov space can be used in nonlinear approximation. To be more precise, assume that, in the setting described in Section 2.1, all $\mathcal{P}_m$ are regular partitions of Ω, that is, for each cell $I \in \mathcal{P}_m$ there exist balls B_{r_1}, B_{r_2} of radii r_1, r_2 such that $B_{r_1} \subset I \subset B_{r_2}$ and $r_2 \leq cr_1$ with c a constant. It is not hard to be seen that if for some $\alpha > 0$, $\|\varphi_\theta\|_{B^{d\alpha}_\tau(L_\tau(\Omega))} \leq c\|\varphi_\theta\|_p$ for all $\varphi_\theta \in \Phi$, then $B^\alpha_\tau(\mathcal{M}) \subset B^{d\alpha}_\tau(L_\tau(\Omega))$ and $\|f\|_{B^{d\alpha}_\tau(L_\tau(\Omega))} \leq c\|f\|_{B^\alpha_\tau(\mathcal{M})}$ (see [16], [23] for the spline case).

In anisotropic setups, when basis functions of strongly elongated supports are involved, the Besov spaces are no longer suitable for characterization of the rates of nonlinear approximation whereas the B-space concept still applies.

B-spaces have been used implicitly or explicitly elsewhere, see, e.g., [25], [2]. They are systematically developed and used in the case of anisotropic MRAs generated by piecewise polynomials in [16], [23], [24], [26], [27].

4. Best Nonlinear n-Term Approximation

Our primary goal in this section is to characterize the approximation spaces generated by nonlinear n-term approximation from the scaling functions of an MRA.

We let Σ_n denote the nonlinear set consisting of all functions g of the form

$$g = \sum_{\theta \in \Lambda} a_\theta \varphi_\theta,$$

where $\Lambda \subset \Theta$, $\#\Lambda \le n$, and Λ is allowed to vary with g. We denote by $\sigma_n(f)_p$ the error of best L_p-approximation to $f \in L_p(\Omega)$ from Σ_n:

$$\sigma_n(f)_p := \inf_{g \in \Sigma_n} \|f - g\|_p.$$

To characterize the approximation spaces generated by $(\sigma_n(f)_p)$, we shall use the machinery of Jackson–Bernstein estimates combined with interpolation (see, e.g., [19], [28]).

As elsewhere, our standing assumption is that $0 < p \le \infty$ and $\alpha \ge 1$ for $p = \infty$ and $\alpha > 0$ if $p < \infty$; in both cases we set $1/\tau := \alpha + 1/p$.

Theorem 4.1 (Jackson Estimate). *If $f \in \mathcal{B}^\alpha_\tau(\mathcal{M})$, then*

$$\sigma_n(f)_p \le c n^{-\alpha} \|f\|_{\mathcal{B}^\alpha_\tau(\mathcal{M})}, \tag{4.1}$$

where c depends only on α, p, and the parameters of the MRA.

Estimate (4.1) follows from the basic estimates of the error of the "Push-the-Error" algorithm ($p = \infty$) and the "Threshold" algorithm ($0 < p < \infty$), stated in Theorems 5.6 and 6.1 below.

Theorem 4.2 (Bernstein Estimate). *If $g \in \Sigma_n$, then*

$$\|g\|_{\mathcal{B}^\alpha_\tau(\mathcal{M})} \le c n^\alpha \|g\|_p, \tag{4.2}$$

where c depends only on α, p, and the parameters of the MRA.

To avoid a major diversion from the presentation of our central results we postpone the proof of this theorem to the Appendix.

One can now follow the standard lines to obtain "regularity-free error estimates." To this end, denote by $K(f,t)_p := K(f, t; L_p(\Omega), \mathcal{B}^\alpha_\tau(\mathcal{M}))$ ($L_\infty(\Omega) := C(\Omega)$) the K-functional defined by

$$K(f,t)_p := \inf_{g \in \mathcal{B}^\alpha_\tau} \|f - g\|_p + t\|g\|_{\mathcal{B}^\alpha_\tau}, \qquad t > 0.$$

The real interpolation space $(L_p, \mathcal{B}^\alpha_\tau)_{\lambda,q}$ is defined as the set of all functions $f \in L_p$ such that

$$\|f\|_{(L_p,\mathcal{B}^\alpha_\tau)_{\lambda,q}} := \left(\int_0^\infty (t^{-\lambda} K(f,t)_p)^q \frac{dt}{t} \right)^{1/q} < \infty.$$

(For more details see, e.g., [1].)

By standard arguments (see, e.g., [28]), the Jackson and Bernstein estimates (4.1)–(4.2) imply the following direct and inverse estimates: for $f \in L_p(\Omega)$ one has

$$\sigma_n(f)_p \le cK(f, n^{-\alpha})_p \tag{4.3}$$

and

$$K(f, n^{-\alpha})_p \le cn^{-\alpha}\left(\left[\sum_{\nu=1}^{n} \frac{1}{\nu}(\nu^\alpha \sigma_\nu(f)_p)^{\tau^*}\right]^{1/\tau^*} + \|f\|_p\right), \tag{4.4}$$

where $\tau^* := \min\{\tau, 1\}$.

We define the approximation space $A_q^\gamma = A_q^\gamma(\Phi, L_p)$ to be the set of all functions $f \in L_p(\Omega)$ such that

$$\|f\|_{A_q^\gamma} := \|f\|_p + \left(\sum_{n=1}^{\infty}(n^\gamma \sigma_n(f)_p)^q \frac{1}{n}\right)^{1/q} < \infty \tag{4.5}$$

with the usual modification when $q = \infty$.

The following characterization of the approximation spaces A_q^γ in terms of the above defined interpolation spaces is immediate from estimates (4.3)–(4.4).

Theorem 4.3. *If* $0 < \gamma < \alpha$ *and* $0 < q \le \infty$, *then*

$$A_q^\gamma(\Phi, L_p) = (L_p(\Omega), \mathcal{B}_\tau^\alpha(\mathcal{M}))_{\gamma/\alpha, q}$$

with equivalent norms.

In one specific case the approximation space $A_q^\alpha(L_p)$ can be identified as a B-space.

Theorem 4.4. *Assuming that* $\alpha > 0$ *if* $p < \infty$ *and* $\alpha > 1$ *if* $p = \infty$, *and* $1/\tau := \alpha + 1/p$ *in both cases*, *we have*

$$A_\tau^\alpha(\Phi, L_p) = \mathcal{B}_\tau^\alpha(\mathcal{M}) \tag{4.6}$$

with equivalent norms.

The proof of this theorem is a mere repetition of the proof of Theorem 3.4 in [16] and will be omitted. We next turn to a constructive realization of best n-term approximation.

5. "Push-the-Error" Algorithm

5.1. *Description of the Algorithm*

For a given function $f \in C(\Omega)$, we use the decomposition scheme from (2.33) to represent f in the form

$$f = \sum_{\theta \in \Theta} b_\theta(f)\varphi_\theta = \sum_{m \in \mathbf{N}_0} \sum_{\theta \in \Theta_m} b_\theta(f)\varphi_\theta, \tag{5.1}$$

where the coefficients $b_\theta := b_\theta(f)$ depend linearly on f and the series converges uniformly on Ω. As elsewhere in this paper, the basis functions φ_θ are normalized in L_∞, i.e., $\|\varphi_\theta\|_\infty = 1$. Whenever f has a wavelet expansion (see (2.16)), we rewrite the wavelets in terms of scaling functions to obtain (5.1). We shall drop the reference to f at times when this is clear from the context.

For the purpose of designing an algorithm capable of achieving the rates of the best n-term approximation from $\{\varphi_\theta\}$ in the uniform norm, the initial decomposition (5.1) should provide an efficient representation of f. In our case, this means that the terms in (5.1) should characterize the norm in $\mathcal{B}^\alpha_\tau(\mathcal{M})$, $\alpha \geq 1$, $\tau := 1/\alpha$, as in (3.7)–(3.8), which we achieve by employing simple projectors onto the spaces (V_m) (see Section 2.4 and Section 3).

To describe the “Push-the-Error” algorithm we need a few preliminaries that help us to develop substitutes for simple thresholding concepts that would work in L_p, $p < \infty$.

For any $\eta, \theta \in \Theta$ with $l(\eta) > l(\theta)$, we say that η is *connected* with θ via sets from Θ if there exists a sequence of elements $\eta =: \eta_0, \eta_1, \dots, \eta_k := \theta$ with $k := l(\eta) - l(\theta)$ such that

(i) $l(\eta_i) = l(\eta_{i+1}) + 1$, $i = 0, \dots, k-1$;
(ii) η_i overlaps η_{i+1}, $i = 0, \dots, k-1$, i.e., $\eta_i^\circ \cap \eta_{i+1}^\circ \neq \emptyset$.

The notion of being *connected* is closely related to the notion of being *connected by n-stars*, introduced in Section 2.3. The relevant Lemma 2.2 will play a vital role in the following.

Given $\theta \in \Theta$, we define

$$\mathcal{U}'_\theta := \{\eta \in \Theta : l(\eta) > l(\theta),\ \eta \text{ is connected with } \theta\} \tag{5.2}$$

and

$$\mathcal{U}_\theta := \mathcal{U}'_\theta \cup \{\theta\}. \tag{5.3}$$

Note that $\eta \in \mathcal{U}_\theta$ implies that $\mathcal{U}_\eta \subseteq \mathcal{U}_\theta$ and, hence, by Lemma 2.2 and (2.17),

$$\eta \in \mathcal{U}_\theta \quad \Rightarrow \quad \eta \subset \mathrm{Star}^{(m)}_{N_*}(\theta), \qquad N_* := 2\tilde{\nu}\nu_*, \quad m := l(\theta). \tag{5.4}$$

In order to compress the representation (5.1), it would not be reasonable to threshold the coefficients $b_\theta(f)$, due to the lack of stability across levels. Therefore we need more subtle indicators and introduce local error terms by

$$E(f,\theta) = E(\theta) := |b_\theta(f)| + \Bigg\| \sum_{\eta \in \mathcal{U}'_\theta} b_\eta(f)\varphi_\eta \Bigg\|_\infty . \tag{5.5}$$

Remark 5.1. Since by (2.9) and (2.31), one has, for $\theta \in \Theta_m$,

$$|b_\theta(f)| \leq \beta_2 \|(Q_m - Q_{m-1})f\|_\infty \leq \beta_2(\|Q_m(f) - f\|_\infty + \|f - Q_{m-1}(f)\|_\infty), \tag{5.6}$$

and

$$\Bigg\| \sum_{\eta \in \mathcal{U}'_\theta} b_\eta(f)\varphi_\eta \Bigg\|_\infty \leq \|Q_m(f) - f\|_\infty, \tag{5.7}$$

the assumed uniform convergence of (5.1) and (2.30) ensure that, for each $f \in C(\Omega)$ and every $\varepsilon > 0$, there exists an $M \in \mathbf{N}$ such that $E(f, \theta) < \varepsilon$ for $\theta \in \Theta_m$, $m > M$.

For each $\theta \in \Theta$, we define its *"concrete"* Ω_θ by

(5.8) $$\Omega_\theta := \mathrm{Star}^{(m)}_{N_*+4\nu_*}(\theta), \qquad m := l(\theta),$$

where ν_* is from (2.17) and N_* is from (5.4).

Also, for a given $\theta \in \Theta$, we define

(5.9) $$\mathcal{X}_\theta := \{\eta \in \Theta_m : \eta^\circ \cap \Omega_\theta^\circ \neq \emptyset\} \qquad \text{with} \quad m := l(\theta).$$

We shall call the elements of $\mathcal{X}_\theta$ the *neighbors* of θ. By (2.17) and (5.8),

(5.10) $$\eta \in \mathcal{X}_\theta \quad \Rightarrow \quad \eta \subset \mathrm{Star}^{(m)}_{N_*+5\nu_*}(\theta).$$

We are now prepared to describe the "Push-the-Error" algorithm which, with a slight abuse of terminology, will play two different roles. On one hand, it will be used as a theoretical tool that offers a constructive way of identifying n-term approximations realizing optimal rates. In this role it will be applied to an arbitrary *infinite* expansion of the form (5.1), although the error terms $E(f, \theta)$ would then not be practically accessible. In a practical context the scheme should be thought of as applied to some initial approximation consisting of a *finite* expansion of the form (5.1). We shall briefly discuss corresponding practical ramifications later, and work here first with the conceptual version of the first form.

PTE$[\varepsilon, f] \to \mathcal{A}_\varepsilon(f)$ produces, for a given function $f \in C(\Omega)$ and any target accuracy $\varepsilon > 0$, an approximation

$$\mathcal{A}_\varepsilon(f) = \mathcal{A}_\varepsilon(f) = \sum_{\theta \in \Lambda(f,\varepsilon)} d_\theta(f)\varphi_\theta$$

by the following steps:

Step 1 (*Decomposition*). We represent f in the form (5.1) (see also (2.33)).

Step 2 ("*Prune the Shrubs*"). We discard all terms $b_\theta\varphi_\theta$ such that

(5.11) $$E(f, \eta) \le \varepsilon \qquad \forall\, \eta \in \mathcal{U}_\theta.$$

We denote by $\Gamma = \Gamma(f, \varepsilon)$ the set of all elements of Θ which have not been discarded and write

(5.12) $$f_\Gamma := \sum_{\theta \in \Gamma} b_\theta \varphi_\theta.$$

From Remark 5.1 we know that there exists some $M \in \mathbf{N}$ such that

(5.13) $$E(f, \theta) < \varepsilon \qquad \forall\, \theta \in \Theta_m, \quad m > M,$$

i.e., Γ is a finite set.

Step 3 ("*Push the Error*"). This step is a variation of Step 3 of the "Push-the-Error" algorithm described in [24].

Let $\tilde{\Lambda}_0$ be the set of all $\theta \in \Theta_0 \cap \Gamma$ such that $|b_\theta(f)| > \varepsilon$ and set $\Lambda_0 := (\bigcup_{\theta \in \tilde{\Lambda}_0} \mathcal{X}_\theta) \cap \Gamma$. We define

$$\mathcal{A}_0 := \sum_{\theta \in \Lambda_0} b_\theta \varphi_\theta.$$

Using the refinement equations (2.13), we represent (rewrite) each of the remaining terms $b_\theta \varphi_\theta$, $\theta \in (\Theta_0 \cap \Gamma) \backslash \Lambda_0$, as a linear combination of $\{\varphi_\eta\}_{\eta \in \Theta_1}$ and add to the resulting terms the existing terms $b_\theta \varphi_\theta$, $\theta \in \Theta_1 \cap \Gamma$. As a result we obtain a representation of f_Γ in the form

$$f_\Gamma = \mathcal{A}_0 + \sum_{\theta \in \Theta_1 \backslash \Gamma} d_\theta \varphi_\theta + \sum_{\theta \in \Theta_1 \cap \Gamma} d_\theta \varphi_\theta + \sum_{m=2}^{M} \sum_{\theta \in \Theta_m \cap \Gamma} b_\theta \varphi_\theta.$$

Further, we define $\tilde{\Lambda}_1$ as the set of all $\theta \in \Theta_1 \cap \Gamma$ such that $|d_\theta| > \varepsilon$ and set $\Lambda_1 := (\bigcup_{\theta \in \tilde{\Lambda}_1} \mathcal{X}_\theta) \cap \Gamma$. Then we define

$$\mathcal{A}_1 := \sum_{\theta \in \Lambda_1} d_\theta \varphi_\theta.$$

Similarly as above, we rewrite all remaining terms $d_\theta \varphi_\theta$, $\theta \in (\Theta_1 \cap \Gamma) \backslash \Lambda_1$, at the next level and add to them the existing terms $b_\theta \varphi_\theta$, $\theta \in \Theta_2 \cap \Gamma$. We obtain

$$f_\Gamma = \mathcal{A}_0 + \mathcal{A}_1 + \sum_{\theta \in \Theta_1 \backslash \Gamma} d_\theta \varphi_\theta + \sum_{\theta \in \Theta_2 \backslash \Gamma} d_\theta \varphi_\theta + \sum_{\theta \in \Theta_2 \cap \Gamma} d_\theta \varphi_\theta + \sum_{m=3}^{M} \sum_{\theta \in \Theta_m \cap \Gamma} b_\theta \varphi_\theta.$$

We process in the same way all other levels until we reach the finest level Θ_M. We define $\tilde{\Lambda}_M$, Λ_M, and $\mathcal{A}_M$ as above.

We obtain as an output the set $\tilde{\Lambda}(f, \varepsilon) := \bigcup_{m=0}^{M} \tilde{\Lambda}_m$ of the ε-significant indices (with $|d_\theta(f)| > \varepsilon$), the set $\Lambda(f, \varepsilon) := \bigcup_{m=0}^{M} \Lambda_m$ containing also the neighbors of the elements in $\tilde{\Lambda}(f, \varepsilon)$ identified by the concrete Ω_θ, and the approximation

$$\mathcal{A}_\varepsilon = \mathcal{A}_\varepsilon(f) := \sum_{m=0}^{M} \mathcal{A}_m = \sum_{\theta \in \Lambda(f, \varepsilon)} d_\theta \varphi_\theta.$$

Lemma 5.2. *We have*

$$\|f - f_\Gamma\|_\infty \le \nu_2 \varepsilon \tag{5.14}$$

with ν_2 the constant from property (β) of the elements of Θ, Section 2.1.

Proof. To see this, let $x \in \Omega$ and set $\mathcal{C}(x, \Gamma) := \{\theta \notin \Gamma : x \in \theta^\circ, l(\theta) \text{ is minimal}\}$. Notice that the θ's in $\mathcal{C}(x, \Gamma)$ are from one and the same level in Θ. If $\mathcal{C}(x, \Gamma) = \emptyset$, then $f_\Gamma(x) = f(x)$. Suppose $\mathcal{C}(x, \Gamma) \neq \emptyset$. By property (β), $\#\mathcal{C}(x, \Gamma) \le \nu_2$. Then, for any $\theta' \in \mathcal{C}(x, \Gamma)$,

$$|f(x) - f_\Gamma(x)| \le \sum_{\theta \in \mathcal{C}(x,\Gamma)} |b_\theta| + \left\| \sum_{\eta \in \mathcal{U}'_{\theta'}} b_\eta \varphi_\eta \right\|_\infty \le \sum_{\theta \in \mathcal{C}(x,\Gamma)} E(\theta) \le \nu_2 \varepsilon,$$

which confirms the claim. ∎

Lemma 5.3. *We have*

$$\|f_\Gamma - \mathcal{A}_\varepsilon(f)\|_\infty \le c\varepsilon \tag{5.15}$$

with $c = 2\nu_2^2\beta_1$, *where* β_1 *is a bound of the coefficients from* (2.13).

Proof. Fix $x \in \Omega$ and let $\mathcal{C}(x,\Gamma) := \{\theta \notin \Gamma : x \in \theta^\circ, l(\theta) \text{ is minimal}\}$ as in the proof of the previous lemma.

Suppose first that $\mathcal{C}(x,\Gamma) \ne \emptyset$. Let $\theta' \in \mathcal{C}(x,\Gamma)$ and set $m := l(\theta')$. Since $x \in \theta'$ and $\theta' \notin \Gamma$, then $\mathcal{U}'_{\theta'} \cap \Gamma = \emptyset$ and, therefore, there is no contribution to f_Γ at x from levels $> m$. Then

$$\begin{aligned} f_\Gamma(x) &= \mathcal{A}_\varepsilon(f)(x) + \sum_{\theta\in\mathcal{C}(x,\Gamma)} d_\theta\varphi_\theta(x) + \sum_{\theta\in\Theta_m\cap\Gamma: x\in\theta} r_\theta\varphi_\theta(x) \\ &=: \mathcal{A}_\varepsilon(f)(x) + F_1(x) + F_2(x). \end{aligned} \tag{5.16}$$

Here the terms $d_\theta\varphi_\theta$, $\theta \in \mathcal{C}(x,\Gamma)$, are obtained from the rewriting of some terms $d_\eta\varphi_\eta$, $\eta \in \Theta_{m-1}$ so that $x \in \eta$ and $|d_\eta| \le \varepsilon$. Denote by $\mathcal{K}(x,m)$ the set of their indices. By (2.13), $\varphi_\eta = \sum_{\theta\in\Theta_m,\,\theta\subset\eta} a_{\eta,\theta}\varphi_\theta$ with $|a_{\eta,\theta}| \le \beta_1$, and hence

$$\sum_{\eta\in\mathcal{K}(x,m)} d_\eta\varphi_\eta = \sum_{\theta\in\mathcal{C}(x,\Gamma)} \left(\sum_{\eta\in\mathcal{K}(x,m)} a_{\eta,\theta}d_\eta\right)\varphi_\theta,$$

which yields

$$d_\theta = \sum_{\eta\in\mathcal{K}(x,m)} a_{\eta,\theta}d_\eta.$$

Therefore,

$$|d_\theta| \le \sum_{\eta\in\mathcal{K}(x,m)} |a_{\eta,\theta}||d_\eta| \le \nu_2\beta_1\varepsilon, \qquad \theta \in \mathcal{C}(x,\Gamma),$$

and, hence,

$$|F_1(x)| \le \sum_{\theta\in\mathcal{C}(x,\Gamma)} |d_\theta| \le \nu_2^2\beta_1\varepsilon, \tag{5.17}$$

where ν_2 is from property (β), Section 2.1.

The terms $r_\theta\varphi_\theta$ (if any) in the second sum in (5.16) have indices $\theta \in \Theta_m \cap \Gamma$ such that $x \in \theta$ and $|r_\theta| \le \varepsilon$ since they have not been selected in $\mathcal{A}_m$. Therefore,

$$|F_2(x)| \le \sum_{\theta\in\Theta_m\cap\Gamma: x\in\theta} |r_\theta| \le \nu_2\varepsilon.$$

Combining this with (5.17) yields (5.15).

It remains to consider the case when $\mathcal{C}(x,\Gamma) = \emptyset$. Now, we have

$$f_\Gamma(x) = \mathcal{A}_\varepsilon(f)(x) + \sum_{\theta\in\Theta_M\cap\Gamma: x\in\theta} d_\theta\varphi_\theta(x),$$

where $d_\theta\varphi_\theta$ are terms which have not been selected in the approximant. Therefore, $|d_\theta| \le \varepsilon$ and (5.15) follows as above. ■

Remark 5.4. Combining the estimates from Lemmas 5.2 and 5.3, we obtain the following error bound for the "Push-the-Error" algorithm with target accuracy $\varepsilon > 0$:

$$\|f - \mathcal{A}_\varepsilon(f)\|_\infty \le \hat{c}\varepsilon \tag{5.18}$$

with $\hat{c} < 3\nu_2^2\beta_1$.

5.2. *Error Analysis of "Push-the-Error"*

Assuming that "Push-the-Error" is applied to a function $f \in C(\Omega)$ with $\varepsilon > 0$ and $\mathcal{A}_\varepsilon(f)$ is the approximant obtained, we denote

$$N(\varepsilon) = N_f(\varepsilon) := \#\Lambda(f, \varepsilon), \qquad A_{N(\varepsilon)}(f) := \|f - \mathcal{A}_\varepsilon(f)\|_\infty,$$

and

$$A_n(f) := \inf_{\varepsilon>0}\{A_{N(\varepsilon)}(f) : N(\varepsilon) \le n\}.$$

The main conceptual tool is the following weak quasi-sub-additivity of the counting functional $N(\varepsilon)$. We shall point out later in which sense this may be regarded as a weak stability property.

Theorem 5.5. *There exist constants c_* and $\tilde{c}$ depending only on the parameters of the MRA such that if $f = f_0 + f_1$, $f_j \in C(\Omega)$, and the "Push-the-Error" algorithm is applied to f_j with $\varepsilon_j > 0$ $(j = 0, 1)$ and to f with $\varepsilon := c_*(\varepsilon_0 + \varepsilon_1)$, then*

$$N_f(\varepsilon) \le \tilde{c}(N_{f_0}(\varepsilon_0) + N_{f_1}(\varepsilon_1)). \tag{5.19}$$

The proof of this theorem is rather involved and will be postponed to Section 7.

We shall now make precise in which sense the "Push-the-Error" scheme gives rise to an optimal approximation scheme.

Theorem 5.6. *If $f \in \mathcal{B}_\tau^\alpha(\mathcal{M})$, $\alpha \ge 1$, $\tau := 1/\alpha$, then, for each $\varepsilon > 0$,*

$$A_{N(\varepsilon)}(f) \le c\varepsilon \qquad \textit{and} \qquad N(\varepsilon) \le c\varepsilon^{-\tau}\|f\|^\tau_{\mathcal{B}_\tau^\alpha(\mathcal{M})} \tag{5.20}$$

and, therefore,

$$A_n(f) \le cn^{-\alpha}\|f\|_{\mathcal{B}_\tau^\alpha(\mathcal{M})}, \qquad n = 1, 2, \ldots. \tag{5.21}$$

Moreover, for $f \in C(\Omega)$,

$$A_{N(\varepsilon)}(f)_\infty \le c\min\{\varepsilon, \|f\|_\infty\}. \tag{5.22}$$

Here the constants depend only on α and the parameters of the MRA.

The proof of this theorem is closely related to that of the previous theorem and will also be deferred to Section 7.

We can now address the program outlined in Section 4. Let us denote by $K(f, t)_\infty$ the K-functional generated by the spaces $C(\Omega)$ and $\mathcal{B}_\tau^\alpha(\mathcal{M})$ with $\tau := 1/\alpha$.

Theorem 5.7. *Suppose that $f \in C(\Omega)$ and $\alpha \geq 1$. Then one has*

$$A_n(f)_\infty \leq cK(f, n^{-\alpha})_\infty \tag{5.23}$$

and, therefore,

$$\sigma_n(f)_\infty \leq A_n(f)_\infty \leq cn^{-\alpha}\left(\left[\sum_{\nu=1}^{n} \frac{1}{\nu}(\nu^\alpha \sigma_\nu(f)_\infty)^\tau\right]^{1/\tau} + \|f\|_\infty\right), \tag{5.24}$$

where c depends on α, and the parameters of the MRA.

Proof. We need only prove (5.23), since (5.24) follows by (5.23) and (4.4). Suppose $g \in \mathcal{B}^\alpha_\tau(\mathcal{M})$ with $\|g\|_{\mathcal{B}^\alpha_\tau} \neq 0$ and $\|f - g\|_\infty \neq 0$. Choose $\varepsilon_0 := 4\beta_2 c^\flat \|f - g\|_\infty$, where β_2 and $c^\flat$ are the constants from (2.9) and (2.29). Further, choose $\varepsilon_1 := n^{-\alpha}\|g\|_{\mathcal{B}^\alpha_\tau}$. Let $\varepsilon := c_*(\varepsilon_0 + \varepsilon_1)$, where c_* is the constant from Theorem 5.5. By (5.18) and Theorem 5.5, applied with $f_0 := f - g$, $f_1 := g$, we have

$$A_{N_f(\varepsilon)}(f) \leq c(\varepsilon_0 + \varepsilon_1) \leq c(\|f - g\|_\infty + n^{-\alpha}\|g\|_{\mathcal{B}^\alpha_\tau}) \tag{5.25}$$

and

$$N_f(\varepsilon) \leq c(N_{f_0}(\varepsilon_0) + N_{f_1}(\varepsilon_1)),$$

where c depends here on the constant $\hat{c}$ in (5.18) and on the constants $c_*, \tilde{c}$ in Theorem 5.5.

We next show that $N_f(\varepsilon) \leq cn$. Similarly, as in Remark 5.1, using (2.9) and (2.29), we have, for $\theta \in \Theta_m$,

$$|b_\theta(f_0)| \leq \beta_2(\|Q_m(f_0)\|_\infty + \|Q_{m-1}(f_0)\|_\infty) \leq 2\beta_2 c^\flat \|f_0\|_\infty$$

and

$$\left\|\sum_{\eta \in \mathcal{U}'_\theta} b_\eta(f_0)\varphi_\eta\right\|_\infty \leq \|f_0 - Q_m(f_0)\|_\infty \leq 2c^\flat \|f_0\|_\infty$$

and, hence,

$$E(f_0, \theta) \leq 4\beta_2 c^\flat \|f_0\|_\infty.$$

Now, since $\varepsilon_0 := 4\beta_2 c^\flat \|f_0\|_\infty$, then $\mathcal{A}_{\varepsilon_0}(f_0) = 0$ and $N_{f_0}(\varepsilon_0) = 0$, due to Step 2 of the algorithm. On the other hand, by Theorem 5.6, $N_{f_1}(\varepsilon) \leq c\varepsilon_1^{-\tau}\|g\|^\tau_{\mathcal{B}^\alpha_\tau} \leq cn$, where we have expressed ε_1 in terms of n according to the above choice, and hence $N_f(\varepsilon) \leq cn$.

Since g was selected arbitrarily in $\mathcal{B}^\alpha_\tau(\mathcal{M})$, (5.25) yields $A_{cn}(f) \leq cK(f, n^{-\alpha})_\infty$, which implies (5.23) (with a different constant c). ■

The following result is an immediate consequence of Theorem 5.7.

Theorem 5.8. *For $f \in C(\Omega)$ and $\gamma > 0$, $A_n(f) = O(n^{-\gamma})$ if and only if $\sigma_n(f)_\infty = O(n^{-\gamma})$.*

More generally, let $A_q^\gamma(\sigma, L_\infty)$ be the approximation spaces generated by the nonlinear n-term approximation from the scaling functions of the MRA, defined in (4.5). Let $A_q^\gamma(\mathcal{A})$ be the set of all functions $f \in C(\Omega)$ such that

$$\|f\|_{A_q^\gamma(\mathcal{A})} := \|f\|_\infty + \left(\sum_{n=1}^{\infty}(n^\gamma A_n(f))^q \frac{1}{n}\right)^{1/q} < \infty \tag{5.26}$$

with the usual modification when $q = \infty$.

Theorem 5.7 yields the following more general result.

Theorem 5.9. *For any $\gamma > 0$ and $0 < q \le \infty$, we have $A_q^\gamma(\mathcal{A}) = A_q^\gamma(\sigma, L_\infty)$ and $\|f\|_{A_q^\gamma(\mathcal{A})} \approx \|f\|_{A_q^\gamma(\sigma, L_\infty)}$ for $f \in A_q^\gamma(\mathcal{A}) = A_q^\gamma(\sigma, L_\infty)$.*

5.3. *Practical Aspects of* "*Push-the-Error*"

From a practical perspective the "Push-the-Error" algorithm can be applied only to finite expansions (5.1) since otherwise Step 2 is not feasible. Therefore it can be viewed as a *coarsening* procedure that turns some initial (nonoptimal) approximation into a (nearly) optimal one. To make this more precise, suppose that f belongs to some space V_M so that the decomposition Step 1 of the scheme yields a representation

$$f = \sum_{\theta\in\Theta'} b_\theta(f)\varphi_\theta,$$

where $\Theta' \subseteq \bigcup_{m\le M}\Theta_m$ and thus $N := \#\Theta' \le \dim V_M < \infty$. Suppose, furthermore, that f is an approximation to the (ideal) function $f^* \in C(\Omega)$ and that

$$\|f - f^*\|_\infty \le \varepsilon. \tag{5.27}$$

From the proof of Theorem 5.7 we infer that there exist constants $c_1, c_2 \ge 1$ such that the (theoretical version of the) "Push-the-Error" scheme yields that, for every $n \in \mathbf{N}$ there exists $\varepsilon^* > 0$ such that

$$A_{N(\varepsilon^*)}(f^*) \le c_2 K(f^*, n^{-\alpha})_\infty, \qquad N(\varepsilon^*) \le n, \quad \varepsilon^* \le c_1 K(f^*, n^{-\alpha})_\infty. \tag{5.28}$$

Now let n be the smallest positive integer for which $c_2K(f^*, n^{-\alpha})_\infty \le \varepsilon$. One easily confirms that then

$$\varepsilon \le c_2 2^\alpha K(f^*, n^{-\alpha})_\infty \le 2^\alpha\varepsilon. \tag{5.29}$$

Setting $f_n^* := \mathcal{A}_{\varepsilon^*}(f^*)$, one therefore clearly has $\|f - f_n^*\|_\infty \le 2\varepsilon$. Now we write $f = (f - f_n^*) + f_n^*$ and set

$$f_0 := f - f_n^*, \qquad f_1 := f_n^*, \quad \varepsilon_0 := 8\beta_2 c^\flat \varepsilon.$$

Next note that

$$K(f_1, n^{-\alpha})_\infty \le \|f_1 - f^*\|_\infty + K(f^*, n^{-\alpha})_\infty \le (1 + c_2)K(f^*, n^{-\alpha})_\infty.$$

Hence, by the same reasoning as above, there exists $\varepsilon^{**} > 0$ such that $\varepsilon^{**} \le c_1 K(f_1, n^{-\alpha})_\infty \le c_1(1 + c_2)\varepsilon$,

$$A_{N_{f_1}(\varepsilon^{**})}(f_1) \le c_2 K(f_1, n^{-\alpha})_\infty \le c_1 c_2 (1 + c_2)\varepsilon, \qquad N_{f_1}(\varepsilon^{**}) \le n. \tag{5.30}$$

Choose $\varepsilon_1 = \varepsilon^{**}$. We now apply Theorem 5.5 with the above selection of f_0, f_1, ε_0, and ε_1 to conclude that

$$N_f(c_*(\varepsilon_0 + \varepsilon_1)) \le \tilde{c}(N_{f_0}(\varepsilon_0) + N_{f_1}(\varepsilon_1))$$

and, using (5.29),

$$\|f - \mathcal{A}_{c^*(\varepsilon_0+\varepsilon_1)}(f)\|_\infty \le c(\varepsilon_0 + \varepsilon_1) \le c'\varepsilon \le cK(f^*, n^{-\alpha})_\infty.$$

But as in the proof of Theorem 5.7 one confirms that, by (5.28), Step 2 of the algorithm returns $\Gamma_{f_0} = \emptyset$ and hence $\mathcal{A}_{\varepsilon_0}(f_0) = N_{f_0}(\varepsilon_0) = 0$. Therefore, using (5.30),

$$N_f(c_*(\varepsilon_0 + \varepsilon_1)) \le \tilde{c} N_{f_1}(\varepsilon^{**}) \le \tilde{c} n.$$

Consequently,

$$\|\mathcal{A}_{c^*(\varepsilon_0+\varepsilon_1)}(f) - f^*\|_\infty \le cK(f^*, n^{-\alpha})_\infty, \qquad N_f(c^*(\varepsilon_0 + \varepsilon_1)) \le cn,$$

where $K(f^*, n^{-\alpha})_\infty \approx \varepsilon$. Thus a proper coarsening of f, obtained through the (practical version of the) "Push-the-Error" scheme, yields a near optimal approximation to the ideal f^* whenever an initial error bound (5.27) is given. Such situations arise in the context of adaptive schemes. One also derives from the above considerations that, when $f^* \in \mathcal{B}^\alpha_\tau(\mathcal{M})$, one has

$$\|\mathcal{A}_{c^*(\varepsilon_0+\varepsilon_1)}(f) - f^*\|_\infty \le c\varepsilon, \qquad N_f(c^*(\varepsilon_0 + \varepsilon_1)) \le c\varepsilon^{-\tau} \|f^*\|^\tau_{\mathcal{B}^\alpha_\tau(\mathcal{M})},$$

which explains in which sense the scheme deserves to be termed stable in L_∞.

Complexity. Assume now that the function f (a surface or multidimensional data) has an initial representation (approximation) in some "finest" space V_M of an MRA involving $O(N)$ terms. Let us assume that the "Push-the-Error" algorithm (as described in Section 5.1) is applied to this f. The decomposition Step 1 of "Push-the-Error" will run in $O(N)$ flops. Step 2 ("Prune the Shrubs") of the algorithm can evidently be realized in $O(N \log N)$ flops by rewriting all terms of interest at the finest level. Step 3 ("Push the Error") works in $O(N)$ flops. Finally, the reconstruction step also runs in $O(N)$ flops. Therefore, the "Push-the-Error" algorithm appears to be an attractive coarsening scheme from a practical point of view. Our next goal is to propose an even more economical version of the second step of the algorithm.

Scalable Second Version of Step 2 ("*Prune the Shrubs*"). We define a new local error term $\widetilde{E}(f, \theta)$ by

$$\widetilde{E}(f, \theta) := |b_\theta(f)| + \max_{v \in \theta} \sum_{\eta \in \mathcal{U}'_\theta : v \in \eta} |b_\eta(f)|. \tag{5.31}$$

Now, the condition $E(f, \eta) \le \varepsilon$ in (5.11) is replaced by the condition $\widetilde{E}(f, \eta) \le \varepsilon$ (see (5.5)) which is practically easier to be verified. The new version of Step 2 of the algorithm can be realized in $O(N)$ flops by employing the well-known principle of Dynamic Programming. We use the coefficient $\{b_\theta(f)\}$ obtained in Step 1 to compute

$$M(f, \theta) := \max_{v \in \theta} \sum_{\eta \in \mathcal{U}'_\theta : v \in \eta} |b_\eta(f)| \qquad \text{for every} \quad \theta \in \Theta.$$

To this end, we proceed from finer to coarser levels and compute each $M(f, \theta)$ by using the outcome of the previous steps.

It is easy to see that for this new version of "Push-the-Error," Theorem 5.6 remains valid with a slight modification of the same proof. However, it is impossible for us to establish Theorem 5.5 in this case, which makes this version less attractive from a theoretical point of view. In particular, we fail to have estimates like (5.23).

Further Observations and Practical Modifications. As already mentioned at the beginning of Section 5.1, for an optimal performance of the "Push-the-Error" algorithm it is important to have an initial sparse representation of the function f being approximated. To this end, the dual functionals $\{c_\theta(\cdot)\}$ should be bounded in L_q for some $q < \infty$ (see Remark 3.1). In turn, this means that decomposition methods based on interpolatory schemes do not provide efficient representations and should be avoided.

In the description of Step 3 of "Push-the-Error," the *neighbors* of a given $\theta' \in \Theta$ are described as all θ's from the same level which overlap with the concrete $\Omega_{\theta'}$ of θ'; all terms $\{d_\theta \varphi_\theta\}$ with such indices are taken in the approximation whenever $|d_{\theta'}| > \varepsilon$. For practical implementations much smaller concretes should be used and one can even consider realizations where the neighbors are not included at all.

Finally, one can run the "Push-the-Error" algorithm without executing Step 2 at all. An algorithm consisting of only Steps 1 and 2 is also reasonable in some situations. Other modifications are also possible. However, one should be aware of the existence of several traps which may defeat such modifications of the algorithms (see [24]).

6. "Threshold" Algorithm in L_p ($p < \infty$)

Here we show that the usual threshold scheme used in nonlinear n-term approximation from wavelets in L_p ($1 < p < \infty$) can be successfully utilized for n-term approximation from the scaling functions of MRA in L_p ($0 < p < \infty$) (see also [24]).

We begin with a description of the algorithm.

Step 1 (*Decomposition*). We represent the function f being approximated by using the decomposition (2.33) if $1 < p \le \infty$ and (2.34) with $0 < q < p$ if $0 < p \le 1$. So, in both cases,

$$f = \sum_{\theta \in \Theta} b_\theta(f) \varphi_\theta \quad \text{in } L_p(\Omega). \tag{6.1}$$

Step 2 ("*Threshold*"). We first order the terms $\{b_\theta \varphi_\theta\}_{\theta \in \Theta}$ in a sequence $(b_{\theta_j} \varphi_{\theta_j})_{j \in \mathbf{N}}$ so that

$$\|b_{\theta_1} \varphi_{\theta_1}\|_p \ge \|b_{\theta_2} \varphi_{\theta_2}\|_p \ge \cdots.$$

Then we define the approximant by $\mathcal{A}_n(f)_p := \sum_{j=1}^n b_{\theta_j} \varphi_{\theta_j}$.

We now turn to the error analysis of the "Threshold" algorithm. We define the error of the algorithm by

$$A_n^T(f)_p := \|f - \mathcal{A}_n(f)_p\|_{L_p(\Omega)}.$$

As elsewhere we assume that $\alpha > 0$, $0 < p < \infty$, and $\tau := (\alpha + 1/p)^{-1}$.

Theorem 6.1. *If $f \in \mathcal{B}_\tau^\alpha(\mathcal{M})$, then*

$$A_n^T(f)_p \le cn^{-\alpha}\|f\|_{\mathcal{B}_\tau^\alpha(\mathcal{M})}. \tag{6.2}$$

Furthermore,

$$A_{2n}^T(f)_p \le cn^{-\alpha}\left(\sum_{j=n+1}^{\infty}\|b_{\theta_j}\varphi_{\theta_j}\|_p^\tau\right)^{1/\tau}. \tag{6.3}$$

Here c depends only on α, p, and the parameters of the MRA.

Proof. Estimate (6.2) follows immediately by the general direct estimate of Theorem 3.4 in [23] and the equivalence $\|f\|^Q_{\mathcal{B}_\tau^\alpha(\mathcal{M})} \approx \|f\|_{\mathcal{B}_\tau^\alpha(\mathcal{M})}$ established in Theorem 3.4. To prove (6.3) we again apply Theorem 3.4 from [23] but this time to the sequence $(b_{\theta_j}\varphi_{\theta_j})_{j=n+1}^\infty$. ∎

We next show that in a sense the "Threshold" algorithm captures the rates of the best nonlinear n-term approximation in L_p, $0 < p < \infty$. For this denote by $A_\tau^\alpha(\sigma, L_p) := A_\tau^\alpha(\Phi, L_p)$ the approximation space defined in (4.5) and by $A_q^\gamma(\mathcal{A}^T, L_p)$ the set of all functions $f \in L_p(\Omega)$ such that

$$\|f\|_{A_q^\gamma(\mathcal{A}^T,\, L_p)} := \|f\|_p + \left(\sum_{n=1}^{\infty}(n^\gamma A_n^T(f)_p)^q\frac{1}{n}\right)^{1/q} < \infty \tag{6.4}$$

with the usual modification when $q = \infty$ (see also (5.26)).

Theorem 6.2. *For any $\alpha > 0$ and $1/\tau = \alpha + 1/p$, we have $A_\tau^\alpha(\mathcal{A}^T,\ L_p) = \mathcal{B}_\tau^\alpha(\mathcal{M}) = A_\tau^\alpha(\sigma,\ L_p)$ and for each f in this space*

$$\|f\|_{A_\tau^\alpha(\mathcal{A}^T, L_p)} \approx \|f\|_{\mathcal{B}_\tau^\alpha(\mathcal{M})} \approx \|f\|_{A_\tau^\alpha(\sigma, L_p)}. \tag{6.5}$$

Proof. The right-hand side equivalence in (6.5) is the statement of Theorem 4.4 when $p < \infty$. Clearly, to complete the proof we need only show that

$$\mathbf{A} := \left(\sum_{\nu=0}^{\infty}[2^{\nu\alpha}A_{2^\nu}^T(f)_p]^\tau\right)^{1/\tau} \le c\|f\|_{\mathcal{B}_\tau^\alpha(\mathcal{M})}. \tag{6.6}$$

Choose α_1 so that $0 < \alpha_1 < \alpha$ and set $\tau_1 := (\alpha_1 + 1/p)^{-1}$. By (6.3) applied with α replaced by α_1, it follows that

$$A_{2^{\nu+1}}^T(f)_p \le c2^{-\nu\alpha_1}\left(\sum_{k=\nu}^{\infty}2^k\|b_{\theta_{2^k}}\varphi_{\theta_{2^k}}\|_p^{\tau_1}\right)^{1/\tau_1}. \tag{6.7}$$

Denote briefly $\beta_k := 2^{k/\tau_1}\|b_{\theta_{2^k}}\varphi_{\theta_{2^k}}\|_p$. Then by (6.7) for $\nu \ge 0$ and (6.2) with $n = 1$, we obtain

$$\mathbf{A}^\tau \le c\sum_{\nu=0}^{\infty}\left[2^{\nu(\alpha-\alpha_1)}\left(\sum_{k=\nu}^{\infty}\beta_k^{\tau_1}\right)^{1/\tau_1}\right]^\tau \le c\sum_{k=0}^{\infty}(2^{k(\alpha-\alpha_1)}\beta_k)^\tau,$$

where we used the well-known Hardy inequality (see, e.g., Lemma 3.4 from [19]). Using now that $\alpha - \alpha_1 + 1/\tau_1 = 1/\tau$, we have

$$\mathbf{A}^\tau \le c\sum_{k=0}^{\infty}2^{k(\alpha-\alpha_1+1/\tau_1)\tau}\|b_{\theta_{2^k}}\varphi_{\theta_{2^k}}\|_p^\tau = c\sum_{k=0}^{\infty}2^k\|b_{\theta_{2^k}}\varphi_{\theta_{2^k}}\|_p^\tau \le c\sum_{\nu=1}^{\infty}\|b_{\theta_\nu}\varphi_{\theta_\nu}\|_p^\tau,$$

and (6.6) follows. ■

Several remarks are in order. We first observe that the "Threshold" algorithm, in principle, cannot be applied for approximation in the uniform norm because of the "piling up" effect: there can be a huge number of terms $b_\theta\varphi_\theta$ with small coefficients and with significant contribution to the norm of f at a certain location, which the algorithm will fail to anticipate.

As for the "Push-the-Error" algorithm, it is critical to have an efficient initial decomposition of the function f being approximated, i.e., representation (6.1) should provide a decomposition of the norm in $\mathcal{B}_\tau^\alpha(\mathcal{M})$, $1/\tau = \alpha + 1/p$. For the "Threshold" algorithm this is guaranteed by employing the decompositions from (2.33)–(2.34) with $q < p$.

The estimate $A_n^T(f)_p \le c\|f\|_p$ fails to be true in general (even if $1 < p < \infty$) since the convergence in the representation of the function f being approximated that is used (see (2.33)–(2.34)) is not assumed to be unconditional. (This problem does not arise in the case when wavelets exist.) Consequently, we are unable to prove the analog of estimate (5.23) and the right-hand side of (5.24) for the "Threshold" algorithm. This is why the result from Theorem 6.2 is somewhat weaker than the result from Theorem 5.9.

It is possible to extend the "Push-the-Error" algorithm to approximation in L_p ($p < \infty$). However, the resulting algorithm is very close to the "Threshold" algorithm. Therefore, the "Threshold" algorithm should be considered as a natural generalization of "Push-the-Error" in L_p.

7. Proof of the Main Results

Proof of Theorem 5.5. Our strategy will be to find for each index from $\tilde\Lambda := \tilde\Lambda(f, c_*\varepsilon)$ a reference index η in $\tilde\Lambda_0 \cup \tilde\Lambda_1$ with $\tilde\Lambda_i = \tilde\Lambda(f_i, \varepsilon_i)$, so that $\eta \in \tilde\Lambda_0 \cup \tilde\Lambda_1$ serves as a reference index for at most a uniformly bounded number of indices in $\tilde\Lambda$.

In what follows the "Push-the-Error" algorithm is applied to $g \in \{f, f_0, f_1\}$. We shall adhere to all the notation established in the previous sections, in particular, β_1, β_2, $\tilde\nu$, ν_*, $\nu_0, \dots, \nu_5$ (all of them ≥ 1) denote the parameters of the underlying MRA (Section 2.1), and we recall that $N_* := 2\tilde\nu\nu_*$, $K := \nu_*\nu_5$.

Our main tools are criteria for identifying indices in $\tilde\Lambda(g, \varepsilon)$. The simplest one is based on a sufficiently large threshold for the coefficients $b_\theta(g)$.

Lemma 7.1. *If* $|b_\theta(g)| > \tilde{c}\varepsilon$ *where* $\tilde{c} := 2\beta_1\nu_2$, *then* $\theta \in \tilde{\Lambda}(g,\varepsilon)$. *Equivalently*, *if* $\theta \notin \tilde{\Lambda}(g,\varepsilon)$, *then* $|b_\theta(g)| \le \tilde{c}\varepsilon$.

Proof. Since $\tilde{c} > 1, \theta$ cannot be discarded in the pruning Step 2 of the "Push-the-Error" algorithm. Suppose that after processing all levels $\le m-1$ with $m := l(\theta)$ in Step 3, the current approximation to g has the form

$$g_{m-1} := \sum_{l \le m-1} \mathcal{A}_l(g) + \sum_{\eta \in (\Theta_{m-1} \cap \Gamma) \setminus \Lambda_{m-1}} r_\eta \varphi_\eta + \sum_{l(\eta) \ge m} b_\eta(g) \varphi_\eta,$$

where $|r_\eta| \le \varepsilon$ for every $\eta \in (\Theta_{m-1} \cap \Gamma) \setminus \Lambda_{m-1}$, and $\sum_{l \le m-1} \mathcal{A}_l(g)$ is the approximation generated so far. Since by (2.13) we have $\varphi_\eta = \sum_{l(\xi)=l(\eta)+1} a_{\eta,\xi}\varphi_\xi$, where $|a_{\eta,\xi}| \le \beta_1$. Hence

$$\sum_{l(\eta)=m-1} r_\eta \varphi_\eta = \sum_{l(\xi)=m} \left(\sum_\eta a_{\eta,\xi} r_\eta \right) \varphi_\xi,$$

so that we can rewrite g_{m-1} as

$$g_{m-1} := \sum_{l \le m-1} \mathcal{A}_l(g) + \sum_{\theta \in \Theta_m \cap \Gamma} d_\theta \varphi_\theta + \sum_{l(\eta) > m, \eta \in \Gamma} b_\eta(g) \varphi_\eta.$$

This implies

$$d_\theta(g) = \sum_{l(\eta)=m-1: \theta \subset \eta} a_{\eta,\theta} r_\eta + b_\theta(g).$$

Therefore,

$$|d_\theta(g)| \ge |b_\theta(g)| - \sum_{l(\eta)=m-1: \theta \subset \eta} |a_{\eta,\theta}||r_\eta| \le |b_\theta(g)| - \beta_1 \sum_\eta |r_\eta| > \tilde{c}\varepsilon - \beta_1\nu_2\varepsilon > \varepsilon$$

and hence $\theta \in \tilde{\Lambda}(g,\varepsilon)$. ∎

Next we have to take into account that through rewriting small terms in Step 3 of the algorithm new significant terms may build up. The identification of those terms will be based on certain subsets of indices in Θ, which we call *segments*. For a given $v \in \Omega$ and integers $k_1 \ge k_0 \ge 0$, we define the segment $\mathcal{S}(v, k_0, k_1)$ by

$$\mathcal{S}(v, k_0, k_1) := \{\eta \in \Theta : v \in \eta^\circ \text{ and } k_0 \le l(\eta) \le k_1\}. \tag{7.1}$$

It is an important observation that for each $v \in \Omega$ and $\theta \in \Theta$ the set $\{\eta \in \mathcal{U}'_\theta : v \in \eta^\circ\}$ is a segment or is empty.

We call $\mathcal{S} = \mathcal{S}(v, k_0, k_1)$ an *ε-segment* for a given function g, if

$$F_{\mathcal{S}}(g) := \left| \sum_{\eta \in \mathcal{S}} b_\eta(g) \varphi_\eta(v) \right| > \varepsilon. \tag{7.2}$$

Large segments imply the existence of significant coefficients in a certain neighborhood which is quantified by the following lemma:

Lemma 7.2. *Let $L \geq 1$ and $c^{\diamond} := 7L\nu_2^2\beta_1$. Suppose that the "Push-the-Error" algorithm has been applied to g with threshold $\varepsilon > 0$ and let $\mathcal{S} = \mathcal{S}(v, k_0, k_1)$ be a $c^{\diamond}\varepsilon$-segment for g. Then there exists $\theta^* \in \tilde{\Lambda}(g, \varepsilon)$ with the following properties*:

(a) $k_0 \leq l(\theta^*) \leq k_1$.
(b) $v \in \theta^*$ *and* $k_0 \leq l(\theta^*) < k_0 + L$, *or* $v \in \mathrm{Star}_{N_*+5\nu_*}(\theta^*)$ *and* $k_0 + L \leq l(\theta^*) \leq k_1$.

Proof. If $\mathcal{S} \cap \tilde{\Lambda} \neq \emptyset$ with $\tilde{\Lambda} := \tilde{\Lambda}(g, \varepsilon)$, then the assertion of the lemma obviously holds.

Suppose now that $\mathcal{S} \cap \tilde{\Lambda} = \emptyset$. Let m be the minimum of k_1 and the lowest level so that all indices η with $v \in \eta^\circ$, $l(\eta) > m$ have been discarded in Step 2. Denote by $\mathcal{D}_m$ the set of all indices $\eta \in \mathcal{S}$ with $l(\eta) = m$, which have been discarded in Step 2 as well. Thus $\mathcal{D}_m = \mathcal{S} \cap \Theta_m \cap \Gamma^c$, where Γ^c is the complement of $\Gamma := \Gamma(g, \varepsilon)$. By our assumption, $(\mathcal{S} \cap \Theta_m)\backslash\mathcal{D}_m \neq \emptyset$ and all $\theta \in \mathcal{S}$ with $l(\theta) < m$ belong to Γ. Now, since by assumption $\mathcal{S} \cap \tilde{\Lambda} = \emptyset$, we have for each $\theta \in \mathcal{D}_m$ that $|b_\theta(g)| + \|\sum_{\eta \in \mathcal{U}'_\theta} b_\eta(g)\varphi_\eta\|_\infty \leq \varepsilon$. Using this and the fact that, by the hypotheses of the lemma,

$$F_{\mathcal{S}}(g) := \left|\sum_{\eta \in \mathcal{S}} b_\eta(g)\varphi_\eta(v)\right| > c^{\diamond}\varepsilon, \tag{7.3}$$

we obtain

$$\begin{aligned} \left|\sum_{\eta \in \mathcal{S}\cap\Gamma} b_\eta(g)\varphi_\eta(v)\right| &\geq \left|\sum_{\eta \in \mathcal{S}} b_\eta(g)\varphi_\eta(v)\right| - \sum_{\eta \in \mathcal{D}_m} |b_\eta(g)| - \left\|\sum_{\eta \in \mathcal{U}'_\theta} b_\eta(g)\varphi_\eta\right\|_\infty \\ &\geq (c^{\diamond} - \nu_2 - 1)\varepsilon, \end{aligned} \tag{7.4}$$

where θ is an arbitrary index from $\mathcal{D}_m$.

Since $\mathcal{S} \cap \tilde{\Lambda} = \emptyset$, by Lemma 7.1, $|b_\eta(g)| \leq \tilde{c}\varepsilon$ for $\eta \in \mathcal{S}$ and hence

$$\sum_{\eta \in \mathcal{S},\, l(\eta) < k_0 + L} |b_\eta(g)| \leq L\nu_2\tilde{c}\varepsilon. \tag{7.5}$$

Since $c^{\diamond} - \nu_2 - 1 > L\nu_2\tilde{c}$, it follows by (7.4)–(7.5) that $k_0 + L \leq m \leq k_1$.

From (7.4)–(7.5), we obtain

$$\begin{aligned} \left|\sum_{\eta \in \mathcal{S}\cap\Gamma,\, l(\eta) \geq k_0 + L} b_\eta(g)\varphi_\eta(v)\right| &\geq \left|\sum_{\eta \in \mathcal{S}\cap\Gamma} b_\eta(g)\varphi_\eta(v)\right| - \sum_{\eta \in \mathcal{S},\, k_0 \leq l(\eta) < k_0 + L} |b_\eta(g)| \\ &\geq (c^{\diamond} - \nu_2 - 1 - L\nu_2\tilde{c})\varepsilon. \end{aligned} \tag{7.6}$$

Suppose now that after having processed all levels $< k_0 + L$ in Step 3 of "Push-the-Error," we have

$$\begin{aligned} g_{k_0+L-1}(v) &= \sum_{l < k_0 + L} \mathcal{A}(g; v) + \sum_{l(\eta) = k_0 + L : v \in \eta^\circ} r_\eta\varphi_\eta(v) + \sum_{\eta \in \mathcal{S}\cap\Gamma,\, l(\eta) \geq k_0 + L} b_\eta(g)\varphi_\eta(v) \\ &=: \mathcal{A}(v) + g_1(v) + g_2(v), \end{aligned}$$

where $\mathcal{A}(v)$ is the approximation generated so far. As before, the r_η arise from rewriting small lower level terms and can thus be estimated as $|r_\eta| \le \beta_1 \nu_2 \varepsilon$. Hence, the second sum can be bounded by $|g_1(v)| \le \beta_1 \nu_2^2 \varepsilon$. Using this and (7.6), we obtain

$$(7.7)\quad |g_1(v) + g_2(v)| \ge |g_2(v)| - |g_1(v)| \ge (c^\diamond - 1 - \nu_2(1 + \beta_1\nu_2 + L\tilde{c}))\varepsilon.$$

Suppose now that none of the indices $\eta \in \mathcal{S} \cap \Gamma$, $l(\eta) \ge k_0 + L$, has a neighbor in $\tilde{\Lambda}$. Then we can write

$$g_1(v) + g_2(v) = \sum_{\eta \in \mathcal{S} \cap \Theta_m} d_\eta(g)\varphi_\eta(v)$$

and, hence,

$$|g_1(v) + g_2(v)| \le \left\| \sum_{\eta \in \mathcal{S} \cap \Theta_m} d_\eta(g)\varphi_\eta \right\|_\infty \le \nu_2 \max_{\eta \in \mathcal{S} \cap \Theta_m} |d_\eta(g)|.$$

This together with (7.7) yields

$$(7.8)\qquad \max_{\eta \in \mathcal{S} \cap \Theta_m} |d_\eta(g)| \ge \nu_2^{-1}(c^\diamond - 1 - \nu_2(1 + \beta_1\nu_2 + L\tilde{c}))\varepsilon > \varepsilon,$$

because, recalling the definition of $\tilde{c}$ from Lemma 7.1, $\nu_2^{-1}(c^\diamond - 1 - \nu_2(1 + \beta_1\nu_2 + L\tilde{c})) > 1$ if and only if

$$c^\diamond > 1 + \nu_2(2 + \beta_1\nu_2 + L\tilde{c}) = 1 + 2\nu_2 + \beta_1\nu_2^2(2L + 1),$$

and $1 + 2\nu_2 + \beta_1\nu_2^2(2L + 1) \le 6L\beta_1\nu_2^2$. Since therefore (7.8) contradicts the assumption $\mathcal{S} \cap \tilde{\Lambda} = \emptyset$, there exists $\eta \in \mathcal{S} \cap \Gamma$, $l(\eta) \ge k_0 + L$, with a neighbor $\theta^* \in \tilde{\Lambda}$. Then, using (5.10), $v \in \mathrm{Star}_{N_* + 5\nu_*}(\theta^*)$ and the assertion of the lemma holds. ∎

We have now collected the necessary tools for detecting reference elements in $\tilde{\Lambda}(f_0, \varepsilon_0) \cup \tilde{\Lambda}(f_1, \varepsilon_1)$ from $\tilde{\Lambda}(f, c^*(\varepsilon_0 + \varepsilon_1))$. We shall verify the claim for

$$(7.9)\quad c_* := 14\beta_1\beta_2\tilde{\nu}\nu_2^2(N_* + 7\nu_*)K \qquad \text{where} \quad N_* := 2\tilde{\nu}\nu_*, \quad K := \nu_*\nu_5,$$

which is certainly far from being optimal (and we make no attempt at determining optimal constants here). For the rest of the proof, we assume that the hypotheses of Theorem 5.5 are fulfilled.

It is an important observation that the coefficients $b_\theta(f)$ from the decomposition of Step 1 of the algorithm (see (5.1)) are linear functionals and hence $b_\theta(f) = b_\theta(f_0) + b_\theta(f_1)$.

In what follows we shall use the abbreviations $\tilde{\Lambda} := \tilde{\Lambda}(f, c_*(\varepsilon_0 + \varepsilon_1))$ and $\tilde{\Lambda}_i = \tilde{\Lambda}(f_i, \varepsilon_i)$, $i = 0, 1$.

We shall use two detection devices. The following first one says that one can for any $\theta \in \tilde{\Lambda}$ always find an element θ^* in $\tilde{\Lambda}_0 \cup \tilde{\Lambda}_1$ which is spatially located near θ but has possibly a higher level. This device is, for instance, useful for the leaves in $\tilde{\Lambda}$.

Lemma 7.3. *For any $\theta \in \tilde{\Lambda}$ there exists an index $\theta^* \in \tilde{\Lambda}_0 \cup \tilde{\Lambda}_1$ such that*

$$\theta^* \subset \mathrm{Star}_{N_* + \nu_*}(\theta) \qquad \textit{and} \qquad l(\theta^*) \ge l(\theta).$$

Proof. From Step 3 of the algorithm, $\tilde{\Lambda} \subset \Gamma(f, c_*(\varepsilon_0 + \varepsilon_1))$. Then, by Step 2, for every $\theta \in \tilde{\Lambda}$ there is $\eta \in \mathcal{U}_\theta$ such that $E(f, \eta) > c_*(\varepsilon_0 + \varepsilon_1)$. Since $E(f, \eta) \leq E(f_0, \eta) + E(f_1, \eta)$, we must have either $E(f_0, \eta) > c_*\varepsilon_0$ or $E(f_1, \eta) > c_*\varepsilon_1$. Suppose that the first inequality is true, so that $\theta \in \Gamma(f_0, c_*\varepsilon_0)$. Then either $|b_\eta(f_0)| > c_*\varepsilon_0/2$ or $\|\sum_{\xi \in \mathcal{U}'_\eta} b_\xi(f_0)\varphi_\xi\|_\infty > c_*\varepsilon_0/2$. If the first happens to be true, we set $\theta^* := \eta$. We use Lemma 7.1 and the fact that $c_*/2 \geq \tilde{c}$ to conclude that $\theta^* \in \tilde{\Lambda}(f_0, \varepsilon_0)$. By (5.4), we know that $\theta^* \subset \mathrm{Star}_{N_*}(\theta)$. Thus θ^* has the claimed properties.

Consider now the second case $\|\sum_{\xi \in \mathcal{U}'_\eta} b_\xi(f_0)\varphi_\xi\|_\infty > c_*\varepsilon_0/2$ of a significant segment. Then for some point v and $\mathcal{S}(v) := \{\xi \in \mathcal{U}'_\eta : v \in \xi^\circ\}$, we have $F_{\mathcal{S}}(f) > c_*\varepsilon_0/2$. Choose $L := (N_* + 6\nu_*)K$ with $K := \nu_*\nu_5$. One easily verifies that $c_*/2 \geq c^\diamond$ with $c^\diamond := 7L\nu_2^2\beta_1$. This allows us to apply Lemma 7.2 to f_0 with the above segment $\mathcal{S}(v)$ to find $\theta^* \in \tilde{\Lambda}_0$ such that $l(\theta^*) \geq l(\theta)$ and either $v \in \theta^*$ or $v \in \mathrm{Star}_{N_*+5\nu_*}(\theta^*)$ and $l(\theta^*) \geq l(\theta) + L$. If we denote $m := l(\theta)$ and $m^* := l(\theta^*)$, then from the above choice of L, we have $m^* \geq m + L = m + (N_* + 6\nu_*)K$. Employing Lemma 2.3, we obtain

$$\theta^* \subset \mathrm{Star}^{(m^*)}_{N_*+6\nu_*}(v) \subset \mathrm{Star}^{(m)}_{\nu_4}(v) \subset \mathrm{Star}^{(m)}_{N_*+\nu_*}(\theta),$$

which completes the proof. ∎

We need a second somewhat refined device for elements in $\tilde{\Lambda}$ whose neighborhood is hit by some higher level elements in $\tilde{\Lambda}$. In this case we need to cap the reference element from above.

Lemma 7.4. *Suppose $\theta_0, \theta_1 \in \tilde{\Lambda}$ satisfy the following: $l(\theta_0) < l(\theta_1)$ and $\theta_1 \subset \mathrm{Star}_j(\theta_0)$, where $j \leq N_* + 2\nu_*$. Then there exists $\theta^* \in \tilde{\Lambda}_0 \cup \tilde{\Lambda}_1$ such that*

$$\theta^* \subset \mathrm{Star}_{j+2\nu_*}(\theta_0) \qquad \text{and} \qquad l(\theta_0) \leq l(\theta^*) \leq l(\theta_1). \tag{7.10}$$

Proof. Let $l_0 := l(\theta_0)$, $l_1 := l(\theta_1)$, and consider the set

$$\mathcal{T} := \{\eta \in \Theta \cap \Gamma : \eta \subset \mathrm{Star}_{j+2\nu_*}(\theta_0) \text{ and } l_0 < l(\eta) \leq l_1\}$$

of indices which are sandwiched by θ_0 and θ_1, and where we have to search for θ^*.

If $|b_\eta(f)| > \tilde{c}(\varepsilon_0 + \varepsilon_1)$ for some $\eta \in \mathcal{T}$, then since $b_\eta(f) = b_\eta(f_0) + b_\eta(f_1)$ either $|b_\eta(f_0)| > \tilde{c}\varepsilon_0$ or $|b_\eta(f_1)| > \tilde{c}\varepsilon_1$. Applying Lemma 7.1, it follows that $\eta \in \tilde{\Lambda}_0 \cup \tilde{\Lambda}_1$ and the lemma holds.

Suppose

$$|b_\eta(f)| \leq \tilde{c}(\varepsilon_0 + \varepsilon_1) \qquad \text{for} \quad \eta \in \mathcal{T}. \tag{7.11}$$

Choose

$$\hat{L} := 2\tilde{\nu}(N_* + 6\nu_*)K \tag{7.12}$$

and split $\mathcal{T}$ into a lower part $\mathcal{T}^- := \{\eta \in \mathcal{T} : l(\eta) < l_0 + \hat{L}\}$ and an upper part $\mathcal{T}^+ := \mathcal{T} \backslash \mathcal{T}^-$.

We first show that, under the assumption (7.11), the lower part $\mathcal{T}^-$ cannot intersect $\tilde{\Lambda}$ so that $l_1 \geq l_0 + \hat{L}$. To this end, fix $\theta \in \mathcal{T}^-$ and denote

$$\mathcal{T}_\theta := \{\eta \in \Theta : \theta \subset \eta \text{ and } l_0 < l(\eta) < l(\theta)\}.$$

From the definition of Γ in Step 2, it follows that $\mathcal{T}_\theta \subset \Gamma$. Moreover, if $\theta \in \Gamma$, then all $\eta \in \Theta$, with $l(\eta) < l(\theta)$ which are connected to θ via sets from Θ, belong to Γ.

Now a possible source of significant coefficients $d_\theta(f)$ in $\mathcal{T}^-$ is through rewriting small lower level terms in Step 3. However, the important point here is that, since $\theta \subset \mathrm{Star}_{N_*+2\nu_*}(\theta_0)$ (it suffices to have $\theta \subset \mathrm{Star}_{N_*+3\nu_*}(\theta_0)$) and the "concrete" of $\theta_0 \in \tilde{\Lambda}$ is $\Omega_{\theta_0} := \mathrm{Star}_{N_*+4\nu_*}(\theta_0)$, there are no contributions to $d_\theta(f)$ (obtained in Step 3) from levels $\leq l_0$. (Since $\theta_0 \in \tilde{\Lambda}$, all neighbors of θ_0 are taken in the approximant.) Therefore, a significant coefficient $d_\theta(f)$ could only be fed from $\mathcal{T}_\theta$ which, however, turns out to be prevented by the bound (7.11). In fact, using (2.9), (7.11), and property (β) of Θ, we obtain

$$
\begin{aligned}
|d_\theta(f)| &\leq |b_\theta(f)| + \beta_2 \left\| \sum_{\eta \in \mathcal{T}_\theta : \theta \subset \eta} b_\eta(f)\varphi_\theta \right\|_\infty \\
&\leq \beta_2 \left(|b_\theta(f)| + \sum_{m=l_0+1}^{l(\theta)-1} \sum_{\eta \in \mathcal{T}_\theta \cap \Theta_m : \theta \subset \eta} |b_\eta(f)| \right) \\
&\leq \beta_2 \hat{L} \nu_2 \tilde{c}(\varepsilon_0 + \varepsilon_1) < c_*(\varepsilon_0 + \varepsilon_1),
\end{aligned}
$$

where we have used $c_* > \beta_2 \hat{L} \nu_2 \tilde{c}$ (see (7.9)). Therefore, $\theta \notin \tilde{\Lambda}$ and $\mathcal{T}^- \cap \tilde{\Lambda} = \emptyset$.

Thus, under the assumption (7.11), it suffices to search in the upper part $\mathcal{T}^+$. For a given $\theta \in \mathcal{T}^+$, we distinguish again an upper section

$$
\mathcal{T}_\theta^+ := \{\eta \in \Theta : \theta \subset \eta \ \text{ and } \ l_0 + \hat{L} \leq l(\eta) < l(\theta)\}
$$

and a lower section

$$
\mathcal{T}_\theta^- := \{\eta \in \Theta : \theta \subset \eta \ \text{ and } \ l_0 < l(\eta) < l_0 + \hat{L}\},
$$

which may build up $d_\theta(f)$. Notice that, by the same reasoning as above, $\mathcal{T}_\theta^\pm \subset \Gamma$.

We next show that there exists $\theta^\diamond \in \tilde{\Lambda}$ with the following properties:

(P1) $l_0 + \hat{L} \leq l(\theta^\diamond) \leq l_1$, $\theta^\diamond \subset \mathrm{Star}_{j+\nu_*}(\theta_0)$; and
(P2) neither $\eta \in \mathcal{T}_{\theta^\diamond}^+$ has a neighbor in $\tilde{\Lambda}$.

Indeed, if none of the $\eta \in \mathcal{T}_{\theta_1}^+$ has a neighbor in $\tilde{\Lambda}$, then $\theta^\diamond := \theta_1$ has the claimed properties since (P1) holds by assumption. Otherwise, using (5.10) there is $\theta^1 \in \mathcal{T}^+ \cap \tilde{\Lambda}$ with $l(\theta^1) < l_1$ such that $\theta_1 \subset \mathrm{Star}_{n_*}(\theta^1)$, where $n_* := N_* + 5\nu_*$. If none of the $\eta \in \mathcal{T}_{\theta^1}^+$ has a neighbor in $\tilde{\Lambda}$, i.e., (P2) holds, we set $\theta^\diamond := \theta^1$. If (P2) is not true we proceed further in the same way and find indices $\theta^2, \theta^3, \ldots$ with strictly decreasing levels. After finitely many steps, this process will therefore terminate and we find an index $\theta^r \in \mathcal{T}^+ \cap \tilde{\Lambda}$ such that either each $\eta \in \mathcal{T}_{\theta^r}^+$ has no neighbor in $\tilde{\Lambda}$, thus satisfying (P2), or $l(\theta^r) = l_0 + \hat{L}$. In this latter case $\mathcal{T}_{\theta^r}^+ = \emptyset$ so that (P2) is trivially satisfied. We define $\theta^\diamond := \theta^r$ and show next that $\theta^\diamond$ also satisfies (P1). To this end, note that $\theta^\diamond$ (as well as every other θ^j, $j = 1, 2, \ldots, r-1$) is n_*-star connected with θ_1 and, hence, by Lemma 2.2, $\theta_1 \subset \mathrm{Star}_{2\tilde{\nu}n_*}^{(m^\diamond)}(\theta^\diamond)$, where $m^\diamond := l(\theta^\diamond)$. Now, using (2.17), we have $\theta^\diamond \subset \mathrm{Star}_{2\tilde{\nu}n_*+\nu_*}^{(m^\diamond)}(\theta_1)$. Further, taking into account that $m^\diamond \geq l_0 + \hat{L} \geq l_0 + (2\tilde{\nu}n_* + \nu_*)K$

(see (7.12)), we apply Lemma 2.3 (see (2.25)) to obtain

$$\theta^\diamond \subset \mathrm{Star}^{(m^\diamond)}_{2\tilde{\nu}n_*+\nu_*}(\theta_1) \subset \mathrm{Star}^{(l_0)}_{\nu_4}(\theta_1) \subset \mathrm{Star}^{(l_0)}_{j+\nu_*}(\theta_0) \qquad (\nu_* \geq \nu_4).$$

Thus $\theta^\diamond$ satisfies (P1) as well and thus $\theta^\diamond$ has the desired properties.

Consider first the case when $m^\diamond := l(\theta^\diamond) > l_0 + \hat{L}$. As was argued above, since $\theta^\diamond \subset \mathrm{Star}_{N_*+3\nu_*}(\theta_0)$ then there are no contributions to $d_{\theta^\diamond}(f)$ (obtained in Step 3) from levels $\leq l_0$. Then using (2.9), property (β) of Θ, and (7.11), we obtain

$$\text{(7.13)} \qquad \begin{aligned} |d_{\theta^\diamond}(f)| &\leq \beta_2\left(\sum_{\eta\in\mathcal{T}^-_{\theta^\diamond}} |b_\eta(f)| + \left\|\sum_{\eta\in\mathcal{T}^+_{\theta^\diamond}} b_\eta(f)\varphi_\eta\right\|_\infty + |b_{\theta^\diamond}(f)|\right) \\ &\leq \beta_2\left(\hat{L}\nu_2\tilde{c}(\varepsilon_0+\varepsilon_1) + \left\|\sum_{\eta\in\mathcal{T}^+_{\theta^\diamond}} b_\eta(f)\varphi_\eta\right\|_\infty + |b_{\theta^\diamond}(f)|\right). \end{aligned}$$

This will allow us to find a large coefficient $b_\theta(f)$ or a significant segment and either case will lead to a θ^*. In fact, since $\theta^\diamond \in \tilde{\Lambda}$, $|d_{\theta^\diamond}(f)| \geq c_*(\varepsilon_0+\varepsilon_1)$. Combining with (7.13), we obtain

$$\left\|\sum_{\eta\in\mathcal{T}^+_{\theta^\diamond}} b_\eta(f)\varphi_\eta\right\|_\infty + |b_{\theta^\diamond}(f)| \geq (c_*\beta_2{}^{-1} - \hat{L}\nu_2\tilde{c})(\varepsilon_0+\varepsilon_1) =: c^\natural(\varepsilon_0+\varepsilon_1).$$

If $|b_{\theta^\diamond}(f)| \geq (c^\natural/2)(\varepsilon_0+\varepsilon_1)$, then either $|b_{\theta^\diamond}(f_0)| \geq (c^\natural/2)\varepsilon_0$ or $|b_{\theta^\diamond}(f_1)| \geq (c^\natural/2)\varepsilon_1$. Using that $c^\natural/2 > \tilde{c}$ and Lemma 7.1, we infer that $\theta^\diamond \in \tilde{\Lambda}_0 \cap \tilde{\Lambda}_1$ and the lemma follows.

If $\|\sum_{\eta\in\mathcal{T}^+_{\theta^\diamond}} b_\eta(f)\varphi_\eta\|_\infty \geq (c^\natural/2)(\varepsilon_0+\varepsilon_1)$, then

$$\left\|\sum_{\eta\in\mathcal{T}^+_{\theta^\diamond}} b_\eta(f_0)\varphi_\eta\right\|_\infty \geq \frac{c^\natural}{2}\varepsilon_0 \qquad \text{or} \qquad \left\|\sum_{\eta\in\mathcal{T}^+_{\theta^\diamond}} b_\eta(f_1)\varphi_\eta\right\|_\infty \geq \frac{c^\natural}{2}\varepsilon_1.$$

Therefore, there exists a $(c^\natural/2)\varepsilon_i$-segment ($i=0$ or 1) $\mathcal{S}(v)$ for f_0 or f_1 with $v \in \theta^\diamond$. Now applying Lemma 7.2 with $L=1$, there exists $\theta^* \in \tilde{\Lambda}_0 \cup \tilde{\Lambda}_1$ such that $l_0+\hat{L} \leq l(\theta^*) \leq l_1$ and either $(\theta^*)^\circ \cap (\theta^\diamond)^\circ \neq \emptyset$ or $\mathrm{Star}_{N_*+5\nu_*}(\theta^*) \cap (\theta^\diamond)^\circ \neq \emptyset$. In the latter case, we obtain as above, using that $\hat{L} \geq (N_*+6\nu_*)K$,

$$\theta^* \subset \mathrm{Star}^{(m^*)}_{N_*+6\nu_*}(\theta^\diamond) \subset \mathrm{Star}^{(l_0)}_{\nu_4}(\theta^\diamond) \subset \mathrm{Star}^{(l_0)}_{j+2\nu_*}(\theta_0), \qquad m^* := l(\theta^*).$$

The proof of Lemma 7.4 is complete. ■

Finally, we are in a position to complete the proof of Theorem 5.5. An important vehicle for proving this theorem will be the coloring property of the extended cells (Section 2.1). We begin with some additional coloring type preprocessing of the subsets $\{\mathcal{O}^j\}_{j=1}^J$ of $\mathcal{O}$. By Lemma 2.3, for each $\theta \in \Theta$ there exists $\omega \in \mathcal{O}$ such that

$$\text{(7.14)} \qquad \Omega_\theta := \mathrm{Star}_{N_*+4\nu_*}(\theta) \subset \omega \quad \text{and} \quad l(\omega) = l(\theta) - \tilde{K} \quad \text{with} \quad \tilde{K} := (N_*+4\nu_*)K,$$

whenever $l(\theta) \geq \tilde{K}$. We associate ω with θ. Note that each $\omega \in \mathcal{O}$ can be associated in this way with no more than $\tilde{N} := \nu_3^{\tilde{K}}$ indices $\theta \in \Theta$. In fact, recall that, by property (ζ)

of Θ (Section 2.1), there is $\theta^\diamond$ such that $\omega \subset \theta^\diamond$, and $\tilde{N}$ is a rough upper bound for the number of elements $\theta \in \Theta$ at level $l(\theta) = l(\omega) + \tilde{K}$ which are contained in any $\theta^\diamond$ with $l(\theta^\diamond) = l(\omega)$. We take $\tilde{N}$ copies of each class $\mathcal{O}^j$, denoting them by

$$\mathcal{O}^{j,n}, \qquad n = 1, 2, \ldots, \tilde{N}; \quad j = 1, 2, \ldots, J.$$

From above, it is clear that we can establish a one-to-one correspondence between $\Theta' := \bigcup_{m \geq \tilde{K}} \Theta_m$ and a subset of $\bigcup_{j,n} \mathcal{O}^{j,n}$.

The set $\Theta \backslash \Theta'$ is finite with $\#\Theta \backslash \Theta' \leq (\#\Theta_0) \cdot \nu_3^{\tilde{K}}$, which is a constant that can be absorbed by the constant c in (5.19) and hence $\Theta \backslash \Theta'$ can be ignored.

To simplify the notation, we denote by $\mathcal{O}^\diamond$ an arbitrary class $\mathcal{O}^{j,n}$ and also we denote by $\Theta^\diamond$ the corresponding subset of Θ' which is in one-to-one correspondence with a subset of $\mathcal{O}^\diamond$. Thus we can associate with each $\theta \in \Theta^\diamond$ an $\omega_\theta \in \mathcal{O}^\diamond$ such that $\mathrm{Star}_{N_*+4\nu_*}(\theta) \subset \omega_\theta$ and $l(\omega) = l(\theta) - \tilde{K}$. In addition, if $\theta', \theta'' \in \mathcal{O}^\diamond$, $\theta' \neq \theta''$, and $\omega_{\theta'} \subset \omega_{\theta''}$, then $l(\theta') > l(\theta'')$.

Clearly $\mathcal{O}^\diamond$ inherits the tree structure of the corresponding $\mathcal{O}^j$. Setting $\tilde{\Lambda}^\diamond := \tilde{\Lambda}(f, \varepsilon) \cap \Theta^\diamond$, the theorem will be proved if we show that $\#\tilde{\Lambda}^\diamond \leq c(\#\tilde{\Lambda}_0 + \#\tilde{\Lambda}_1)$.

We now introduce a partial order $(\prec)$ in $\tilde{\Lambda}^\diamond$: $\theta_1 \prec \theta_2$ if $\omega_{\theta_1} \subset \omega_{\theta_2}$. With this partial order $\tilde{\Lambda}^\diamond$ becomes a tree as well.

We next introduce several subsets of $\tilde{\Lambda}^\diamond$. We denote by $\tilde{\Lambda}^\diamond_\ell$ the set of all *leaves* in $\tilde{\Lambda}^\diamond$ ($\theta \in \tilde{\Lambda}^\diamond_\ell$ if θ has no children in $\tilde{\Lambda}^\diamond$) and by $\tilde{\Lambda}^\diamond_{\mathrm{b}}$ the set of all *branching* elements in $\tilde{\Lambda}^\diamond$ (elements in $\tilde{\Lambda}^\diamond$ with at least two children in $\tilde{\Lambda}^\diamond$). Also, we denote $\tilde{\Lambda}^\diamond_{\mathrm{ch}} := \tilde{\Lambda}^\diamond \backslash (\tilde{\Lambda}^\diamond_\ell \cup \tilde{\Lambda}^\diamond_{\mathrm{b}})$ which is the set of all *chain* elements in $\tilde{\Lambda}^\diamond$ (elements of $\tilde{\Lambda}^\diamond$ with exactly one child in $\tilde{\Lambda}^\diamond$).

After this ground work, we proceed with estimating $\#\tilde{\Lambda}^\diamond_\ell$, $\#\tilde{\Lambda}^\diamond_{\mathrm{b}}$, and $\#\tilde{\Lambda}^\diamond_{\mathrm{ch}}$. By Lemma 7.3, for each $\theta \in \tilde{\Lambda}^\diamond_\ell$ there exists $\theta^* \in \tilde{\Lambda}_0 \cap \tilde{\Lambda}_1$ such that $\theta^* \subset \mathrm{Star}_{N_*+\nu_*}(\theta) \subset \omega_\theta$. We assign such θ^* as a reference index for θ. Clearly, the extended cells ω_θ associated with leaves $\theta \in \tilde{\Lambda}^\diamond_\ell$ are leaves in the corresponding subtree of $\mathcal{O}^\diamond$ and hence are with disjoint interiors. Therefore, the θ^*'s which are associated to indices in $\tilde{\Lambda}^\diamond_\ell$ are distinct and hence

$$\#\tilde{\Lambda}^\diamond_\ell \leq \#\tilde{\Lambda}_0 + \#\tilde{\Lambda}_1. \tag{7.15}$$

Evidently, in any tree the number of the branching elements does not exceed the number of the leaves. Therefore,

$$\#\tilde{\Lambda}^\diamond_{\mathrm{b}} \leq \#\tilde{\Lambda}^\diamond_\ell \leq \#\tilde{\Lambda}_0 + \#\tilde{\Lambda}_1. \tag{7.16}$$

It remains to show that $\#\tilde{\Lambda}^\diamond_{\mathrm{ch}} \leq c(\#\tilde{\Lambda}_0 + \#\tilde{\Lambda}_1)$. To this end, decompose $\#\tilde{\Lambda}^\diamond_{\mathrm{ch}}$ into at most $\tilde{K}$ subsets $\#\tilde{\Lambda}^\diamond_{\mathrm{ch},i}$ such that for each $i \leq \tilde{K}$, $\theta' \prec \theta$ implies $l(\theta') \geq l(\theta) + \tilde{K}$. It suffices to show that $\#\tilde{\Lambda}^\diamond_{\mathrm{ch},i} \leq c(\#\tilde{\Lambda}_0 + \#\tilde{\Lambda}_1)$, $i \leq \tilde{K}$. Fix $\theta \in \tilde{\Lambda}^\diamond_{\mathrm{ch},i}$ and let $\theta' \prec \theta$ be the only descendent of θ in the tree $\tilde{\Lambda}^\diamond \cap \tilde{\Lambda}^\diamond_{\mathrm{ch},i}$ and hence $\theta' \in \tilde{\Lambda}$. Let $m := l(\theta)$. Then $\omega_{\theta'} \subset \omega_\theta$ and $l(\theta') \geq m + \tilde{K}$. Two cases present themselves here:

Case 1: $\theta' \subset \mathrm{Star}^{(m)}_{N_*+2\nu_*}(\theta)$. Then by Lemma 7.4 and (7.14), there exists $\theta^* \in \tilde{\Lambda}_0 \cap \tilde{\Lambda}_1$ such that

$$\theta^* \subset \mathrm{Star}^{(m)}_{N_*+4\nu_*}(\theta) \subset \omega_\theta \qquad \text{and} \qquad l(\theta) \leq l(\theta^*) \leq l(\theta').$$

We assign θ^* as a reference index to θ.

Case 2: $\theta' \not\subset \mathrm{Star}^{(m)}_{N_*+2\nu_*}(\theta)$. By Lemma 7.3, there exists $\theta^* \in \tilde{\Lambda}_0 \cap \tilde{\Lambda}_1$ such that

$$\theta^* \subset \mathrm{Star}^{(m)}_{N_*+\nu_*}(\theta) \subset \omega_\theta \qquad \text{and} \qquad l(\theta^*) \geq l(\theta).$$

We assign θ^* as a reference index to θ. Since $\theta' \not\subset \mathrm{Star}^{(m)}_{N_*+2\nu_*}(\theta)$, there exists a point $v \in \theta' \cap (\mathrm{Star}^{(m)}_{N_*+2\nu_*}(\theta))^c$, and hence

$$(\mathrm{Star}^{(m)}_{\nu_*}(v))^\circ \cap (\mathrm{Star}^{(m)}_{N_*+\nu_*}(\theta))^\circ = \emptyset.$$

Further, using (2.22), we have $\theta' \subset \omega_{\theta'} \subset \mathrm{Star}^{(l(\theta')-\tilde{K})}_{\nu_*}(v) \subset \mathrm{Star}^{(m)}_{\nu_*}(v)$. Therefore, $\theta^* \subset \omega_\theta \backslash \omega_{\theta'}$.

To summarize, we have assigned to each $\theta \in \tilde{\Lambda}^\diamond_{\mathrm{ch},i}$ (with descendent θ' in $\tilde{\Lambda}^\diamond$) an index $\theta^* \in \tilde{\Lambda}_0 \cap \tilde{\Lambda}_1$ such that either $\theta^* \subset \omega_\theta$ and $l(\theta) \leq l(\theta^*) \leq l(\theta')$ or $\theta^* \subset \omega_\theta \backslash \omega_{\theta'}$ and $l(\theta^*) \geq l(\theta)$. Recalling that the ω_θ's are from a tree with respect to the inclusion relation, it follows that each $\theta^* \in \tilde{\Lambda}_0 \cap \tilde{\Lambda}_1$ can be a reference index to at most two indices from $\tilde{\Lambda}^\diamond_{\mathrm{ch},i}$ and hence

$$\#\tilde{\Lambda}^\diamond_{\mathrm{ch}} \leq 2\tilde{K}(\#\tilde{\Lambda}_0 + \#\tilde{\Lambda}_1).$$

Combining this with (7.15)–(7.16), gives $\#\tilde{\Lambda}^\diamond \leq 2(1+\tilde{K})(\#\tilde{\Lambda}_0+\#\tilde{\Lambda}_1)$, which completes the proof of Theorem 5.5. ■

Proof of Theorem 5.6. We shall follow the scheme of the proof of Theorem 5.5, but everything will be much easier. We adopt all necessary notation from the proof of Theorem 5.5. Denote briefly $\tilde{\Lambda} := \tilde{\Lambda}(f, \varepsilon)$.

The following two trivial lemmas can be considered as analogues of Lemmas 7.3 and 7.4.

Lemma 7.5. *For any $\theta \in \tilde{\Lambda}$ there exists a segment $\mathcal{S} = \mathcal{S}(v, k_0, k_1)$ such that*

$$\bar{\mathcal{S}} := \bigcup \{\eta : \eta \in \mathcal{S}(v, k_0, k_1)\} \subset \mathrm{Star}_{N_*}(\theta), \quad k_0 \geq l(\theta),$$

and

$$\sum_{\eta \in \mathcal{S}} |b_\eta(f)| > \varepsilon/2.$$

Proof. If $\theta \in \tilde{\Lambda}$, then $\theta \in \Gamma(f, \varepsilon)$ and hence there exists $\theta' \in \mathcal{U}_\theta$ such that $E(f, \theta') > \varepsilon$. This immediately implies that either $|b_{\theta'}(f)| > \varepsilon/2$ or there exists a segment $\bar{\mathcal{S}} \subset \mathrm{Star}_{N_*}\theta'$ such that $F_{\mathcal{S}}(f) > \varepsilon/2$, which yields $\sum_{\eta \in \mathcal{S}} |b_\eta(f)| \geq F_{\mathcal{S}}(f) > \varepsilon/2$. ■

Lemma 7.6. *Let $\theta_0, \theta_1 \in \tilde{\Lambda}$ be such that $l(\theta_0) < l(\theta_1)$ and $\theta_1 \subset \mathrm{Star}_{N_*+3\nu_*}(\theta_0)$. Let $l_0 := l(\theta_0)$, $l_1 := l(\theta_1)$, and let*

$$\mathcal{S} = \mathcal{S}(\theta_1, l_0, l_1) := \{\eta \in \Theta : \theta_1 \subset \eta \text{ and } l_0 < l(\eta) \leq l_1\}.$$

Then

$$\sum_{\eta\in\mathcal{S}} |b_\eta(f)| > \varepsilon/\beta_2.$$

Proof. Since $\theta_1 \subset \mathrm{Star}_{N_*+3\nu_*}(\theta_0)$, there is no contribution to $d_{\theta_1}(f)$ from levels $\le j_0$. Denote by $\mathcal{S}'$ the set of all terms $b_\theta(f)\varphi_\theta$ which contribute to $d_{\theta_1}(f)$. Clearly, $\mathcal{S}' \subset \mathcal{S}$ and, using (2.9),

$$\varepsilon < |d_{\theta_1}(f)| \le \beta_2 \left\| \sum_{\eta\in\mathcal{S}':\theta_1\subset\eta} b_\eta(f)\varphi_\eta \right\|_\infty \le \beta_2 \sum_{\eta\in\mathcal{S}} |b_\eta(f)|$$

and the lemma follows. ■

To complete the proof of Theorem 5.6 we shall utilize the coloring construction from the proof of Theorem 5.5. According to this construction (with a slight change of notation), $\tilde\Lambda$ can be represented as a disjoint union $\tilde\Lambda = (\bigcup_{j=1}^J \tilde\Lambda^j) \cup \mathring{\Lambda}$ and there exists a collection $\{\mathcal{O}^j\}_{j=1}^J$ of subsets of $\mathcal{O}$ with the following properties:

(i) $\#\mathring{\Lambda} \le$ constant.
(ii) There is a one-to-one correspondence between $\tilde\Lambda^j$ and $\mathcal{O}^j$ $(1 \le j \le J)$. If we denote by ω_θ the extended cell in $\mathcal{O}^j$ which corresponds to $\theta \in \tilde\Lambda^j$, then $\mathrm{Star}_{N_*+\nu_*}(\theta) \subset \omega_\theta$.
(iii) Each set $\mathcal{O}^j$ is a tree with respect to the inclusion relation which we often indicate by writing $(\mathcal{O}^j, \subset)$. Moreover, if $\theta', \theta'' \in \tilde\Lambda^j$, $\theta' \ne \theta''$, and $\omega_{\theta'} \subset \omega_{\theta''}$, then $l(\theta') > l(\theta'')$.

We introduce a partial order $(\prec)$ in $\tilde\Lambda^j$: $\theta' \prec \theta''$ if $\omega_{\theta'} \subset \omega_{\theta''}$. Since $\mathcal{O}^j$ $(\subset)$ is a tree, then $\tilde\Lambda^j$ $(\prec)$ becomes a tree as well.

As in the proof of Theorem 5.5, we introduce the following subsets of $\tilde\Lambda^j$: $\tilde\Lambda^j_\ell$ is the set of all *leaves* in $\tilde\Lambda^j$, $\tilde\Lambda^j_{\mathrm{b}}$ is the set of all *branching elements* in $\tilde\Lambda^j$, and $\tilde\Lambda^j_{\mathrm{ch}} := \tilde\Lambda^j \backslash (\tilde\Lambda^j_\ell \cup \tilde\Lambda^j_{\mathrm{b}})$ is the set of all *chain elements* in $\tilde\Lambda^j$.

We denote briefly (see (3.8))

$$\|f\|_{\mathcal{B}^\alpha_\tau} := \left(\sum_{\theta\in\Theta} |b_\theta(f)|^\tau\right)^{1/\tau}. \tag{7.17}$$

Here $\tau = 1/\alpha$, $\alpha \ge 1$, and $\|f\|_{\mathcal{B}^\alpha_\tau} = \|f\|^{Q}_{\mathcal{B}^\alpha_\tau(\mathcal{M})} \approx \|f\|_{\mathcal{B}^\alpha_\tau(\mathcal{M})}$, using (3.7) and Theorem 3.4.

We first estimate $\#\tilde\Lambda^j_\ell$. By Lemma 7.5, for each $\theta \in \tilde\Lambda^j_\ell$ there is a segment $\mathcal{S}_\theta$ such that $\mathcal{S}_\theta \subset \mathrm{Star}_{N_*}(\theta) \subset \omega_\theta$ and

$$\sum_{\eta\in\mathcal{S}_\theta} |b_\eta(f)| > \varepsilon/2, \quad \text{and since } \tau \le 1, \quad (\varepsilon/2)^\tau \le \sum_{\eta\in\mathcal{S}_\theta} |b_\eta(f)|^\tau.$$

Clearly, the extended cells ω_θ associated with leaves $\theta \in \tilde\Lambda^j_\ell$ are leaves in $\mathcal{O}^j$ and hence are with disjoint interiors. As a consequence, the segments $\{\mathcal{S}_\theta\}_{\theta\in\tilde\Lambda^j_\ell}$ are disjoint. From this and (7.17),

$$\|f\|_{\mathcal{B}_\tau^\alpha}^\tau \geq \sum_{\theta\in\tilde{\Lambda}_\ell^j}\sum_{\theta\in\mathcal{S}_\theta}|b_\theta(f)|^\tau \geq (\#\tilde{\Lambda}_\ell^j)(\varepsilon/2)^\tau$$

and, therefore,

$$\#\tilde{\Lambda}_\ell^j \leq c\varepsilon^{-\tau}\|f\|_{\mathcal{B}_\tau^\alpha}^\tau \qquad \text{and} \qquad \#\tilde{\Lambda}_{\mathrm{b}}^j \leq \#\tilde{\Lambda}_\ell^j \leq c\varepsilon^{-\tau}\|f\|_{\mathcal{B}_\tau^\alpha}^\tau. \tag{7.18}$$

It remains to estimate $\#\tilde{\Lambda}_{\mathrm{ch}}^j$. To this end, we shall associate with the indices $\theta \in \tilde{\Lambda}_{\mathrm{ch}}^j$ segments $\mathcal{S}_\theta$ which essentially do not overlap and have significant ($\geq c\varepsilon^\tau$) contribution to $\|f\|_{\mathcal{B}_\tau^\alpha}^\tau$. For a given $\theta \in \tilde{\Lambda}_{\mathrm{ch}}^j$, let θ' be the only child of θ in $\tilde{\Lambda}^j$. Set $m := l(\theta)$. Two cases are to be considered here:

Case 1: $\theta' \subset \mathrm{Star}_{N_*+\nu_*}(\theta)$. Then we associate with θ the segment $\mathcal{S}_\theta := \{\theta', l(\theta'), l(\theta)\}$. By Lemma 7.6,

$$\sum_{\eta\in\mathcal{S}_\theta}|b_\theta(f)| \geq \varepsilon/\beta_2, \quad \text{and since } \tau \leq 1, \quad (\varepsilon/\beta_2)^\tau \leq \sum_{\eta\in\mathcal{S}_\theta}|b_\theta(f)|^\tau. \tag{7.19}$$

Case 2: $\theta' \not\subset \mathrm{Star}_{N_*+\nu_*}(\theta)$. Then by Lemma 7.3, there exists a segment $\mathcal{S}_\theta = \mathcal{S}(v, k_0, k_1)$ such that $\mathcal{S}_\theta \subset \mathrm{Star}_{N_*}(\theta)$, $k_0 \geq l(\theta)$, and

$$\sum_{\eta\in\mathcal{S}}|b_\eta(f)| > \varepsilon/2, \qquad \text{and hence} \qquad (\varepsilon/2)^\tau \leq \sum_{\eta\in\mathcal{S}_\theta}|b_\theta(f)|^\tau. \tag{7.20}$$

Choose a point $v \in \theta' \backslash \mathrm{Star}_{N_*+\nu_*}(\theta)$. Then, using (2.22), $\theta' \subset \omega_{\theta'} \subset \mathrm{Star}_{\nu_*}(v)$ and

$$(\mathrm{Star}_{\nu_*}(v))^\circ \cap (\mathrm{Star}_{N_*}(\theta))^\circ = \emptyset.$$

Therefore, $\mathcal{S}_\theta \subset \omega_\theta \backslash \omega_{\theta'}$.

Taking into account that $\mathcal{O}^j$ is a tree with respect to the inclusion relation, it is easy to see that the set of all segments $\mathcal{S}_\theta$, which were associated with indices $\theta \in \tilde{\Lambda}_{\mathrm{ch}}^j$ has the property that any two segments may have a common element only if one is obtained from Case 1 followed immediately by the other obtained from Case 2. Using this and (7.19)–(7.19), we infer

$$\|f\|_{\mathcal{B}_\tau^\alpha}^\tau \geq \tfrac{1}{2}\sum_{\theta\in\tilde{\Lambda}_{\mathrm{ch}}^j}\sum_{\theta\in\mathcal{S}_\theta}|b_\theta(f)|^\tau \geq \tfrac{1}{2}(\#\tilde{\Lambda}_{\mathrm{ch}}^j)(\varepsilon/2\beta_2)^\tau$$

and hence $\#\tilde{\Lambda}_{\mathrm{ch}}^j \leq c\varepsilon^\tau\|f\|_{\mathcal{B}_\tau^\alpha}^\tau$.

Combining this with (7.18), yields $\#\tilde{\Lambda}^j \leq c\varepsilon^\tau\|f\|_{\mathcal{B}_\tau^\alpha}^\tau$, which implies $N(\varepsilon) \leq c\varepsilon^\tau\|f\|_{\mathcal{B}_\tau^\alpha}^\tau$. The latter estimate, in turn, coupled with (5.18), establishes (5.20).

For the proof of (5.22), denote $\varepsilon_0 := 4\beta_2 c^\flat\|f\|_\infty$. Exactly as in the proof of Theorem 5.7 $E(f,\theta) \leq \varepsilon_0$ for each $\theta \in \Theta$ and hence $\mathcal{A}_{\varepsilon_0}(f) = 0$. Consequently, $\|f - \mathcal{A}_{\varepsilon_0}(f)\|_\infty = \|f\|_\infty$, which, coupled with the left-hand side estimate in (5.20), implies (5.22). The proof of Theorem 5.6 is complete. ∎

8. Appendix

Proof of Theorem 3.4. We shall consider only the case when $1 < p \le \infty$. The proof in the case $0 < p \le 1$ is similar to the proof of the corresponding results in [16], [23].

Evidently, if $\|f\|^{Q}_{\mathcal{B}^\alpha_\tau} < \infty$, then

$$\|f\|^{A}_{\mathcal{B}^\alpha_\tau} \le \|f\|^{Q}_{\mathcal{B}^\alpha_\tau}. \tag{A.1}$$

Our second step is to prove that if $\|f\|^{\mathcal{E}_q}_{\mathcal{B}^\alpha_\tau} < \infty$, then

$$\|f\|^{Q}_{\mathcal{B}^\alpha_\tau} \le c\|f\|^{\mathcal{E}_q}_{\mathcal{B}^\alpha_\tau}. \tag{A.2}$$

To this end, we first observe that, by (2.11) and (2.10),

$$(\|f\|^{Q}_{\mathcal{B}^\alpha_\tau})^\tau \approx \sum_{\theta\in\Theta_0} \|c_\theta(f)\varphi_\theta\|_p^\tau + \sum_{m=1}^{\infty}\sum_{I\in\mathcal{P}_m} (|I|^{-\alpha-1+1/\tau}\|q_m(f)\|_{L_1(I)})^\tau, \tag{A.3}$$

where $q_m(f)$ is defined in (2.31). By (2.8) it follows that

$$\sum_{\theta\in\Theta_0} \|c_\theta(f)\varphi_\theta\|_p^\tau \le c\|f\|_p, \qquad c = c(\#\Omega_0, \tau, p). \tag{A.4}$$

On the other hand, for $m \ge 1$, by Hölder's inequality, and Lemma 2.4,

$$\begin{aligned}
\|q_m(f)\|_{L_1(I)} &\le |I|^{1-1/q}\|q_m(f)\|_{L_q(I)} \\
&\le |I|^{1-1/q}(\|f - Q_m(f)\|_{L_q(I)} + \|f - Q_{m-1}(f)\|_{L_q(I^\star)}) \\
&\le c|I|^{1-1/q}(\mathcal{E}(f, \widehat{I})_q + \mathcal{E}(f, \widehat{I^\star})_q),
\end{aligned}$$

where $I^\star$ is the only parent of I in $\mathcal{P}_{m-1}$ ($I \subset I^\star$). Using this, and (A.4) in (A.3), we obtain (A.2).

We next prove that if $\|f\|^{A}_{\mathcal{B}^\alpha_\tau} < \infty$, then

$$\|f\|^{\mathcal{E}_q}_{\mathcal{B}^\alpha_\tau} \le c\|f\|^{A}_{\mathcal{B}^\alpha_\tau}, \qquad 1 \le q < p. \tag{A.5}$$

By Hölder's inequality, $\|f\|^{\mathcal{E}_q}_{\mathcal{B}^\alpha_\tau} \le c\|f\|^{\mathcal{E}_\tau}_{\mathcal{B}^\alpha_\tau}$ if $1 \le q \le \tau$. So, it suffices to prove (A.5), only when $\max\{1, \tau\} < q < p$. By Theorem 3.2, $\|f\|_p \le c\|f\|^{A}_{\mathcal{B}^\alpha_\tau}$. Since $1/\tau := \alpha + 1/p$ we have, by (3.2),

$$|f|^{\mathcal{E}_q}_{\mathcal{B}^\alpha_\tau} := \left(\sum_{I\in\mathcal{P}} |I|^{\tau(1/p-1/q)}\mathcal{E}(f, \widehat{I})_q^\tau\right)^{1/\tau}. \tag{A.6}$$

Evidently, $\mathcal{E}(g, \widehat{I})_q = 0$ for $I \in \mathcal{P}_m$ if $g \in V_m$, and $\mathcal{E}(f, \widehat{I})_q \le \|f\|_{L_q(\widehat{I})}$ if $f \in L_q$. For $I \in \mathcal{P}_m$, denote $\widetilde{I} := \bigcup\{\theta \in \Theta_m : \theta^\circ \cap \widehat{I}^\circ \ne \emptyset\}$. Let $f = \sum_{\theta\in\Theta} a_\theta\varphi_\theta$ be any representation of f in L_1 (and hence in L_p) such that $(\sum_{\theta\in\Theta}\|a_\theta\varphi_\theta\|_p^\tau)^{1/\tau} \le c\|f\|^A_{\mathcal{B}^\alpha_\tau}$. Then using the above, Theorem 3.2, and (2.11), we obtain

$$\begin{aligned}\mathcal{E}(f,\widehat{I})_q^\tau &\le \left\|\sum_{j=m+1}^\infty\sum_{\theta\in\Theta_j} a_\theta\varphi_\theta\right\|^\tau_{L_q(\widehat{I})} \le c\sum_{j=m+1}^\infty\sum_{\theta\in\Theta_j,\,\theta\subset\widetilde{I}}\|a_\theta\varphi_\theta\|_q^\tau\\ &\le c\sum_{j=m+1}^\infty\sum_{\theta\in\Theta_j,\,\theta\subset\widetilde{I}}|\theta|^{\tau(1/q-1/p)}\|a_\theta\varphi_\theta\|_p^\tau.\end{aligned}$$

Substituting this in (A.6) gives

$$\begin{aligned}(|f|^{\mathcal{E}_q}_{\mathcal{B}^\alpha_\tau})^\tau &\le c\sum_{I\in\mathcal{P}}|I|^{\tau(1/p-1/q)}\sum_{\theta\in\Theta,\,\theta\subset\widetilde{I}}|\theta|^{\tau(1/q-1/p)}\|a_\theta\varphi_\theta\|_p^\tau\\ &\le c\sum_{\theta\in\Theta}\|a_\theta\varphi_\theta\|_p^\tau\sum_{I\in\mathcal{P}:\theta\subset\widetilde{I}}(|\theta|/|I|)^{\tau(1/q-1/p)},\end{aligned}$$

where we have switched the order of summation once. By the properties of cells and supports of bases functions, $\#\{I\in\mathcal{P}_\nu : \theta\subset\widetilde{I}\}\le c<\infty$ and $|\theta|\le c\rho^\nu|I|$ if $\theta\subset\widetilde{I}$, $\theta\in\Theta_j$, and $I\in\mathcal{P}_{j-\nu}$. Using this and that $1/q-1/p>0$, we obtain

$$\sum_{I\in\mathcal{P}:\theta\subset\widetilde{I}}(|\theta|/|I|)^{\tau(1/q-1/p)}\le c\sum_{j=0}^\infty\rho^{\tau(1/q-1/p)}\le c<\infty.$$

Therefore, $(|f|^{\mathcal{E}_q}_{\mathcal{B}^\alpha_\tau})^\tau\le c\sum_{\theta\in\Theta}\|a_\theta\varphi_\theta\|_p^\tau$, which completes the proof of (A.5).

In view of (3.8) and (3.9), the equivalence of $\|\cdot\|^W_{\mathcal{B}^\alpha_\tau}$ and $\|\cdot\|_{\mathcal{B}^\alpha_\tau}$ is an immediate consequence of the relations

$$|b_\theta(f)|\le C\max\{|c_\lambda(f)| : \lambda^\circ\cap\theta\ne\emptyset\},\qquad |c_\lambda(f)|\le C\max\{|b_\theta(f)| : \theta^\circ\cap\lambda^\circ\ne\emptyset\},$$

which follow by taking scalar products of both sides of the relation

$$\sum_{\theta\in\Theta_{m+1}}b_\theta(f)\varphi_\theta=\sum_{\lambda\in\mathcal{L}_m}c_\lambda(f)\psi_\lambda$$

with the dual functions $\tilde{\psi}_{\lambda'}$ or applying the dual functionals $c_{\theta'}$. ■

Proof of Theorem 4.2 (Bernstein Estimate). We shall give the proof of estimate (4.2) only in the case $p=\infty$. We shall utilize the idea of the proof of the Bernstein estimates in [16], [24], where the case of piecewise polynomials is treated. The proof in the case $p<\infty$ can be carried out in a similar way (see the proofs of the Bernstein estimates in [16], [23]) and will be omitted.

Suppose $g\in\Sigma_n$ and $g=:\sum_{\theta\in\Lambda}a_\theta\varphi_\theta$, where $\Lambda\subset\Theta$ and $\#\Lambda\le n$. Let $\mathcal{K}_0$ be the set of all cells in $\mathcal{P}$ which are involved (covered) in all sets $\theta\in\Lambda$. Then $g=\sum_{I\in\mathcal{K}_0}g_I$, where $g_I=:\mathbf{1}_I\cdot v_I$, $v_I\in V_m$ with $m:=\text{level}(I)$. Evidently, $\#\mathcal{K}_0\le\nu_1\#\Lambda\le cn$.

The proof of (4.2) hinges on the tree structure in $\mathcal{P}$ induced by the inclusion relation: Each $I\in\mathcal{P}_m$ has at most ν_0 children in $\mathcal{P}_{m+1}$ and one parent in $\mathcal{P}_{m-1}$, if $m\ge 1$. We denote by $\mathcal{T}_0$ the set of all cells $I \in \mathcal{P}$ for which there exists $J \in \mathcal{K}_0$ such that $J \subset I$, which is the minimal subtree of $\mathcal{P}$ containing $\mathcal{K}_0$ with its root(s) in $\mathcal{P}_0$. We denote by $\mathcal{T}_{\rm b}$ the set of all *branching* cells in $\mathcal{T}_0$ (cells in $\mathcal{T}_0$ with at least two children in $\mathcal{T}_0$) and by $\mathcal{T}_{\rm b}^+$ the set of all *children of branching* cells in $\mathcal{P}$ (which may or may not belong to $\mathcal{T}_0$). We define $\mathcal{T} := \mathcal{T}_0 \cup \mathcal{T}_{\rm b}^+$, which is again a subtree of $\mathcal{P}$.

We next introduce several subsets of $\mathcal{T}$ which will be needed later on. We denote by $\mathcal{T}_\ell$ the set of all *leaves* in $\mathcal{T}$ ($I \in \mathcal{T}_\ell$ if I does not contain any other $J \in \mathcal{T}$) and $\mathcal{T}_{\rm ch} := \mathcal{T}\backslash(\mathcal{K}_0 \cup \mathcal{T}_{\rm b} \cup \mathcal{T}_{\rm b}^+ \cup \mathcal{T}_\ell)$ the set of all *chain cells* in $\mathcal{T}$. (Notice that each $I \in \mathcal{T}_{\rm ch}$ has only one child which belongs to $\mathcal{T}$.)

Let μ be the smallest positive integer such that $\rho^\mu \le \delta$, where $0 < \delta < 1$ is the constant from (2.14). For each $I \in \mathcal{T}_{\rm ch}$, we denote by $I^\diamond$ the unique cell $I^\diamond \in \mathcal{K}_0 \cup \mathcal{T}_{\rm b} \cup \mathcal{T}_\ell$ such that $I^\diamond \subset I$ and $I^\diamond$ is of the least possible level. Let $\mathcal{T}_{\rm ch}^1$ be the set of all $I \in \mathcal{T}_{\rm ch}$ such that $\mathrm{level}(I^\diamond) - \mathrm{level}(I) \le \mu$, and $\mathcal{T}_{\rm ch}^0 := \mathcal{T}_{\rm ch}\backslash\mathcal{T}_{\rm ch}^1$.

Clearly, $\#\mathcal{T}_{\rm b} \le \#(\mathcal{T}_0)_\ell \le \#\mathcal{K}_0 \le cn$, which implies $\#\mathcal{T}_{\rm b}^+ \le M_0\#\mathcal{T}_{\rm b} \le cn$, $\#\mathcal{T}_\ell \le \#\mathcal{K}_0 + \#\mathcal{T}_{\rm b}^+ \le cn$, and $\#\mathcal{T}_{\rm ch}^1 \le \mu\#(\mathcal{K}_0 \cup \mathcal{T}_{\rm b} \cup \mathcal{T}_\ell) \le cn$. Notice that $\#\mathcal{T}_{\rm ch}^0$ can be huge.

We now extend $\mathcal{K}_0$ to $\mathcal{K} := \mathcal{K}_0 \cup \mathcal{T}_{\rm b} \cup \mathcal{T}_{\rm b}^+ \cup \mathcal{T}_{\rm ch}^1$. From above $\#\mathcal{K} \le \#\mathcal{K}_0 + \#\mathcal{T}_{\rm b} + \#\mathcal{T}_{\rm b}^+ + \#\mathcal{T}_{\rm ch}^1 \le cn$. Evidently, g can be represented in the form $g = \sum_{I\in\mathcal{K}} g_I$ with g_I similar to the g_I's from above.

After this ground work, we next estimate $|g|_{B_\tau^\alpha}^\tau := \sum_{I\in\mathcal{P}} |I|^{-\tau}\mathcal{E}(g, \widehat{I})_1^\tau$, where $\tau := 1/\alpha$ (see (3.1)) and $\widehat{I}$ is defined in (2.26). We denote

$$g_m := \sum_{\theta\in\Lambda,\ \mathrm{level}(\theta)\le m} a_\theta\varphi_\theta, \qquad m \ge 0.$$

A key fact is that

$$\mathcal{E}(g, \widehat{I})_1 = \mathcal{E}(g - g_m, \widehat{I})_1 \le \|g - g_m\|_{L_1(\widehat{I})}, \qquad I \in \mathcal{P}_m. \tag{A.7}$$

We also have $\mathcal{E}(g, \widehat{I})_1 \le \|g\|_{L_1(\widehat{I})}$.

Let $\mathcal{L} := \{I \in \mathcal{P}_m : I \subset \widehat{J} \quad \text{for some} \quad J \in \mathcal{K} \cap \mathcal{P}_m\}$ and let $\mathcal{L} := \bigcup_{m\ge0} \mathcal{L}_m$. Evidently, $\#\mathcal{L} \le \nu_1\#\mathcal{K} \le cn$.

We shall split up the sum in the definition of $|g|_{B_\tau^\alpha}^\tau$ above into two sums: over $I \in \mathcal{L}$ and over $I \in \mathcal{P}\backslash\mathcal{L}$.

(a) If $I \in \mathcal{L}_m$, then there is $J \in \mathcal{K} \cap \mathcal{P}_m$ such that $I \subset \widehat{J}$ and (see (2.26)) $|I|^{-\tau}\mathcal{E}(g, \widehat{I})_1^\tau \le c|I|^{-\tau}\|g\|_{L_1(\widehat{I})}^\tau \le c\|g\|_\infty^\tau$. Therefore, we have

$$\begin{aligned}\sum_{I\in\mathcal{L}} |I|^{-\tau}\mathcal{E}(g, \widehat{I})_1^\tau &\le \sum_{m\ge0}\sum_{I\in\mathcal{L}_m} \mathcal{E}(g, \widehat{I})_1^\tau \le c\|g\|_\infty^\tau \sum_{m\ge0} \#\mathcal{L}_m \\ &= c\|g\|_\infty^\tau \#\mathcal{L} \le cn\|g\|_\infty^\tau.\end{aligned} \tag{A.8}$$

(b) Let $I \in \mathcal{P}_m\backslash\mathcal{L}_m$. Then $\widehat{I} = \bigcup_{i=1}^{\nu_I} J_i$ for some $J_i \in (\mathcal{T}_{\rm ch}^0 \cap \mathcal{P}_m) \cup (\mathcal{P}_m\backslash\mathcal{T})$, where $\nu_I \le \nu_1$ (see Section 2.1). We have, by (A.7),

$$\mathcal{E}(g, \widehat{I})_1^\tau \le \sum_{i=1}^{\nu_I} \|g - g_m\|_{L_1(J_i)}^\tau \qquad (\tau \le 1).$$

Clearly, if $J_i \in \mathcal{P}_m\backslash\mathcal{T}$, then $g|_{J_i} = g_m|_{J_i}$ and hence $\|g - g_m\|_{L_1(J_i)} = 0$. Suppose $J_i \in \mathcal{T}_{\rm ch}^0 \cap \mathcal{P}_m$ and let $J_i^\diamond$ be the unique largest element of $\mathcal{K}$ contained in J_i (see the definition of $\mathcal{T}_{\text{ch}}^0$ above). We have $g|_{J_i \setminus J_i^\diamond} = g_m|_{J_i \setminus J_i^\diamond} = \mathbf{1}_{J_i \setminus J_i^\diamond} \cdot v_i$ for some $v_i \in V_m$. On the other hand, $\text{level}(J_i^\diamond) - \text{level}(J_i) > \mu$ and hence $|J_i^\diamond| \le \rho^\mu |J_i| \le \delta |J_i|$. Therefore, using (2.14),

$$\|v_i\|_{L_\infty(J_i^\diamond)} \le \|v_i\|_{L_\infty(J_i)} \le \|v_i\|_{L_\infty(J_i \setminus J_i^\diamond)} \le c\|g\|_\infty.$$

We use the above to obtain

$$\|g - g_m\|_{L_1(J_i)} = \|g - g_m\|_{L_1(J_i^\diamond)} \le |J_i^\diamond|(\|g\|_\infty + \|v_i\|_{L_\infty(J_i^\diamond)}) \le c|J_i^\diamond| \|g\|_\infty.$$

Therefore,

$$|I|^{-\tau} \mathcal{E}(g, \widehat{I})_1^\tau \le c\|g\|_\infty \sum_{1 \le i \le v_I,\, J_i \in \mathcal{T}_{\text{ch}}^0 \cap \mathcal{P}_m} (|J_i^\diamond|/|J_i|)^\tau$$

and, hence,

$$\sum_{I \in \mathcal{P}_m \setminus \mathcal{L}_m} |I|^{-\tau} \mathcal{E}(g, \widehat{I})_1^\tau \le c\|g\|_\infty^\tau \sum_{J \in \mathcal{T}_{\text{ch}}^0 \cap \mathcal{P}_m} (|J^\diamond|/|J|)^\tau.$$

Summing over $m \ge 0$, we obtain

$$\sum_{I \in \mathcal{P} \setminus \mathcal{L}} |I|^{-\tau} \mathcal{E}(g, \widehat{I})_1^\tau \;\le\; c\|g\|_\infty^\tau \sum_{J \in \mathcal{T}_{\text{ch}}^0} (|J^\diamond|/|J|)^\tau \le c\|g\|_\infty^\tau \sum_{J' \in \mathcal{K}} \sum_{J \in \mathcal{P}:\, J' \subset J} (|J'|/|J|)^\tau$$

$$\le\; c\|g\|_\infty^\tau \sum_{J' \in \mathcal{K}} \sum_{m=0}^\infty \rho^{\tau m} \le c\|g\|_\infty^\tau \# \mathcal{K} \le cn\|g\|_\infty^\tau,$$

where we used (2.2). The above estimates and (A.8) imply (4.2). ■

Acknowledgments. This work has been supported in part by the European Community's Human Potential Programme under contract HPRN-CT-202-00286 (BREAKING COMPLEXITY), and by the National Science Foundation grant DMS-0200665.

References

1. J. BERGH, J. LÖFSTRÖM (1976): Interpolation Spaces: An Introduction. Grundlehren der Mathematischen Wissenschaften, Vol. 223. Berlin: Springer-Verlag.
2. YU. BRUDNYI (2002): *Nonlinear approximation by refinable functions and piecewise polynomials*. In: Approximation Theory X (C. Chui, L. Schumaker, J. Stoeckler, eds.). Nashville, TN: Vanderbilt University Press, pp. 91–112.
3. A. CANUTO, A. TABACCO, K. URBAN (1999): *The wavelet element method*. I. *Construction and analysis*. Appl. Comput. Harmon. Anal., **6**:1–52.
4. A. CANUTO, A. TABACCO, K. URBAN (2000): *The wavelet element method*. II. *Realization and additional features in* 2*D and* 3*D*. Appl. Comput. Harmon. Anal., **8**:123–165.
5. A. COHEN (2003): Numerical Analysis of Wavelet Methods. Studies in Mathematics and its Applications. Amsterdam: Elsevier.
6. A. COHEN, R. MASSON (1997): *Wavelet adaptive methods for second order elliptic problems, boundary conditions and domain decomposition*. Numer. Math., **8**:21–47.
7. A. COHEN, I. DAUBECHIES, J.-C. FEAUVEAU (1992): *Biorthogonal bases of compactly supported wavelets*. Comm. Pure Appl. Math., **45**:485–560.
8. W. DAHMEN (1997): Wavelet and Multiscale Methods for Operator Equations. Acta Numerica, Vol. 6. Cambridge: Cambridge University Press, pp. 55–228.

9. W. DAHMEN, A. KUNOTH, K. URBAN (1999): *Biorthogonal spline-wavelets on the interval—Stability and moment conditions*. Appl. Comput. Harmon. Anal., **6**:132–196.
10. W. DAHMEN, C. A. MICCHELLI (1985): *On the local linear independence of translates of a box spline*. Studia Math., **82**:243–263.
11. W. DAHMEN, R. SCHNEIDER (1999): *Composite wavelet bases for operator equations*. Math. Comp., **68**:1533–1567.
12. W. DAHMEN, R. SCHNEIDER (1999): *Wavelets on manifolds* I: *Construction and domain decomposition*. SIAM J. Math. Anal., **31**:184–230.
13. W. DAHMEN, R. STEVENSON (1999): *Element-by-element construction of wavelets—Stability and moment conditions*. SIAM J. Numer. Anal., **37**:319–325.
14. I. DAUBECHIES (1992): Ten Lectures on Wavelets. CBMS-NSF Regional Conference Series in Applied Math., Vol. 61. Philadelphia: SIAM.
15. O. DAVYDOV (2001): *Stable local bases for multivariate spline spaces*. J. Approx. Theory, **111**:267–297.
16. O. DAVYDOV, P. PETRUSHEV (2004): *Nonlinear approximation from differentiable piecewise polynomials*. SIAM J. Math. Anal., **35**:708–758.
17. R. DEVORE (1998): Nonlinear Approximation. Acta Numerica, Vol. 7. Cambridge: Cambridge University Press, pp. 51–150.
18. R. DEVORE, B. JAWERTH, B. LUCIER (1992): *Surface compression*. Comput. Aided Geom. Design, **9**:219–239.
19. R. DEVORE, G. G. LORENTZ (1993): Constructive Approximation. Grundlehren der Mathematischen Wissenschaften, Vol. 303. Heidelberg: Springer-Verlag.
20. R. DEVORE, P. PETRUSHEV, X. YU (1992): *Nonlinear wavelet approximation in the space* $C(R^d)$. In: Progress in Approximation Theory (A. A. Gonchar, E. B. Saff, eds.). New York: Springer-Verlag, pp. 261–283.
21. R. DEVORE, V. POPOV (1988): *Interpolation of Besov spaces*. Trans. Amer. Math. Soc., **305**:397–414.
22. R. DEVORE, V. POPOV (1988): *Interpolation spaces and non-linear approximation*. In: Function Spaces and Applications (M. Cwikel, J. Peetre, Y. Sagher, H. Wallin, eds.). Lecture Notes in Mathematics, Vol. 1302. Berlin: Springer-Verlag, pp. 191–205.
23. B. KARAIVANOV, P. PETRUSHEV (2003): *Nonlinear piecewise polynomial approximation beyond Besov spaces*. Appl. Comput. Harmon. Anal., **15**:177–223.
24. B. KARAIVANOV, P. PETRUSHEV, R. C. SHARPLEY (2003): *Algorithms for nonlinear piecewise polynomial approximation*. Trans. Amer. Math. Soc., **355**:2585–2631.
25. P. OSWALD (1994): Multilevel Finite Element Approximation. Stuttgart: Teubner.
26. P. PETRUSHEV (2003): *Multivariate n-term rational and piecewise polynomial approximation*. J. Approx. Theory, **121**:158–197.
27. P. PETRUSHEV (2003): *Nonlinear n-term approximation from hierarchical spline bases*. In: Constructive Theory of Functions, Varna 2002 (B. Bojanov, ed.). Sofia: DARBA, pp. 33–85.
28. P. PETRUSHEV, V. POPOV (1987): Rational Approximation of Real Functions. Cambridge: Cambridge University Press.
29. V. N. TEMLYAKOV (2003): *Nonlinear methods of approximation*. Found. Comput. Math., **3**:33–107.

W. Dahmen
Institut für Geometrie und Praktische Mathematik
RWTH Aachen
Templergraben 55
52056 Aachen
Germany
dahmen@igpm.rwth-aachen.de
http://www.igpm.rwth-aachen.de/∼dahmen/

P. Petrushev
Department of Mathematics
University of South Carolina
Columbia, SC 29208
USA
pencho@math.sc.edu
http://www.math.sc.edu/∼pencho/