CONSTRUCTIVE THEORY OF FUNCTIONS, Varna 2002
(B. Bojanov, Ed.), DARBA, Sofia, 2003, pp. 33-85.

Nonlinear n-term Approximation from Hierarchical Spline Bases

Pencho Petrushev *

This article is a survey of some recent developments which concern two multilevel approximation schemes: (a) Nonlinear n-term approximation from piecewise polynomials generated by anisotropic dyadic partitions in $\mathbb{R}^d$, and (b) Nonlinear n-term approximation from sequences of hierarchical spline bases generated by multilevel triangulations in $\mathbb{R}^2$. A construction is given of sequences of bases consisting of differentiable (in C^r with $r \geq 1$) piecewise polynomials (splines) over multilevel triangulations, which allow triangles with arbitrarily sharp angles. Both schemes are based on multiscale decompositions that are defined through multilevel nested partitions and their common features are best captured by the term "multiresolution", which also relates them to wavelets. In contrast to the wavelet case, these are highly nonlinear approximation methods from redundant systems with a great deal of flexibility. It is shown that the rates of nonlinear n-term spline approximation, when using the above schemes with an arbitrary but fixed multilevel partition or triangualtion, are governed by certain smoothness spaces, called B-spaces. Unlike the commonly used Besov spaces, the B-spaces allow to characterize all rates of approximation, which gives more complete results in the isotropic case as well. An effective algorithm is described for finding, for a given function, an anisotropic dyadic partition which minimizes the corresponding B-norm of the function and thus provides an optimal rate of highly nonlinear approximation using the first scheme above. Scalable algorithms are given for nonlinear approximation which both capture the rate of the best approximation and provide the basis for numerical implementation.

1. Introduction

Nonlinear approximation of functions in dimensions $d > 1$ is a challenging area, specially when one moves away from regular (i.e., tensor product type) schemes in order to more adequately approximate functions with singularities along curves and with other anisotropies. One of the most natural tools

---

*This research was supported by the National Science Foundation Grant DMS-0200665.

for approximation is piecewise polynomials over triangulations, and a fundamental problem is to characterize the rate of nonlinear approximation in L_p $(0 < p \le \infty)$ in terms of properly defined global smoothness conditions. This problem is disheartening if one allows the nonlinear approximation to be from any piecewise polynomial over an arbitrary triangulation. The difficulty stems from the highly nonlinear nature of piecewise polynomials in dimensions $d > 1$. For instance, if s_1 and s_2 are two piecewise polynomials over n triangles in $\mathbb{R}^2$ each, then $s_1 + s_2$ is in general a piecewise polynomial over many more than n (even $> n^2$) pieces. Therefore, the well-known Jackson-Bernstein estimate machinery is not applicable in this case. In addition, if we are interested in differentiable piecewise polynomials (splines), which is an important requirement for surface modeling, for example, then the problem becomes much harder. Indeed, even for a general triangulation of a polygonal domain in $\mathbb{R}^2$ consisting of a finite number of triangles, the corresponding space of all piecewise polynomials of degree $< k$ and smoothness $r \ge 1$ can be very complicated. For example, the dimension of this space is not known and stable local bases are impossible, in general, if $k \le 3r + 2$ [4].

A reasonable alternative to "spline approximation with free triangulations" is "nonlinear n-term approximation from hierarchical spline bases, associated with mutiresolution structures". These are highly nonlinear approximation methods from large redundant systems which allow a great deal of flexibility. This article will focus on two multilevel approximation schemes:

(a) Nonlinear n-term approximation from piecewise polynomials generated by dyadic partitions in $\mathbb{R}^d$.

(b) Nonlinear n-term approximation from hierarchical spline bases associated with spline multiresolution over multilevel nested triangulations in $\mathbb{R}^2$. We next explain these two classes of highly nonlinear approximation methods linked by the concept of "multiresolution".

First, let $(\mathcal{P}_m)_{m\in\mathbb{Z}}$ be an arbitrary sequence of dyadic partitions of $\mathbb{R}^d$ $(d > 1)$ such that each level $\mathcal{P}_m$ is a partition of $\mathbb{R}^d$ into disjoint dyadic boxes I of volume $|I| = 2^{-m}$, $(\mathcal{P}_m)$ are nested, and the boxes in $\mathcal{P}$ form a single tree with respect to the inclusion relation. Set $\mathcal{P} = \bigcup_{m\in\mathbb{Z}} \mathcal{P}_m$. Evidently, each dydic box in $\mathcal{P}_m$ can be subdivided in d $(d > 1)$ different ways (with children in $\mathcal{P}_{m+1}$) and hence there is a huge variety of anisotropic dyadic partitions. We denote by $\mathcal{S}^{k,-1}(\mathcal{P}_m)$ the set of all (discontinuous) piecewise polynomials of degree $< k$ over the boxes of $\mathcal{P}_m$. Evidently, we have

$$\cdots \subset \mathcal{S}^{k,-1}(\mathcal{P}_{-1}) \subset \mathcal{S}^{k,-1}(\mathcal{P}_0) \subset \mathcal{S}^{k,-1}(\mathcal{P}_1) \subset \cdots . \tag{1.1}$$

Consider now the problem for nonlinear (n-term) approximation in L_p from the set $\Sigma_n^k(\mathcal{P})$ of all piecewise polynomials of the form $s = \sum_{I\in\Lambda_n} \mathbb{1}_I \cdot P_I$, where $\mathbb{1}_I$ is the characteristic function of I, P_I is a polynomial of degree $< k$, $\Lambda_n \subset \mathcal{P}$ with cardinality $\#\Lambda_n \le n$, and Λ_n may very. When the partition $\mathcal{P}$ is allowed to vary, the approximation from $(\Sigma_n^k(\mathcal{P}))$ is highly nonlinear with a lot of flexibility.

Of course, much more flexibility is achieved in the second approximation scheme, where the dyadic partitions are replaced by multilevel nested triangulations. Consider a sequence $(\mathcal{T}_m)_{m\in\mathbb{Z}}$ of partitions of $\mathbb{R}^2$ (here $d = 2$ for simplicity only) into triangles with disjoint interiors such that each level $\mathcal{T}_m$ is a refinement of the previous level $\mathcal{T}_{m-1}$. Let $\mathcal{T} := \bigcup_{m\in\mathbb{Z}} \mathcal{T}_m$. We assume that the partitions $(\mathcal{T}_m)_{m\in\mathbb{Z}}$ satisfy certain natural mild conditions which prevent them from deterioration, but still allow the triangles to change in size, shape, and orientation quickly when moving around at a given level or through the levels. In particular, triangles with arbitrarily sharp angles are allowed in any location and at any level. We denote by $\mathcal{S}^{k,r}(\mathcal{T}_m)$ the set of all r-times differentiable piecewise polynomials of degree $< k$ over the triangles of $\mathcal{T}_m$. Assume that there exists a **multiresolution** consisting of a ladder of spaces

$$\cdots \subset \mathcal{S}_{-1} \subset \mathcal{S}_0 \subset \mathcal{S}_1 \subset \cdots, \qquad \mathcal{S}_m \subset \mathcal{S}^{k,r}(\mathcal{T}_m) \tag{1.2}$$

with bases Φ_m of $\mathcal{S}_m$, $m \in \mathbb{Z}$. Set $\Phi_{\mathcal{T}} := \bigcup_{m\in\mathbb{Z}} \Phi_m$.

This structure is quite similar to a wavelet multiresolution analysis with $\Phi_{\mathcal{T}}$ playing the role of the set of wavelet scaling functions. It lacks the explicit orthogonal (or biorthogonal) structure of wavelets but has much more flexibility.

We now consider the problem for nonlinear (n-term) approximation from the set $\Sigma_n(\Phi_{\mathcal{T}})$ of all piecewise polynomials of the form $s = \sum_{j=1}^{n} c_j\varphi_j$, where $\varphi_j \in \Phi_{\mathcal{T}}$ may come from arbitrary levels and locations.

For both approximation schemes, the first primary goal is to characterize the rates of approximation in L_p from $(\Sigma_n^k(\mathcal{P}))$ or $(\Sigma_n(\Phi_{\mathcal{T}}))$ (with $\mathcal{P}$ or $\mathcal{T}$ fixed) in terms of certain global smoothness conditions of the function f which is being approximated. Secondly, in both cases, it is highly desirable to construct algorithms which capture the rate of the best n-term approximation. Thirdly, it is natural to add another degree of nonlinearity by allowing $\mathcal{P}$ or $\mathcal{T}$ to vary with f. Then an important problem is to find , for a given function f, an optimal partition or triangulation in which f exhibits the most smoothness. Of course, the ultimate problem here is to characterize the rates of nonlinear n-term approximation, when the partition or triangulation is allowed to very. Evidently, these goals are easier to achieve when developing the approximation theory of the first scheme above (approximation from piecewise polynomials over dyadic partitions). So, it is natural that every advancement in the theory of the second approximation method be preceded by a similar result in the first theory.

We next give in more specific terms the program that will lead us in developing the theory and algorithms of nonlinear spline approximation in this article:

(i) Construct hierarchical sequences of bases $(\Phi_m)_{m\in\mathbb{Z}}$ on multilevel triangulations satisfying certain natural requirements of local regularity but allowing triangles with arbitrarily sharp angles. For the first approximation method, there is no alternative but to approximate from discontinuous piecewise polynomials.

(ii) To quantify the approximation process, introduce and develop collections of smoothness spaces (B-spaces) depending on $\Phi_{\mathcal{T}}$ or $\mathcal{P}$. So, the idea is to measure the smoothness of the functions using in each case a family of smoothness space scales which vary with $\Phi_{\mathcal{T}}$ or $\mathcal{P}$, instead of a single space scale like the scale of Besov spaces.

(iii) Develop a coherent theory of nonlinear n-term approximation from $\Phi_{\mathcal{T}}$ or from piecewise polynomials over an arbitrary (fixed) dyiadic partition based on the idea of proving Jackson and Bernstein estimates using B-spaces and interpolation.

(iv) Utilize this theory in the development of algorithms for nonlinear piecewise polynomial (spline) approximation which capture the rate of the best approximation.

(v) Develop an algorithm which, for a given function f, finds an optimal partition or triangulation in which f exhibits the most smoothness with respect to the corresponding B-spaces. Characterize the nonlinear n-term approximation from any collection $\{\Phi_{\mathcal{T}}\}_{\mathcal{T}}$ ($\mathcal{T}$ is allowed to very with f) and similarly for n-term approximation from piecewise polynomials over anisotropic dyadic partitions.

The above program was suggested in [21, 28] and implemented in [11, 21, 22, 28]. The first step in the program was developed in [11], where the construction from [10] was adapted. The idea for constructing B-spaces and applying them to nonlinear approximation was first utilized in [28] in the case of nonlinear approximation from piecewise polynomials over anisotropic dyadic partitions of $\mathbb{R}^d$. This idea was further implemented in [21] and fully developed in [11]. Note that the B-spaces can be viewed as a generalization of the approximation spaces considered in §3.4 of [25]. More precisely, in the specific setting of [25], the approximation spaces there are B-spaces. The characterization of nonlinear n-term approximation from piecewise polynomials over a fixed anisotropic dyadic partition $\mathcal{P}$ was established in [28], and from a single family of basis functions Φ in [11, 21]. In [22], three algorithms for nonlinear n-term approximation in L_p ($0 < p \leq \infty$) from Courant bases are developed, which both capture the rate of the best approximation and provide the basis for numerical implementation. These algorithms can be immediately implemented for nonlinear n-term approximation from differentiable spline bases as well. Naturally, the last step in the above program presents the most challenging problems. The problem for finding, for a given function f, an optimal dyadic partition in which f exhibits the most smoothness has a complete and efficient solution (see [28]) and it remains open in the case of approximation from piecewise polynomials over multilevel triangulations. The more delicate problem for characterization of highly nonlinear n-term approximation from collections of basis families, i.e., when the partition or triangualtion may very, is open as well.

The theory of nonlinear n-term approximation from box splines (uniform triangulations) has been developed in [15] ($p < \infty$) and [18] ($p = \infty$) (see also [6, 7] and the references therein, and [23]). In these articles, the rates of nonlinear spline approximation are characterized by using certain Besov spaces.

However, even in this isotropic case, the results which utilize B-spaces (in place of Besov spaces) are more complete since (unlike Besov spaces) they characterize nonlinear n-term box spline approximation for all rates of approximation.

The situation in the univariate nonlinear piecewise polynomial approximation is quite unique, since the scale of Besov spaces $B_\tau^{\alpha,k}(L_\tau)$ $(1/\tau = \alpha + 1/p)$ governs all rates of approximation in this case (see [27]).

Nonlinear spline approximation in the norms of Besov spaces has been considered in [24]. In the present article, we do not go beyond approximation in L_p, $0 < p \le \infty$.

There is an apparent connection between the developments presented in this article and multilevel finite element methods for PDEs, see, e.g., [25]. It seems reasonable to develop finite element algorithms for solving PDEs which achieve the rate of the best nonlinear n-term spline approximation to the solution.

The idea of using B-spaces is not limited only to nonlinear piecewise polynomial approximation. It can be implemented immediately to nonlinear n-term approximation from refinable functions, in general, and to wavelet approximation, in particular. This is, however, beyond the scope of this article.

There is a strong relationship between rational and nonlinear spline approximation which we shall not bring forward here (see, e.g., [28, 29]).

The outline of the article is the following. In §2, we explain some basic principles of nonlinear n-term approximation. In §3, we introduce and discuss the B-spaces needed for the characterization of nonlinear n-term approximation from families of basis functions over multilevel nested tiangulations in $\mathbb{R}^2$. In §4, we introduce the B-spaces over dyadic partitions in $\mathbb{R}^d$. In §5, we clarify the relations between B-spaces and Besov spaces in the univariate case. In §6, we consider nonlinear approximation from piecewise polynomials generated by dyadic partitions of $\mathbb{R}^d$. In §7, we present the theory of nonlinear n-term approximation from piecewise polynomials over multilevel nested tiangulations in $\mathbb{R}^2$. In §8, we present the basic results of nonlinear spline approximation in dimension $d = 1$. In §9, we consider algorithms for nonlinear n-term spline approximation. In §10, we give some concluding remarks and open problems.

Throughout this article, the positive constants are denoted by $c, c_1, \dots$ and they may vary at every occurrence, $A \approx B$ means $c_1 A \le B \le c_2 A$, and $A := B$ or $B =: A$ stands for "A is by definition equal to B"; Π_k denotes the set of all algebraic polynomials in d variables of total degree $< k$ (usually $d = 2$). For a set $E \subset \mathbb{R}^d$, $\mathbb{1}_E$ denotes the characteristic function of E, and $|E|$ denotes the Lebesgue measure of E. Since we systematically work with quasi-normed spaces such as L_p, $0 < p < 1$, "norm" will stand for "norm" or "quasi-norm".

2. Nonlinear n-term Approximation: The Principles

In this section, we give some of the general principles of the theory of nonlinear n-term approximation which will guide us throughout this article.

Let X be a normed or quasi-normed function space, where the approximation takes place (in this article, $X = L_p(E)$, $0 < p \le \infty$). Suppose $\Phi = \{\varphi_\theta\}_{\theta\in\Theta}$ is a collection of elements in X which is, in general, redundant, and we are interested in nonlinear n-term approximation from Φ, which we describe in what follows. Let Σ_n denote the nonlinear set of all function s of the form

$$s = \sum_{\theta\in\Lambda_n} a_\theta \varphi_\theta,$$

where $\Lambda_n \subset \Theta$, $\#\Lambda_n \le n$, and Λ_n may vary with s. The error of n-term approximation to $f \in X$ from Φ is defined by

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

Our primary goal is to describe all rates of nonlinear n-term approximation from Φ. More precisely, we want to characterize the approximation space generated by nonlinear n-term approximation from Φ.

Approximation spaces. We define the approximation space $A_q^\gamma := A_q^\gamma(\Phi, X)$, $\alpha > 0$, $0 < q \le \infty$, as the set of all functions $f \in X$ such that

$$\|f\|_{A_q^\gamma} := \|f\|_X + \left(\sum_{n=1}^{\infty} (n^\gamma \sigma_n(f))^q \frac{1}{n}\right)^{1/q} < \infty \tag{2.1}$$

with the ℓ_q-norm replaced by the sup-norm if $q = \infty$. Thus A_∞^γ is the set of all $f \in X$ such that $\sigma_n(f) = O(n^{-\gamma})$.

An important feature of nonlinear n-term approximation is that (no matter what the family Φ is) there is no saturation, which means that the approximation space $A_q^\gamma(\Phi, X)$ is nontrivial for every $0 < \gamma < \infty$. Therefore, we are interested in characterizing the approximation spaces $A_q^\gamma(\Phi, X)$ in the whole range of the parameters γ and q. To achieve this goal, we shall use the machinery of Jackson and Bernstein estimates plus interpolation spaces.

Interpolation spaces. We recall some basic definitions from the real interpolation method. We refer the reader to [1] and [2] as general references for interpolation theory. Let X and B be two quasi-normed speces such that $B \subset X$ (B is continuously embedded in X) and suppose that $\|\cdot\|_B$ satisfies the τ-triangle inequality: $\|f+g\|_B^\tau \le \|f\|_B^\tau + \|g\|_B^\tau$ with $0 < \tau \le 1$. We shall also assume that $\Phi \subset B$. The K-functional is defined for each $f \in X$ by

$$K(f,t) := K(f,t;X,B) := \inf_{g\in B}(\|f-g\|_X + t\|g\|_B), \qquad t > 0.$$

The real interpolation space $(X,B)_{\lambda,q}$ with $0 < \lambda \le 1$ and $0 < q \le \infty$ is defined as the set of all $f \in X$ such that

$$\|f\|_{(X,B)_{\lambda,q}} := \|f\|_X + \left(\int_0^\infty (t^{-\lambda}K(f,t))^q \frac{dt}{t}\right)^{1/q} < \infty,$$

where the L_q-norm is replaced by the sup-norm if $q = \infty$. It is readily seen that

$$\|f\|_{(X,B)_{\lambda,q}} \approx \|f\|_X + \Big(\sum_{m=0}^{\infty}[2^{m\lambda}K(f,2^{-m})]^q\Big)^{1/q}.$$

A pair of companion **Jackson and Bernstein estimates** yields direct and inverse estimates for n-term approximation from Φ, which involve the K-functional.

Theorem 1. (a) *Suppose $\alpha > 0$ and for each $f \in B$ the following Jackson inequality holds:*

$$\sigma_n(f) \le cn^{-\alpha}\|f\|_B, \qquad n = 1, 2, \dots \tag{2.2}$$

Then, for $f \in X$,

$$\sigma_n(f) \le cK(f, n^{-\alpha}), \qquad n = 1, 2, \dots$$

(b) *Suppose $\alpha > 0$ and the following Bernstein inequality holds:*

$$\|s\|_B \le cn^{\alpha}\|S\|_X, \qquad \textit{for } s \in \Sigma_n,\ n = 1, 2, \dots \tag{2.3}$$

Then, for $f \in X$,

$$K(f, n^{-\alpha}) \le cn^{-\alpha}\Bigg[\|f\|_X^{\tau} + \sum_{\nu=1}^{n}\frac{1}{\nu}(\nu^{\alpha}\sigma_\nu(f))^{\tau}\Bigg]^{1/\tau}, \quad n = 1, 2, \dots$$

For the proof of this theorem, see, e.g., [29].

An immediate consequence of Theorem 1 is that if the Jackson and Bernstein inequalities (2.2) and (2.3) hold, then $\sigma_n(f) = O(n^{-\gamma})$, $0 < \gamma < \alpha$, if and only if $K(f, n^{-\alpha}) = O(n^{-\gamma})$. In general, Theorem 1 readily implies the following characterization of the approximation spaces $A_q^{\gamma}(\Phi)$:

Theorem 2. *If the Jackson and Bernstein inequalities* (2.2) *and* (2.3) *hold, then*

$$A_q^{\gamma}(\Phi, X) = (X, B)_{\frac{\gamma}{\alpha},q}, \qquad 0 < \gamma < \alpha,\ 0 < q \le \infty,$$

with equivalent norms.

General embedding theorem and Jackson estimate. Jackson estimates are easy to prove when approximating in L_p, $0 < p < \infty$ from families Φ consisting of well localized functions. We first give an embedding theorem which is crucial for the whole development.

Theorem 3. *Suppose (Φ_m) is a sequence of functions in $L_p(\mathbb{R}^d)$, $d \ge 1$, $0 < p < \infty$, which satisfies the following additional properties when $1 < p < \infty$:*

(i) $\Phi_m \in L_\infty(\mathbb{R}^2)$, supp $\Phi_m \subset E_m$ *with* $0 < |E_m| < \infty$, *and*

$$\|\Phi_m\|_\infty \le c_1|E_m|^{-1/p}\|\Phi_m\|_p.$$

(ii) *If* $x \in E_m$, *then*

$$\sum_{E_j \ni x,\ |E_j| \geq |E_m|} (|E_m|/|E_j|)^{1/p} \leq c_1,$$

where the summation is over all indices j *for which* E_j *satisfies the indicated conditions. Denote* $f := \sum_m \Phi_m$ *and assume that for some* $0 < \tau < p$

$$N(f) := (\sum_m \|\Phi_m\|_p^\tau)^{1/\tau} < \infty. \tag{2.4}$$

Then $\sum_m |\Phi_m(\cdot)| < \infty$ *a.e. on* $\mathbb{R}^d$, *and hence,* f *is well-defined on* $\mathbb{R}^d$, $f \in L_p(\mathbb{R}^d)$, *and*

$$\|f\|_p \leq \| \sum_m |\Phi_m(\cdot)| \|_p \leq cN(f),$$

where $c = c(\alpha, p, c_1)$.

Furthermore, if $1 \leq p < \infty$, *condition* (2.4) *can be replaced by the weaker condition*

$$N(f) := \|(\|\Phi_m\|_p)\|_{w\ell_\tau} < \infty, \tag{2.5}$$

where $\|(x_m)\|_{w\ell_\tau}$ *denotes the weak* ℓ_τ*-norm of the sequence* (x_m)*:*

$$\|(x_m)\|_{w\ell_\tau} := \inf\{M : \#\{m : |x_m| > Mn^{-1/\tau}\} \leq n \text{ for } n = 1, 2, \dots\}.$$

Theorem 4 (Jackson estimate). *Under the hypothesis of* Theorem 3, *suppose* $(\Phi_m^*)_{j=1}^\infty$ *is a rearrangement of the sequence* (Φ_m) *such that* $\|\Phi_1^*\|_p \geq \|\Phi_2^*\|_p \geq \dots$. *Set* $s_n := \sum_{j=1}^n \Phi_j^*$. *Then*

$$\|f - s_n\|_p \leq cn^{-\alpha} N(f) \quad \text{with} \quad \alpha = 1/\tau - 1/p,$$

where $c = 1$ *if* $0 < p \leq 1$ *and* $c = c(\alpha, p, c_1)$ *if* $1 < p < \infty$. *Furthermore, the estimate remains valid if condition* (2.4) *is replaced by* (2.5) *when* $1 \leq p < \infty$.

For the proof of Theorems 3-4, see [21, 28]. Note that there is no simple recipe for proving Jackson estimates in the uniform norm. More sophisticated techniques are needed in this case (see §9). There is no simple recipe for proving **Bernstein estimates** as well (see [11, 21, 28]).

3. B-spaces Generated by Spline Multiresolution over Multilevel Nested Triangulations in $\mathbb{R}^2$

In this section, we introduce and discuss the smoothness spaces (B-spaces) needed for the characterization of nonlinear n-term approximation generated by sequences of hierarchical (multiscale) spline bases over triangulations. We include all necessary tools for B-spaces and, in particular, three types of triangulations, a description of hierarchical spline bases in general and a construction of concrete differentiable spline bases, etc. We refer the reader to [11, 21, 22] as references for this section.

3.1. Multilevel Triangulations

Here we introduce several types of multilevel nested triangulations which will be needed for the definition of sequences of spline bases and B-spaces. Each triangulation exists in two versions, namely, for $E = \mathbb{R}^2$ and for any compact polygonal domain $E \subset \mathbb{R}^2$. Since there is no substantial difference between them we shall present in detail only the triangulations on $\mathbb{R}^2$.

Weak locally regular (WLR) triangulations. We call $\mathcal{T} = \bigcup_{m\in\mathbb{Z}} \mathcal{T}_m$ a *weak locally regular* (WLR) *triangulation* of $\mathbb{R}^2$ with levels $(\mathcal{T}_m)$ if the following conditions are fulfilled:

(a) Every level $\mathcal{T}_m$ is a partition of $\mathbb{R}^2$, that is, $\mathbb{R}^2 = \bigcup_{\Delta\in\mathcal{T}_m} \Delta$ and $\mathcal{T}_m$ consists of closed triangles with disjoint interiors.

(b) The levels $(\mathcal{T}_m)$ of $\mathcal{T}$ are nested, i.e., $\mathcal{T}_{m+1}$ is a refinement of $\mathcal{T}_m$.

(c) Each triangle $\Delta \in \mathcal{T}_m$ $(m \in \mathbb{Z})$ has at least two and at most M_0 children (subtriangles) in $\mathcal{T}_{m+1}$, where $M_0 \geq 4$ is a constant.

(d) For any compact $K \subset \mathbb{R}^2$ and any fixed $m \in \mathbb{Z}$, there is a finite collection of triangles from $\mathcal{T}_m$ which cover K.

(e) There exist constants $0 < r < \rho < 1$ $(r \leq \frac{1}{4})$ such that for each $\Delta \in \mathcal{T}_m$ $(m \in \mathbb{Z})$ and any child $\Delta' \in \mathcal{T}_{m+1}$ of Δ

$$r|\Delta| \leq |\Delta'| \leq \rho|\Delta|. \tag{3.1}$$

We denote by $\mathcal{V}_m$ and $\mathcal{E}_m$ the sets of all vertices and edges of $\mathcal{T}_m$, respectively. We also set $\mathcal{V} = \bigcup_{m\in\mathbb{Z}} \mathcal{V}_m$ and $\mathcal{E} = \bigcup_{m\in\mathbb{Z}} \mathcal{E}_m$.

Locally regular (LR) triangulations. We call $\mathcal{T} = \bigcup_{m\in\mathbb{Z}} \mathcal{T}_m$ a *locally regular* (LR) *triangulation* of $\mathbb{R}^2$ if $\mathcal{T}$ is a WLR-triangulation of $\mathbb{R}^2$ and satisfies the following additional conditions:

(f) *No hanging vertices condition*: No vertex of any triangle $\Delta \in \mathcal{T}_m$ lies in the interior of an edge of another triangle from $\mathcal{T}_m$.

(g) The valence N_v of each vertex v of any triangle $\Delta \in \mathcal{T}_m$ (the number of the triangles from $\mathcal{T}_m$ which share v as a vertex) is at most N_0, where $N_0 \geq 3$ is a constant.

(h) There exists a constant $0 < \delta_1 \leq 1$ independent of m such that for any $\triangle', \triangle'' \in \mathcal{T}_m$ $(m \in \mathbb{Z})$ with a common edge

$$\delta_1 \leq |\triangle'|/|\triangle''| \leq \delta_1^{-1}. \tag{3.2}$$

Strong locally regular (SLR) triangulations. We call $\mathcal{T} = \bigcup_{m\in\mathbb{Z}} \mathcal{T}_m$ a *strong locally regular* (SLR) *triangulation* of $\mathbb{R}^2$ if $\mathcal{T}$ is an LR-triangulation of $\mathbb{R}^2$ and satisfies the following additional condition:

(i) There exists a constant $0 < \delta_2 \leq 1/2$ such that for any $\triangle', \triangle'' \in \mathcal{T}_m$ $(m \in \mathbb{Z})$ sharing an edge,

$$|\mathrm{conv}\,(\triangle' \cup \triangle'')|/|\triangle'| \leq \delta_2^{-1}, \tag{3.3}$$

where conv(G) denotes the convex hull of $G \subset \mathbb{R}^2$.

Obviously, (3.3) implies (3.2) with $\delta_1 = \delta_2$.

Regular (R) triangulations. By definition, $\mathcal{T} = \bigcup_{m\in\mathbb{Z}} \mathcal{T}_m$ is a *regular* (R) *triangulation* if $\mathcal{T}$ is an LR-triangulation and $\mathcal{T}$ satisfies the following condition:

(k) There exists a constant $\beta = \beta(\mathcal{T}) > 0$ such that the minimal angle of each triangle $\triangle \in \mathcal{T}$ is $\geq \beta$.

Evidently, every regular triangulation is an SLR-triangulation.

Triangulations on compact polygonal domains in $\boldsymbol{R^2}$. A set $E \subset \mathbb{R}^2$ is said to be a compact polygonal domain if E can be represented as the union of a finite set $\mathcal{T}_0$ of closed triangles with disjoint interiors: $E = \bigcup_{\Delta\in\mathcal{T}_0} \Delta$. Weak locally regular, locally regular, etc. triangulations $\mathcal{T} = \bigcup_{m\geq 0}^{\infty} \mathcal{T}_m$ of such domain $E \subset \mathbb{R}^2$ are defined similarly as when $E = \mathbb{R}^2$. The only essential distinctions are that the levels $(\mathcal{T}_m)_{m\geq 0}$ now are consecutive refinements of the initial coarse level $\mathcal{T}_0$ and, if a vertex $v \in \mathcal{V}_m$ is on the boundary, we should include in $\mathcal{V}_m$ as many copies of v as is its multiplicity.

Remarks. It is a key observation that the collection of all SLR-triangulations with given (fixed) parameters is invariant under affine transforms. The same is true for LR-triangulations.

It is easy to see that condition (i) (see (3.3)) is equivalent to the following

Affine transform angle condition: There exists a constant $\beta = \beta(\mathcal{T})$, $0 < \beta \leq \pi/3$, such that if $\triangle_0 \in \mathcal{T}_m$ ($m \in \mathbb{Z}$) and $A : \mathbb{R}^2 \to \mathbb{R}^2$ is an affine transform that maps $\triangle_0$ one-to-one onto an equilateral reference triangle, then for every $\triangle \in \mathcal{T}_m$ which has at least one common vertex with $\triangle_0$, we have

$$\min \operatorname{angle}(A(\triangle)) \geq \beta, \tag{3.4}$$

where $A(\triangle)$ is the image of $\triangle$ by the affine transform A.

It is important to know how $|\Delta|$, min angle(Δ), and max edge(Δ) of a triangle $\Delta \in \mathcal{T}$ may change as Δ moves away from a fixed triangle $\Delta^\diamond$ within the same level or through the nested refinements. If $\mathcal{T}$ is a WLR-triangulation, then it may happen that $\Delta', \Delta'' \in \mathcal{T}_m$ share an edge and $|\Delta'|/|\Delta''|$ is arbitrarily small (or large). Because of condition (h), this is impossible if $\mathcal{T}$ is an LR-triangulation.

Consider now the case when $\mathcal{T}$ is an LR-triangulation. Then conditions (e) and (h) suggest a geometric rate of change of $|\Delta|$ (at the same level). In fact, the rate is polynomial. Furthermore, if $\Delta', \Delta'' \in \mathcal{T}_m$ have a common vertex and are also children of some $\Delta \in \mathcal{T}_{m-1}$, then, it is possible that Δ' be equilateral (or close to such), but Δ'' have an uncontrollably sharp angle.

If $\mathcal{T}$ is an SLR-triangulation, the above configuration is impossible, but the triangles from $\mathcal{T}$ still may have uncontrollably sharp angles. In this case, min angle(Δ) changes gradually from one triangle to the adjacent ones.

For more details about multilevel triangulations, see [21, 22].

Some additional notation will be needed. For a triangle $\triangle \in \mathcal{T}_m$ ($m \in \mathbb{Z}$), we define level$(\triangle) := m$. For any vertex $v \in \mathcal{V}_m$, we let star$(v) =$ star$^1(v)$

denote the *star* of v, i.e., the union of all triangles $\triangle \in \mathcal{T}_m$ attached to v. Moreover, for $\ell \ge 2$, we denote by $\operatorname{star}^{\ell}(v)$ the union of $\operatorname{star}^{\ell-1}(v)$ and the stars of the vertices of $\operatorname{star}^{\ell-1}(v)$. We also set

$$\Omega_{\triangle}^{\ell} := \bigcup \{\operatorname{star}^{\ell}(v) : v \in \mathcal{V}_m,\ \triangle \subset \operatorname{star}^{\ell}(v)\}, \qquad \triangle \in \mathcal{T}_m. \tag{3.5}$$

3.2. Local Polynomial Approximation and Moduli of Smoothness

For a function $f \in L_q(G)$ with $G \subset \mathbb{R}^d$, $d \ge 1$, and $0 < q \le \infty$, we denote by $E_k(f,G)_q$ the error of L_q-approximation of f from Π_k on G, i.e.,

$$E_k(f,G)_q := \inf_{P \in \Pi_k} \|f - P\|_{L_q(G)}. \tag{3.6}$$

We also denote by $\omega_k(f,G)_q$ the k-th local modulus of smoothness of f on G:

$$\omega_k(f,G)_q := \sup_{h \in \mathbb{R}^d} \|\Delta_h^k(f,\cdot)\|_{L_q(G)}, \tag{3.7}$$

where $\Delta_h^k(f,x) = \Delta_h^k(f,x,G) := \sum_{j=0}^{k} (-1)^{k+j} \binom{k}{j} f(x+jh)$ if $[x, x+rh]$ is entirely contained in G and $\Delta_h^k(f,x) := 0$ otherwise.

Whitney's theorem is an important tool in piecewise polynomial approximation. We shall give it in the form we need it. Suppose $G = \triangle$ or $G = \Omega_{\triangle} := \Omega_{\triangle}^1$ with $\triangle$ a triangle from an SLR-triangulation or G is a box in $\mathbb{R}^d$. If $f \in L_q(G)$, $0 < q \le \infty$, then

$$E_k(f,G)_q \le c\,\omega_k(f,G)_q, \tag{3.8}$$

where c depends only on the corresponding parameters. Note that this estimate holds for much more general regions G, but then the constant $c = c(G)$ may become hard to control. For this reason we restrict ourselves to using (3.8) only on simple regions G as above.

The (global) k-th modulus of smoothness of f in $L_q(G)$ is defined by

$$\omega_k(f,t)_q = \omega_k(f,t,G)_q := \sup_{|h| \le t} \|\Delta_h^k(f,\cdot)\|_{L_q(G)}, \qquad t > 0. \tag{3.9}$$

Another important technical tool is the averaged modulus of smoothness which is defined by

$$\mathrm{w}_k(f,t)_q^q = \mathrm{w}_r(f,t,G)_q^q := \frac{1}{t^d} \int_{[0,t]^d} \int_G |\Delta_h^k(f,x,G)|^q \, dx \, dh, \qquad G \subset \mathbb{R}^d. \tag{3.10}$$

It is well known that $\mathrm{w}_k(f,t)_q$ is equivalent to $\omega_k(f,t)_q$:

$$c_1 \mathrm{w}_k(f,t)_q \le \omega_k(f,t)_q \le c_2 \mathrm{w}_k(f,t)_q, \quad t > 0, \tag{3.11}$$

where $c_1, c_2 > 0$ depend only on q, r, and d. (see, e.g., [29] for the proof of this in the univariate case; the same proof applies in the multivariate case as well).

We shall often use the equivalence of different norms of polynomials. for instance, if $P \in \Pi_k$ and G is a box or $G = \Omega_\triangle$ with $\triangle$ a triangle from a SLR-triangualtion, then

$$\|P\|_{L_p(G)} \approx |G|^{1/p-1/q}\|P\|_{L_q(G)} \tag{3.12}$$

with constants of equivalence independent of P and G.

3.3. Hierarchical Families of Spline Bases: The General Setting

Let $\mathcal{T} = \bigcup_{m\in\mathbb{Z}} \mathcal{T}_m$ be a locally regular (or better) triangulation of $\mathbb{R}^2$. For $r \geq 0$, and $k \geq 1$, we denote by $\mathcal{S}_m^{k,r} = \mathcal{S}^{k,r}(\mathcal{T}_m)$ the set of all r times differentiable piecewise polynomial functions of degree $< k$ over $\mathcal{T}_m$, i.e., $s \in \mathcal{S}_m^{k,r}$ if $s \in C^r(\mathbb{R}^2)$ and $s = \sum_{\triangle\in\mathcal{T}_m} \mathbb{1}_\triangle \cdot P_\triangle$ with $P_\triangle \in \Pi_k$. Naturally, $\mathcal{S}_m^{k,-1}$ will denote the set of all piecewise polynomials of degree $< k$ over $\mathcal{T}_m$ which are, in general, discontinuous across the edges from $\mathcal{E}_m$.

Spline multiresolution. We assume that for each $m \in \mathbb{Z}$ there exist a subspace $\mathcal{S}_m$ of $\mathcal{S}_m^{k,r}$ ($r \geq 0$, $k \geq 2$) and a family $\Phi_m = \{\varphi_\theta : \theta \in \Theta_m\} \subset \mathcal{S}_m$ satisfying the following conditions:

- $\mathcal{S}_m \subset \mathcal{S}_{m+1}$.
- $\Pi_{\tilde{k}} \subset \mathcal{S}_m$, for some $1 \leq \tilde{k} \leq k$ ($\tilde{k}$ independent of m).
- For any $s \in \mathcal{S}_m$ there exists a unique sequence of real coefficients $\{a_\theta(s)\}$, $\theta \in \Theta_m$, such that
$$s = \sum_{\theta\in\Theta_m} a_\theta(s)\varphi_\theta.$$
(Thus Φ_m is a basis for $\mathcal{S}_m$ and $\{a_\theta(\cdot)\}_{\theta\in\Theta_m}$ are the dual functionals.)
- For each $\theta \in \Theta_m$ there is a vertex $v = v_\theta \in \mathcal{V}_m$ such that
$$\begin{gathered}\operatorname{supp}\varphi_\theta \subset \operatorname{star}^\ell(v) =: E_\theta, \text{ (see (3.5))} \\ \|\varphi_\theta\|_{L_\infty(\mathbb{R}^2)} = \|\varphi_\theta\|_{L_\infty(E_\theta)} \leq M_1, \\ |a_\theta(s)| \leq M_2\|s\|_{L_\infty(E_\theta)}, \qquad s \in \mathcal{S}_m,\end{gathered}$$
where $\ell \geq 1$ and $M_1, M_2 > 0$ are constants, independent of θ and m.

We denote $\mathcal{S} := (\mathcal{S}_m)_{m\in\mathbb{Z}}$, $\Phi := \bigcup_{m\in\mathbb{Z}} \Phi_m$ and $\Theta := \bigcup_{m\in\mathbb{Z}} \Theta_m$. We shall call $\mathcal{S}$ a *spline multiresolution* over $\mathcal{T}$ with a family of basis functions Φ.

A simple example of spline multiresolution is the sequence $(\mathcal{S}_m)_{m\in\mathbb{Z}}$ of all continuous piecewise linear functions ($r = 0$, $k = 2$) on the levels $(\mathcal{T}_m)_{m\in\mathbb{Z}}$ of a given LR-triangulation $\mathcal{T}$ of $\mathbb{R}^2$. A basis for each space $\mathcal{S}_m$ is given by the set Φ_m of the Courant elements φ_θ, supported on the cells θ of $\mathcal{T}_m$ (θ is the union of all triangles of $\mathcal{T}_m$ attached to a vertex, say, v_θ). The function φ_θ takes the value 1 at v_θ and the value 0 at all other vertices.

Box splines with the corresponding ladder of spline spaces provide another example of a spline multiresolution.

Concrete constructions of differentiable spline basis functions (from C^r, $r \ge 1$) associated with spline multiresolution over general triangulations will be discussed in §3.4.

Note that Θ and Θ_m ($m \in \mathbb{Z}$) above are simply index sets, which in the case of Courant elements can be identified as sets of cells (supports of basis functions). In general, several basis functions of Φ_m may have the same support. However, the supports of only $\le$ *constant* of them may overlap.

It follows from the above conditions that each basis Φ_m is L_q-stable for all $0 < q \le \infty$, i.e., if $g := \sum_{\theta \in \Theta_m} b_\theta \varphi_\theta$, where $\{b_\theta\}_{\theta \in \Theta_m}$ is an arbitrary sequence of real numbers, then

$$\|g\|_q \approx \Big(\sum_{\theta \in \Theta_m} \|b_\theta \varphi_\theta\|_q^q \Big)^{1/q}$$

with constants of equivalence independent of $(b_\theta)_{\theta \in \Theta_m}$ and m.

Quasi-interpolant. For $0 < q \le \infty$ and an arbitrary triangle $\triangle$, we let $P_{\triangle,q} : L_q(\triangle) \to \Pi_k$ be a projector such that

$$\|f - P_{\triangle,q}(f)\|_{L_q(\triangle)} \le cE_k(f,\triangle)_q, \quad \text{for } f \in L_q(\triangle). \tag{3.13}$$

Note that $P_{\triangle,q}$ can be realized as a linear projector if $q \ge 1$.

We define a linear operator $Q_m : \mathcal{S}^{k,-1}(\mathcal{T}_m) \to \mathcal{S}_m$ as follows. For each $\theta \in \Theta_m$, let $\lambda_\theta : \mathcal{S}^{k,-1}(\mathcal{T}_m)|_{E_\theta} \to \mathbb{R}$ be a linear functional such that

$$\lambda_\theta(s|_{E_\theta}) = a_\theta(s), \ \text{if } s \in \mathcal{S}_m, \ \text{and}$$
$$|\lambda_\theta(f)| \le M_2 \|f\|_{L_\infty(E_\theta)}, \quad f \in \mathcal{S}^{k,-1}(\mathcal{T}_m)|_{E_\theta}.$$

Such linear functionals always exist according to the Hahn-Banach theorem. We set

$$Q_m(s) := \sum_{\theta \in \Theta_m} \lambda_\theta(s|_{E_\theta}) \varphi_\theta, \qquad s \in \mathcal{S}^{k,-1}(\mathcal{T}_m). \tag{3.14}$$

Clearly, $Q_m(s) = s$ if $s \in \mathcal{S}_m$, and thus Q_m is a linear projector of $\mathcal{S}^{k,-1}(\mathcal{T}_m)$ into $\mathcal{S}_m$. Moreover, Q_m is a bounded projector: For any $s \in \mathcal{S}^{k,-1}(\mathcal{T}_m)$, $0 < q \le \infty$ and $\triangle \in \mathcal{T}_m$,

$$\|Q_m(s)\|_{L_q(\triangle)} \le c\|s\|_{L_q(\Omega_\triangle^\ell)},$$

with a constant c independent of m, $\triangle$, and s.

We now extend Q_m to $L_q^{\rm loc}(\mathbb{R}^2)$, $0 < q \le \infty$ Let $P_{\triangle,q} : L_q(\triangle) \to \Pi_k$ be a projector satisfying (3.13). We define

$$p_{m,q}(f) := \sum_{\triangle \in \mathcal{T}_m} \mathbb{1}_\triangle \cdot P_{\triangle,q}(f), \quad \text{for } f \in L_q^{\rm loc},$$

and the quasi-interpolant that we need is defined by

$$Q_{m,q}(f) := Q_m(p_{m,q}(f)), \quad \text{for } f \in L_q^{\rm loc}, \tag{3.15}$$

which is a projector of $L_q^{\rm loc}$ into $\mathcal{S}_m$.

We next show that $Q_{m,q}$ provides a good local L_q-approximation from $\mathcal{S}_m$. We let $\mathbb{S}_\triangle(f)_q$ denote the error of L_q-approximation from $\mathcal{S}_m$ on $\Omega_\triangle^\ell$, i.e.,

$$\mathbb{S}_\triangle(f)_q := \inf_{s\in\mathcal{S}_m} \|f-s\|_{L_q(\Omega_\triangle^\ell)}, \qquad \triangle\in\mathcal{T}_m. \tag{3.16}$$

The good local approximation properties of $Q_{m,q}$ can be described as follows:

(a) If $f\in L_q^{\rm loc}$, $0<q\le\infty$ ($f\in C$ if $q=\infty$), then

$$\|f-Q_{m,q}(f)\|_{L_q(\triangle)} \le c\mathbb{S}_\triangle(f)_q, \qquad \triangle\in\mathcal{T}_m\ (m\in\mathbb{Z}),$$

with c independent of f, m, and $\triangle$.

(b) If $f\in L_q^{\rm loc}$, $0<q\le\infty$, then for every compact $K\subset\mathbb{R}^2$,

$$\|f-Q_{m,q}(f)\|_{L_q(K)}\to 0 \quad\text{as}\quad m\to\infty.$$

We denote $\mathcal{S}_{-\infty} := \bigcap_{m\in\mathbb{Z}}\mathcal{S}_m$. Note that if $s\in\mathcal{S}_{-\infty}$, $s\neq$ constant, and $|\{x\in\mathbb{R}^2 : |s(x)|>t\}|<\infty$ for some $t>0$, then $s\equiv 0$ and, in particular, if $s\in\mathcal{S}_{-\infty}\cap L_p$ ($p<\infty$), then $s\equiv 0$.

3.4. Construction of Differentiable Spline Bases

In this section, we present a concrete construction of bases for the spline spaces $\mathcal{S}_m^{k,r} := \mathcal{S}^{k,r}(\mathcal{T}_m)$, $r\ge 1$, $k>4r+1$, $m\in\mathbb{Z}$, provided $\mathcal{T}=\bigcup_{m\in\mathbb{Z}}\mathcal{T}_m$ is an SLR-triangulation of $\mathbb{R}^2$. This construction provides bases which satisfy all requirements for such bases from §3.3.

Nodal functionals. As before, let $\mathcal{V}_m$ and $\mathcal{E}_m$ be the sets of all vertices and all edges of $\mathcal{T}_m$. We shall construct basis functions for $\mathcal{S}_m=\mathcal{S}^{k,r}(\mathcal{T}_m)$, $k>4r+1$, (see §3.3) by the so called *nodal functionals* defined on $\mathcal{S}^{k,r}(\mathcal{T}_m)$, which involve the values of the splines and their derivatives at specific points in $\mathbb{R}^2$. The functional corresponding to the simple evaluation of the splines at $\xi\in\mathbb{R}^2$ will be denoted by δ_ξ. Of particular interest as evaluation points are the vertices $v\in\mathcal{V}_m$ of $\mathcal{T}_m$, where we also need the derivative evaluation functionals of type $\delta_v D_e^\alpha$ with e being any edge in $\mathcal{E}_m$ attached to v, and $\delta_v D_{e_1}^\alpha D_{e_2}^\beta$, where e_1,e_2 are adjacent edges attached to v. Here $D_e^\alpha s$ denotes the derivative of s of order α in the direction of e weighted by the length of e.

We shall need the following additional notation. For an interval $e:=[v_1,v_2]$, we denote by $|e|$ the length of e. Throughout we assume that the vertices v_1,v_2,v_3 of any triangle $[v_1,v_2,v_3]$ are ordered counterclockwise. We let star(e) denote the union of the two triangles attached to $e\in\mathcal{E}_m$.

Let $\triangle_1,\triangle_2\in\mathcal{T}_m$ share an edge e. Since every $s\in\mathcal{S}_m^{k,r}$ is continuous, the two polynomial patches $s|_{\triangle_1}$ and $s|_{\triangle_2}$ coincide along e. Therefore, $\delta_v D_e^\alpha s$ may be computed for any $\alpha=0,1,\dots$ as either $\delta_v D_e^\alpha(s|_{\triangle_1})$ or $\delta_v D_e^\alpha(s|_{\triangle_2})$ with the same result. Similarly, if $e_1,e_2\in\mathcal{E}_m$ are two edges of a triangle $\triangle\in\mathcal{T}_m$ with a common vertex v, then $\delta_v D_{e_1}^\alpha D_{e_2}^\beta s$ denotes the mixed derivative of s at v in the directions of e_1 and e_2 away from v. If $\alpha+\beta\le r$, this derivative is uniquely defined. If $\alpha+\beta>r$, the result may depend on the choice of the polynomial patch of s attached to v. We follow the convention to always take $\delta_v D^\alpha_{e_1} D^\beta_{e_2} s := \delta_v D^\alpha_{e_1} D^\beta_{e_2}(s|_\triangle)$, where $\triangle$ is the above triangle formed by e_1, e_2.

We shall also need functionals evaluating at some points on an edge e the derivatives of the spline in an affine invariant direction not parallel to e. Let $e=[v_1,v_2]\in\mathcal{E}_m$ and let $\triangle_e=[v_1,v_2,v_3]\in\mathcal{T}_m$ be a triangle attached to e. Denote by $\mu(e,\triangle)$ the median of $\triangle$ connecting the middle point $(v_1+v_2)/2$ of e with the third vertex v_3 of $\triangle$. For any point $\xi\in e$, $\delta_\xi D_{\mu(e,\triangle)}$ will denote the derivative at ξ in the direction pointing into the half-plane containing $\triangle$ parallel to $\mu(e,\triangle)$, weighted with the length of $\mu(e,\triangle)$. For each edge $e\in\mathcal{E}_m$, we choose one of the two triangles attached to e and denote it by $\triangle_e$. (Note that this selection of $\triangle_e$ is not unique but it will cause no problems to the basis construction.)

Characterization of differentiability. Let L be a straight line splitting $\mathbb{R}^2$ into two halfplanes $H, \tilde{H}$. Given $p,\tilde{p}\in\Pi_k$, we define a piecewise polynomial function s by setting $s|_H=p$, $s|_{\tilde{H}}=\tilde{p}$. To verify whether s is differentiable across L, we choose two points u,v on L as well as two points $w,\tilde{w}$ in the interiors of H and $\tilde{H}$, respectively. We set $\triangle:=[u,v,w]$, $\tilde{\triangle}:=[u,v,\tilde{w}]$, $e:=[u,v]$, $\mu:=[u,w]$, $\tilde{\mu}:=[u,\tilde{w}]$, $\theta:=\angle e\mu$, $\tilde{\theta}:=\angle\tilde{\mu}e$.

It is readily seen that if $0\le r<k$, then $s\in C^r(\mathbb{R}^2)$ if and only if

$$\delta_u D^\alpha_{\tilde{\mu}} D^{q-\alpha}_e \tilde{p} = \sum_{\beta=0}^{\alpha}(-1)^\beta \binom{\alpha}{\beta}\sin^{\alpha-\beta}(\theta+\tilde{\theta})\Big(\frac{|e|\sin\tilde{\theta}}{|\mu|}\Big)^\beta\Big(\frac{|e|\sin\theta}{|\tilde{\mu}|}\Big)^{-\alpha}\delta_u D^\beta_\mu D^{q-\beta}_e p, \tag{3.17}$$

for all $\alpha=0,\dots,r$ and $q=\alpha,\dots,k-1$ (see, e.g. [10]).

Construction of basis splines. Consider the following set $\mathcal{N}_m$ of nodal functionals on $\mathcal{S}^{k,r}_m$,

$$\mathcal{N}_m := \Big(\bigcup_{v\in\mathcal{V}_m}\mathcal{N}^v_m\Big)\cup\Big(\bigcup_{e\in\mathcal{E}_m}\mathcal{N}^e_m\Big)\cup\Big(\bigcup_{\triangle\in\mathcal{T}_m}\mathcal{N}^\triangle_m\Big), \quad \text{where}$$

(i) for each triangle $\triangle=[v_1,v_2,v_3]\in\mathcal{T}_m$,

$$\mathcal{N}^\triangle_m := \{\eta^\triangle_\xi := \delta_\xi \;:\; \xi\in\Xi_\triangle\},$$

$$\Xi_\triangle := \Big\{\frac{i_1v_1+i_2v_2+i_3v_3}{k-1} \;:\quad i_1+i_2+i_3=k-1,\quad i_1,i_2,i_3>r\Big\}\subset\triangle;$$

(ii) for each edge $e=[v_1,v_2]\in\mathcal{E}_m$,

$$\mathcal{N}^e_m := \{\eta^e_{q,\xi} := \delta_\xi D^q_{\mu(e,\triangle_e)} \;:\; q=0,\dots,r,\quad \xi\in\Xi_{e,q}\},$$

$$\Xi_{e,q} := \Big\{\frac{i_1v_1+i_2v_2}{k-q-1} \;:\; i_1+i_2=k-q-1,\quad i_1,i_2>2r-q\Big\}\subset e;$$

(iii) for each vertex $v\in\mathcal{V}_m$,

$$\mathcal{N}^v_m := \bigcup_{q=0}^{2r}\mathcal{N}^{v,q}_m,$$

with $\mathcal{N}_m^{v,q}$, $q = 0, \ldots, 2r$, being defined as follows. Let $\triangle^{[i]} = [v, v_i, v_{i+1}]$, $i = 1, \ldots, N_v$, be the triangles in $\mathcal{T}_m$ attached to v in counterclockwise order, $v_{N_v+\ell} = v_\ell$, and let $e_i = [v, v_i]$. We set

$$\mathcal{N}_m^{v,0} := \{\eta^{v,0} := \delta_v\},$$
$$\mathcal{N}_m^{v,q} := \{\eta_{i,\alpha}^{v,q} := \delta_v D_{e_i}^{q-\alpha} D_{e_{i+1}}^{\alpha} \ : \ i = 1, \ldots, N_v,\ \alpha = 0, \ldots, q-1\}, \quad q \geq 1.$$

In view of (3.17), the functionals in $\mathcal{N}_m^{v,q}$ are not linearly independent on $\mathcal{S}_m^{k,r}$ if $q \geq 1$. Namely, the following conditions hold for all $s \in \mathcal{S}_m^{k,r}$, $v \in \mathcal{V}_m$, $q = 1, \ldots, 2r$:

$$\eta_{i,\alpha}^{v,q}(s) - \sum_{\beta=0}^{\alpha} (-1)^\beta \tbinom{\alpha}{\beta} \sin^{\alpha-\beta}(\theta_{i-1} + \theta_i) \Big(\tfrac{|e_i| \sin \theta_i}{|e_{i-1}|}\Big)^\beta \Big(\tfrac{|e_i| \sin \theta_{i-1}}{|e_{i+1}|}\Big)^{-\alpha} \eta_{i-1,q-\beta}^{v,q}(s)$$
$$= 0, \qquad \alpha = 1, \ldots, \min\{r, q\}, \quad i = 1, \ldots, N_v, \tag{3.18}$$

where $\theta_i := \angle e_i e_{i+1}$, $\eta_{i,q}^{v,q} := \eta_{i+1,0}^{v,q}$.

The following key lemma is basic in constructing the basis functions.

Lemma 1. *There is a unique spline $s \in \mathcal{S}_m^{k,r}$ such that*

$$\begin{aligned}
&\eta_\xi^\triangle(s) = a_\xi^\triangle, && \xi \in \Xi_\triangle,\ \triangle \in \mathcal{T}_m,\\
&\eta_{q,\xi}^e(s) = a_{q,\xi}^e, && \xi \in \Xi_{e,q},\ q = 0, \ldots, r,\ e \in \mathcal{E}_m,\\
&\eta^{v,0}(s) = a^{v,0}, && v \in \mathcal{V}_m,\\
&\eta_{i,\alpha}^{v,q}(s) = a_{i,\alpha}^{v,q}, && i = 1, \ldots, N_v,\ \alpha = 0, \ldots, q-1,\ q = 1, \ldots, 2r,\ v \in \mathcal{V}_m,
\end{aligned}$$

for any given $a_\xi^\triangle, a_{q,\xi}^e, a^{v,0} \in \mathbb{R}$ and any $a_{i,\alpha}^{v,q} \in \mathbb{R}$ satisfying

$$a_{i,\alpha}^{v,q} - \sum_{\beta=0}^{\alpha} (-1)^\beta \tbinom{\alpha}{\beta} \sin^{\alpha-\beta}(\theta_{i-1} + \theta_i) \Big(\tfrac{|e_i| \sin \theta_i}{|e_{i-1}|}\Big)^\beta \Big(\tfrac{|e_i| \sin \theta_{i-1}}{|e_{i+1}|}\Big)^{-\alpha} a_{i-1,q-\beta}^{v,q} = 0,$$
$$\alpha = 1, \ldots, \min\{r, q\}, \quad i = 1, \ldots, N_v.$$

Moreover, for each $\triangle \in \mathcal{T}_m$,

$$\|s|_\triangle\|_{L_\infty(\triangle)} \leq c\, \delta_2^{-2r} \max_{\eta \in \mathcal{N}_m(\triangle)} |\eta(s)|, \tag{3.19}$$

where c is a constant depending only on k, and

$$\mathcal{N}_m(\triangle) := \Big(\bigcup_{v \in \mathcal{V}_m \cap \triangle} \mathcal{N}_m^v \Big) \cup \Big(\bigcup_{\substack{e \in \mathcal{E}_m \\ e \subset \triangle}} \mathcal{N}_m^e \Big) \cup \mathcal{N}_m^\triangle.$$

For each $v \in \mathcal{V}_m$ and $q = 1, \ldots, 2r$, we denote by $R_m^{v,q}$ the $(\min\{r, q\} N_v \times qN_v)$-matrix of differentiability conditions (3.18). Let the vectors

$$a^{v,q,j}, \qquad j = 1, \ldots, \rho_{v,q} := qN_v - \operatorname{rank}(R_m^{v,q}),$$

form an orthonormal basis for the null space of $R_m^{v,q}$:

$$\text{null}(R_m^{v,q}) := \{a \in \mathbb{R}^{qN_v} \ : \ R_m^{v,q} a = 0\}.$$

For convenience, we shall use the double indices introduced in the definition of $\mathcal{N}_m^{v,q}$ also for the components of $a^{v,q,j}$: $a_{i,\alpha}^{v,q,j}$, $i = 1, \dots, N_v$, $\alpha = 0, \dots, q-1$. We set

$$\begin{aligned}
\eta^{v,q,j} &:= \sum_{i=1}^{N_v} \sum_{\alpha=0}^{q-1} a_{i,\alpha}^{v,q,j} \eta_{i,\alpha}^{v,q}, \qquad j = 1, \dots, \rho_{v,q}, \\
\tilde{\mathcal{N}}_m^{v,q} &:= \{\eta^{v,q,j} \ : \ j = 1, \dots, \rho_{v,q}\}, \qquad q = 1, \dots, 2r, \\
\tilde{\mathcal{N}}_m^{v} &:= \mathcal{N}_m^{v,0} \cup \bigcup_{q=1}^{2r} \tilde{\mathcal{N}}_m^{v,q}, \qquad v \in \mathcal{V}_m, \\
\tilde{\mathcal{N}}_m &:= \Big(\bigcup_{v \in \mathcal{V}_m} \tilde{\mathcal{N}}_m^{v} \Big) \cup \Big(\bigcup_{e \in \mathcal{E}_m} \mathcal{N}_m^{e} \Big) \cup \Big(\bigcup_{\triangle \in \mathcal{T}_m} \mathcal{N}_m^{\triangle} \Big),
\end{aligned}$$

and define the set

$$\Phi_m := \{\varphi_\eta \ : \ \eta \in \tilde{\mathcal{N}}_m\}$$

of the *basis functions* for $\mathcal{S}_m^{k,r}$ by the duality condition,

$$\mu(\varphi_\eta) = \begin{cases} 1, & \text{if } \mu = \eta \\ 0, & \text{if } \mu \in \tilde{\mathcal{N}}_m \setminus \{\eta\}. \end{cases}$$

Properties of basis splines. It follows by Lemma 1 that every spline $s \in \mathcal{S}^{k,r}(\mathcal{T}_m)$ is uniquely determined by the sequence $(\eta(s))_{\eta \in \tilde{\mathcal{N}}_m}$. Furthermore, (3.19) implies that $\|\varphi_\eta\|_{L_\infty(\mathbb{R}^2)} \le c\,\delta_2^{-2r}$ and $\operatorname{supp}\varphi_\eta$ is contained in $\text{star}(v)$, $\text{star}(e)$, or $\triangle$, whenever η belongs to $\tilde{\mathcal{N}}_m^{v}$, $\mathcal{N}_m^{e}$, or $\mathcal{N}_m^{\triangle}$, respectively. Also, by Markov's inequality, $|\eta(s)|$ is bounded above by a constant multiple of $\|s\|_{L_\infty(\text{star}\,(v))}$, $\|s\|_{L_\infty(\text{star}\,(e))}$, or $\|s\|_{L_\infty(\triangle)}$, respectively.

From this, it follows that $\Phi_m = \{\varphi_\eta \ : \ \eta \in \tilde{\mathcal{N}}_m\}$ satisfies all requirements of §3.3 with $\mathcal{S}_m = \mathcal{S}^{k,r}(\mathcal{T}_m)$ and $\tilde{k} = k$. (Obviously, $\Pi_k \subset \mathcal{S}_m$ and $\mathcal{S}^{k,r}(\mathcal{T}_m) \subset \mathcal{S}^{k,r}(\mathcal{T}_{m+1})$ if $\mathcal{T}_{m+1}$ is a refinement of $\mathcal{T}_m$.)

Remarks. The above construction of spline bases is given in [11] and follows the scheme of [10] with appropriate modifications.

A key property of the basis families constructed above is that they are invariant under affine transforms.

In [11] (see Example 4.7), there is an example which shows that for the construction of differentable spline bases the assumption that $\mathcal{T}$ is an SLR-triangulation cannot be omitted.

The construction from this section is extendable to the spaces $\mathcal{S}_m^{k,r}$, $k > r2^d + 1$, in dimensions $d > 2$. To this end the algorithm given in [10] should be extended to strong locally regular triangulations in $\mathbb{R}^d$.

If the triangulation only covers a compact domain E, then usual modifications of basis functions corresponding to boundary edges or vertices (see [10]) lead to the desired stable local bases.

For complete review of the spline basis constructions which can be adapted for SLR-triangulations, see [11].

Spline bases on special triangulations. There are several constructions of differentiable spline bases fitting into the scheme of §3.3, which are only available for specific multilevel triangulations. The box splines provide the best known example of such bases (see [5]). (Note that only box splines are available for dimensions $d > 2$.) For a complete review of the spline bases on uniform triangulations, see §5 of [11].

3.5. Slim B-spaces on $\mathbb{R}^2$

Suppose $\mathcal{T}$ is an LR(or better)-triangulation of $\mathbb{R}^2$ and let $\mathcal{S} = \mathcal{S}_{\mathcal{T}}$ be a spline multiresolution over $\mathcal{T}$ with a hierarchical family of basis functions $\Phi = \Phi_{\mathcal{T}}$, as described in §3.3. For the characterization of nonlinear n-term L_p-approximation from Φ, we need the *slim B-spaces* $B^\alpha_\tau(\mathcal{S})$ induced by $\mathcal{S}$. As will be shown below this spaces have atomic decompositions using Φ. For this reason and to simplify the notation, we shall primary use the notation $B^\alpha_\tau(\Phi) := B^\alpha_\tau(\mathcal{S})$. We shall need the slim B-spaces $B^\alpha_\tau(\mathcal{S})$ in two cases: (a) $0 < p < \infty$ and $\alpha > 0$, or (b) $p = \infty$ and $\alpha \geq 1$. In both cases, $1/\tau := \alpha + 1/p$ $(1/\infty := 0)$.

Definition of $B^\alpha_\tau(\Phi) = B^\alpha_\tau(\mathcal{S})$ via local approximation. We define the B-space $B^\alpha_\tau(\Phi)$ as the set of all functions $f \in L_\tau(\mathbb{R}^2)$ such that

$$\|f\|_{B^\alpha_\tau(\Phi)} := \Big(\sum_{\triangle\in\mathcal{T}} (|\triangle|^{-\alpha} \mathbb{S}_\triangle(f)_\tau\Big)^\tau)^{1/\tau} < \infty, \tag{3.20}$$

where $\mathbb{S}_\triangle(f)_\tau$ is the error of L_τ-approximation to f from $\mathcal{S}_m$ on $\Omega^\ell_\triangle$, if $\triangle \in \mathcal{T}_m$ (see (3.16)).

From the properties of $\mathcal{S}_{-\infty}$ (see §3.3), it follows that $\|f\|_{B^\alpha_\tau(\Phi)} = 0$ implies $f = 0$ a.e. and now it is easy to see that $\|\cdot\|_{B^\alpha_\tau(\Phi)}$ is a norm if $\tau \geq 1$ and a quasi-norm if $\tau < 1$. We shall call it a “norm” in both cases. We next introduce several other equivalent norms in $B^\alpha_\tau(\Phi)$.

First, for $f \in L_\eta$, $0 < \eta < p$, we define

$$N_{\Phi,\mathbb{S},\eta}(f) := \Big(\sum_{\triangle\in\mathcal{T}} (|\triangle|^{1/p-1/\eta} \mathbb{S}_\triangle(f)_\eta)^\tau \Big)^{1/\tau}. \tag{3.21}$$

Clearly, $N_{\Phi,\mathbb{S},\tau}(f) = \|f\|_{B^\alpha_\tau(\Phi)}$.

Definition of a norm in $B^\alpha_\tau(\Phi)$ via basis functions (atomic decomposition). For $f \in L_\tau$, we define

$$N_\Phi(f) := \inf_{f=\sum_{\theta\in\Theta} c_\theta\varphi_\theta} \Big(\sum_{\theta\in\Theta} (|E_\theta|^{-\alpha} \|c_\theta\varphi_\theta\|_\tau)^\tau \Big)^{1/\tau}, \tag{3.22}$$

where the infimum is over all representations of f: $f = \sum_{\theta\in\Theta} c_\theta\varphi_\theta$ in L_τ (an absolute convergence a.e. and in L_p follows immediately if $N_\Phi(f) < \infty$). It follows that

$$N_\Phi(f) \approx \inf_{f=\sum_{\theta\in\Theta} c_\theta\varphi_\theta} \Big(\sum_{\theta\in\Theta} \|c_\theta\varphi_\theta\|_p^\tau\Big)^{1/\tau}.$$

Definition of norms in $B^\alpha_\tau(\Phi)$ via projections. For $f \in L_\eta$, we set

$$q_{m,\eta}(f) := Q_{m,\eta}(f) - Q_{m-1,\eta}(f) \in \mathcal{S}_m, \tag{3.23}$$

where $Q_{m,\eta}$ is from (3.15), and let $\{b_{\theta,\eta}(f)\}_{\theta\in\Theta_m}$ be defined by the identity

$$q_{m,\eta}(f) = \sum_{\theta\in\Theta_m} b_{\theta,\eta}(f)\varphi_\theta \text{ , i.e.,} \quad b_{\theta,\eta}(f) := a_\theta(q_{m,\eta}(f)), \qquad \theta \in \Theta_m.$$

We define

$$N_{\Phi,Q,\tau}(f) := \Big(\sum_{\theta\in\Theta}(|E_\theta|^{-\alpha}\|b_{\theta,\tau}(f)\varphi_\theta\|_\tau)^\tau\Big)^{1/\tau}$$

and, more generally (see (3.21)), for $0 < \eta < p$,

$$N_{\Phi,Q,\eta}(f) := \Big(\sum_{\theta\in\Theta}(|E_\theta|^{1/p-1/\eta}\|b_{\theta,\eta}(f)\varphi_\theta\|_\eta)^\tau\Big)^{1/\tau} \approx \Big(\sum_{\theta\in\Theta}\|b_{\theta,\eta}(f)\varphi_\theta\|_p^\tau\Big)^{1/\tau}. \tag{3.24}$$

The following embedding theorem plays a crucial role in the proof of the equivalence of the norms, introduced above, and in the overall development.

Theorem 5. *If $f \in L_\eta$, $0 < \eta < p \le \infty$, and $N_{\Phi,Q,\eta}(f) < \infty$, then $f \in L_p$,*

$$f = \sum_{m\in\mathbb{Z}} q_{m,\eta}(f) = \sum_{\theta\in\Theta} b_{\theta,\eta}(f)\varphi_\theta \tag{3.25}$$

with the series converging absolutely a.e., and

$$\|f\|_p \le c\|\sum_{m\in\mathbb{Z}}|q_{m,\eta}(f)(\cdot)|\|_p \le c\|\sum_{\theta\in\Theta}|b_{\theta,\eta}(f)\varphi_\theta(\cdot)|\|_p \le cN_{\Phi,Q,\eta}(f).$$

with c independent of f.

Theorem 6. *The norms $\|\cdot\|_{B^\alpha_\tau(\Phi)}$, $N_{\Phi,\mathbb{S},\eta}(\cdot)$ $(0 < \eta < p)$, $N_\Phi(\cdot)$, and $N_{\Phi,Q,\eta}(\cdot)$ $(0 < \eta < p)$, defined in (3.20)–(3.22) and (3.24), are equivalent with constants of equivalence depending only on p, α, η, and the parameters of $\mathcal{T}$ and Φ.*

Since the B-spaces can be considered as sequence spaces, the embedding of one of them into another is quite easy to establish. Also, it is easy to interpolate them: Suppose $0 < p < \infty$ and $\alpha_0, \alpha_1 > 0$ or $p = \infty$ and $\alpha_0, \alpha_1 \ge 1$. Let $\tau_j := (\alpha_j + 1/p)^{-1}$, $j = 0, 1$. Then

$$(B^{\alpha_0}_{\tau_0}(\Phi), B^{\alpha_1}_{\tau_1}(\Phi))_{\lambda,\tau} = B^\alpha_\tau(\Phi)$$

with equivalent norms, provided $\alpha = (1-\lambda)\alpha_0 + \lambda\alpha_1$ with $0 < \lambda < 1$ and $\tau := (\alpha + 1/p)^{-1}$.

Remarks. The slim B-spaces are introduced and utilized in nonlinear spline approximation in [11, 21, 22].

If $p = \infty$, then the B-space $B^\alpha_\tau(\Phi)$ ($\tau := 1/\alpha$) is useful for our goals only if $\alpha \ge 1$. The reason for this is that $B^\alpha_\tau(\Phi)$ is not embedded in C if $\alpha < 1$.

We introduced the B-norms $N_{\Phi,\mathbb{S},\eta}(\cdot)$ and $N_{\Phi,Q,\eta}(\cdot)$ with $0 < \eta < p$ (see (3.21) and (3.24)) for the following reason. As we shall see in §7, normally $\alpha > 1$ and hence $\tau < 1$ which compels us to work in L_τ with $\tau < 1$ that is not a very friendly space. At the same time, if $p > 1$ we can choose $1 \le \eta < p$ and work in L_η instead.

We also want to explain why we introduced the slim B-spaces over locally regular (or better) triangulations but not over more general ones. The reason for this is that if we relax the main conditions (3.1)–(3.2) in the definition of LR-triangulations, then we can hardly work with the B-spaces. In particular, the equivalence of the norms (see Theorem 6) fails to exist which makes it impossible to prove all approximation results from §7.

Given a spline multiresolution $\mathcal{S}_\mathcal{T}$ over an LR(or better)-tiangulation $\mathcal{T}$ and an associated family of basis functions $\Phi = \Phi_\mathcal{T}$, as described in §3.3, we define the more general slim B-space $B^\alpha_{pq}(\Phi)$, $\alpha > 0$, $0 < p, q \le \infty$, as the set of all $f \in L_p$ such that

$$\|f\|_{B^\alpha_{pq}(\Phi)} := \Bigg(\sum_{m\in\mathbb{Z}} \Bigg[2^{m\alpha}\Bigg(\sum_{\triangle\in\mathcal{T},\, 2^{-m}\le|\triangle|<2^{-m+1}} \mathbb{S}_\triangle(f)_p^p\Bigg)^{1/p}\Bigg]^q\Bigg)^{1/q} < \infty,$$

with the ℓ_q-norm replaced by the sup-norm if $q = \infty$. Evidently, $B^\alpha_p(\Phi) = B^\alpha_{pp}(\Phi)$. Here, we do not explore the B-spaces in such generality because the space scale $B^\alpha_\tau(\Phi)$ is sufficient for our goal of characterizing the rates of nonlinear n-term spline approximation.

3.6. Skinny B-spaces on $\mathbb{R}^2$

In this subsection, we define a family of B-spaces which is needed for the characterization of nonlinear approximation from (discontinuous) piecewise polynomials generated by sequences of nested triangulations. Throughout this subsection, we assume that $\mathcal{T}$ is an arbitrary weak locally regular triangulation of $\mathbb{R}^2$ (see §3.1). As elsewhere in this article, we assume that $k \ge 1$, $0 < p < \infty$ and $\alpha > 0$, or $p = \infty$ and $\alpha \ge 1$, and in both cases $1/\tau := \alpha + 1/p$.

The *skinny B-space* $\mathcal{B}^{\alpha k}_\tau(\mathcal{T})$ is defined as the set of all $f \in L_\tau$ such that

$$\|f\|_{\mathcal{B}^{\alpha k}_\tau(\mathcal{T})} := \Big(\sum_{\Delta\in\mathcal{T}} (|\Delta|^{-\alpha}\omega_k(f,\Delta)_\tau)^\tau\Big)^{1/\tau} < \infty, \tag{3.26}$$

where $\omega_k(f, \triangle)_\tau$ is the k-th local modulus of smoothness of f on $\triangle$ (see (3.7)).

Whitney's inequality (3.8) implies

$$\|f\|_{\mathcal{B}_\tau^{\alpha k}(\mathcal{T})} \approx \Big(\sum_{\Delta\in\mathcal{T}} (|\Delta|^{-\alpha} E_k(f,\Delta)_\tau)^\tau\Big)^{1/\tau},$$

where $E_k(f,\Delta)_\tau$ is the error of $L_\tau(\Delta)$-approximation to f from Π_k (see (3.6)). As in Theorem 6, one can prove that, for $0<\eta<p$,

$$\|f\|_{\mathcal{B}_\tau^{\alpha k}(\mathcal{T})} \approx \Big(\sum_{\Delta\in\mathcal{T}} (|\Delta|^{1/p-1/\eta}\omega_k(f,\Delta)_\eta)^\tau\Big)^{1/\tau} \approx \inf_{f=\sum_{\Delta\in\mathcal{P}} P_\Delta} \Big(\sum_{\Delta\in\mathcal{T}} \|P_\Delta\|_p^\tau\Big)^{1/\tau},$$

where $P_\Delta \in \Pi_k$.

Let $P_{\triangle,\eta} : L_\eta(\triangle) \to \Pi_k$ be a projector satisfying (3.13) with $q=\eta$. Set $P_{m,\eta}(f) := \sum_{\Delta\in\mathcal{T}_m} \mathbb{1}_\Delta \cdot P_{\Delta,\eta}(f)$. Clearly, $P_{m,\eta}(f)$ is a projector into $\mathcal{S}^{k,-1}(\mathcal{T}_m)$. We define

$$p_{m,\eta}(f) := p_{m,\eta}(f,\mathcal{T}) := P_{m,\eta}(f) - P_{m-1,\eta}(f) \in \mathcal{S}^{k,-1}(\mathcal{T}_m),$$

and set $p_{\Delta,\eta}(f) := \mathbb{1}_\Delta \cdot p_{m,\eta}(f)$ for $\Delta\in\mathcal{T}_m$. If $f\in\mathcal{B}_\tau^{\alpha k}(\mathcal{T})$, then $f\in L_p$, $f=\sum_{\Delta\in\mathcal{T}} p_{\Delta,\eta}(f)$, and

$$\|f\|_{\mathcal{B}_\tau^{\alpha k}(\mathcal{T})} \approx \Big(\sum_{\Delta\in\mathcal{T}} (|\Delta|^{1/p-1/\eta}\|p_{\Delta,\eta}(f)\|_\eta)^\tau\Big)^{1/\tau} \approx \Big(\sum_{\Delta\in\mathcal{T}} \|p_{\Delta,\eta}(f)\|_p^\tau\Big)^{1/\tau}.$$

For more details, see [21].

3.7. Fat B-spaces on $\mathbb{R}^2$: The Link to Besov Spaces

Let $\mathcal{T}$ be an arbitrary strong locally regular triangulation of $\mathbb{R}^2$. We again assume that $k\ge 1$, $0<p<\infty$ and $\alpha>0$, or $p=\infty$ and $\alpha\ge 1$, and in both cases $1/\tau := \alpha + 1/p$.

The *fat B-space* $\mathbb{B}_\tau^{\alpha k}(\mathcal{T})$ is defined as the set of all functions $f\in L_\tau$ such that

$$\|f\|_{\mathbb{B}_\tau^{\alpha k}(\mathcal{T})} := \Big(\sum_{\Delta\in\mathcal{T}} (|\Delta|^{-\alpha}\omega_k(f,\Omega_\Delta)_\tau)^\tau\Big)^{1/\tau} < \infty, \tag{3.27}$$

where $\Omega_\Delta := \Omega_\Delta^1$ is the union of all triangles in $\mathcal{T}_m$ which have a common vertex with Δ, if $\Delta\in\mathcal{T}_m$ (see (3.5)). The principle difference between the skinny B-space $\mathcal{B}_\tau^{\alpha k}(\mathcal{T})$ and fat B-space $\mathbb{B}_\tau^{\alpha k}(\mathcal{T})$ is that the Ω_Δ's in definition (3.27) overlap substantially. One can prove that, for $0<\eta<p$,

$$\|f\|_{\mathbb{B}_\tau^{\alpha k}(\mathcal{T})} \approx \Big(\sum_{\Delta\in\mathcal{T}} (|\Delta|^{1/p-1/\eta}\omega_k(f,\Omega_\Delta)_\eta)^\tau\Big)^{1/\tau}$$

Also, similarly as for skinny B-spaces, one can introduce equivalent norms via local polynomial projections. For more details, see [21].

3.8. B-spaces on Compact Polygonal Domains in $\mathbb{R}^2$

Slim, skinny, and fat B-spaces can be introduced on an arbitrary compact polygonal domain $E \subset \mathbb{R}^2$ quite similarly as on $\mathbb{R}^2$. We shall only define the slim B-spaces on such domain E, placing the emphasis on the distinctions from the case $E = \mathbb{R}^2$.

Suppose $\mathcal{T} = \bigcup_{m=0}^{\infty} \mathcal{T}_m$ is an LR(or better)-triangualation on E (see §3.1) and let $\mathcal{S} = (\mathcal{S}_m)_{m\ge 0}$ be a spline multiresolution over $\mathcal{T}$ with a hierarchical family of basis functions $\Phi := \Phi_{\mathcal{T}}$ (see §3.3-§3.4). Assuming that p, α, and τ are as in §3.5 and elsewhere, we define the *slim B-space* $B_\tau^\alpha(\mathcal{S}) = B_\tau^\alpha(\Phi)$ on E as the set of all functions $f \in L_\tau(E)$ such that

$$|f|_{B_\tau^\alpha(\Phi)} := \Big(\sum_{\Delta\in\mathcal{T}} (|\Delta|^{-\alpha} \mathbb{S}_\Delta(f)_\tau)^\tau \Big)^{1/\tau} < \infty,$$

where $\mathbb{S}_\Delta(f)_\tau$ is the error of L_τ-approximation to f from $\mathcal{S}_m$ on Ω_Δ^ℓ, if $\Delta \in \mathcal{T}_m$ (see (3.16)). Clearly, $|\cdot|_{B_\tau^\alpha}$ is a semi-norm if $\tau \ge 1$ and a semi-quasi-norm if $\tau < 1$. Since $B_\tau^\alpha(\Phi)$ is continuously embedded in $L_p(E)$, it is natural to define a (quasi-)norm in $B_\tau^\alpha(\mathcal{T})$ by

$$\|f\|_{B_\tau^\alpha(\Phi)} := \|f\|_p + |f|_{B_\tau^\alpha(\mathcal{T})}.$$

Equivalent norms similar to the ones from (3.21), (3.22), and (3.24) can be defined. The only difference would be that, in this case, $\Theta = \bigcup_{m=0}^{\infty} \Theta_m$ and the operators $Q_{m,\eta}(\cdot)$ and $q_{m,\eta}(\cdot)$ should be defined on E accordingly with the natural modification $Q_{-1,\eta}(f) := 0$.

For more details, we refer the reader to [22].

3.9. B-spaces over Triangulations and Besov Spaces: Miscellaneous

Comparison between different B-spaces over the same triangulation. Suppose $\mathcal{S}_{\mathcal{T}}$ is a spline multiresolution over $\mathcal{T}$ an SLR-triangualation of $\mathbb{R}^2$ and $\Phi_{\mathcal{T}}$ is a family of basis functions for $\mathcal{S}$ (see §3.3). Let $\Pi_k \subset \mathcal{S}_m \subset \mathcal{S}_m^{k,r}(\mathcal{T})$ ($m \in \mathbb{Z}$). Evidently, for $f \in L_\tau$ and $\Delta \in \mathcal{T}_m$, we have

$$E_k(f,\Delta)_\tau \le \mathbb{S}_\Delta(f,\mathcal{T})_\tau \le E_k(f,\Omega_\Delta^\ell)_\tau \le c \sum_{\Delta'\in\mathcal{T}_m,\, \Delta'\subset\Omega_\Delta^\ell} E_k(f,\Omega_{\Delta'})_\tau,$$

which implies

$$\|f\|_{\mathcal{B}_\tau^{\alpha k}(\mathcal{T})} \le c\|f\|_{B_\tau^\alpha(\Phi_{\mathcal{T}})} \le c\|f\|_{\mathbb{B}_\tau^{\alpha k}(\mathcal{T})}.$$

Comparison of regular B-spaces with Besov spaces. We first recall the definition of the classical Besov spaces on a set $E \subset \mathbb{R}^d$ with moduli of smoothness. The *Besov space* $B_q^s(L_p) := B_q^s(L_p(E))$, $s > 0$, $1 \le p,q \le \infty$, is defined as the set of all functions $f \in L_p(E)$ such that

$$|f|_{B_q^s(L_p)} := \left(\int_0^\infty (t^{-s}\omega_k(f,t)_p)^q \frac{dt}{t} \right)^{1/q} < \infty \tag{3.28}$$

with the L_q-norm replaced by the sup-norm if $q = \infty$, where $k := [s] + 1$ and $\omega_k(f,t)_p$ is the k-th modulus of smoothness of f in $L_p(E)$ (see (3.9)). The norm in $B^s_q(L_p)$ is usually defined by $\|f\|_{B^s_q(L_p)} := \|f\|_p + |f|_{B^s_q(L_p)}$. It is well-known that if in (3.28) k is replaced by any other $k > s$, then the resulting space would be the same with an equivalent norm. However, the situation is totally different when $p < 1$ (for more details, see §5) and this is a reason for introducing k as a parameter of the Besov spaces with the next definition. We define the space

$$B^{s,k}_q(L_p) := B^{s,k}_q(L_p(E)), \qquad 0 < p, q \le \infty,\ s > 0,\ k \ge 1,$$

as the Besov space $B^s_q(L_p(E))$ from above, where the parameters k and s are already set *independent* of each other.

For the theory of nonlinear (regular) spline approximation in $L_p(E)$, $0 < p \le \infty$, one can use the Besov space

$$B^{d\alpha,k}_\tau(L_\tau) := B^{d\alpha,k}_\tau(L_\tau(E))$$

with parameters set as elsewhere in this article: $k \ge 1$ and $1/\tau := \alpha + 1/p$, according to two specific choices of p and α: (a) $0 < p < \infty$ and $0 < \alpha < \infty$, or (b) $p = \infty$ and $1 \le \alpha < \infty$. Since $B^{d\alpha,k}_\tau(L_\tau)$ is embedded in L_p, it is natural to define the (quasi-)norm in $B^{d\alpha,k}_\tau(L_\tau)$ by

$$\|f\|_{B^{d\alpha,k}_\tau(L_\tau)} := \|f\|_p + |f|_{B^{d\alpha,k}_\tau(L_\tau)}.$$

However, if $E = \mathbb{R}^d$ and $|f|_{B^{d\alpha,k}_\tau(L_\tau)} < \infty$, then $\|f\|_p \le c|f|_{B^{d\alpha,k}_\tau(L_\tau)}$ and hence $\|f\|_{B^{d\alpha,k}_\tau(L_\tau)} \approx |f|_{B^{d\alpha,k}_\tau(L_\tau)}$. In the following, we shall restrict our attention to the case $E = \mathbb{R}^d$.

Notice that the smoothness parameters of the B-spaces and Besov spaces are normalized differently. For instance, the B-space $B^\alpha_\tau(\Phi)$ in dimension $d = 2$ corresponds to the Besov space $B^{2\alpha,k}_\tau(L_\tau)$.

It is often convenient to use the equivalence

$$\|f\|_{B^{d\alpha,k}_\tau(L_\tau)} \approx \Big(\sum_{m\in\mathbb{Z}} (2^{d\alpha m}\omega_k(f, 2^{-m})_\tau)^\tau\Big)^{1/\tau}, \tag{3.29}$$

which follows by the properties of $\omega_k(f,t)_\tau$.

Next, we give an equivalent norm of the Besov space $B^{d\alpha,k}_\tau(L_\tau)$ in terms of local polynomial approximation over dyadic boxes. We let D'_m denote the set of all dyadic boxes $I \subset \mathbb{R}^d$ of the form $I = \prod_{j=1}^d [\frac{\nu_j - 1}{2^m}, \frac{\nu_j}{2^m})$, $\nu_j \in \mathbb{Z}$, and let D''_m be the set of all shifts of boxes $I \in D'_m$ by the vector $e := (2^{-m-1}, \dots, 2^{-m-1})$, i.e., $D''_m := \{I + e : I \in D'_m\}$. We set $D_m := D'_m \bigcup D''_m$ and $D := \bigcup_{m\in\mathbb{Z}} D_m$. We now introduce the following norm

$$\mathbf{N}(f) := \Big(\sum_{I\in D} (|I|^{-\alpha}\omega_k(f, I)_\tau)^\tau\Big)^{1/\tau} \approx \Big(\sum_{I\in D} (|I|^{-\alpha}E_k(f, I)_\tau)^\tau\Big)^{1/\tau}, \tag{3.30}$$

where $E_k(f,I)_\tau$ is the error of L_τ-approximation to f on I from Π_k (see (3.6)). We have

$$\mathbf{N}(f) \approx \|f\|_{B_\tau^{d\alpha,k}(L_\tau)}, \tag{3.31}$$

which easily follows using (3.11).

As in the case of B-spaces (see §3.5-§3.7), the norm $\mathbf{N}(\cdot)$ from (3.30) can be modified as follows. We define

$$\mathbf{N}_\eta(f) := \Big(\sum_{I\in D}(|I|^{1/p-1/\eta}\omega_k(f,I)_\eta)^\tau\Big)^{1/\tau} \approx \Big(\sum_{I\in D}(|I|^{1/p-1/\eta}E_k(f,I)_\eta)^\tau\Big)^{1/\tau} \tag{3.32}$$

which in integral form gives

$$\mathbf{N}_\eta(f) \approx \left(\int_0^\infty \int_{\mathbb{R}^d} [t^{d(1/p-1/\eta)}\omega_k(f,B_t(x))_\eta]^\tau t^{-d-1}\, dxdt\right)^{1/\tau}, \tag{3.33}$$

where $B_t(x) := \{y \in \mathbb{R}^d : \|y-x\|_2 \le t\}$ or $B_t(x) := \{y \in \mathbb{R}^d : \|y-x\|_\infty \le t\}$. We have, for $0<\eta<p$,

$$\mathbf{N}_\eta(f) \approx \|f\|_{B_\tau^{d\alpha,k}(L_\tau)}. \tag{3.34}$$

Thus the semi-norm $\mathbf{N}_\eta(f)$ enables us to work in L_η with $\eta \ge 1$ if $p>1$ for all $\alpha>0$, while normally $\tau<1$.

Next, we compare different B-spaces over an arbitrary regular triangulation $\mathcal{T}^*$ of $\mathbb{R}^2$ (§3.1) with the corresponding Besov spaces on $\mathbb{R}^2$.

(a) Using (3.11), it easily follows that

$$\mathbb{B}_\tau^{\alpha k}(\mathcal{T}^*) = B_\tau^{2\alpha,k}(L_\tau), \qquad 0<\alpha<\infty, \tag{3.35}$$

with equivalent norms.

(b) Suppose $\Phi_{\mathcal{T}^*} = \{\varphi_\theta\}$ is a family of basis functions over $\mathcal{T}^*$ as in §3.3 such that $\Pi_k \subset \mathcal{S}_m \subset \mathcal{S}_m^{k,r}$ $(m\in\mathbb{Z})$, where $r\ge 0$ and $k>r$. We have

$$B_\tau^\alpha(\Phi_{\mathcal{T}^*}) = B_\tau^{2\alpha,k}(L_\tau), \quad 0<\alpha<r+1+1/p, \tag{3.36}$$

with equivalent norms. Furthermore, if a single basis function $\varphi_\theta \in \Phi_{\mathcal{T}^*}$ does not belong to C^{r+1}, then

$$\omega_k(\varphi_\theta,t)_\tau^\tau \approx \begin{cases} |E_\theta|^{\frac{1}{2}(1-(r+1)\tau)}\cdot t^{1+(r+1)\tau}, & \text{if } 0<t<|E_\theta|^{1/2} \\ |E_\theta|, & \text{if } t \ge |E_\theta|^{1/2}, \end{cases}$$

which implies $\|\varphi_\theta\|_{B_\tau^{2\alpha,k}(L_\tau)} = \infty$ if $\alpha \ge r+1+1/p$, while at the same time $\|\varphi_\theta\|_{B_\tau^\alpha(\Phi_{\mathcal{T}^*})} \approx \|\varphi_\theta\|_p$. Therefore, (3.36) is no longer valid if $\alpha \ge r+1+1/p$.

An interesting situation occurs when $p=\infty$ and $r=0$. Then there is no α for which (3.36) holds, since α must be at least one in this case. This is the case when $\Phi_{\mathcal{T}^*}$ is the set of all Courant elements generated by $\mathcal{T}^*$ (a regular triangulation).

(c) We have

$$\mathcal{B}_\tau^{\alpha k}(\mathcal{T}^*) = B_\tau^{2\alpha,k}(L_\tau), \qquad 0 < \alpha < 1/p, \tag{3.37}$$

with equivalent norms, which is no longer true if $\alpha \ge 1/p$. Moreover, for every $\Delta \in \mathcal{T}^*$ and $\alpha \ge 1/p$, we have $\|\mathbb{1}_\Delta\|_{B_\tau^{2\alpha,k}(L_\tau)} = \infty$, while $\|\mathbb{1}_\Delta\|_{\mathcal{B}_\tau^{\alpha k}(\mathcal{T}^*)} \approx \|\mathbb{1}_\Delta\|_p$.

We refer the reader to [11, 21] for more details on the connections between B-spaces and Besov spaces.

Comparison between B-spaces over different triangulations and Besov spaces. If $\mathcal{T}$ is an SLR-triangulation of $\mathbb{R}^2$, then the relationship between slim (or skinny) and fat B-spaces over $\mathcal{T}$ is quite similar to the one between slim B-spaces over regular triangulations and Besov spaces. For instance, if $\Phi_\mathcal{T}$ is the set of all Courant elements generated by $\mathcal{T}$, there exists a constant $\alpha_0 > 0$, depending on p and the parameters of $\mathcal{T}$ such that

$$\mathbb{B}_\tau^{\alpha,2}(\mathcal{T}) = B_\tau^\alpha(\Phi_\mathcal{T}), \quad \text{for } 0 < \alpha < \alpha_0,$$

with equivalent norms, and $B_\tau^\alpha(\Phi_\mathcal{T})$ is substantially larger than $\mathbb{B}_\tau^{\alpha,2}(\mathcal{T})$ when $\alpha \ge \alpha_0$.

If one compares a B_τ^α-space over an arbitrary triangulation with the corresponding Besov space $B_\tau^{2\alpha,k}(L_\tau)$ (or two B-spaces over different triangulations with each other), then everything changes dramatically. As we already mentioned in §3.1, there exist strong locally regular triangulations with extremely (uncontrollably) “skinny” triangles which cause problems to Besov spaces. More precisely, suppose φ_θ is the Courant element associated with a cell $\theta \in \Theta$ which is convex, and has length $l > 0$ and width ε with $0 < \varepsilon < l$. Simple calculations show that $\omega_2(\varphi_\theta, t)_\tau^\tau \approx \min\{l\varepsilon^{-\tau}t^{1+\tau}, l\varepsilon\}$, which readily implies $\|\varphi_\theta\|_{B_\tau^{2\alpha,2}(L_\tau)} \approx (l/\varepsilon)^\alpha \|\varphi_\theta\|_p$ if $0 < \alpha < 1 + 1/p$ and $\|\varphi_\theta\|_{B_\tau^{2\alpha,2}(L_\tau)} = \infty$ if $\alpha \ge 1 + 1/p$. At the same time, $\|\varphi_\theta\|_{B_\tau^\alpha(\Phi_\mathcal{T})} \approx \|\varphi_\theta\|_p$ for all $\alpha > 0$ with constants of equivalence independent of ε and l. Therefore, even for small α the Besov norm of a Courant element can be huge in comparison to its L_p-norm. This is why the Besov spaces are completely unsuitable for the theory of n-term spline approximation in the case of nonregular triangulations.

B-spaces in dimensions $d > 2$. Multilevel nested triangulations and B-spaces can be introduced much in the same way in dimensions $d > 2$. Naturally, the triangles should be replaced by simplices, making some of the geometric argumentation of this section more involved.

4. B-spaces Generated by Dyadic Partitions of $\mathbb{R}^d$

In this short section, we define the B-spaces needed in §6 for characterization of the rates of nonlinear piecewise polynomial approximation generated by dyadic partitions of $\mathbb{R}^d$ $(d > 1)$ or a box $\Omega \subset \mathbb{R}^d$. These spaces are quite similar to the skinny B-spaces from §3.6 and we shall skip some details.

Anisotropic dyadic partitions of $\mathbb{R}^d$ or Ω. We call $\mathcal{P} = \bigcup_{m\in\mathbb{Z}} \mathcal{P}_m$ a dyadic partition of $\mathbb{R}^d$ with levels $(\mathcal{P}_m)$ if the following conditions are fulfilled:

(a) Every level $\mathcal{P}_m$ is a partition of $\mathbb{R}^d$: $\mathbb{R}^d = \bigcup_{I\in\mathcal{P}_m} I$ and $\mathcal{P}_m$ consists of disjoint dyadic boxes of the form $I = \mathcal{I}_1 \times \cdots \times \mathcal{I}_d$, where each $\mathcal{I}_j$ is a semi-open dyadic interval $(\mathcal{I}_j = [(\nu-1)2^\mu, \nu 2^\mu))$, and $|I| = 2^{-m}$.

(b) The levels of $\mathcal{P}$ are nested, i.e., $\mathcal{P}_{m+1}$ is a refinement of $\mathcal{P}_m$.

(c) For any boxes $I', I'' \in \mathcal{P}$ there exists a box $I \in \mathcal{P}$ such that $I' \cup I'' \subset I$.

Also, we call $\mathcal{P} = \bigcup_{m\ge 0} \mathcal{P}_m$ a dyadic partition of Ω (Ω a dyadic box with $|\Omega| = 1$) if $\mathcal{P}_0 := \{\Omega\}$ and the levels $(\mathcal{P}_m)_{m\ge 1}$ satiffy conditions (a)-(b) from above with $\mathbb{R}^d$ replaced by Ω.

We note that the two children, say, $J_1, J_2 \in \mathcal{P}_{m+1}$ of any $I \in \mathcal{P}_m$ can be obtain by splitting I in two equal subboxes in d $(d > 1)$ different ways. Therefore, there is a huge variety of anisotropic dyadic partitions.

A typical property of the anisotropic dyadic partitions is that each level $\mathcal{P}_m$ of a partition $\mathcal{P}$ consists of dyadic boxes I with $|I| = 2^{-m}$ and at the same time there could be extremely (uncontrolably) long and narrow boxes in $\mathcal{P}_m$.

We denote by $\mathbb{L}_p := \mathbb{L}_p(\mathcal{P}, k)$ the closed in L_p span of the set $\{\mathbb{1}_I \cdot \mathcal{P}_I : P_I \in \Pi_k, I \in \mathcal{P}\}$. Evidently, $\mathbb{L}_p \subset L_p$ but it may happen that $\mathbb{L}_p \ne L_p$.

The B-spaces $\mathcal{B}^{\alpha k}_\tau(\mathcal{P})$ on $\mathbb{R}^d$. For the purposes of nonlinear approximation in L_p $(0 < p < \infty)$ from (discontinuous) piecewise polynomials generated by dyadic partitions of $\mathbb{R}^d$, we need the space $\mathcal{B}^{\alpha k}_\tau(\mathcal{P})$, where $k \ge 1$, $\alpha > 0$, and $1/\tau := \alpha + 1/p$. This space is defined as the set of all $f \in L_\tau(\mathbb{R}^d)$ such that

$$\|f\|_{\mathcal{B}^{\alpha k}_\tau(\mathcal{P})} := \Big(\sum_{I\in\mathcal{P}} (|I|^{-\alpha}\omega_k(f, I)_\tau)^\tau\Big)^{1/\tau} < \infty. \tag{4.1}$$

Whitney's theorem (see (3.8)) implies

$$\|f\|_{\mathcal{B}^{\alpha k}_\tau(\mathcal{P})} \approx \Big(\sum_{I\in\mathcal{P}} (|I|^{-\alpha} E_k(f, I)_\tau)^\tau\Big)^{1/\tau},$$

where $E_k(f, I)_\tau$ is the error of $L_\tau(I)$-approximation to f from Π_k.

As in §3.6, we have, for $0 < \eta < p$,

$$\|f\|_{\mathcal{B}^{\alpha,k}_\tau(\mathcal{P})} \approx \Big(\sum_{I\in\mathcal{P}} (|I|^{1/p-1/\eta}\omega_k(f, I)_\eta)^\tau\Big)^{1/\tau} \approx \inf_{f=\sum_{I\in\mathcal{P}} P_I} \Big(\sum_{I\in\mathcal{P}} \|P_I\|_p^\tau\Big)^{1/\tau}, \tag{4.2}$$

where $P_I \in \Pi_k$.

Let $P_{I,\eta} : L_\eta(I) \to \Pi_k$ be a projector (linear if $\eta \ge 1$) such that

$$\|f - P_{I,\eta}(f)\|_{L_\eta(I)} \le c E_k(f, I)_\eta, \quad \text{for } f \in L_\eta(I).$$

Set $P_{m,\eta}(f) := \sum_{I\in\mathcal{P}_m} \mathbb{1}_\Delta \cdot P_{I,\eta}(f)$. Clearly, $P_{m,\eta}(f)$ is a projector into $\mathcal{S}^{k,-1}(\mathcal{P}_m)$, the set of all piecewise polynomials of degree $< k$ over boxes of $\mathcal{P}_m$. We define

$$p_{m,\eta}(f) := p_{m,\eta}(f,\mathcal{P}) := P_{m,\eta}(f) - P_{m-1,\eta}(f) \in \mathcal{S}^{k,-1}(\mathcal{P}_m),$$

and set $p_{I,\eta}(f) := \mathbb{1}_I \cdot p_{m,\eta}(f)$ for $I \in \mathcal{P}_m$. We have, for $0 < \eta < p$,

$$\|f\|_{\mathcal{B}^{\alpha k}_\tau(\mathcal{P})} \approx \Big(\sum_{I\in\mathcal{P}} (|I|^{1/p-1/\eta}\|p_{I,\eta}(f)\|_\eta)^\tau\Big)^{1/\tau} \approx \Big(\sum_{I\in\mathcal{P}} \|p_{I,\eta}(f)\|_p^\tau\Big)^{1/\tau}. \tag{4.3}$$

The B-spaces $\mathcal{B}^{\alpha k}_\tau(\mathcal{P})$ on Ω. Only for convenience, we assume that $\Omega \subset \mathbb{R}^d$ is a dyadic box with $|\Omega| = 1$. We again assume that $0 < p < \infty$, $\alpha > 0$, $k \ge 1$, and $1/\tau := \alpha + 1/p$. Let $\mathcal{P} = \bigcup_{m\ge 0} \mathcal{P}_m$ be an arbitrary dyadic partition of Ω. We define the space $\mathcal{B}^{\alpha k}_\tau(\mathcal{P})$ on Ω as the set of all $f \in L_\tau(\Omega)$ such that

$$|f|_{B^{\alpha k}_\tau(\mathcal{P})} := \Big(\sum_{I\in\mathcal{P}} (|I|^{-\alpha}\omega_k(f,I)_\tau)^\tau\Big)^{1/\tau} < \infty. \tag{4.4}$$

Evidently, $|f + P|_{B^\alpha_\tau} = |f|_{B^\alpha_\tau}$ for $P \in \Pi_k$ and hence $|\cdot|_{B^\alpha_\tau}$ is a semi-norm if $\tau \ge 1$ and a semi-quasi-norm if $\tau < 1$. Since $\mathcal{B}^{\alpha k}_\tau(\mathcal{P})$ is continuously embedded in $L_p(\Omega)$, it is natural to define a norm in $\mathcal{B}^{\alpha k}_\tau(\mathcal{P})$ by

$$\|f\|_{\mathcal{B}^{\alpha k}_\tau(\mathcal{P})} := \|f\|_{L_p(\Omega)} + |f|_{\mathcal{B}^{\alpha k}_\tau(\mathcal{P})}.$$

Equivalent norms in $\mathcal{B}^{\alpha k}_\tau(\mathcal{P})$ similar to the ones from (4.2)-(4.3) can be introduced and utilized as well.

Remark. The relationship between the B-space $\mathcal{B}^{\alpha k}_\tau(\mathcal{P})$ and Besov space $B^{d\alpha,k}_\tau(L_\tau)$ is similar as the relationship between the corresponding skinny B-space and $B^{d\alpha,k}_\tau(L_\tau)$ (see §3.9). For more details, see [28].

5. B-spaces and Besov Spaces in Dimension $d = 1$

In this subsection, we show that the univariate B-spaces on $\mathbb{R}$ (with one exception) coincide with the corresponding Besov spaces and, therefore, they only provide additional equivalent norms for Besov spaces. Thus, in dimension $d = 1$, there is only one (super) scale of smoothness spaces which governs nonlinear spline approximation. This is a fundamental distinction between the univariate and multivariate spline approximation.

Weak locally regular (WLR) partitions of $\mathbb{R}^1$. We call $\mathcal{P} = \bigcup_{m\in\mathbb{Z}} \mathcal{P}_m$ a *multilevel weak locally regular partition* of $\mathbb{R}$ with levels $(\mathcal{P}_m)$ if the following conditions are fulfilled:

(a) Every level $\mathcal{P}_m$ is a partition of $\mathbb{R}$, i.e., $\mathbb{R} = \bigcup_{I\in\mathcal{P}_m} I$, and $\mathcal{P}_m$ consists of compact intervals with disjoint interiors.

(b) The levels $(\mathcal{P}_m)$ of $\mathcal{P}$ are nested, i.e., $\mathcal{P}_{m+1}$ is a refinement of $\mathcal{P}_m$.

(c) Each interval $I \in \mathcal{P}_m$ has at least two and at most M_0 children in $\mathcal{P}_{m+1}$, where $M_0 \geq 2$ is a constant.

(d) There exist constants $0 < r < \rho < 1$ such that for each $I \in \mathcal{P}_m$ and any child $I' \in \mathcal{P}_{m+1}$ of I

$$r|I| \leq |I'| \leq \rho|I|. \tag{5.1}$$

Locally regular (LR) partitions of $\mathbb{R}^1$. We call $\mathcal{P} = \bigcup_{m\in\mathbb{Z}} \mathcal{P}_m$ a *locally regular partition* of $\mathbb{R}$ if $\mathcal{P}$ is a WLR-partition of $\mathbb{R}$ and in addition to this $\mathcal{P}$ satisfies the following property:

(e) There exists a constant $0 < \delta \leq 1$ such that for each $I', I'' \in \mathcal{P}_m\,(m \in \mathbb{Z})$ with a common end point $\delta \leq |I'|/|I''| \leq \delta^{-1}$.

The set $\mathcal{D}$ of all dyadic intervals on $\mathbb{R}$ is an example of an LR-partition.

We next define the univariate slim, skinny, and fat B-spaces $B_\tau^{\alpha k}(\mathcal{P})$, $\mathcal{B}_\tau^{\alpha k}(\mathcal{P})$, and $\mathbb{B}_\tau^{\alpha k}(\mathcal{P})$ with parameters $k \geq 1$, $1/\tau := \alpha + 1/p$, and p and α according to the specific choices (as elsewhere): (a) $0 < p < \infty$ and $0 < \alpha < \infty$, or (b) $p = \infty$ and $1 \leq \alpha < \infty$.

The slim B-space $B_\tau^{\alpha k}(\mathcal{P})$. Let $\mathcal{P}$ be an LR-partition of $\mathbb{R}$ and let $\cdots < x_{-1}^{(m)} < x_0^{(m)} < x_1^{(m)} < \cdots$ be the end points of the intervals from $\mathcal{P}_m$. Fix $I \in \mathcal{P}_m$ and let $I =: [x_j^{(m)}, x_{j+1}^{(m)}]$ $(j \in \mathbb{Z})$. We denote by φ_I the B-spline of degree $k-1$ with knots $x_j^{(m)}, x_{j+1}^{(m)}, \dots, x_{j+k}^{(m)}$ (see [3] or [30]). Denote by $\mathcal{S}_m$ the spline space spanned by $\{\varphi_I\}_{I\in\mathcal{P}}$ $(\mathcal{S}_m \subset C^{k-2})$. For $I \in \mathcal{P}_m$, we denote $\Omega_I := [x_{j-k+1}^{(m)}, x_{j+k}^{(m)}]$ and by $\mathbb{S}_I(f)_\tau$ the error of $L_\tau(\Omega_I)$-approximation (local) to f from $\mathcal{S}_m$ (similarly as in (3.16)).

We define the *slim B-space* $B_\tau^{\alpha,k}(\mathcal{P})$ as the set of all functions $f \in L_\tau$ such that

$$\|f\|_{B_\tau^{\alpha k}(\mathcal{P})} := \Big(\sum_{I\in\mathcal{P}} (|I|^{-\alpha}\mathbb{S}_I(f)_\tau)^\tau\Big)^{1/\tau} < \infty.$$

Similarly as in §3.5, one can show that, for $0 < \eta < p$,

$$\|f\|_{B_\tau^{\alpha k}(\mathcal{P})} \approx \Big(\sum_{I\in\mathcal{P}} (|I|^{1/p-1/\eta}\mathbb{S}_I(f)_\eta)^\tau\Big)^{1/\tau} \approx \inf_{f=\sum_{I\in\mathcal{P}} c_I\varphi_I} \Big(\sum_{I\in\mathcal{P}} \|c_I\varphi_I\|_p^\tau\Big)^{1/\tau}.$$

Also, equivalent norms in $B_\tau^{\alpha k}(\mathcal{P})$ can be introduced via projections (quasi-interpolants) as in §3.5.

The skinny B-space $\mathcal{B}_\tau^{\alpha k}(\mathcal{P})$. Assuming that $\mathcal{P}$ is an WLR-partition of $\mathbb{R}$, we define the *skinny B-space* $\mathcal{B}_\tau^{\alpha k}(\mathcal{P})$ as the set of all $f \in L_\tau$ such that

$$\|f\|_{\mathcal{B}_\tau^{\alpha k}(\mathcal{P})} := \Big(\sum_{I\in\mathcal{P}} (|I|^{-\alpha}\omega_k(f,I)_\tau)^\tau\Big)^{1/\tau} < \infty.$$

Again as in §3.6 and §4, we have, for $0 < \eta < p$,

$$\|f\|_{\mathcal{B}_\tau^{\alpha k}(\mathcal{P})} \approx \Big(\sum_{I\in\mathcal{P}} (|I|^{1/p-1/\eta}\omega_k(f,I)_\eta)^\tau\Big)^{1/\tau} \approx \inf_{f=\sum_{I\in\mathcal{P}} P_I} \Big(\sum_{I\in\mathcal{P}} \|P_I\|_p^\tau\Big)^{1/\tau}, \tag{5.2}$$

where $P_I \in \Pi_k$.

The fat B-space $\mathbb{B}_\tau^{\alpha k}(\mathcal{P})$. We now assume that $\mathcal{P}$ is an LR-partition of $\mathbb{R}$ and define the *fat B-space* $\mathbb{B}_\tau^{\alpha,k}(\mathcal{P})$ as the set of all $f \in L_\tau$ such that

$$\|f\|_{\mathbb{B}_\tau^{\alpha k}(\mathcal{P})} := \Big(\sum_{I \in \mathcal{P}} (|I|^{-\alpha} \omega_k(f, \Omega_I)_\tau)^\tau \Big)^{1/\tau} < \infty,$$

where Ω_I is defined as above.

Besov spaces in dimension $d = 1$. The univariate Besov space $B_\tau^{\alpha,k}(L_\tau)$ on $\mathbb{R}$ is needed for the characterization of the approximation spaces of nonlinear spline approximation. We recall that (see §3.9) $B_\tau^{\alpha,k}(L_\tau) := B_\tau^{\alpha,k}(L_\tau(\mathbb{R}))$ is defined as the set of all functions $f \in L_\tau(\mathbb{R})$ such that

$$\|f\|_{B_\tau^{\alpha,k}(L_\tau)} := \left(\int_0^\infty (t^{-\alpha} \omega_k(f,t)_\tau)^\tau \frac{dt}{t} \right)^{1/\tau} < \infty, \tag{5.3}$$

where k and α are independent of each other. As elsewhere, we shall only use the spaces $B_\tau^{\alpha,k}(L_\tau)$ with parameters k, α, p, and τ set as above.

We next clarify why in the definition of $B_\tau^{\alpha,k}(L_\tau)$ above the parameters α and k are set to be independent. It is easily seen that the set of all $f \in L_\tau$ such that $\omega_k(f,t)_\tau = O(t^\alpha)$ is nontrivial if and only if $0 < \alpha \le \max\{k, k-1+1/\tau\}$ (see, e.g., [29]). Hence, if $1/\tau := \alpha + 1/p$, this condition imposes no restriction on α. Therefore, by choosing $1/\tau := \alpha + 1/p$ and allowing $\tau < 1$, it follows that $B_\tau^{\alpha,k}(L_\tau)$ is nontrivial for all $k \ge 1$. This property of Besov spaves is of fundamental importance for the theory of nonlinear spline approximation in dimension $d = 1$ (see §8).

Equivalence of B-spaces and Besov spaces ($d = 1$). The following theorem gives the precise conditions under which the B-spaces and Besov spaces from above coincide:

Theorem 7. *Suppose $0 < p < \infty$ and $0 < \alpha < \infty$, or $p = \infty$ and $\alpha \ge 1$, $1/\tau := \alpha + 1/p$, and $k \ge 1$.*

(a) *If $\mathcal{P}$ is a WLR partition of $\mathbb{R}$ and $0 < p < \infty$, then $\mathcal{B}_\tau^{\alpha k}(\mathcal{P}) = B_\tau^{\alpha,k}(L_\tau)$ with equivalent norms.*

(b) *If $\mathcal{P}$ is an LR partition of $\mathbb{R}$, then $\mathbb{B}_\tau^{\alpha k}(\mathcal{P}) = B_\tau^{\alpha,k}(L_\tau)$ with equivalent norms.*

(c) *If $\mathcal{P}$ is an LR partition of $\mathbb{R}$ and $k \ge 2$, then $B_\tau^{\alpha k}(\mathcal{P}) = B_\tau^{\alpha,k}(L_\tau)$ with equivalent norms.*

In all cases, the constants of equivalence depend on p, α, k, and the parameters of $\mathcal{P}$.

Remark. Claim (a) of Theorem 7 is no longer true when $p = \infty$. Indeed, it is easily seen that for any interval I, $\omega_k(\mathbb{1}_I, t)_\tau \approx \min\{t^{1/\tau}, |I|^{1/\tau}\}$. This yields $\|\mathbb{1}_I\|_{B_\tau^{\alpha,k}(L_\tau)} = \infty$, if $p = \infty$ ($\tau = 1/\alpha$), while $\|\mathbb{1}_I\|_{\mathcal{B}_\tau^{\alpha k}(\mathcal{P})} \approx 1$ if $I \in \mathcal{P}$.

Since we do not have a reference for this important theorem, we shall prove it in the following.

Proof of Theorem 7. We shall only prove part (a) of the theorem. The proof of (b)-(c) is similar.

Let $f \in \mathcal{B}_\tau^{\alpha k}(\mathcal{P})$, where $\mathcal{P}$ is a WLR-partition of $\mathbb{R}$. Then f can be represented in the form $f = \sum_{I\in\mathcal{P}} P_I \cdot \mathbb{1}_I$ with $P_I \in \Pi_k$ and (see (5.2))

$$\|f\|_{\mathcal{B}_\tau^{\alpha k}(\mathcal{P})} \approx \Big(\sum_{I\in\mathcal{P}} \|P_I\|_{L_p(I)}^\tau \Big)^{1/\tau}. \tag{5.4}$$

We denote $\mathcal{X}_j := \{I \in \mathcal{P} : 2^{-j-1} < |I| \le 2^{-j}\}$ and $f_j := \sum_{I\in\mathcal{X}_j} P_I \cdot \mathbb{1}_I$ ($j \in \mathbb{Z}$). It follows by (5.1) that each $x \in \mathbb{R}$ may belong to $\le (\log_2 \frac{1}{\rho})^{-1}$ intervals from $\mathcal{X}_j$, i.e., only finitely many intervals from $\mathcal{X}_j$ may overlap at a time.

Now, we fix $j \in \mathbb{Z}$. The above property of $\mathcal{X}_j$ implies the following obvious estimate ($t > 0$)

$$\omega_k(f_j, t)_\tau^\tau \le c \sum_{I\in\mathcal{X}_j} \|P_I\|_{L_\tau(I)}^\tau \le c \sum_{I\in\mathcal{X}_j} |I|^{\alpha\tau} \|P_I\|_{L_p(I)}^\tau \le c 2^{-j\alpha\tau} \sum_{I\in\mathcal{X}_j} \|P_I\|_{L_p(I)}^\tau, \tag{5.5}$$

where we used the norm equivalence of polynomials and that $1/\tau := \alpha + 1/p$. On the other hand, since $\Delta_h^k(P, x) = 0$ for any polynomial $P \in \Pi_k$, we readily obtain (see [27])

$$\int_{\mathbb{R}} |\Delta_h^k(f_j, x)|^\tau \, dx \le c \sum_{I\in\mathcal{X}_j} \min\{|h|, |I|\} \|P_I\|_{L_\infty(I)}^\tau, \quad h \in \mathbb{R}. \tag{5.6}$$

Using (5.6) and again the norm equivalence of polynomials, we obtain, for $m \ge j$,

$$\omega_k(f_j, 2^{-m})_\tau^\tau \le c 2^{-m} \sum_{I\in\mathcal{X}_j} |I|^{-\tau/p} \|P_I\|_{L_p(I)}^\tau \le c 2^{-m} 2^{j\tau/p} \sum_{I\in\mathcal{X}_j} \|P_I\|_{L_p(I)}^\tau. \tag{5.7}$$

We set $\lambda := \min\{\tau, 1\}$ and use (5.5) and (5.7) to obtain

$$\begin{aligned} \omega_k(f, 2^{-m})_\tau^\lambda &\le \textstyle\sum_{j\in\mathbb{Z}} \omega_k(f_j, 2^{-m})_\tau^\lambda \le c \sum_{j=m+1}^\infty 2^{-j\alpha\lambda} \Big(\sum_{I\in\mathcal{X}_j} \|P_I\|_{L_p(I)}^\tau \Big)^{\lambda/\tau} \\ &\quad + c 2^{-m\lambda/\tau} \textstyle\sum_{j=-\infty}^m 2^{j\lambda/p} \Big(\sum_{I\in\mathcal{X}_j} \|P_I\|_{L_p(I)}^\tau \Big)^{\lambda/\tau}. \end{aligned}$$

Inserting this in (3.29), we find

$$
\begin{aligned}
\|f\|^\tau_{B^{\alpha,k}_\tau(L_\tau)} &\le c\sum_{m\in\mathbb{Z}}\Big[\sum_{j=m+1}^{\infty}2^{-(j-m)\alpha\lambda}\Big(\sum_{I\in\mathcal{X}_j}\|P_I\|^\tau_{L_p(I)}\Big)^{\lambda/\tau}\Big]^{\tau/\lambda}\\
&+ c\sum_{m\in\mathbb{Z}}\Big[\sum_{j=-\infty}^{m}2^{-(m-j)\lambda/p}\Big(\sum_{I\in\mathcal{X}_j}\|P_I\|^\tau_{L_p(I)}\Big)^{\lambda/\tau}\Big]^{\tau/\lambda}\\
&\le c\sum_{I\in\mathcal{P}}\|P_I\|^\tau_{L_p(I)}\le c\|f\|^\tau_{\mathcal{B}^{\alpha k}_\tau(\mathcal{P})},
\end{aligned}
$$

where we used the well-known Hardy inequalities (see, e.g., Lemma 3.10 in [29]) as well as (5.4) and $1/\tau := \alpha + 1/p$.

To prove the opposite estimate we shall make use of (3.10)-(3.11) and the sets $(\mathcal{X}_j)$ from above. We have

$$
\begin{aligned}
\|f\|^\tau_{\mathcal{B}^{\alpha k}_\tau(\mathcal{P})} &= \sum_{I\in\mathcal{P}}|I|^{-\alpha\tau}\omega_k(f,I)^\tau_\tau\le c\sum_{j\in\mathbb{Z}}2^{j\alpha\tau}\sum_{I\in\mathcal{X}_j}\omega_k(f,I)^\tau_\tau\\
&\le c\sum_{j\in\mathbb{Z}}2^{j\alpha\tau}\sum_{I\in\mathcal{X}_j}\frac{1}{|I|}\int_0^{|I|}\int_I|\Delta^k_h(f,x,I)|^\tau\,dx\,dh\\
&\le c\sum_{j\in\mathbb{Z}}2^{j(\alpha\tau+1)}\int_0^{2^{-j}}\Big(\sum_{I\in\mathcal{X}_j}\int_I|\Delta^k_h(f,x,I)|^\tau\,dx\Big)\,dh\\
&\le c\sum_{j\in\mathbb{Z}}2^{j\alpha\tau}\omega_k(f,2^{-j})^\tau_\tau\le c\|f\|^\tau_{B^{\alpha,k}_\tau(L_\tau)}.
\end{aligned}
$$

Therefore, $\|f\|_{\mathcal{B}^{\alpha k}_\tau(\mathcal{P})}\approx\|f\|_{B^{\alpha,k}_\tau(L_\tau)}$.

Remark. A theorem similar to Theorem 7 holds on $[a,b]$ as well.

6. Nonlinear Piecewise Polynomial Approximation Generated by Dyadic Partitions in $\mathbb{R}^d$

Piecewise polynomials generated by a single dydic partition. Here, we shall utilize the B-spaces introduced in §4 to characterize the rates of nonlinear approximation in L_p from piecewise polynomials generated by an arbitrary anisotropic dyadic partition $\mathcal{P}$ of $\mathbb{R}^d$ $(d>1)$. The same results with almost identical proofs hold on any box $\Omega\subset\mathbb{R}^d$.

We let $\Sigma^k_n(\mathcal{P})$ denote the nonlinear set consisting of all piecewise polynomial functions of the form

$$s=\sum_{I\in\Lambda_n}\mathbb{1}_I\cdot P_I,$$

where $P_I \in \Pi_k$, $\Lambda_n \subset \mathcal{P}$, and $\#\Lambda_n \le n$ (Λ_n may very). We denote by $\sigma_n^k(f,\mathcal{P})_p$ the error of L_p-approximation to $f \in L_p(\mathbb{R}^d)$ from $\Sigma_n^k(\mathcal{P})$:

$$\sigma_n^k(f,\mathcal{P})_p := \inf_{s\in\Sigma_n^k(\mathcal{P})} \|f-s\|_p.$$

To characterize the approximation spaces generated by $(\sigma_n^k(f,\mathcal{P})_p)$, we follow the general scheme described in §2. Namely, we first establish Jackson and Bernstein estimates and then the desired characterization of the approximation spaces follows immediately by interpolation. Throughout this section, we assume that $\mathcal{P}$ is an arbitrary dyadic partition of $\mathbb{R}^d$ (§4), $0 < p < \infty$, $\alpha > 0$, $k \ge 1$, and $1/\tau := \alpha + 1/p$. We shall use the B-spaces $B_\tau^{\alpha k}(\mathcal{P})$ defined in §4.

Theorem 8 (Jackson estimate). *If $f \in B_\tau^{\alpha k}(\mathcal{P})$, then*

$$\sigma_n^k(f,\mathcal{P})_p \le cn^{-\alpha}\|f\|_{B_\tau^{\alpha k}(\mathcal{P})}$$

with $c = c(\alpha,p,k,d)$.

Theorem 9 (Bernstein estimate). *If $s \in \Sigma_n^k(\mathcal{P})$, then*

$$\|s\|_{B_\tau^{\alpha k}(\mathcal{P})} \le cn^{\alpha}\|\varphi\|_p$$

with $c = c(\alpha,p,k,d)$.

We denote by $A_q^\gamma(L_p,\mathcal{P})$ the approximation space generated by $(\sigma_n(f,\mathcal{P})_p)$ (see (2.1)). The Jackson and Bernstein inequalities from Theorems 8-9, combined with Theorems 1-2 of §2, imply the following characterization of the approximation spaces $A_q^\gamma(L_p,\mathcal{P})$:

Theorem 10. *If $0 < \gamma < \alpha$ and $0 < q \le \infty$, then*

$$A_q^\gamma(L_p,\mathcal{P}) = (\mathbb{L}_p(\mathcal{P},k), B_\tau^{\alpha k}(\mathcal{P}))_{\gamma/\alpha,q}$$

with equivalent norms (for the definition of $\mathbb{L}_p(\mathcal{P},k)$, see §4).

In one specific case the interpolation space as well as the corresponding approximation space can be identified as a B-space.

Theorem 11. *Suppose $\mathcal{P}$ is a dyadic partition of $\mathbb{R}^d$, $k \ge 1$, $1 \le p < \infty$, and $1/\tau := \alpha + 1/p$. Let $0 < \alpha < \beta$ and $1/\lambda := \beta + 1/p$. We have*

$$\Big(\mathbb{L}_p(\mathcal{P},k), B_\lambda^{\beta k}(\mathcal{P})\Big)_{\alpha/\beta,\tau} = B_\tau^{\alpha k}(\mathcal{P}) = A_\tau^\alpha(L_p,\mathcal{P})$$

with equivalent norms.

The analogue of this result for Besov spaces is well-known (see [17]).

Nonlinear approximation from the library $\{\Sigma_n^k(\mathcal{P})\}_{\mathcal{P}}$. We denote

$$\sigma_n^k(f)_p := \inf_{\mathcal{P}} \sigma_n^k(f,\mathcal{P})_p,$$

where the infimum is taken over all dyadic partitions $\mathcal{P}$. The following theorem is immediate from the Jackson estimate in Theorem 8.

Theorem 12. *If* $\inf_{\mathcal{P}} \|f\|_{B^{\alpha k}_{\tau}(\mathcal{P})} < \infty$, *then*

$$\sigma^k_n(f)_p \le cn^{-\alpha} \inf_{\mathcal{P}} \|f\|_{B^{\alpha k}_{\tau}(\mathcal{P})}$$

with $c = c(\alpha, k, p, d)$.

As we show later in this section, in a natural discrete setting, there exists an effective algorithm for finding a partition $\mathcal{P}^*$ which minimizes $\|f\|_{B^{\alpha k}_{\tau}(\mathcal{P})}$, for a given f, over all dyadic partitions $\mathcal{P}$.

It is an *open problem* to characterize the approximation spaces generated by $(\sigma^k_n(f)_p)$. See §10 for further discussion of this and other related open problems for approximation from libraries of basis families.

Remarks. There exists another technique that can be employed for the proof of Theorem 8. This method is called "split and merge" and was introduced in [8] and used for nonlinear approximation of functions from the space $BV(\mathbb{R}^2)$. It was further used in [23]. Also, the modulus $\mathcal{W}(f,t)_{\sigma,p}$, used in [23] which is a generalization of a characteristic from [26] ($d = 1$), can be generalized and utilized for anisotropic dyadic partitions $\mathcal{P}$.

Nonlinear n-term approximation from anisotropic Haar bases. An anisotropic Haar basis is naturally associated with each anisotropic dyadic partition $\mathcal{P}$ of a box Ω in $\mathbb{R}^d$ (or $\mathbb{R}^d$). For the sake of simplicity, we shall consider Haar bases only on a dyadic box Ω with sides parallel to the coordinate axes and $|\Omega| = 1$. Then any dyadic partition of Ω is of the form $\mathcal{P} = \bigcup_{m=0}^{\infty} \mathcal{P}_m$. Let $I \in \mathcal{P}$ and $I =: \mathcal{I}_1 \times \cdots \times \mathcal{I}_d$. Suppose I is split (in $\mathcal{P}$) by dividing in half the νth ($1 \le \nu \le d$) side of I. Then we define $H_I := \mathbb{1}_{\mathcal{I}_1} \times \cdots \times H_{\mathcal{I}_\nu} \times \cdots \times \mathbb{1}_{\mathcal{I}_d}$, where $H_{\mathcal{I}_\nu}$ is the univariate Haar function supported on $\mathcal{I}_\nu$ and normalized in L_∞. We need to add the characteristic function of Ω to the collection of the above defined Haar functions. To this end we denote $I^0 := I_0 := \Omega$ and include both I^0 and I_0 in $\mathcal{P}_0$ and $\mathcal{P}$. We define $H_{I^0} := \mathbb{1}_{I^0}$.

Thus $\mathcal{H}_{\mathcal{P}} := \{H_I : I \in \mathcal{P}\}$ is the Haar basis associated with $\mathcal{P}$. We let $\mathcal{H} := \{\mathcal{H}_{\mathcal{P}}\}_{\mathcal{P}}$ denote the collection (library) of all anisotropic Haar bases on Ω.

The most important properties of the Haar bases $\{\mathcal{H}_{\mathcal{P}}\}_{\mathcal{P}}$ are the following:

(a) For each dyadic partition $\mathcal{P}$ of Ω the Haar basis $\mathcal{H}_{\mathcal{P}}$ is an unconditional basis for $\mathbb{L}_p(\mathcal{P})$ (the linear span of $\mathcal{H}_{\mathcal{P}}$ in $L_p(\Omega)$), $1 < p < \infty$.

(b) The B-norm of $f \in B^{\alpha,1}_{\tau}(\mathcal{P})$ ($1 < p < \infty$, $\alpha > 0$, $1/\tau := \alpha + 1/p$) can be characterized by means of its Haar coefficients using $\mathcal{H}_{\mathcal{P}}$: Every $f \in B^{\alpha,1}_{\tau}(\mathcal{P})$ can be represented uniquely in the form $f = \sum_{I\in\mathcal{P}} c_I(f) H_I$ with $c_I(f) := |I|^{-1} \int_I f H_I$ and

$$\|f\|_{B^{\alpha,1}_{\tau}(\mathcal{P})} \approx \mathcal{N}(f, \mathcal{H}_{\mathcal{P}}) := \Big(\sum_{I\in\mathcal{P}} |I|^{-\alpha\tau} \|c_I(f)H_I\|_\tau^\tau\Big)^{1/\tau} = \Big(\sum_{I\in\mathcal{P}} \|c_I(f)H_I\|_p^\tau\Big)^{1/\tau}. \tag{6.1}$$

For a given partition $\mathcal{P}$, we denote by $\widehat{\Sigma}_n(\mathcal{P})$ the nonlinear set of all functions s of the form

$$s = \sum_{I\in\Lambda_n} a_I H_I,$$

where $\Lambda_n \subset \mathcal{P}$ and $\#\Lambda_n \le n$. The error $\widehat{\sigma}_n(f, \mathcal{H}_\mathcal{P})_p$ of nonlinear L_p-approximation to f from $\widehat{\Sigma}_n(\mathcal{P})$ is defined by

$$\widehat{\sigma}_n(f, \mathcal{H}_\mathcal{P})_p := \inf_{s \in \widehat{\Sigma}_n(\mathcal{P})} \|f - s\|_{L_p(\Omega)}.$$

Clearly, $\widehat{\Sigma}_n(\mathcal{P}) \subset \Sigma^1_{2n}(\mathcal{P})$ and hence $\sigma^1_{2n}(f, \mathcal{P})_p \le \widehat{\sigma}_n(f, \mathcal{H}_\mathcal{P})_p$. We denote by $\widehat{A}^\gamma_q := \widehat{A}^\gamma_q(L_p, \mathcal{H}_\mathcal{P})$ the approximation spaces generated by nonlinear n-term approximation from $\mathcal{H}_\mathcal{P}$ (see (2.1)). Now, the main goal is to characterize the approximation spaces $\widehat{A}^\gamma_q$, which reduces to establishing Jackson and Bernstein inequalities and interpolation (see §2).

Theorem 13. *Suppose $\mathcal{P}$ is an arbitrary dyadic partition of Ω and let $1 < p < \infty$, $\alpha > 0$, and $1/\tau := \alpha + 1/p$. Then the following Jackson and Bernstein inequalities hold:*

$$\begin{aligned} \widehat{\sigma}_n(f, \mathcal{H}_\mathcal{P})_p &\le cn^{-\alpha}\|f\|_{B^{\alpha,1}_\tau(\mathcal{P})}, \qquad f \in B^{\alpha,1}_\tau(\mathcal{P}), \\ \|s\|_{B^{\alpha,1}_\tau(\mathcal{P})} &\le cn^{\alpha}\|s\|_{L_p(\Omega)}, \qquad s \in \widehat{\Sigma}_n(\mathcal{P}), \quad c = c(\alpha, p, d). \end{aligned} \tag{6.2}$$

Therefore, if $0 < \gamma < \alpha$ and $0 < q \le \infty$, then

$$\widehat{A}^\gamma_q(L_p, \mathcal{H}_\mathcal{P}) = (\mathbb{L}_p(\mathcal{P}), B^{\alpha,1}_\tau(\mathcal{P}))_{\gamma/\alpha, q} = A^\gamma_q(L_p, \mathcal{H}_\mathcal{P})$$

with equivalent norms (*see* Theorem 10).

Nonlinear n-term approximation from the library $\mathcal{H} := \{\mathcal{H}_\mathcal{P}\}$. We denote by $\widehat{\sigma}_n(f)_p$ the error of n-term approximation to $f \in L_p$ from the best basis in $\mathcal{H}$, i.e.,

$$\widehat{\sigma}_n(f)_p := \inf_{\mathcal{P}} \widehat{\sigma}_n(f, \mathcal{H}_\mathcal{P})_p.$$

The following theorem is immediate from the Jackson estimate (6.2):

Theorem 14. *If $\inf_\mathcal{P} \|f\|_{B^{\alpha,1}_\tau(\mathcal{P})} < \infty$, then*

$$\widehat{\sigma}_n(f)_p \le cn^{-\alpha} \inf_{\mathcal{P}} \|f\|_{B^{\alpha,1}_\tau(\mathcal{P})}$$

with $c = c(p, \alpha, d)$.

The scheme for nonlinear n-term approximation of a given function $f \in L_p(\Omega)$ from the library $\mathcal{H} := \{\mathcal{H}_\mathcal{P}\}$ of all anisotropic Haar bases consists of two steps:

(i) Find a basis $\mathcal{H}(f) \in \mathcal{H}$ which minimizes the $B^{\alpha,1}_\tau$-norm of f.

(ii) Run a threshold algorithm for near best n-term approximation from $\mathcal{H}(f)$.

The most significant fact in this part is that, in a natural discrete setting, there is an effective algorithm for best Haar basis selection, which we present below.

The above approximation scheme requires a priori information (or an estimate) for the smoothness order $\alpha > 0$ of the function f (which is being approximated) with respect to the optimal $B^{\alpha,1}_{\tau}$-scale. How to determine this smoothness is an *open problem.*

Best Haar basis or best B-space selection. We next describe a fast algorithm for best anisotropic Haar basis or best B-space selection in the discrete case of dimension $d = 2$. This algorithm is well-known (see, e.g., [19] and the references therein).

We consider the set $\mathcal{X}_n$ of all functions $f : [0,1)^2 \to \mathbb{R}$ which are constants on each of the $2^n \times 2^n$ "pixels"

$$I = [(i-1)2^{-n}, i2^{-n}) \times [(j-1)2^{-n}, j2^{-n}), \qquad 1 \le i, j \le 2^n.$$

Denote by $\mathcal{D}_n$ the set of all such pixels on $[0,1)^2$. We let $\mathbb{P}_n$ denote the set of all dyadic partitions $\mathcal{P}$ of $[0,1)^2$ such that $\mathcal{P}_{2n} = \mathcal{D}_n$ and we shall consider $\mathcal{P}$ terminated at level $2n$. Thus $\mathcal{P} = \bigcup_{m=0}^{2n} \mathcal{P}_m$.

Motivated by the result from Theorem 14, our next goal is to find, for a given $f \in \mathcal{X}_n$, a dyadic partition $\mathcal{P}^* := \mathcal{P}^*(f) \in \mathbb{P}_n$ which minimizes the B-norm $\mathcal{N}(f, \mathcal{P}) := \mathcal{N}(f, \mathcal{H}_{\mathcal{P}})$ from (6.1). Evidently, for $\mathcal{P} \in \mathbb{P}_n$, $\mathcal{H}_{\mathcal{P}}$ is an orthogonal basis for the linear space $\mathcal{X}_n$ and, therefore,

$$f = \sum_{I \in \mathcal{P}} c_I(f) H_I \quad \text{with} \quad c_I(f) := |I|^{-1} \int_I f H_I.$$

We denote $d(I, \mathcal{P}) := |I|^{-\alpha\tau+1} |c_I(f)|^{\tau}$. Also, we set $d_0(I) := d(I, \mathcal{P})$ if I is subdivided, say, horizontally in $\mathcal{P}$, and $d_1(I) := d(I, \mathcal{P})$ if I is subdivided vertically in $\mathcal{P}$. We have, for the B-norm from (6.1),

$$\mathcal{N}(f, \mathcal{P})^{\tau} = \sum_{I \in \mathcal{P}} d(I, \mathcal{P}) =: D(\mathcal{P}).$$

For a given dyadic box J, we denote by $\mathbb{P}_J$ the set of all dyadic partitions $\mathcal{P}_J$ of J which are subpartitions of partitions from $\mathbb{P}_n$. Similarly as above, we set

$$D(\mathcal{P}_J) := \sum_{I \in \mathcal{P}_J} d(I, \mathcal{P}_J).$$

We next describe a fast algorithm for finding a partition $\mathcal{P}^* \in \mathbb{P}_n$ which minimizes the B-norm $\mathcal{N}(f, \mathcal{P})$. The idea of this construction is to proceed from finer to coarser levels minimizing $D(\mathcal{P}_J)$ for every dyadic box J at every step. More precisely, we use the following recursive procedure. We first consider all dyadic boxes J with $|J| = 2^{-2n+1}$. Each box J like this is the union of two adjacent pixels and, hence, it can be subdivided in exactly one way. Thus $\mathcal{P}_J^*$ is uniquely determined. Now, suppose that we have already found all partitions $\mathcal{P}_J^*$ of all dyadic boxes J with $|J| \le 2^{-\mu}$ $(0 < \mu < 2n)$ which minimize $D(\mathcal{P}_J)$ over all partitions $\mathcal{P}_J \in \mathbb{P}_J$. Let J be an arbitrary dyadic box such that $|J| = 2^{-\mu+1}$. There are two cases to be considered.

Case I: One of the sides of J is of length 2^{-n}. Then there is only one way to subdivide J and, hence, $\mathcal{P}_J^*$ and $\min D(\mathcal{P}_J) = D(\mathcal{P}_J^*)$ are uniquely determined.

Case II: Both sides of J are of length $> 2^{-n}$. Then J can be subdivided in two possible ways: horizontally or vertically and, therefore, J has two sets of children. Let us denote by J_1° and J_2° the children of J obtain when dividing J horizontally and J_1' and J_2' the children of J obtain when dividing J vertically. The key observation is that

$$\min_{\mathcal{P}_J} D(\mathcal{P}_J) = \min\left\{D(\mathcal{P}_{J_1^\circ}^*) + D(\mathcal{P}_{J_2^\circ}^*) + d_0(I), D(\mathcal{P}_{J_1'}^*) + D(\mathcal{P}_{J_2'}^*) + d_1(I)\right\}.$$

Therefore, if $\min_{\mathcal{P}_J} D(\mathcal{P}_J)$ is attained when J is (first) subdivided horizontally, then $\mathcal{P}_J^* = \mathcal{P}_{J_1^\circ}^* \cup \mathcal{P}_{J_2^\circ}^* \cup \{J\}$ will be an optimal partition of J and $\mathcal{P}_J^* = \mathcal{P}_{J_1'}^* \cup \mathcal{P}_{J_2'}^* \cup \{J\}$ will be optimal in the other case. We process like this every dyadic box of area $2^{-\mu+1}$ and this completes the recursive procedure. After finitely many steps, we find a partition $\mathcal{P}^*$ of Ω which minimizes $D(\mathcal{P}) = \mathcal{N}(f,\mathcal{P})^\tau$.

Every $f \in X_n$ belongs to any (discrete) space $B_\tau^{\alpha,1}(\mathcal{P})$ and we have, by Theorem 14,

$$\widehat{\sigma}_m(f)_p \le c m^{-\alpha} \inf_{\mathcal{P}\in\mathbb{P}_n} \|f\|_{B_\tau^{\alpha,1}(\mathcal{P})}, \qquad m = 1, 2, \ldots$$

Once the smoothness parameter $\alpha > 0$ is fixed, the above algorithm provides a dyadic partition which minimizes the $B_\tau^\alpha(\mathcal{P})$-norm of f. It is a problem to find the optimal smoothness order α of a given function f.

Several observations are in order. Fix $f \in \mathcal{X}_n$. Evidently, the number of all coefficients $c_I(f)$ (or Haar functions H_I) that participate in the representations of f using all anisotropic Haar bases is $\le 2N$, where $N := 2^{2n}$ is the number of the pixels. Moreover, these coefficients can be found by $O(N)$ operations. For fixed indices α and τ, only $O(N)$ operations are needed to find a Haar basis $\mathcal{H}(f)$ which minimizes the $B_\tau^{\alpha,1}$-norm $\mathcal{N}(f,\mathcal{P})$. Another $O(N \ln N)$ operations (mainly for ordering the coefficients) are needed for finding a near best n-term approximation to f in L_p $(1 < p < \infty)$ from the best Haar basis $\mathcal{H}(f)$.

Remark. The above idea for best basis selection applies immediately for best B-space selection, namely, for the selection of a partition $\mathcal{P}^*$ which minimizes the B-norm $\|f\|_{B_\tau^{\alpha k}(\mathcal{P})}$ of a given function f, when $k > 1$.

For the results of this section, we refer the reader to [28].

7. Nonlinear n-term Approximation from Hierarchical Sequences of Spline Bases over Triangulations in $\mathbb{R}^2$

Nonlinear n-term approximation from a single hierarchical sequence of bases in $\mathbb{R}^2$. Let $\mathcal{T}$ be a locally regular (or better) triangulation of $\mathbb{R}^2$. Suppose $\Phi := \Phi_{\mathcal{T}}$ is a family of basis functions associated with a spline multiresolution over $\mathcal{T}$ (see §3.3-§3.4). Notice that Φ is not a basis; Φ is redundant.

We consider nonlinear n-term approximation from Φ in $L_p(\mathbb{R}^2)$ $(0 < p \le \infty)$, where we identify $L_\infty(\mathbb{R}^2)$ as $C_0(\mathbb{R}^2)$. We let $\Sigma_n(\Phi)$ denote the nonlinear set consisting of all splines s of the form

$$s = \sum_{\theta \in \mathcal{M}} a_\theta \varphi_\theta,$$

where $\mathcal{M} \subset \Theta(\mathcal{T})$, $\#\mathcal{M} \le n$, and $\mathcal{M}$ may vary. We denote by $\sigma_n(f, \Phi)_p$ the error of L_p-approximation to $f \in L_p(\mathbb{R}^2)$ from $\Sigma_n(\Phi)$:

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

The primary goal is to characterize the approximation spaces $A^\gamma_q(\Phi, L_p)$ (see (2.1)) generated by nonlinear n-term approximation from Φ. To this end we next establish a pair of companion Jackson and Bernstein estimates which utilize the slim B-spaces $B^\alpha_\tau(\Phi)$, introduced in §3.5. As elsewhere in this article, we shall assume that $0 < p < \infty$ and $\alpha > 0$ or $p = \infty$ and $\alpha \ge 1$. In both cases, $1/\tau := \alpha + 1/p$.

Theorem 15 (Jackson estimate). *If* $f \in B^\alpha_\tau(\Phi)$, *then*

$$\sigma_n(f, \Phi)_p \le cn^{-\alpha} \|f\|_{B^\alpha_\tau(\Phi)} \tag{7.1}$$

with c *depending only on* p, α, *and the parameters of* $\mathcal{T}$ *and* $\Phi_\mathcal{T}$.

Theorem 16 (Bernstein estimate). *If* $s \in \Sigma_n(\Phi)$, *then*

$$\|s\|_{B^\alpha_\tau(\Phi)} \le cn^\alpha \|s\|_p \tag{7.2}$$

with c *depending only on* p, α, *and the parameters of* $\mathcal{T}$ *and* $\Phi_\mathcal{T}$.

The following characterization of the approximation spaces follows by the Jackson-Bernstein estimates (7.1)-(7.2), using Theorems 1-2 of §2:

Theorem 17. *If* $0 < \gamma < \alpha$ *and* $0 < q \le \infty$, *then*

$$A^\gamma_q(\Phi, L_p) = (L_p, B^\alpha_\tau(\Phi))_{\frac{\gamma}{\alpha}, q}$$

with equivalent norms.

As in §6 in one specific case, the interpolation space as well as the corresponding approximation space can be identified as a B-space.

Theorem 18. *Suppose* $0 < p < \infty$ *and* $\alpha > 0$ *or* $p = \infty$ *and* $\alpha > 1$, *and let* $\tau := (\alpha + 1/p)^{-1}$. *Then*

$$A^\alpha_\tau(\Phi, L_p) = B^\alpha_\tau(\Phi)$$

with equivalent norms.

The following interpolation result is immediate from Theorem 17 and Theorem 18.

Corollary 1. *Suppose p, α, and $\tau =: \tau(\alpha)$ are as in the hypothesis of Theorem 18, and let $\beta > \alpha$ and $\tau(\beta) := (\beta + 1/p)^{-1}$. Then*

$$(L_p, B^{\beta}_{\tau(\beta)}(\Phi))_{\frac{\alpha}{\beta}, \tau(\alpha)} = B^{\alpha}_{\tau(\alpha)}(\Phi)$$

with equivalent norms.

Approximation from the library $\{\Phi_{\mathcal{T}}\}_{\mathcal{T}}$. An important element of our concept for nonlinear spline approximation is the introduction of another level of nonlinearity by allowing the triangulation $\mathcal{T}$ to vary. For a given SRL(or LR)-triangulation $\mathcal{T}$, let $\Phi_{\mathcal{T}}$ be a family of spline basis functions associated with a spline multiresolution, like the ones considered in §3.3-§3.4. Now, without changing the nature of the basis elements from $\Phi_{\mathcal{T}}$, we let $\mathcal{T}$ vary and obtain a collection (library) of basis families $\{\Phi_{\mathcal{T}}\}_{\mathcal{T}}$. We denote

$$\sigma_n(f)_p := \inf_{\mathcal{T}} \sigma_n(f, \Phi_{\mathcal{T}})_p,$$

where the infimum is taken over all SLR-triangulations $\mathcal{T}$ with fixed parameters and we also assume that the parameters of $\Phi_{\mathcal{T}}$ are fixed. The following theorem is immediate from the Jackson estimate in Theorem 15.

Theorem 19. *If $\inf_{\mathcal{T}} \|f\|_{B^{\alpha}_{\tau}(\Phi_{\mathcal{T}})} < \infty$, then*

$$\sigma_n(f)_p \le cn^{-\alpha} \inf_{\mathcal{T}} \|f\|_{B^{\alpha}_{\tau}(\Phi_{\mathcal{T}})}$$

with c depending only on p, α, and the parameters of $\mathcal{T}$ and $\Phi_{\mathcal{T}}$.

This theorem gives rise to several interesting problems and, in particular, to the problem for finding a triangulation $\mathcal{T}^* := \mathcal{T}^*_f$ which minimizes the B-norms $B^{\alpha}_{\tau}(\Phi_{\mathcal{T}})$ of a given function f. For further and more complete discussion of this and other related problems, see §10.

Nonlinear approximation from discontinuous piecewise polynomials over multilevel triangulations. Similarly as in §6-§7, one can consider nonlinear n-term approximation from discontinuous piecewise polynomials over WLR-triangulations and results similar to Theorems 15-19 hold true. The only difference is that the slim B-spaces $B^{\alpha k}_{\tau}(\Phi_{\mathcal{T}})$ should be replaced by the skinny B-spaces $\mathcal{B}^{\alpha k}_{\tau}(\mathcal{T})$. For more details, see [21].

Approximation on polygonal domains of $\mathbb{R}^2$. The results of this section hold for nonlinear spline approximation on polygonal domains of $\mathbb{R}^2$ with a natural and obvious adaptation of the setting and the spaces (see [22]).

Remarks. For the results of this section, we refer the reader to [11, 21, 22]. In the case $0 < p < \infty$, Theorem 15 follows by Theorem 4 of §2. The proof of this theorem when $p = \infty$ can be carried out as the proofs of Theorems 24-25 of §9 (for more details, see Theorem 4.1 of [22]). In the case of regular triangulations and approximation from box-splines, the results from this section imply the results from [15, 18, 23] which involve Besov spaces, but are more complete since they do not impose any restrictions on the approximation rates.

8. Nonlinear Univariate Spline Approximation

In this section, we consider three types of piecewise polynomial (spline) approximation in dimension $d = 1$ and show that in all cases (with one exception only) the corresponding approximation spaces are the same and can be characterized by means of Besov spaces and interpolation. We begin by introducing the necessary notation.

(a) Nonlinear univariate n-term spline approximation. In this part, we assume that $\mathcal{P}$ is an LR-partition of $\mathbb{R}$ (§5) and $\Phi_{\mathcal{P}}^k = \{\varphi_I\}_{I\in\mathcal{P}}$ is the set of all B-splines of degree $< k$ ($k \geq 2$) as in §5. Notice that $\Phi_{\mathcal{P}}^k$ is not a basis; it is redundant. We consider nonlinear n-term approximation from $\Phi_{\mathcal{P}}^k$ in $L_p(\mathbb{R})$ ($L_\infty(\mathbb{R}) := C_0(\mathbb{R})$). We let $\Sigma_n^k(\mathcal{P})$ denote the nonlinear set consisting of all splines s of the form $s = \sum_{I\in\Lambda_n} a_I\varphi_I$, where $\Lambda_n \subset \mathcal{P}$ and $\#\Lambda_n \leq n$ (Λ_n may vary). We denote by $\sigma_n^k(f,\mathcal{P})_p$ the error of $L_p(\mathbb{R})$-approximation to f from $\Sigma_n^k(\mathcal{P})$:

$$\sigma_n^k(f,\mathcal{P})_p := \inf_{s\in\Sigma_n^k(\mathcal{P})} \|f - s\|_p.$$

(b) Nonlinear univariate n-term approximation from discontinuous piecewise polynomials. Suppose that $\mathcal{P}$ is a WLR-partition of $\mathbb{R}$ (§5) and $\tilde{\Sigma}_n^k(\mathcal{P})$ is the set of all piecewise polynomials s of the form $s = \sum_{I\in\Lambda_n} P_I$, where $\mathcal{P}_I \in \Pi_k$, $\Lambda_n \subset \mathcal{P}$, and $\#\Lambda_n \leq n$. We let $\tilde{\sigma}_n^k(f,\mathcal{P})_p$ denote the error of L_p-approximation to $f \in L_p(\mathbb{R})$ from $\tilde{\Sigma}_n^k(\mathcal{P})$ (replacing the uniform norm by the L_∞-norm when $p = \infty$):

$$\tilde{\sigma}_n^k(f,\mathcal{P})_p := \inf_{s\in\tilde{\Sigma}_n^k(\mathcal{P})} \|f - s\|_p.$$

(c) Free knot piecewise polynomial approximation. We now consider nonlinear approximation from (discontinuous) free knot piecewise polynomials in $L_p(\mathbb{R})$. We denote by $S(k,n)$ the set of all piecewise polynomial functions of degree $< k$ on $\mathbb{R}$ with $n+1$ free knots, i.e., $s \in S(k,n)$ if there exist points (knots) $x_0 < x_1 < \cdots < x_n$ such that $s = \sum_{j=1}^n \mathbb{1}_{I_j} \cdot P_j$, where $I_j := [x_{j-1}, x_j)$ and $P_j \in \Pi_k$. (The values of s at the knots $\{x_j\}$ are insignificant since we use the L_∞ norm instead of the uniform norm if s is discontinuous.) We denote by $s_n^k(f)_p$ the error of L_p-approximation to $f \in L_p(\mathbb{R})$ from $S(k,n)$:

$$s_n^k(f)_p := \inf_{s\in S(k,n)} \|f - s\|_p.$$

Evidently, $\Sigma_n^k(\mathcal{P}) \subset S(k,2n)$ and $\tilde{\Sigma}_n^k(\mathcal{P}) \subset S(k,2n)$ and hence, for $f \in L_p$,

$$s_{2n}^k(f)_p \le \sigma_n^k(f,\mathcal{P})_p \quad \text{and} \quad s_{2n}^k(f)_p \le \tilde{\sigma}_n^k(f,\mathcal{P})_p. \tag{8.1}$$

Now, we denote by $A_q^\gamma(\sigma^k(\mathcal{P}), L_p)$, $A_q^\gamma(\tilde{\sigma}^k(\mathcal{P}), L_p)$, and $A_q^\gamma(s^k, L_p)$ the approximation spaces generated by $(\sigma_n^k(f,\mathcal{P})_p)$, $(\tilde{\sigma}_n^k(f,\mathcal{P})_p)$, and $(s_n^k(f)_p)$, respectively (see (2.1)).

The main goal again is to characterize these approximation spaces. To this end we next establish Jackson and Bernstein estimates, which utilize the Besov spaces $B_\tau^{\alpha,k}(L_\tau)$, defined in §5. For the remainder of this section, we assume that $k \ge 1$, $0 < p < \infty$ and $\alpha > 0$ or $p = \infty$ and $\alpha \ge 1$. In both cases, $1/\tau := \alpha + 1/p$.

Theorem 20 (Jackson estimates). *If $f \in B_\tau^{\alpha,k}(L_\tau)$, then*

$$\sigma_n^k(f,\mathcal{P})_p \le cn^{-\alpha}\|f\|_{B_\tau^{\alpha,k}(L_\tau)} \qquad (k \ge 2), \tag{8.2}$$

where $\mathcal{P}$ is any LR-partition of $\mathbb{R}$,

$$\tilde{\sigma}_n^k(f,\mathcal{P})_p \le cn^{-\alpha}\|f\|_{B_\tau^{\alpha,k}(L_\tau)}, \tag{8.3}$$

where $\mathcal{P}$ is any WLR-partition of $\mathbb{R}$, and hence

$$s_n^k(f)_p \le cn^{-\alpha}\|f\|_{B_\tau^{\alpha,k}(L_\tau)}, \tag{8.4}$$

where each c is independent of f and n.

Theorem 21 (Bernstein estimates). (a) *If $0 < p < \infty$ and $s \in S(k,n)$, then*

$$\|s\|_{B_\tau^{\alpha,k}(L_\tau)} \le cn^\alpha\|s\|_p, \qquad c = c(p,\alpha,k). \tag{8.5}$$

(b) *If $p = \infty$ and $s \in \Sigma_n^k(\mathcal{P})$, where $k \ge 2$ and $\mathcal{P}$ is an LR-partition of $\mathbb{R}$, then*

$$\|s\|_{B_\tau^{\alpha,k}(L_\tau)} \le cn^\alpha\|s\|_\infty \tag{8.6}$$

where c depends only on p, α, k and the parameters of $\mathcal{P}$ (§5).

Remark. Estimate (8.5) is not valid when $p = \infty$ even when s is continuous. Indeed, consider a spline s defined by $s(x) := 1$ on $[0,1]$, $s(x) := 0$ on $(-\infty,-\varepsilon]\cup[1+\varepsilon,\infty)$, and with maximal smoothness and $\|s\|_\infty = 1$. It is easily seen that $\|s\|_{B_\tau^{\alpha,k}(L_\tau)} \to \infty$ as $\varepsilon \to 0$, if $\tau = 1/\alpha$, and hence (8.5) fails when $p = \infty$.

In view of Theorems 1-2 of §2, the Jackson-Bernstein estimates from Theorems 20-21 imply the following characterization of the approximation spaces:

Theorem 22. *Suppose $\mathcal{P}'$ is an LR-partition and $\mathcal{P}''$ is a WLR-partition of $\mathbb{R}$.*

(a) *Let $0 < p < \infty$ and $\alpha > 0$. If $0 < \gamma < \alpha$ and $0 < q \le \infty$, then*

$$A_q^\gamma(\sigma^k(\mathcal{P}'), L_p) = A_q^\gamma(\tilde{\sigma}^k(\mathcal{P}''), L_p) = A_q^\gamma(s^k, L_p) = (L_p, B_\tau^{\alpha,k}(L_\tau))_{\frac{\gamma}{\alpha},q}$$

with equivalent norms.

(b) *Let $p = \infty$, $\alpha \ge 1$ and $k \ge 2$. If $0 < \gamma < \alpha$ and $0 < q \le \infty$, then*

$$A_q^\gamma(\sigma^k(\mathcal{P}'), L_\infty) = (L_\infty, B_\tau^{\alpha,k}(L_\tau))_{\frac{\gamma}{\alpha},q}$$

with equivalent norms ($L_\infty := C_0$).

Results similar to the ones from Theorem 11 of §6, and Theorem 18 and Corollary 1 of §7 hold true for the above types of approximation. We skip the details.

Remarks. We first note that the approximation spaces $A_q^\gamma(s^k, L_\infty)$ cannot be characterized by Besov spaces as in Theorem 22. The reason for this is that the Bernstein estimate (8.5) is no longer valid when $p = \infty$ (see the remark after Theorem 21). However, as Theorem 22 shows, in all other cases of nonlinear univariate spline approximation the corresponding approximation spaces are the same and are governed by the scale of Besov spaces $B_\tau^{\alpha,k}(L_\tau)$, where $1/\tau := \alpha + 1/p$. In view of the results from §6-§7, the situation in the univariate case is quite unique: A scale of Besov spaces governs all rates of all kinds of nonlinear spline approximation (except for one case). The key role here is played by the Bernstein estimate (8.5).

Estimates (8.2)-(8.3) and (8.6) can be proved similarly as estimate (7.1)-(7.2) of §7 are proved (see [11, 21, 22]), using additionally the equivalence of the B-spaces and Besov spaces from Theorem 7 of §5. Estimate (8.4) is immediate from (8.2) or (8.3), taking into account (8.1). For the proof of (8.5), see [27]. (See also [15] and [18].)

9. Algorithms for n-term Spline Approximation

In this section, we present three algorithms for n-term Courant element approximation, which are call "Threshold", "Trim and Cut", and "Push the Error" algorithms. We consider approximation in $L_p(E)$ ($0 < p \le \infty$), where E a compact polygonal domain in $\mathbb{R}^2$. Throughout this subsection, we assume that $\mathcal{T} = \bigcup_{m=0}^\infty \mathcal{T}_m$ is a LR-triangulation of E (see §3.1) and $\Phi_\mathcal{T} = \{\varphi_\theta\}_{\theta\in\Theta}$ is the set of all Courant element generated by $\mathcal{T}$. Note that $\Theta = \bigcup_{m=0}^\infty \Theta_m$.

The primary goal is, by using the B-spaces and the related techniques, to develop (or refine) algorithms for nonlinear n-term approximation so that the new algorithms be capable of achieving the rate of the best approximation.

From the description of the three algorithms below, it will become clear that they can be applied immediately to nonlinear n-term approximation from an arbitrary hierarchical family of basis functions Φ associated with a spline multiresolution. We do not consider the general case here only for simplicity.

Decomposition step for all approximation algorithms. The first step of each of the three approximation algorithms that we consider in this section is a decomposition step. This step is not trivial since the set $\Phi_{\mathcal{T}} := \{\varphi_\theta\}_{\theta\in\Theta}$ of all Courant elements is redundant. For each algorithm, it is crucial to have a sufficiently efficient initial representation of the function f which is being approximated. This means that the representation of f should allow a realization of the corresponding B-norm. As we indicated in §3.8, all theorems from §3.5 and, in particular, Theorems 5-6 have almost identical analogies for a polygonal domain E. We use a representation of f similar to (3.25) from Theorem 5 of §3.5, which utilizes the operators $Q_{m,\eta}(\cdot)$ and $q_{m,\eta}(\cdot)$ from (3.15) and (3.23). Thus we have an initial desirable sparse representation of f of the form

$$f = \sum_{\theta\in\Theta} b_\theta \varphi_\theta, \qquad b_\theta = b_\theta(f), \tag{9.1}$$

which allows a realization of the B-norm:

$$\|f\|_{B^\alpha_\tau(\Phi_{\mathcal{T}})} \approx \Big(\sum_{\theta\in\Theta} \|b_\theta(f)\varphi_\theta\|_p^\tau\Big)^{1/\tau}$$

(see (3.24) and Theorem 6 of §3.5). Without loss of generality we may assume (when needed) that there is a final level Θ_L ($L < \infty$) in (9.1).

9.1. "Threshold" Algorithm ($p < \infty$ only)

This algorithm utilizes the usual thresholding strategy used for n-term approximation from a basis in L_p ($1 < p < \infty$). The resulting procedure performs very well due to the sparse representation realized by the first step.

Description of the algorithm.

Step 1. (*Decompose*) We use the decomposition of $f \in L_p(E)$ from (9.1).

Step 2. (*Select the n largest terms*) We order the terms $\{b_\theta\varphi_\theta\}_{\theta\in\Theta}$ in a sequence $(b_{\theta_j}\varphi_{\theta_j})_{j=1}^\infty$ so that

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

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

Error estimation for the "Threshold" algorithm.
We denote the corresponding error of approximation of this threshold algorithm by $\mathbb{A}_n^T(f)_p := \|f - A_n^T(f)_p\|_p$. The following theorem is immediate from Theorem 4 of §2.

Theorem 23. *If $f \in B^\alpha_\tau(\mathcal{T})$, $\alpha > 0$, $1/\tau := \alpha + 1/p$ $(0 < p < \infty)$, then*

$$\mathbb{A}^T_n(f)_p \le cn^{-\alpha}\|f\|_{B^\alpha_\tau(\mathcal{T})}, \tag{9.2}$$

where c depends on α, p, and the parameters of $\mathcal{T}$.

Remark. The main drawback of the "Threshold" algorithm is that it is not applicable to approximation in the uniform norm since the constant $c = c(\alpha, p)$ in (9.2) tends to infinity as $p \to \infty$ and the performance of the algorithm deteriorates as p gets large. The obvious reason for this behavior is that f can be built out of many terms $\{b_\theta\varphi_\theta\}$ which have small coefficients and are supported at the same location. These terms can pile up to an essential contribution but the algorithm will fail to anticipate their future significance.

9.2. "Trim & Cut" Algorithm

The idea of this algorithm has its origins in the proof of the Jackson estimate in [18] (see §5). The approximation considered there is by wavelets or splines over a uniform partition in the uniform norm.

Description of the "Trim & Cut" algorithm in the case $p = \infty$.

Step 1. (*Decompose*) We use the common decomposition of $f \in L_p(E)$ given in (9.1).

Step 2. (*Organize the cells of Θ into manageable trees Θ^ν*) This is a procedure for coloring the elements of Θ with K colors ν, so that no two Courant elements of the same color from the same level have supports that intersect, in fact corresponding cells of the same color will have a tree structure with set inclusion as the order relation. This allows us to partition Θ into a disjoint union of sets Θ^ν $(1 \le \nu \le K)$, and correspondingly organize f as the sum $f = \sum_{\nu=1}^K f_\nu$, where $f_\nu := \sum_{\theta\in\Theta^\nu} b_\theta\varphi_\theta$.

Lemma 2 (Coloring lemma). *For any multilevel-triangulation $\mathcal{T}$ of E, the set $\Theta := \Theta(\mathcal{T})$ of all cells generated by $\mathcal{T}$ can be represented as a finite disjoint union of its subsets $(\Theta^\nu)_{\nu=1}^K$ with $K = K(N_0, M_0)$ (N_0 is the maximal valence and M_0 is the maximal number of children of a triangle in $\mathcal{T}$), such that each Θ^ν has a tree structure with respect to set inclusion, i.e., if θ', $\theta'' \in \Theta^\nu$ with $(\theta')^\circ \cap (\theta'')^\circ \ne \emptyset$, then either $\theta' \subset \theta''$ or $\theta'' \subset \theta'$.*

Fix $\varepsilon > 0$ and let $\varepsilon^* := \frac{\varepsilon}{2K}$, where K is from the above lemma.

Step 3. (*"Trimming" of Θ^ν $(1 \le \nu \le K)$ with ε^**) We trim with ε^* each Θ^ν, starting from the finest level Θ^ν_J and proceeding to the coarsest level. Namely, we remove from Θ^ν every cell $\theta^\diamond$ such that

$$\sum_{\theta\subset\theta^\diamond} |b_\theta| \le \varepsilon^*.$$

We denote by Γ^ν the set of all $\theta \in \Theta^\nu$ which have been retained after completing this procedure.

Step 4. (*Partition the remaining trees into "segments"*) We cut each Γ^ν into a set Σ^ν of disjoint segments σ of the form $(\theta_j)_{j=i}^{i+\mu}$, $\mu \geq 0$, so that each segment satisfies exactly one of the following conditions:

(a) σ consists of a single "significant cell": $|b_{\theta_i}| > \varepsilon^*$ (case of $\mu = 0$),
(b) σ is a "significant segment":

$$\sum_{j=i}^{i+\mu-1} |b_{\theta_j}| \leq \varepsilon^*, \text{ but } \sum_{j=i}^{i+\mu} |b_{\theta_j}| > \varepsilon^*, \text{ (case of } \mu > 0),$$

(c) σ is an "remnant segment": $\sum_{j=i}^{l} |b_{\theta_j}| \leq \varepsilon^*$.

Step 5. (*Rewriting elements from certain segments of* Σ^ν) Let $\sigma = (\theta_j)_{j=1}^{\mu}$ be any segment from Σ^ν and suppose that the finest cell θ_μ of σ belongs to Θ_m. We rewrite the Courant elements $(\sum_{j=1}^{\mu} b_{\theta_j}\varphi_{\theta_j})$ of the segment at its finest (m-th) level, finding coefficients (c_θ) such that

$$\sum_{\theta\in\Theta_m, \theta^\circ\cap\theta_\mu\neq\emptyset} c_\theta\varphi_\theta = \sum_{j=1}^{\mu} b_{\theta_j}\varphi_{\theta_j} \quad \text{on} \quad \theta_\mu.$$

We denote $\mathcal{X}_\sigma := \{\theta \in \Theta_m : \theta^\circ \cap \theta_\mu \neq \emptyset \text{ and } \theta \subset \theta_1\}$. Obviously, if $\mu = 1$ (i.e., the segment consists of a single cell), then the coefficient remains unchanged and $\mathcal{X}_\sigma = \sigma = \{\theta_1\}$. Observe in any case that $\#\mathcal{X}_\sigma \leq N_0 + 1$ and $\cup_{\theta\in\mathcal{X}_\sigma}\theta \subset \theta_1$. Finally, set $\Sigma := \cup_{\nu=1}^{K}\Sigma^\nu$, and correspondingly define

$$A_\varepsilon^{TC}(f) := \sum_{\sigma\in\Sigma}\sum_{\theta\in\mathcal{X}_\sigma} c_\theta\varphi_\theta \tag{9.3}$$

as our approximant produced by the "Trim & Cut" algorithm.

Error estimation for the "Trim & Cut" algorithm (Case $p = \infty$).

Suppose that the "Trim & Cut" procedure has been applied to a function f with $\varepsilon > 0$, and $A_\varepsilon^{TC}(f) = \sum_{\theta\in\Lambda_\varepsilon} c_\theta\varphi_\theta$ is the resulting approximant from (9.3), where $\Lambda_\varepsilon = \cup_{\sigma\in\Sigma}\mathcal{X}_\sigma$. We denote

$$n(\varepsilon) := n_f(\varepsilon) := \#\Lambda_\varepsilon, \qquad \mathbb{A}_{n(\varepsilon)}^{TC}(f)_\infty := \|f - A_\varepsilon^{TC}(f)\|_\infty,$$

and

$$\mathbb{A}_n^{TC}(f)_\infty := \inf\left\{\mathbb{A}_{n(\varepsilon)}^{TC}(f)_\infty : n(\varepsilon) \leq n\right\}.$$

Note that each of these quantities depends implicitly on $\mathcal{T}$.

The following theorem indicates that the "Trim & Cut" algorithm provides optimal rate of convergence.

Theorem 24. *If $f \in B^\alpha_\tau(\mathcal{T})$, $\alpha \ge 1$, $\tau := 1/\alpha$, then for each $\varepsilon > 0$*

$$\mathbb{A}^{TC}_{n(\varepsilon)}(f)_\infty \le \varepsilon \quad \text{and} \quad n(\varepsilon) \le c\,\varepsilon^{-\tau}\|f\|^\tau_{B^\alpha_\tau(\mathcal{T})},$$

where $c = c(N_0, M_0, \alpha)$. Therefore,

$$\mathbb{A}^{TC}_n(f)_\infty \le c\,n^{-\alpha}\|f\|_{B^\alpha_\tau(\mathcal{T})}.$$

Remark. There is a version of the "Trim & Cut" algorithm for approximation in L_p $(0 < p < \infty)$, which is quite similar to the above algorithm $(p = \infty)$ but we do not present here (see [22]).

9.3. "Push the Error" Algorithm in the Uniform Norm

The idea of this algorithm to our knowledge first appeared in [14]. We present here the refined "Push the Error" from [22] which achieves optimal rates of approximation (the rate of convergence of the best approximation).

Description of the algorithm.

Step 1. (*Decompose*) For $f \in C(E)$ initially represented by (9.1), we may assume without loss of generality that there exists a finest level Θ_J $(J > 0)$ such that f is written as

$$f = \sum_{j=0}^{J} \sum_{\theta\in\Theta_j} b_\theta \varphi_\theta. \tag{9.4}$$

Step 2. (*"Prune the shrubs"*) We fix $\varepsilon > 0$ and let $\varepsilon^* := \varepsilon/2$. The goal of this step is to discard small insignificant terms $b_\theta\varphi_\theta$ in the representation of f from (9.4) but insuring that the resulting uniform error is at most ε^*. We shall denote by Γ the set of all retained cells. In addition, we shall construct a set $\Gamma_f \subset \Gamma$, consisting of "final cells" in Γ.

First, we need to introduce a organizational concept as a replacement for the tree structures of §9.2. We shall say (figuratively) that a cell $\theta \in \Theta$ *sits* on another cell $\theta^\diamond \in \Theta$, if θ is at least as fine as $\theta^\diamond$ and its interior (denoted by θ°) intersects the interior of $\theta^\diamond$. Furthermore, for $\theta^\diamond \in \Theta$, we denote the collection of all cells which sit on $\theta^\diamond$ by

$$\mathcal{Y}_{\theta^\diamond} := \{\theta \in \Theta : \theta^\circ \cap \theta^\diamond \ne \emptyset \ \text{ and } \ \text{level}\,(\theta) \ge \text{level}\,(\theta^\diamond)\}.$$

The procedure of Step 2 will begin at the finest level and proceed to the coarsest, level by level, constructing sets Γ_f and Γ. To initialize the procedure we put into Γ_f all *significant* cells $\theta \in \Theta_J$, i.e. such that $|b_\theta| > \varepsilon^*$. We place in Γ any cell from Θ_J which sits on a cell from Γ_f.

The inductive step proceeds as follows. Suppose that all cells from Θ_j with levels $j > m$ $(0 \le m < J)$, have already been processed. We now describe how to process Θ_m. We place into Γ_f all cells $\theta^\diamond \in \Theta_m$ which satisfy

$$\sum_{\theta\in\mathcal{Y}_{\theta^\diamond}} |b_\theta| > \varepsilon^*,$$

and for which there is no $\theta \in \Gamma_f$ from a higher level which sits on $\theta^\diamond$. A cell $\theta^\diamond$ from Θ_m is placed in Γ if there is a cell θ in the current Γ_f which sits on $\theta^\diamond$. Obviously, a cell $\theta^\diamond$ from Θ_m is discarded and not placed in Γ if

$$\sum_{\theta \in \mathcal{Y}_{\theta^\diamond}} |b_\theta| \leq \varepsilon^*,$$

and there is no $\theta \in \Gamma_f$ from level m or finer which sits on $\theta^\diamond$.

The procedure is terminated after Θ_0 is processed and *Step 2* of the algorithm is completed.

We set $f_\Gamma := \sum_{\theta \in \Gamma} b_\theta \varphi_\theta$ and define $a_\theta := \begin{cases} b_\theta, & \text{if } \theta \in \Gamma \\ 0, & \text{if } \theta \in \Theta \setminus \Gamma. \end{cases}$ Then

$$f_\Gamma = \sum_{\theta \in \Theta} a_\theta \varphi_\theta. \tag{9.5}$$

It follows from the construction that

$$\|f - f_\Gamma\|_\infty \leq \varepsilon^*. \tag{9.6}$$

Step 3. (*"Push the Error"*) We now process cells of f_Γ with ε^*, starting from the coarsest level Θ_0 and continuing to finer levels. The outcome of this step will be an approximant $\mathcal{A} := \mathcal{A}_\varepsilon^P(f)$ of the form

$$\mathcal{A} = \sum_{j=0}^{J} \mathcal{A}_j := \sum_{j=0}^{J} \sum_{\theta \in \Lambda_j} d_\theta \varphi_\theta, \tag{9.7}$$

where $\Lambda_j \subset \Theta_j$ and Λ_j will depend on f.

We shall use the notation

$$\mathcal{X}_{\theta^\diamond} := \{\theta \in \Theta : \theta^\circ \cap \theta^\diamond \neq \emptyset \ \text{ and } \ \text{level}\,(\theta) = \text{level}\,(\theta^\diamond)\}.$$

We start from the representation of f_Γ in (9.5). We define $\tilde{\Lambda}_0$ as the set of all $\theta \in \Theta_0$ such that $|a_\theta| > \varepsilon^*$ ($\|\varphi_\theta\|_\infty = 1$) and set $\Lambda_0 := \bigcup_{\theta \in \tilde{\Lambda}_0} \mathcal{X}_\theta$. We denote

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

For each $\theta^\diamond \in \Theta_j$, $\varphi_{\theta^\diamond}$ can be represented as a linear combination of φ_θ's with $\theta \in \Theta_{j+1}$. We use this to rewrite (represent) all remaining terms $a_\theta \varphi_\theta$, $\theta \in \Theta_0 \setminus \Lambda_0$, at the next level and add the resulting terms to the corresponding terms $a_\theta \varphi_\theta$, $\theta \in \Theta_1$. We denote by $d_\theta \varphi_\theta$, $\theta \in \Theta_1$, the new terms and therefore obtain a representation of f in the form

$$f = \mathcal{A}_0 + \sum_{\theta \in \Theta_1} d_\theta \varphi_\theta + \sum_{j=2}^{J} \sum_{\theta \in \Theta_j} a_\theta \varphi_\theta.$$

Continuing with the next level, we define $\tilde{\Lambda}_1$ as the set of all $\theta \in \Theta_1$ such that $|d_\theta| > \varepsilon^*$, set $\Lambda_1 := \bigcup_{\theta \in \tilde{\Lambda}_1} \mathcal{X}_\theta$, and define $\mathcal{A}_1 := \sum_{\theta \in \Lambda_1} d_\theta \varphi_\theta$. As for the previous level, we rewrite the remaining terms $d_\theta \varphi_\theta$, $\theta \in \Theta_1 \setminus \Lambda_1$, at the next level and add the resulting terms to the corresponding terms $a_\theta \varphi_\theta$, $\theta \in \Theta_2$. We obtain the following representation of f:

$$f = \mathcal{A}_0 + \mathcal{A}_1 + \sum_{\theta \in \Theta_2} d_\theta \varphi_\theta + \sum_{j=3}^{J} \sum_{\theta \in \Theta_j} a_\theta \varphi_\theta.$$

We continue in this way until we reach the highest level of cells Θ_J. At level Θ_J, we define $\tilde{\Lambda}_J$, Λ_J, and $\mathcal{A}_J$ as above and discard all terms $d_\theta \varphi_\theta$, $\theta \in \Theta_J \setminus \Lambda_J$. We finally obtain our approximant $\mathcal{A} = \mathcal{A}_\varepsilon^P(f)$ in the form (9.7). We denote $\Lambda := \Lambda_\varepsilon := \bigcup_{j=0}^{J} \Lambda_j$ and so $\mathcal{A} = \sum_{\theta \in \Lambda} d_\theta \varphi_\theta$.

Since we throw away only elements $d_\theta \varphi_\theta$ with $|d_\theta| \le \varepsilon^*$ at the finest level Θ_J, we have the estimate

$$\|f_\Gamma - \mathcal{A}\|_\infty \le \| \sum_{\theta \in \Theta_J \setminus \Lambda_J} d_\theta \varphi_\theta \|_\infty \le \varepsilon^*$$

and hence, using (9.6),

$$\|f - \mathcal{A}\|_\infty \le 2\varepsilon^* = \varepsilon.$$

This completes *Step 3* and with that the description of the algorithm.

Error estimation for the “Push the Error” algorithm.

Suppose “Push the Error” is applied to a function f with $\varepsilon > 0$ and $\mathcal{A}_\varepsilon^P(f)$ is the approximant obtained: $\mathcal{A}_\varepsilon^P(f) := \sum_{\theta \in \Lambda_\varepsilon} d_\theta \varphi_\theta$. As in the “Trim & Cut” method, we use the corresponding notation

$$n(\varepsilon) := \#\Lambda_\varepsilon, \ \mathbb{A}_{n(\varepsilon)}^P(f)_\infty := \mathbb{A}_{n(\varepsilon)}^P(f, \mathcal{T})_\infty := \|f - \mathcal{A}_\varepsilon^P(f)\|_\infty,$$

and

$$\mathbb{A}_n^P(f)_\infty := \mathbb{A}_n^P(f, \mathcal{T})_\infty := \inf\{\mathbb{A}_{n(\varepsilon)}^P(f)_\infty : n(\varepsilon) \le n\}.$$

The following theorem shows that the “Push the Error” algorithm provides the necessary rates of approximation (i.e., the Jackson estimate from Theorem 15 of §7).

Theorem 25. *If $f \in B_\tau^\alpha(\mathcal{T})$, $\alpha \ge 1$, $\tau := 1/\alpha$, then for each $\varepsilon > 0$*

$$\mathbb{A}_{n(\varepsilon)}^P(f)_\infty \le \varepsilon \quad \textit{and} \quad n(\varepsilon) \le c\varepsilon^{-\tau} \|f\|_{B_\tau^\alpha(\mathcal{T})}^\tau,$$

where $c = 6N_0^3$. Furthermore, we have

$$\mathbb{A}_n^P(f)_\infty \le cn^{-\alpha} \|f\|_{B_\tau^\alpha(\mathcal{T})}, \quad n = 1, 2, \dots,$$

with $c = (6N_0^3)^\alpha$ (N_0 is from §3.1).

Remark. There is a version of the “Push the Error” algorithm for approximation in L_p $(0 < p < \infty)$, which we do not consider here.

9.4. Approximation Spaces for Algorithms

The goal in this subsection is to show that the algorithms that we described in §9.1-§9.3 achieve (in a certain sense) the rate of convergence of the best n-term Courant element approximation. We shall utilize the characterization of the approximation spaces

$$A_q^\gamma(L_p, \mathcal{T}; \sigma) := A_q^\gamma(L_p, \mathcal{T}).$$

from Theorem 18 of §7. We shall denote by $A_q^\gamma(L_p, \mathcal{T}; \mathbb{A}^T)$, $A_q^\gamma(L_p, \mathcal{T}; \mathbb{A}^{TC})$, and $A_q^\gamma(L_p, \mathcal{T}; \mathbb{A}^P)$ the approximation spaces generated by the "Threshold", "Trim & Cut", and "Push the Error" algorithms, respectively. Namely, $f \in A_q^\gamma(L_p, \mathcal{T}; \mathbb{A})$, where $\mathbb{A}$ is $\mathbb{A}^T$, $\mathbb{A}^{TC}$ or $\mathbb{A}^P$, if $f \in L_p(E)$ and

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

with the usual modification when $q = \infty$ (this is not quite a norm).

Theorem 26. *Let $\mathcal{T}$ be an LR-triangulation of a compact polygonal domain $E \subset \mathbb{R}^2$.*
(a) *If $p = \infty$, $\alpha > 1$, and $\tau := 1/\alpha$, then*

$$A_\tau^\alpha(L_\infty, \mathcal{T}; \mathbb{A}^P) = A_\tau^\alpha(L_\infty, \mathcal{T}; \mathbb{A}^{TC}) = A_\tau^\alpha(L_\infty, \mathcal{T}; \sigma) = B_\tau^\alpha(\mathcal{T})$$

with equivalent "norms".
(b) *If $0 < p < \infty$, $\alpha > 0$, and $\tau := (\alpha + 1/p)^{-1}$, then*

$$A_\tau^\alpha(L_p, \mathcal{T}; \mathbb{A}^{TC}) = A_\tau^\alpha(L_p, \mathcal{T}; \mathbb{A}^T) = A_\tau^\alpha(L_p, \mathcal{T}; \sigma) = B_\tau^\alpha(\mathcal{T})$$

with equivalent "norms", where "Trim & Cut" is applied with parameter $\tau \leq \varrho < p$ (see [22]).

This theorem (and even more complete results) would follow easily by Theorems 23-25 if the error $\mathbb{A}_n(f)$ of the corresponding method was quasi-semi-additive, namely, if $\mathbb{A}_{cn}(f^0 + f^1) \leq c(\mathbb{A}_n(f^0) + \mathbb{A}_n(f^1))$. Since this is not known (an open problem) the proof employs the following result instead:

Lemma 3. *Let $f = f^0 + f^1$, where $f = \sum_{\theta \in \Theta} b_\theta \varphi_\theta$, $f^j = \sum_{\theta \in \Theta} b_\theta^j \varphi_\theta$ $(j = 0, 1)$, and $b_\theta = b_\theta^0 + b_\theta^1$ (all $\theta \in \Theta$), and suppose*

$$\mathcal{N}_j := \left(\sum_{\theta \in \Theta} |b_\theta^j|^{\tau_j}\right)^{1/\tau_j} < \infty \qquad (j = 0, 1),$$

where $\alpha_0, \alpha_1 \geq 1$ and $\tau_0 := 1/\alpha_0$, $\tau_1 := 1/\alpha_1$. Furthermore, suppose that "Trim & Cut" or "Push the Error" has been applied using the above representation of f, with $\varepsilon := \varepsilon_0 + \varepsilon_1$, where $\varepsilon_0, \varepsilon_1 > 0$. Then we have

$$\mathbb{A}_{n(\varepsilon_0 + \varepsilon_1)}^P(f)_\infty \leq \varepsilon_0 + \varepsilon_1$$

and

$$n(\varepsilon_0 + \varepsilon_1) \le c\varepsilon_0^{-\tau_0}\mathcal{N}_0^{\tau_0} + c\varepsilon_1^{-\tau_1}\mathcal{N}_1^{\tau_1}.$$

Consequently, the estimate

$$\mathbb{A}_n^P(f)_\infty \le cn^{-\alpha_0}\mathcal{N}_0 + cn^{-\alpha_1}\mathcal{N}_1, \qquad n = 1, 2, \dots$$

holds.

Remark. Similar results hold for the "Trim & Cut" algorithm in L_p for $0 < p < \infty$ as well as for the "Threshold" algorithm (see [22]).

9.5. Algorithms: Remarks

The three algorithms that we described in this section capture the rate of the best n-term Courant element approximation in the sense of Theorem 26. A common feature of these algorithms is the first step, a nontrivial decomposition from the redundant collection of all Courant elements from $\Phi_\mathcal{T}$. After this initial step, however, they take three different routes. The "Threshold" algorithm is completely unstructured but easy to implement. The drawback of this procedure is that it is not valid in the uniform case and as a consequence it does not perform well in L_p for p large. The "Trim & Cut" algorithm is valid for L_p, $0 < p \le \infty$, but it is over structured and as a result the performance suffers. The "Push the error" algorithm appears to be the preferred approximation method.

As we already mentioned, the algorithms that we described in this section are not restricted to n-term Courant element approximation only. They can be applied immediately to the approximation from (discontinuous) piecewise polynomials over multilevel triangulations . In this case the role of the B-spaces $B_\tau^\alpha(\mathcal{T})$ should be played by the skinny B-spaces $\mathcal{B}_\tau^{\alpha k}(\mathcal{T})$, introduced in §3.6. The results are similar, but simplify considerably. Furthermore, these algorithms can easily be adapted to nonlinear n-term approximation from smooth piecewise polynomial basis functions such as these considered in §3.3-3.4 and, in particular, from box splines. For more details about these algorithms, we refer the reader to [22].

10. Concluding Remarks

Global smoothness of functions: How to measure it? Here we turn to the fundamental questions in approximation theory (and not only there) of how the global smoothness of the functions should be measured. Thus as shown in §7 of this article, in the case of nonlinear n-term L_p-approximation from a single basis family $\Phi_\mathcal{T}$, a function f should naturally be considered smooth of order $\alpha > 0$ if $\|f\|_{B^\alpha_\tau(\Phi_\mathcal{T})} < \infty$. Then the rate of n-term L_p-approximation of f from $\Phi_\mathcal{T}$ is roughly $O(n^{-\alpha})$.

However, if we consider the highly nonlinear n-term approximation from a given collection (library) of basis families $\{\Phi_\mathcal{T}\}_\mathcal{T}$ ($\mathcal{T}$ can vary), then a function f should be considered smooth of oder $\alpha > 0$ if $\inf_\mathcal{T} \|f\|_{B^\alpha_\tau(\Phi_\mathcal{T})} < \infty$, which means that there exists a triangulation $\mathcal{T}^* := \mathcal{T}^*_f$ such that $\|f\|_{B^\alpha_\tau(\Phi_{\mathcal{T}^*})} < \infty$. Then the rate of n-term L_p-approximation of f from $\{\Phi_\mathcal{T}\}_\mathcal{T}$ is roughly $O(n^{-\alpha})$. It is crystal clear to us that no single super space can do the job in this case.

Now, our approximation scheme proceeds as follows:

(i) For a given function f, find the "right" triangulation $\mathcal{T}^* := \mathcal{T}^*_f$ such that f exhibits the most smoothness when measured via the scale $B^\alpha_\tau(\Phi_{\mathcal{T}^*})$.

(ii) Find an optimal or near optimal representation of f using $\Phi_{\mathcal{T}^*}$. (Note that $\Phi_{\mathcal{T}^*}$ is redundant, i.e., linearly dependent.)

(iii) Using this representation, run an algorithm for n-term L_p-approximation which achieves the rate of the best n-term approximation.

Naturally, the first step presents the most challenging problem in this scheme. This problem has a complete and efficient solution in the simpler case of nonlinear approximation from piecewise polynomials over dyadic partitions (§6) and remains *open* in the case of approximation from piecewise polynomials over nested triangulations. As shown in this article (§7 and §9), the other steps are now well understood and have complete solutions.

Next, we pose some more delicate open problems about highly nonlinear approximation from a library of basis families $\{\Phi_\mathcal{T}\}_\mathcal{T}$.

The ultimate problem is to characterize the approximation spaces generated by $(\sigma_n(f)_p)$, where $\sigma_n(f)_p := \inf_\mathcal{T} \sigma_n(f, \Phi_\mathcal{T})_p$ (see §7).

This problem is intimately connected to the problem for existence of an optimal (near optimal) triangulation:

For a given function $f \in L_p$, does there exist a single triangulation $\mathcal{T}^\diamond := \mathcal{T}^\diamond_f$ such that

$$\sigma_{cn}(f, \Phi_{\mathcal{T}^\diamond})_p \le c\sigma_n(f)_p, \quad \textit{for all } n \ge 1 \textit{ with } c \textit{ independent of } f \textit{ and } n?$$

If the answer of the latter question is "Yes", then the approximation of any $f \in L_p$ from the library $\{\Phi_\mathcal{T}\}_\mathcal{T}$ could be realized by approximation from a single basis family $\Phi_{\mathcal{T}^\diamond}$ and characterized by the interpolation spaces generated by $B^\alpha_\tau(\Phi_{\mathcal{T}^\diamond})$.

Smoothness of the basis functions. Clearly, in nonlinear approximation there is no saturation, which means that the corresponding approximation spaces A^γ_q are nontrivial for all $\gamma > 0$. Therefore, it is highly desirable that the smoothness spaces we use characterize the approximation spaces A^γ_q for all $0 < \gamma < \infty$. This concept immediately leads to the conclusion that the smoothness spaces to be used should naturally be designed so that the basis functions $\{\varphi_\theta\}$ are infinitely smooth with respect to them. This has been one of the guiding principles in constructing B-spaces. Thus each basis function $\varphi_\theta \in \Phi$ is infinitely smooth with respect to the scale of B-spaces $B^\alpha_\tau(\Phi)$, which is reflected in the fact that $\|\varphi_\theta\|_{B^\alpha_\tau(\Phi)} \le c\|\varphi_\theta\|_p$ for $0 < \alpha < \infty$ (see Theorem 16 of §7). This makes it possible that the direct, inverse, and characterization theorems in §6-§7 impose no restrictions on the rate of approximation $\alpha < \infty$.

To make this point more transparent, we shall next briefly compare the approximation results, presented in this article, with previously existing results, which involve Besov spaces. We first note that the situation in the univariate case is quite unique, since the scale of Besov spaces $B^{\alpha,k}_\tau(L_\tau)$ $(1/\tau = \alpha + 1/p)$ governs all rates of nonlinear piecewise polynomial approximation (see §8 and [27]). Therefore, it is no surprise that (as we showed in §5) the B-spaces in dimension $d = 1$ coincide with the corresponding univariate Besov spaces and hence are not needed. Besov spaces are also used in dimensions $d > 1$ (see [15, 18, 23], and also [7] and references therein), but they are not the right smoothness spaces even for nonlinear piecewise polynomial approximation generated by regular partitions. It follows by the discussion in §3.9 and the results from §6-§7 that the Besov spaces $B^{d\alpha,k}_\tau(L_\tau)$ can do the job when $0 < \alpha < r + 1 + 1/p$ and they fail when $\alpha \geq r + 1 + 1/p$, where r is the smoothness of the approximating splines. So, even when working on regular triangulations or partitions, the use of Besov spaces is restricted by the Besov smoothness (regularity) of the basis functions (see §3.9), while B-spaces impose no restrictions on the rates of approximation. Furthermore, if we allow triangulations with arbitrarily sharp angles, we allow very "skinny" basis functions with huge Besov norms compared to their L_p-norms (see §3.9) which precludes the use of Besov spaces in such situations. In a nutshell, the Besov spaces are the right smoothness spaces for characterization of nonlinear piecewise polynomial approximation in dimensions $d > 1$ only for regular partitions and for a limited range of approximation rates, and they are completely unsuitable in the anisotropic case.

Spline wavelets (prewavelets) and frames. It is natural to use bases in nonlinear approximation, and specifically for approximation in L_p $(1 < p < \infty)$. Thus we conveniently have the collection of all anisotropic Haar bases (see §6) which provides an effective tool for nonlinear approximation from piecewise constants over dayadic partitions of $\mathbb{R}^d$.

In the case of uniform triangulations, spline wavelets exist and play a significant role in practical algorithms. It would be desirable to have compactly supported wavelet (prewavelet) bases or frames generated by Courant elements or differentiable spline basis families $\Phi_\mathcal{T}$ over LR or SLR triangulations $\mathcal{T}$. To our knowledge there are no constructions of this type available, as for now. Moreover, there is some evidence that such constructions would be too complicated and impractical for general triangulations, if at all computable. However, continuous spline prewavelets on regular triangulations with uniform dyadic refinements are available from [9, 20, 31]. Evidently, nonlinear n-term approximation from compactly supported spline wavelets or frames, generated by Courant elements or a smoother spline basis family $\Phi_\mathcal{T}$, cannot surpass the rate of convergence of nonlinear n-term approximation from $\Phi_\mathcal{T}$.

References

[1] J. Bergh and J. Löfström, "Interpolation Spaces: An Introduction", Grundlehren der Mathematischen Wissenschaften, No. 223. Springer-Verlag, Berlin-New York, 1976.

[2] C. Bennett and R. Sharpley, "Interpolation of Operators", Pure and Applied Mathematics Vol. **129**, Academic Press, Inc., Boston, MA, 1988.

[3] C. de Boor, "A Practical Guide to Splines", Springer, New York, 1978.

[4] C. de Boor and R. Q. Jia, A sharp upper bound on the approximation order of smooth bivariate pp functions, *J. Approx. Theory* **72** (1993), 24–33.

[5] C. de Boor, K. Höllig, and S. D. Riemenschneider, "Box Splines", Springer-Verlag, New York, 1993.

[6] Yu. Brudnyi, Adaptive approximation of functions with singularities, *Trudy Moskov. Mat. Obshch.* **55** (1994), 149–242 [in Russian]; English transl.: *Trans. Moscow Math. Soc.* **1994**, 123–186 (1995).

[7] Yu. Brudnyi and I. Irodova, Nonlinear spline approximation of functions of several variables and B-spaces, *Algebra i Analiz* **4** (1992), 45–79 [in Russian]; English transl.: *St. Petersburg Math. J.* **4** (1993), 667–694.

[8] A. Cohen, R. DeVore, P. Petrushev, and H. Xu, Nonlinear approximation and the space $BV(\mathbb{R}^2)$, *Amer. J. Math.* **121** (1999), 587–628.

[9] W. Dahmen and R. Stevenson, Element-by-element construction of wavelets satisfying stability and moment conditions, *SIAM J. Numer. Anal.* **37** (1999), 319–352.

[10] O. Davydov, Stable local bases for multivariate spline spaces, *J. Approx. Theory* **111** (2001), 267–297.

[11] O. Davydov and P. Petrushev, Nonlinear approximation from differentiable piecewise polynomials, 2002, preprint.
(http://www.math.sc.edu/˜pencho)

[12] O. Davydov and L. L. Schumaker, On stable local bases for bivariate polynomial spline spaces, *Constr. Approx.* **18** (2002), 87–116.

[13] R. DeVore, Nonlinear approximation, *Acta Numerica*, (1998), 51–150.

[14] R. A. DeVore, B. Jawerth, and B. Lucier, Surface compression, *Comput. Aided Geom. Design* **9** (1992), 219–239.

[15] R. DeVore, B. Jawerth, and V. Popov, Compression of wavelet decompositions, *Amer. J. Math.* **114** (1992), 737–785.

[16] R. DeVore and G. Lorentz, "Constructive Approximation", Springer Grundlehren Vol. **303**, Heidelberg, 1993.

[17] R. A. DeVore and V. Popov, Interpolation of Besov spaces, *Trans. Amer. Math. Soc.* **305** (1988), 397–414.

[18] R. DeVore, P. Petrushev, and X. Yu, Nonlinear wavelet approximation in the space $C(R^d)$, *in* "Progress in Approximation Theory" (A. A. Gonchar, E. B. Saff, Eds.), pp. 261–283, New York, Springer-Verlag, 1992.

[19] D. Donoho, CART and best-ortho-basis: a connection, *Ann. Statist.* **25** (1997), 1870–1911.

[20] M. Floater and E. Quak, Linear independence and stability of piecewise linear prewavelets on arbitrary triangulations, *SIAM J. Numer. Anal.* **38** (2000), 58–79.

[21] B. Karaivanov and P. Petrushev, Nonlinear piecewise polynomial approximation beyond Besov spaces, 2001, preprint. (http://www.math.sc.edu/~pencho).

[22] B. Karaivanov, P. Petrushev and R. C. Sharpley, Algorithms for nonlinear piecewise polynomial approximation, *Trans. Amer. Math. Soc.*, 2003, to appear. (http://www.math.sc.edu/~pencho)

[23] Y. Hu, K. Kopotun, and X. Yu, On multivariate adaptive approximation, *Constr. Approx.* **16** (2000), 449–474.

[24] P. Oswald, On the degree of nonlinear spline approximation in Besov-Sobolev spaces, *J. Approx. Theory* **61** (1990), 131–157.

[25] P. Oswald, "Multilevel Finite Element Approximation", Teubner, Stuttgart, 1994.

[26] A. Pekarskii, Estimates for the derivatives of rational functions in $L_p[-1,1]$, *Mat. Zametki* **39** (1986), 388–394 [in Russian]; English transl.: *Math. Notes* **39** (1986), 212–216.

[27] P. Petrushev, Direct and converse theorems for spline and rational approximation and Besov spaces, *in* "Function Spaces and Applications" (M. Cwikel et. al., Eds), pp. 363-377, Vol. 1302 of Lecture Notes in Mathematics, Springer, Berlin, 1988.

[28] P. Petrushev, Multivariate n-term rational and piecewise polynomial approximation, *J. Approx. Theory*, 2003, to appear. (http://www.math.sc.edu/~pencho)

[29] P. Petrushev and V. Popov, "Rational Approximation of Real Functions", Cambridge University Press, 1987.

[30] L. L. Schumaker, "Spline Functions: Basic Theory", John Wiley and Sons, New York, 1981.

[31] R. Stevenson, Piecewise linear (pre-)wavelets on non-uniform meshes, *in* "Multigrid Methods V" (W. Hackbusch and G. Wittum, Eds.), pp. 306–319, Springer, Berlin, 1998.

Pencho Petrushev

Department of Mathematics
University of South Carolina
Columbia, SC 29208
USA
E-mail: `pencho@math.sc.edu`