Available at
www.ElsevierMathematics.com
POWERED BY SCIENCE DIRECT

Appl. Comput. Harmon. Anal. 15 (2003) 177–223

Applied and Computational Harmonic Analysis

www.elsevier.com/locate/acha

Nonlinear piecewise polynomial approximation beyond Besov spaces ☆

Borislav Karaivanov and Pencho Petrushev

Department of Mathematics, University of South Carolina, Columbia, SC 29208, USA

Received 7 May 2001; revised 7 May 2003; accepted 25 August 2003

Communicated by Ingrid Daubechies

Abstract

We study nonlinear n-term approximation in $L_p(\mathbb{R}^2)$ $(0 < p < \infty)$ from Courant elements or (discontinuous) piecewise polynomials generated by multilevel nested triangulations of $\mathbb{R}^2$ which allow arbitrarily sharp angles. To characterize the rate of approximation we introduce and develop three families of smoothness spaces generated by multilevel nested triangulations. We call them B-spaces because they can be viewed as generalizations of Besov spaces. We use the B-spaces to prove Jackson and Bernstein estimates for n-term piecewise polynomial approximation and consequently characterize the corresponding approximation spaces by interpolation. We also develop methods for n-term piecewise polynomial approximation which capture the rates of the best approximation.

Keywords: Nonlinear approximation; Jackson and Bernstein estimates; Multivariate splines; Nested irregular triangulations; Redundant representations; Courant elements

1. Introduction

Nonlinear approximation from piecewise polynomials and splines is a central theme in nonlinear approximation theory. The ultimate problem is to characterize the rate of approximation in terms of certain smoothness conditions. In the univariate case and in the regular case in d dimensions $(d > 1)$, this problem has found a completely satisfactory solution involving a certain class of Besov spaces and the machinery of Jackson–Bernstein estimates and interpolation (see [6,9,11], and also [2,5]).

☆ This research was supported by ONR/ARO-DEPSCoR Research Contract DAAG55-98-1-0002 and NSF Grant DMS-0200665.

E-mail addresses: karaivan@math.sc.edu (B. Karaivanov), pencho@math.sc.edu (P. Petrushev).

1063-5203/$ – see front matter
doi:10.1016/j.acha.2003.08.002

Our goal in this article is to study nonlinear approximation from piecewise polynomials over triangulations consisting of n pieces. The difficulty of this problem 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 two distinct triangulations of $[0, 1]^2$ consisting of n pieces each, then, in general, $S_1 + S_2$ is a piecewise polynomial over more than n^2 triangles (in the univariate case, the number of pieces is at most $2n$). This makes the idea of using a single smoothness space scale (like Besov spaces) and the recipe of proving Jackson and Bernstein estimates, and interpolation (like in the univariate case) hopeless.

In this article, we take a different approach to this problem. First of all, we modify the problem by considering nonlinear n-term approximation from piecewise polynomials generated by multilevel nested triangulations of $\mathbb{R}^2$. We consider two types of such n-term approximation:

(a) from Courant elements (continuous piecewise linear elements) and
(b) from (discontinuous) piecewise polynomials over triangles.

More precisely, we consider nested triangulations $\{\mathcal{T}_m\}_{m\in\mathbb{Z}}$ such that each level $\mathcal{T}_m$ is a partition of $\mathbb{R}^2$ and a refinement of the previous level $\mathcal{T}_{m-1}$, and define $\mathcal{T} := \bigcup_{m\in\mathbb{Z}} \mathcal{T}_m$. Each nested triangulation $\mathcal{T}$ generates a ladder of spaces $\cdots \subset \mathcal{S}_{-1} \subset \mathcal{S}_0 \subset \mathcal{S}_1 \subset \cdots$ (multiresolution analysis) consisting of piecewise polynomials of a certain degree over the corresponding levels. In the case of continuous piecewise linear functions, $\mathcal{S}_m$ $(m \in \mathbb{Z})$ is spanned by Courant elements φ_θ supported on cells θ at the mth level $\mathcal{T}_m$. We impose some natural mild conditions on the triangulations in order to prevent them from possible deterioration. At the same time, these conditions allow the triangles from $\mathcal{T}$ to have arbitrarily sharp angles and a lot of flexibility. After this preliminary structuring, we consider nonlinear approximation from n-term piecewise linear functions of the form $S = \sum_{j=1}^{n} a_{\theta_j}\varphi_{\theta_j}$ or piecewise polynomials of degree $< k$ of the form $S = \sum_{j=1}^{n} \mathbb{1}_{\triangle_j} \cdot P_{\triangle_j}$, where θ_j and $\triangle_j$ may come from different levels and locations ($\mathbb{1}_\triangle$ denotes the characteristic function of $\triangle$). Note that in both cases we have n-term nonlinear approximation from redundant systems. So, by introducing such a multilevel structure, we make the problem somewhat more accessible and simultaneously preserve a great deal of flexibility.

Although the approximation problem has been tamed to some extent, it still remains highly nonlinear. It is crystal clear to us that such highly nonlinear approximation cannot be governed by a single (super) space scale like the Besov spaces in the univariate case. For instance, it is well known that in presence of functions supported on very "skinny" triangles or long and narrow regions the Besov spaces are completely unsuitable and hence useless (see Section 2.5 below). Thus the second important concept is to quantify the approximation process by using a family of smoothness spaces, say, $B^\alpha(\mathcal{T})$ depending on the triangulations. We called them B-spaces. So, the idea is to measure the smoothness of the functions from a family (library) of space scales $\{B^\alpha(\mathcal{T})\}_\mathcal{T}$ instead of a single smoothness space scale.

The third important issue in our theory is the way we represent the functions. On the one hand, all Courant elements as well as all polynomials restricted to triangles generated by a nested triangulation form redundant systems. On the other hand, there are no good bases available which consist of piecewise polynomials over general triangulations. On top of this, we want to approximate in $L_p(\mathbb{R}^2)$, $0 < p < \infty$. There is, however, a good and well-known means of representing functions by using suitable linear or nonlinear projectors onto the spaces $\{\mathcal{S}_m\}$ (see Sections 2.3 and 2.4). This is our way of representing the functions.

Our approximation scheme is the following:

(i) For a given function f, find the "right" B-space $B^\alpha(\mathcal{T}_f)$ (that means the "right" triangulation $\mathcal{T}_f$) in which f exhibits the highest smoothness (equivalently, in which f has the sparsest representation).
(ii) Find an optimal (or near optimal) representation of f by Courant elements (or piecewise polynomials) generated by $\mathcal{T}_f$.
(iii) Using this representation of f, run an algorithm for n-term approximation that is capable of achieving the rate of the best n-term approximation.

The first step in this scheme is the hardest one and we still do not have a satisfactory algorithm for it. There is, however, an effective scalable algorithm for this step in the case of nonlinear approximation from piecewise polynomials over dyadic partitions, see [12]. Once the triangulation $\mathcal{T}$ is determined, the machinery of Jackson and Bernstein estimates combined with interpolation spaces works perfectly well. As we advance through the implementation of the above program, we shall see that all technological means exist or can be created so that a coherent theory can be developed. The lack of good bases for our spaces is the main obstacle that makes some proofs nonstandard. In particular, the Bernstein inequalities are the most troublesome and require fine analysis. We borrowed a few ideas from [12], where similar results have been obtained in the much simpler setting of nonlinear approximation from piecewise polynomials over dyadic boxes.

The B-spaces from this article can be considered as a generalization of Besov spaces (see Section 2.5 below). They are also a generalization of the approximation spaces from Section 3.4 in [10] (see the references therein).

There are several aspects of our theory that we do not even touch in this article, including nonlinear piecewise polynomial approximation in the uniform norm ($p = \infty$), interpolation of B-spaces and other aspects of the harmonic analysis of B-spaces, n-term approximation from smooth piecewise polynomials, and numerical algorithms for nonlinear piecewise polynomial approximation and their implementation in practice. Some of them will be tackled in a forthcoming article.

The outline of the paper is the following. Section 2 is devoted to the definition and development of B-spaces. In Section 2.1, we introduce and study three types of nested triangulations of $\mathbb{R}^2$, which later serve three different purposes. In Section 2.2, we give all necessary facts about local polynomial and piecewise linear approximation. In Section 2.3, we introduce and develop the first family of B-spaces, the slim B-spaces, which are later utilized for nonlinear n-term Courant element approximation. In Section 2.4, we introduce the skinny B-spaces that are needed for nonlinear n-term approximation from (discontinuous) piecewise polynomial. In Section 2.5, we introduce the fat B-spaces which are the most immediate generalization of Besov spaces. Section 3 contains our main results about nonlinear piecewise polynomial approximation. In Section 3.1, we give some general guiding principles and results for nonlinear n-term approximation. In Section 3.2, we state and prove our main results concerning n-term Courant element approximation except for the proof of the Bernstein inequality. In Section 3.3, we give our results on n-term piecewise polynomial approximation. Section 3.4 is devoted to discussion of some aspects of our theory and open problems. In Appendix A, we prove the Bernstein estimates we need. Appendix B contains the proofs of some auxiliary results.

Throughout the article, the constants are denoted by $c, c_1, \ldots,$ and they may vary at every occurrence. The constants usually depend on some parameters that will be sometimes indicated explicitly. The notation $A \approx B$ means that A and B are equivalent, i.e., there are two constants $c_1, c_2 > 0$ such that $c_1 A \leqslant B \leqslant c_2 A$. For $G \subset \mathbb{R}^2$, $|G|$ denotes the Lebesgue measure of G and $\mathbb{1}_G$ denotes the characteristic (indicator) function of G. We also use the following notation: $\|\cdot\|_q := \|\cdot\|_{L_q(\mathbb{R}^2)}$, $L_q^{\rm loc} := L_q^{\rm loc}(\mathbb{R}^2)$ $(0 < q < \infty)$, and $L_\infty^{\rm loc} := C(\mathbb{R}^2)$.

2. B-spaces over triangulations

In this section, we introduce and explore three collections of smoothness spaces (B-spaces), which will be needed in Section 3 for the characterization of the rates of nonlinear piecewise polynomial approximation. The B-spaces can be defined on $\mathbb{R}^2$ or on any polygonal domain in $\mathbb{R}^2$ as well as in $\mathbb{R}^d$ $(d \neq 2)$. We shall restrict our attention to the case of B-spaces on $\mathbb{R}^2$. The B-spaces are defined using multilevel nested triangulations which we discuss below.

2.1. Multilevel triangulations

Here we introduce several types of multilevel nested triangulations.

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 $\{\mathcal{T}_m\}_{m\in\mathbb{Z}}$ if the following conditions are fulfilled:

(a) Every level $\mathcal{T}_m$ defines a partition of $\mathbb{R}^2$, that is, $\mathbb{R}^2 = \bigcup_{\triangle\in\mathcal{T}_m} \triangle$ and $\mathcal{T}_m$ consists of closed triangles with disjoint interiors.
(b) The levels $\{\mathcal{T}_m\}_{m\in\mathbb{Z}}$ of $\mathcal{T}$ are nested, i.e., $\mathcal{T}_{m+1}$ is a refinement of $\mathcal{T}_m$.
(c) Each triangle $\triangle \in \mathcal{T}_m$ $(m \in \mathbb{Z})$ has at least two and at most M_0 children (subtriangles) in $\mathcal{T}_{m+1}$, where $M_0 \geqslant 2$ 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 covers K.
(e) There exist constants $0 < r < \rho < 1$ $(r \leqslant 1/2)$ such that for each $\triangle \in \mathcal{T}_m$ $(m \in \mathbb{Z})$ and any child $\triangle' \in \mathcal{T}_{m+1}$ of $\triangle$

$$r|\triangle| \leqslant |\triangle'| \leqslant \rho|\triangle|. \tag{2.1}$$

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

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* (*NHV*) *condition*: No vertex of any triangle $\triangle \in \mathcal{T}_m$ 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 $\triangle \in \mathcal{T}_m$ (the number of the triangles from $\mathcal{T}_m$ which share v as a vertex) is at most N_0, where N_0 is a constant.
(h) There exists a constant $0 < \delta \leqslant 1$ independent of m such that for any $\triangle', \triangle'' \in \mathcal{T}_m$ $(m \in \mathbb{Z})$ with a common edge

$$\delta \leqslant |\triangle'|/|\triangle''| \leqslant \delta^{-1}. \tag{2.2}$$

For $v \in \mathcal{V}_m$ ($m \in \mathbb{Z}$), we denote by $\theta_v := \theta_v(\mathcal{T})$ the *cell* associated with v, i.e., θ_v is the union of all triangles $\triangle \in \mathcal{T}_m$ which have v as a common vertex. We denote by $\Theta_m := \Theta_m(\mathcal{T})$ the set of all cells generated by $\mathcal{T}_m$ and set $\Theta := \Theta(\mathcal{T}) := \bigcup_{m\in\mathbb{Z}} \Theta_m$.

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) *Affine transform angle condition* (*ATA-condition*): There exists a constant $\beta = \beta(\mathcal{T})$, $0 < \beta \leqslant \pi/3$, such that if $\triangle_0 \in \mathcal{T}_m$, $m \in \mathbb{Z}$, and $\mathbf{A} : \mathbb{R}^2 \to \mathbb{R}^2$ 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 $\triangle_0$ and for every child $\triangle \in \mathcal{T}_{m+1}$ of $\triangle_0$, we have

$$\min \mathrm{angle}\big(\mathbf{A}(\triangle)\big) \geqslant \beta, \tag{2.3}$$

where $\mathbf{A}(\triangle)$ is the image of $\triangle$ by the affine transform $\mathbf{A}$, and $\min\mathrm{angle}(\triangle')$ denotes the magnitude of the minimal angle of $\triangle'$.

Obviously, (i) implies (2.2) with some $\delta = \delta(\beta)$.

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:

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

Evidently, every regular triangulation is an SLR-triangulation.

Triangulations on compact polygonal domains in $\mathbb{R}^2$. 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: $E = \bigcup_{\triangle\in\mathcal{T}_0} \triangle$. Weak locally regular, locally regular, etc., triangulations $\mathcal{T} = \bigcup_{m\geqslant 0}^{\infty} \mathcal{T}_m$ 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 $\{\mathcal{T}_m\}_{m\geqslant 0}$ now are consecutive refinements of an 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 similar classes of LR-triangulations or WLR-triangulations.

Each type of triangulation depends on several parameters which are not completely independent. For instance, the parameters of an LR-triangulation are M_0, N_0, r, ρ, and δ. We could set, e.g., $M_0 = 1/r$ and $\rho = 1 - r$, and eliminate these as parameters, but this would tend to obscure the actual dependence of the estimates upon given triangulations.

We shall need to know what happens with the levels $\mathcal{T}_m$ of a triangulation $\mathcal{T}$ as $m \to -\infty$. The next lemma answers this question.

Lemma 2.1. *For each WLR-triangulation $\mathcal{T}$ there exists a finite cover $\mathcal{T}_{-\infty}$ of $\mathbb{R}^2$ consisting of sets with disjoint interiors such that each triangle $\triangle \in \mathcal{T}$ and all its ancestors are contained in a set $\triangle_\infty \in \mathcal{T}_{-\infty}$. If*

$\triangle_\infty \in \mathcal{T}_{-\infty}$, then $\triangle_\infty$ must be one of the following: the all of $\mathbb{R}^2$, a half-plane, or an infinite triangle (all points on and between two rays that are not collinear and have a common beginning). The only possible configurations for $\mathcal{T}_{-\infty}$ are the following:

(a) *$\mathbb{R}^2$ only;*
(b) *finitely many infinite triangles with a common vertex;*
(c) *two half-planes;*
(d) *a half-plane and finitely many infinite triangles which cover the other half-plane and have a common vertex lying on the boundary between the two half-planes;*
(e) *two finite families of infinite triangles, each family covering one of two complimentary half-planes, and such that all triangles from the same family have a common vertex lying on the boundary between the two half-planes.*

Moreover, if $\mathcal{T}$ satisfies the NHV-condition, then (a) and (b) are the only possible configurations for $\mathcal{T}_{-\infty}$.

Proof. Let $\triangle \in \mathcal{T}_m$ for some $m \in \mathbb{Z}$. Then there exist unique triangles $\{\triangle_j\}_{j\leqslant m}$, $\triangle_j \in \mathcal{T}_j$, such that $\triangle =: \triangle_m \subset \triangle_{m-1} \subset \cdots$. We let $\triangle_\infty := \bigcup_{j\leqslant m} \triangle_j$. Clearly, if $\triangle', \triangle'' \in \mathcal{T}$ then either $\triangle'_\infty = \triangle''_\infty$ or $\triangle'_\infty$ and $\triangle''_\infty$ have disjoint interiors. To find out which subsets of $\mathbb{R}^2$ can be realized as $\triangle_\infty$, we order the vertices of the triangles $\{\triangle_j\}_{j\leqslant m}$ in a sequence $\{v_k\}$. If $\{v_k\}$ does not have limit points we consider two cases. First, if for every $\triangle_j$ there exists $i < j$ such that $\triangle_j \subset \triangle_i^\circ$, then using condition (d) from the definition of WLR-triangulations one can easily see that $\triangle_\infty$ is all of $\mathbb{R}^2$. Alternatively, if there exists a $\triangle_{j_0}$ which is not contained in $\triangle_j^\circ$ for any $j < j_0$, then each $\triangle_j$, $j \leqslant j_0$, has an edge lying on a given line l. Since $\{v_k\}$ does not have limit points, those edges grow infinitely in both directions, and therefore the whole line l must be contained in $\triangle_\infty$. Hence, since $\triangle_\infty$ is always convex, it must be either a half-plane or a strip. Using that $\{v_k\}$ does not have limit points and condition (d), one can prove that $\sup_{x\in\triangle_\infty} \operatorname{dist}(x,l) = \infty$, which shows that $\triangle_\infty$ cannot be a strip.

If the sequence $\{v_k\}$ has a limit point, say x_0, then using condition (d) we obtain that there exists $j_0 \leqslant m$ such that x_0 is a vertex of all $\triangle_j$ with $j \leqslant j_0$. From condition (d), it follows that a vertex can have only finite valence at any given level. This fact readily implies that $\{v_k\}$ cannot have more than one limit point and also that if $\{v_k\}$ has exactly one limit point then $\triangle_\infty$ is an infinite triangle.

Simple arguments utilizing condition (d) limit the possible configurations for $\mathcal{T}_{-\infty}$ to those described in the lemma. There are straightforward examples showing that each of those configurations can be realized. □

Examples of triangulations and refinement schemes. We begin with the description of a *standard refinement scheme* that can be used to refine a given triangle $\triangle$ infinitely many times. In the first step, we select a point on each edge of $\triangle$ and then join each pair of new points by a line segment. This first step gives us four disjoint triangles, say, $\triangle_1, \triangle_2, \triangle_3, \triangle_4$ which become the first generation of triangles (the children of $\triangle$). In the second step, we subdivide each $\triangle_j$ in the way described in step one and obtain the second generation of triangles. Proceeding inductively, we subdivide each triangle from a given generation in the fashion of step one, thus producing the next generation of triangles. Let $\mathcal{T}_m(\triangle)$ denote the set of all triangles from the mth generation. Then $\mathcal{T}(\triangle) := \bigcup_{m=0}^\infty \mathcal{T}_m(\triangle)$ is a nested triangulation of $\triangle$.

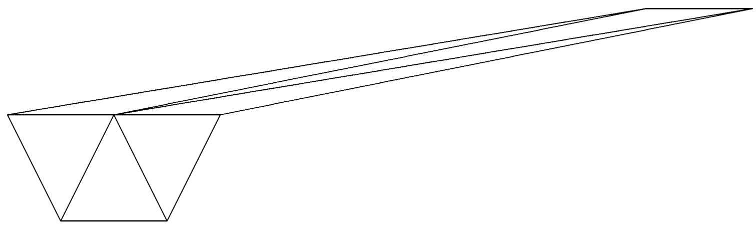

Fig. 1. A skewed cell.

Now, we describe a *standard procedure for constructing triangulations* of $\mathbb{R}^2$. We first cover $\mathbb{R}^2$ by a sequence of growing triangles $\triangle_0 \subset \triangle_1 \subset \triangle_2 \subset \cdots$, where every $\triangle_j$ is a child of $\triangle_{j+1}$, and then refine all children of all $\{\triangle_j\}$ using the standard refinement scheme described above. More precisely, let $\triangle_0$ be any initial triangle. We select a triangle $\triangle_1$ so that $\triangle_0$ is a child of $\triangle_1$. We similarly define $\triangle_2 \supset \triangle_1$ so that $\triangle_1$ is a child of $\triangle_2$, etc. In this way we obtain a growing sequence of triangles. The only additional condition that we impose on $\{\triangle_j\}$, so far, is that $\mathbb{R}^2 = \bigcup_{j=0}^{\infty} \triangle_j$. After having constructed the sequence $\{\triangle_j\}$, we subdivide the children of each $\triangle_j$ ($j = 1, 2, \ldots$) as it was described above. We denote by $\{\mathcal{T}_m\}_{m\in\mathbb{Z}}$ the sets of triangles from each level and by $\mathcal{T} := \bigcup_{m\in\mathbb{Z}} \mathcal{T}_m$ the whole triangulation of $\mathbb{R}^2$. Variety of other refinement schemes can be utilized.

How fast can the elements of triangles change? We investigate how the elements ($|\triangle'|$, $\min \mathrm{angle}(\triangle')$, and $\max \ell(\triangle')$, the longest edge of $\triangle'$) of a triangle $\triangle' \in \mathcal{T}_m$ ($m \in \mathbb{Z}$) can change as $\triangle'$ moves away from a fixed triangle $\triangle'' \in \mathcal{T}_m$, for different types of triangulations $\mathcal{T}$.

First, we consider the case of an arbitrary *weak locally regular triangulation* $\mathcal{T}$. Clearly, even if $\mathcal{T}$ satisfies the NHV-condition of the LR-triangulations, it may happen that $\triangle', \triangle'' \in \mathcal{T}_m$ ($m \in \mathbb{Z}$) are two adjacent triangles and at the same time each of the ratios $|\triangle'|/|\triangle''|$, $(\max \ell(\triangle'))/(\max \ell(\triangle''))$, and $(\min \mathrm{angle}(\triangle'))/(\min \mathrm{angle}(\triangle''))$ is arbitrarily large (or small) independently of the other two. This is possible because the first common ancestor of $\triangle'$ and $\triangle''$ may be at an extremely distant level, or even $\triangle'$ and $\triangle''$ may not have a common ancestor at all (see Lemma 2.1). This fact makes the WLR-triangulations unsuitable for continuous piecewise polynomial approximation.

Secondly, we consider the case of an arbitrary *locally regular triangulation* $\mathcal{T}$. By definition (see (2.2)), if $\triangle', \triangle'' \in \mathcal{T}_m$ and $\triangle'$ and $\triangle''$ have a common edge, then $|\triangle'| \approx |\triangle''|$. However, it may happen that the ratios $(\max \ell(\triangle'))/(\max \ell(\triangle''))$ and $(\min \mathrm{angle}(\triangle'))/(\min \mathrm{angle}(\triangle''))$ are uncontrollably large (or small), see Fig. 1. To show that this situation is possible we shall need the following simple lemma.

Lemma 2.2. *Let $\bigcup_{m=-\infty}^{n} \mathcal{T}_m$, $n \in \mathbb{Z}$, satisfy the conditions of the WLR-triangulations or LR-triangulations or SLR-triangulations with some fixed parameters. Assume also that level $\mathcal{T}_n$ is refined uniformly by introducing the midpoints on the edges of each $\triangle \in \mathcal{T}_n$ and connecting them by line segments* (*see the standard refinement scheme described above*). *Denote by $\mathcal{T}_{n+1}$ the set of all triangles obtained from the refinement of $\mathcal{T}_n$. Then $\bigcup_{m=-\infty}^{n+1} \mathcal{T}_m$ satisfies the conditions of the corresponding type of triangulation with exactly the same parameters.*

Proof. This lemma is fairly obvious and its proof will be omitted. □

Armed with this lemma, one can easily construct the claimed example. We shall give only a sketch of it. We start from a uniform triangulation $\mathcal{T}$ of $\mathbb{R}^2$ generated by an equilateral triangle $\triangle_0$ (see the examples of triangulations above). Let $\mathcal{T}_m$ denote the levels of $\mathcal{T}$ for $m \leqslant 0$. The incomplete triangulation $\bigcup_{m=-\infty}^{0} \mathcal{T}_m$ obviously satisfies the conditions of the LR-triangulations with any parameters $0 < r < \rho < 1$, $r < 1/4$, $\rho > 1/4$. We fix such r and ρ. We now refine $\mathcal{T}_0$. We choose any two triangles $\triangle', \triangle'' \in \mathcal{T}_0$ with a common edge, say e. We may assume that e is horizontal. It is not very hard to see (but it is not obvious) that $\mathcal{T}_0$ can be refined twice so that $\bigcup_{m=-\infty}^{2} \mathcal{T}_m$ satisfies the conditions of the LR-triangulations with the already selected parameters r and ρ, and that there are two grandchildren, say, $\triangle_2'$ and $\triangle_2''$ of $\triangle'$ and $\triangle''$, respectively, with the following properties:

(a) $\triangle_2'$ and $\triangle_2''$ have a common edge, say, $e_2 \subset e$ of length $\ell(e_2) = (1/4)\ell(e)$;
(b) $|\triangle_2'| = |\triangle_2''| = (1/16)|\triangle'|$ $(= (1/16)|\triangle''|)$;
(c) $\triangle_2'$ is equilateral and $\triangle_2''$ is skewed to the right (or left) at $\varepsilon \cdot \ell(e_2)$ with $\varepsilon = \varepsilon(r, \rho) > 0$.

More precisely, the vertex of $\triangle_2''$, which does not belong to e_2, is shifted to the right from the midpoint of e_2 at distance $\varepsilon \cdot \ell(e_2)$. We shall call the above an *angle sharpening procedure*. We next refine $\mathcal{T}_2$ sufficiently many times, by using only midpoints, until we reach a level, say, $\mathcal{T}_{s_1}$ at which there exist two great-grandchildren, say, $\triangle_{s_1}'$ and $\triangle_{s_1}''$ of $\triangle_2'$ and $\triangle_2''$, respectively, such that $\triangle_{s_1}'$ and $\triangle_{s_1}''$ have a common edge, $|\triangle_{s_1}'| = |\triangle_{s_1}''|$, $\triangle_{s_1}'$ is equilateral, $\triangle_{s_1}''$ is similar to $\triangle_2''$, and most importantly the minimal number of edges from $\mathcal{V}_{s_1}$ (edges of triangles in $\mathcal{T}_{s_1}$) which connect an arbitrary vertex of $\triangle_{s_1}'$ or $\triangle_{s_1}''$ with any vertex of $\triangle_2'$ or $\triangle_2''$ is sufficiently large (so, $\triangle_{s_1}'$ and $\triangle_{s_1}''$ are located in the middle of $\triangle_2' \cup \triangle_2''$). By Lemma 2.2, $\bigcup_{m=-\infty}^{s_1} \mathcal{T}_m$ satisfies the conditions of the LR-triangulations with the already fixed parameters ρ and r. Since, in $\mathcal{T}_{s_1}$, $\triangle_{s_1}'$ and $\triangle_{s_1}''$ are surrounded by triangles that are equivalent to $\triangle_{s_1}'$ or $\triangle_{s_1}''$, we can again apply our angle sharpening procedure, followed by sufficiently many midpoint refinements, and keep going on in the same fashion. We use induction to complete the construction of the claimed example.

Let us consider now an arbitrary *strong locally regular triangulation* $\mathcal{T}$. From the definition of SLR-triangulations, it follows that if $\triangle', \triangle'' \in \mathcal{T}_m$, $m \in \mathbb{Z}$, and $\triangle'$ and $\triangle''$ have a common vertex, then $|\triangle'| \approx |\triangle''|$, $\min\text{angle}(\triangle') \approx \min\text{angle}(\triangle'')$, and $\max \ell(\triangle') \approx \max \ell(\triangle'')$. However, this does not mean that $\mathcal{T}$ is regular or close to regular. It may happen that some triangles of $\mathcal{T}$ have arbitrarily small angles, while others are equilateral.

Example of an SLR-triangulation $\mathcal{T}$ with the property

$$\inf_{\triangle \in \mathcal{T}} \min\text{angle}(\triangle) = 0.$$

We shall utilize the idea of the construction from the previous example. As above, we assume that $\mathcal{T} = \bigcup_{m=-\infty}^{0} \mathcal{T}_m$ is an incomplete uniform triangulation generated by an equilateral triangle $\triangle_0$. Clearly, $\mathcal{T}$ satisfies the conditions of the SLR-triangulations for $M_0 = 4$ and an arbitrary $0 < \beta < \pi/3$. We fix β and M_0. Choose $\triangle \in \mathcal{T}_0$. It is readily seen that $\mathcal{T}_0$ can be refined so that $\bigcup_{m=-\infty}^{1} \mathcal{T}_m$ satisfies the conditions of the SLR-triangulations with the fixed parameters β and M_0, and there exists at least one child, say, $\triangle_1 \in \mathcal{T}_1$ of $\triangle$ such that $\min\text{angle}(\triangle_1) < q \cdot \min\text{angle}(\triangle)$ with $q = q(\beta) < 1$. The next step is to refine $\mathcal{T}_1$ several times by using only midpoints until we obtain a great-grandchild, say, $\triangle_{s_1} \in \mathcal{T}_{s_1}$ of $\triangle_1$ which is sufficiently far from the boundary of $\triangle_1$ (in terms of number of edges from $\mathcal{V}_{s_1}$ needed to connect it with the boundary). By Lemma 2.2, $\bigcup_{m=-\infty}^{s_1} \mathcal{T}_m$ satisfies the conditions of the SLR-triangulation with the fixed parameters β and M_0. After that, we apply the above angle sharpening procedure to $\triangle_{s_1}$ and then we again refine by midpoints for sufficiently many levels, etc. Inductively, we obtain the needed triangulation.

We now introduce one more natural condition on triangulations:

Minimal angle condition (*MA-condition*): There exists a constant $\vartheta = \vartheta(\mathcal{T})$, $0 < \vartheta < 1$, such that if $\triangle_0 \in \mathcal{T}_m$, $m \in \mathbb{Z}$, then for every $\triangle \in \mathcal{T}_m$ which has at least one common vertex with $\triangle_0$ and for every $\triangle \in \mathcal{T}_{m+1}$ which is a child of $\triangle_0$,

$$\vartheta \leqslant \frac{\min \operatorname{angle}(\triangle)}{\min \operatorname{angle}(\triangle_0)} \leqslant \vartheta^{-1}. \tag{2.4}$$

Lemma 2.3. *If $\mathcal{T}$ is an SLR-triangulation, then $\mathcal{T}$ satisfies the MA-condition above with $\vartheta = \vartheta(\beta)$. However, the MA-condition is weaker than the ATA-condition.*

Proof. Suppose $\mathcal{T}$ is an SLR-triangulation and let $\triangle_0 \in \mathcal{T}_m$, $m \in \mathbb{Z}$. We may assume that the largest edge of $\triangle_0$ is of length one. We introduce a coordinate system Ox_1x_2 so that the origin O is at the vertex of the sharpest angle of $\triangle_0$ and the largest edge of $\triangle_0$ lies on the positive half of the x_1-axis. Without loss of generality, we can assume that $\triangle_0$ is in the upper right quadrant of Ox_1x_2. We select the equilateral reference triangle $\triangle_0'$ to be in the upper right quadrant of Ox_1x_2 and have one edge coinciding with the longest edge of $\triangle_0$. Evidently, both the affine (linear in this case) transform $\mathbf{A}$ which maps $\triangle_0$ one-to-one onto $\triangle_0'$ and its inverse $\mathbf{A}^{-1}$ have matrices of the form $\left[\begin{smallmatrix} 1 & u_1 \\ 0 & u_2 \end{smallmatrix}\right]$. Suppose that the angle of $\triangle_0'$ with vertex at the origin and magnitude of $\pi/3$ is transformed by $\mathbf{A}^{-1}$ into an angle of magnitude γ, $0 < \gamma < \pi/3$. In this setting, routine (but not trivial) calculations show that $\mathbf{A}^{-1}$ transforms any angle of magnitude $\geqslant \beta$ into an angle of magnitude $\geqslant c\gamma$, where $c = c(\beta)$ is a positive constant. We skip all details and only note that it suffices to prove the above fact only for angles with vertex at the origin because the affine transforms map parallel lines into parallel lines. This result implies that $\mathcal{T}$ satisfies the MA-condition.

The MA-condition does not imply the ATA-condition because the following configuration of triangles is possible: Let $\triangle_1 := [(0,0), (1,0), (\varepsilon/2, \varepsilon\sqrt{3}/2)]$, where $\varepsilon > 0$ is sufficiently small. Denote by $\triangle_2$ the triangle symmetric to $\triangle_1$ with respect to the x_1-axis. Further, let $\triangle_3$ and $\triangle_4$ be the images of $\triangle_1$ and $\triangle_2$ after rotation of $-2\pi/3$ about the origin, and let $\triangle_5$, and $\triangle_6$ be the images of $\triangle_1$ and $\triangle_2$ after rotation of $2\pi/3$ about the origin. A triangulation containing this kind of configuration on one level can be constructed for an arbitrary small ε by starting from some level of a uniform triangulation consisting of equilateral triangles and “sharpening” the angles near a given node in three equiangular directions while refining the rest of the triangulations uniformly, as in the previous example. Obviously, this configuration does not violate the MA-condition but due to the presence of sharp angles in different directions the ATA-condition fails. □

Our next theorem provides estimates for the rate of change of the elements of triangles from a given level of a triangulation when moving away from a fixed triangle. For these estimates, we need the following simple lemma.

Lemma 2.4. *Suppose $\mathcal{T}$ is an LR-triangulation. If $\triangle', \triangle'' \in \mathcal{T}_m$, $m \in \mathbb{Z}$, and $\triangle'$ and $\triangle''$ can be connected by $< 2^\nu$ intermediate triangles* (*with common vertices*) *from $\mathcal{T}_m$, then there exist $\triangle_1, \triangle_2 \in \mathcal{T}_{m-2N_0\nu}$ with a common vertex such that $\triangle' \subset \triangle_1$ and $\triangle'' \subset \triangle_2$, where N_0 is from condition* (*g*) *of LR-triangulations.*

Proof. From conditions (c) and (g) on LR-triangulations (Section 2.1), it follows that every edge of a triangle from $\mathcal{T}_m$ is subdivided at least once after $2N_0$ steps of refinement. From this, we infer that if $G \subset \mathbb{R}^2$, then $\Omega^m(\Omega^m(G)) \subset \Omega^{m-2N_0}(G)$, where $\Omega^l(G) := \bigcup\{\theta \in \Theta_l\colon\ \theta^\circ \cap G \neq \emptyset\}$ (θ° denotes the interior of θ). Applying this fact ν times, we obtain that $\triangle'' \subset \Omega^{m-2N_0\nu}(\{v\})$, where v is an appropriate vertex of $\triangle'$. Then the existence of $\triangle_1$ and $\triangle_2$ follows readily. □

Theorem 2.5. (a) *Let $\mathcal{T}$ be an LR-triangulation with parameters $0 < r < \rho < 1$ and N_0. If $\triangle', \triangle'' \in \mathcal{T}_m$, $m \in \mathbb{Z}$, and $\triangle'$ and $\triangle''$ can be connected by n $(n \geqslant 1)$ intermediate triangles from $\mathcal{T}_m$, then*

$$c_1^{-1} n^{-s} \leqslant \frac{|\triangle'|}{|\triangle''|} \leqslant c_1 n^s \tag{2.5}$$

with $s := 2N_0 \log_2(\rho/r)$ and $c_1 := \delta^{-N_0}(\rho/r)^{2N_0}$.

(b) *Let $\mathcal{T}$ be an SLR-triangulation with parameter $0 < \beta \leqslant \pi/3$. If $\triangle', \triangle'' \in \mathcal{T}_m$, $m \in \mathbb{Z}$, and $\triangle'$ and $\triangle''$ can be connected by n $(n \geqslant 1)$ intermediate triangles from $\mathcal{T}_m$, then*

$$c_2^{-1} n^{-t} \leqslant \frac{\min \operatorname{angle}(\triangle')}{\min \operatorname{angle}(\triangle'')} \leqslant c_2 n^t \tag{2.6}$$

with $t := 4N_0 \log_2(1/\vartheta)$ and $c_2 := \vartheta^{-4N_0-1}$, where $N_0 := [2\pi/\beta]$ and $\vartheta = \vartheta(\beta)$ is the constant from the MA-condition whose existence is established by Lemma 2.3.

Proof. (a) Let $\nu \in \mathbb{Z}$ be such that $2^{\nu-1} \leqslant n < 2^\nu$. By Lemma 2.4, there exist $\triangle_1, \triangle_2 \in \mathcal{T}_{m-2N_0\nu}$ with a common vertex such that $\triangle' \subset \triangle_1$ and $\triangle'' \subset \triangle_2$. By (2.2), $\delta^{N_0} \leqslant |\triangle_1|/|\triangle_2| \leqslant \delta^{-N_0}$, and by (2.1), it follows that $|\triangle'| \leqslant \rho^{2N_0\nu}|\triangle_1|$ and $|\triangle''| \geqslant r^{2N_0\nu}|\triangle_2|$. Combining the above estimates, we obtain (2.5).

(b) The proof of (2.6) is quite similar to the proof of (2.5) and uses Lemma 2.3. We omit it. □

2.2. Local polynomial and piecewise linear approximation

We let Π_k denote the set of all algebraic polynomials in two variables of total degree $< k$. For a function $f \in L_q(G)$, $G \subset \mathbb{R}^2$, $0 < q \leqslant \infty$, and $k \geqslant 1$, we denote by $E_k(f, G)_q$ the error of $L_q(G)$-approximation to f from Π_k, i.e.,

$$E_k(f, G)_q := \inf_{P \in \Pi_k} \|f - P\|_{L_q(G)}. \tag{2.7}$$

Also, we denote by $\omega_k(f, G)_q$ the kth modulus of smoothness of f on G:

$$\omega_k(f, G)_q := \sup_{h \in \mathbb{R}^2} \left\| \Delta_h^k(f, \cdot) \right\|_{L_q(G)}, \tag{2.8}$$

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 the line segment $[x, x + kh]$ is entirely contained in G and $\Delta_h^k(f, x) := 0$ otherwise.

For an LR-triangulation $\mathcal{T}$ and $\triangle \in \mathcal{T}_m$ ($m \in \mathbb{Z}$), we denote by $\Omega_\triangle$ the union of all triangles $\triangle' \in \mathcal{T}_m$ which have a common vertex with $\triangle$, i.e.,

$$\Omega_\triangle := \bigcup\{\triangle' \in \mathcal{T}_m\colon\ \triangle' \cap \triangle \neq \emptyset\}. \tag{2.9}$$

Also, we define

$$\Omega_\triangle^2 := \bigcup\{\triangle' \in \mathcal{T}_m\colon\ \triangle' \cap \Omega_\triangle \neq \emptyset\}. \tag{2.10}$$

Lemma 2.6 (Whitney). *Suppose $G := \triangle$ or $G := \Omega_\triangle$ for some triangle $\triangle \in \mathcal{T}_m$ ($m \in \mathbb{Z}$), where $\mathcal{T}$ is an SLR-triangulation of $\mathbb{R}^2$. If $f \in L_q(G)$, $0 < q \leqslant \infty$, and $k \geqslant 1$, then*

$$E_k(f, G)_q \leqslant c\,\omega_k(f, G)_q \tag{2.11}$$

with $c = c(q, k)$ or $c = c(q, k, \beta)$, where β is the parameter of $\mathcal{T}$ from (2.3).

For the proof of this lemma, see Appendix B.

We shall often use the following lemma, which establishes relations between different norms of polynomials over different sets.

Lemma 2.7. *Let $P \in \Pi_k$, $k \geqslant 1$, and $0 < p, q \leqslant \infty$.*

(a) *Let $\triangle' \subset \triangle$ be two triangles such that $|\triangle| \leqslant c_1|\triangle'|$. Then*

$$\|P\|_{L_p(\triangle)} \leqslant c\|P\|_{L_p(\triangle')} \tag{2.12}$$

with $c = c(p, k, c_1)$.

(b) *Suppose $\triangle' \subset \triangle$ are two triangles such that $|\triangle'| \leqslant \rho|\triangle|$ with $0 < \rho < 1$ or $\triangle' = \emptyset$. Then*

$$\|P\|_{L_p(\triangle)} \leqslant c\|P\|_{L_p(\triangle\setminus\triangle')} \approx |\triangle|^{1/p-1/q}\|P\|_{L_q(\triangle\setminus\triangle')} \tag{2.13}$$

with constants depending only on p, q, k, and ρ.

(c) *If $\mathcal{T}$ is an LR-triangulation and $\triangle \in \mathcal{T}$, then*

$$\|P\|_{L_p(\Omega_\triangle)} \approx |\Omega_\triangle|^{1/p-1/q}\|P\|_{L_q(\Omega_\triangle)} \approx |\triangle|^{1/p-1/q}\|P\|_{L_q(\Omega_\triangle)} \tag{2.14}$$

with constants of equivalence depending only on p, q, k, N_0, and δ.

(d) *If $P \in \Pi_2$ and $\triangle = [x_1, x_2, x_3] \subset \mathbb{R}^2$ is a triangle, then*

$$\|P\|_{L_q(\triangle)} \approx |\triangle|^{1/q} \max_{1\leqslant j\leqslant 3}\bigl|P(x_j)\bigr|, \tag{2.15}$$

with constants of equivalence depending only on q.

Proof. Estimates (2.12)–(2.15) are invariant under affine transforms and hence they follow from the case when $\triangle$ is an equilateral triangle with $|\triangle| = 1$ by change of variables. The details will be omitted. □

We find useful the concept of *near best approximation* which we borrowed from [8]. A polynomial $P_\triangle \in \Pi_k$ is said to be a near best $L_q(\triangle)$-approximation to f from Π_k with constant A if

$$\|f - P_\triangle\|_{L_q(\triangle)} \leqslant AE_k(f, \triangle)_q. \tag{2.16}$$

Note that if $q \geqslant 1$, then a near best $L_q(\triangle)$-approximation $P_\triangle = P_\triangle(f)$ can be easily realized by a linear projector.

Lemma 2.8. *Suppose $0 < q \leqslant p$ and $P_\triangle$ is a near best $L_q(\triangle)$-approximation to $f \in L_p(\triangle)$ from Π_k. Then $P_\triangle$ is a near best $L_p(\triangle)$-approximation to f.*

Proof. See Lemma 3.2 from [8] and also the proof of Lemma 2.12 in Appendix B. □

We next introduce some necessary notation. Let $\mathcal{T} = \bigcup_{m\in\mathbb{Z}} \mathcal{T}_m$ be a WLR-triangulation. For $m \in \mathbb{Z}$ and $k \geqslant 1$, we let $\mathcal{S}_m^k := \mathcal{S}^k(\mathcal{T}_m)$ denote the set of all piecewise polynomial functions over $\mathcal{T}_m$ of degree $< k$, i.e., $S \in \mathcal{S}_m^k$ if $S = \sum_{\triangle\in\mathcal{T}_m} \mathbb{1}_\triangle \cdot P_\triangle$, where $\mathbb{1}_\triangle$ is the characteristic function of $\triangle$ and $P_\triangle \in \Pi_k$.

Now, let $\mathcal{T} = \bigcup_{m\in\mathbb{Z}} \mathcal{T}_m$ be an LR-triangulation. For $v \in \mathcal{V}_m$ ($m \in \mathbb{Z}$), we let θ_v denote the *cell* in $\mathcal{T}_m$ associated with v (Section 2.1). The NHV-condition on LR-triangulations (Section 2.1) guarantees the existence of a *Courant element* φ_{θ_v} supported on θ_v which is a continuous piecewise linear function that takes the value one at v.

For $m \in \mathbb{Z}$, we denote by $\widetilde{\mathcal{S}}_m := \widetilde{\mathcal{S}}(\mathcal{T}_m)$ the set of all continuous piecewise linear functions over $\mathcal{T}_m$, i.e., $\widetilde{\mathcal{S}}_m = \mathcal{S}_m^2 \cap C(\mathbb{R}^2)$. From the NHV-condition on $\mathcal{T}$, each $S \in \widetilde{\mathcal{S}}_m$ has the representation: $S = \sum_{v\in\mathcal{V}_m} S(v)\varphi_{\theta_v}$ and hence $\widetilde{\mathcal{S}}_m = \operatorname{span}\{\varphi_\theta\colon \theta \in \Theta_m\}$.

Throughout the rest of this section, we assume that $\mathcal{T}$ is an LR-triangulation of $\mathbb{R}^2$ with parameters M_0, N_0, r, ρ, and δ (see Section 2.1).

Lemma 2.9. *Suppose $\{a_\theta\}_{\theta\in\Theta_m}$, $m \in \mathbb{Z}$, is a sequence of real numbers and $S := \sum_{\theta\in\Theta_m} a_\theta\varphi_\theta$. Let also $0 < q \leqslant \infty$. Then, for every $\triangle \in \mathcal{T}_m$, we have*

$$\|S\|_{L_q(\triangle)} \approx \left(\sum_{\theta\in\Theta_m\,:\,\triangle\subset\theta} \|a_\theta\varphi_\theta\|_q^q\right)^{1/q} \tag{2.17}$$

and, hence,

$$\|S\|_{L_q(\mathbb{R}^2)} \approx \left(\sum_{\theta\in\Theta_m} \|a_\theta\varphi_\theta\|_q^q\right)^{1/q} \tag{2.18}$$

with constants of equivalence depending only on q, N_0, and δ. In the case $q = \infty$, the ℓ_q-norm above is replaced by the sup*-norm.*

Proof. Clearly, $S(v_\theta) = a_\theta$ (v_θ is the “central point” of θ) and $\|\varphi_\theta\|_q \approx |\theta|^{1/q}$. Therefore, using Lemma 2.7, (d) and the regularity of $\mathcal{T}$, we have, for $\triangle \in \mathcal{T}_m$,

$$\|S\|_{L_q(\triangle)} \approx |\triangle|^{1/q} \max_{\theta\in\Theta_m\colon\,\triangle\subset\theta} |a_\theta| \approx \max_{\theta\in\Theta_m\colon\,\triangle\subset\theta} |a_\theta||\theta|^{1/q} \approx \left(\sum_{\theta\in\Theta_m\colon\,\triangle\subset\theta} \|a_\theta\varphi_\theta\|_q^q\right)^{1/q}. \quad \Box$$

Quasi-interpolant. We shall utilize the following well-known quasi-interpolant for constructing projectors into spaces of continuous piecewise linear functions

$$Q_m(f) = Q_m(f, \mathcal{T}) := \sum_{\theta\in\Theta_m} \langle f, \tilde{\varphi}_\theta\rangle\varphi_\theta, \tag{2.19}$$

where $\langle f, g\rangle := \int_{\mathbb{R}^2} fg$ and $\{\tilde{\varphi}_\theta\}$ are duals of $\{\varphi_\theta\}$ defined by

$$\tilde{\varphi}_\theta := \sum_{\triangle\in\mathcal{T}_m,\,\triangle\subset\theta} \mathbb{1}_\triangle \cdot \tilde{\lambda}_{\triangle,\theta},$$

with $\tilde{\lambda}_{\triangle,\theta}$ the linear polynomial which assumes values $9/(N_{v_\theta}|\triangle|)$ at v_θ (the “central point” of θ) and $-3/(N_{v_\theta}|\triangle|)$ at the other two vertices of $\triangle$ (here N_{v_θ} is the valence of v_θ). Evidently,

$$\langle\varphi_\theta, \tilde{\varphi}_{\theta'}\rangle = \delta_{\theta\theta'}, \quad \theta, \theta' \in \Theta_m.$$

It is easily seen that the quasi-interpolant Q_m satisfies the following:

(a) $Q_m : L_1^{\mathrm{loc}} \to \widetilde{\mathcal{S}}_m$ is a linear operator.
(b) Q_m is a projector into $\widetilde{\mathcal{S}}_m$, i.e., $Q_m(S) = S$ for $S \in \widetilde{\mathcal{S}}_m$.

Other properties will be given in the following.

Lemma 2.10. *If $f \in L_\eta^{\mathrm{loc}}$, $1 \leqslant \eta \leqslant \infty$, and $\triangle \in \mathcal{T}_m$, $m \in \mathbb{Z}$, then*

$$\big\| Q_m(f) \big\|_{L_\eta(\triangle)} \leqslant c \| f \|_{L_\eta(\Omega_\triangle)}$$

with $c = c(\eta, N_0, \delta)$.

Proof. It is readily seen that

$$\big| \langle f, \tilde{\varphi}_\theta \rangle \big| \leqslant \| f \|_{L_\eta(\theta)} \| \tilde{\varphi}_\theta \|_{\eta'} \leqslant c \| f \|_{L_\eta(\theta)} \| \tilde{\varphi}_\theta \|_\infty |\theta|^{1/\eta'} \leqslant c |\theta|^{-1/\eta} \| f \|_{L_\eta(\theta)}$$

and $\|\varphi_\theta\|_\eta \leqslant c|\theta|^{1/\eta}$, where $1/\eta' := 1 - 1/\eta$. Therefore, for every $\triangle \in \mathcal{T}_m$,

$$\big\| Q_m(f) \big\|_{L_\eta(\triangle)} \leqslant \sum_{\theta \in \Theta_m,\, \theta \subset \Omega_\triangle} \big| \langle f, \tilde{\varphi}_\theta \rangle \big| \|\varphi_\theta\|_\eta \leqslant c \sum_{\theta \in \Theta_m,\, \theta \subset \Omega_\triangle} \| f \|_{L_\eta(\theta)} \leqslant c \| f \|_{L_\eta(\Omega_\triangle)}. \quad \square$$

Lemma 2.11. *If $S \in \mathcal{S}_m^2$, $0 < \eta \leqslant \infty$, and $\triangle \in \mathcal{T}_m$, $m \in \mathbb{Z}$, then*

$$\big\| Q_m(S) \big\|_{L_\eta(\triangle)} \leqslant c \| S \|_{L_\eta(\Omega_\triangle)}$$

with $c = c(\eta, N_0, \delta)$.

Proof. If $\eta \geqslant 1$, then the estimate follows by Lemma 2.10. Let $0 < \eta < 1$. We use the estimate $\|\varphi_\theta\|_\eta \leqslant c|\theta|^{1/\eta}$, properties of LR-triangulations (Section 2.1), and Lemma 2.7, (b), to obtain

$$\begin{aligned}
\big\| Q_m(S) \big\|_{L_\eta(\triangle)} &\leqslant c \sum_{\theta \in \Theta_m,\, \theta \subset \Omega_\triangle} \big| \langle S, \tilde{\varphi}_\theta \rangle \big| \|\varphi_\theta\|_\eta \leqslant c \sum_{\theta \in \Theta_m,\, \theta \subset \Omega_\triangle} \|\tilde{\varphi}_\theta\|_\infty \|S\|_{L_1(\theta)} |\theta|^{\frac{1}{\eta}} \\
&\leqslant c \sum_{\theta \in \Theta_m,\, \theta \subset \Omega_\triangle} |\theta|^{-1+\frac{1}{\eta}} \|S\|_{L_1(\theta)} \leqslant c \sum_{\triangle' \in \mathcal{T}_m,\, \triangle' \subset \Omega_\triangle} |\triangle'|^{-1+\frac{1}{\eta}} \|S\|_{L_1(\triangle')} \\
&\leqslant c \sum_{\triangle' \in \mathcal{T}_m,\, \triangle' \subset \Omega_\triangle} \|S\|_{L_\eta(\triangle')} \leqslant c \|S\|_{L_\eta(\Omega_\triangle)}. \quad \square
\end{aligned}$$

Local piecewise linear approximation. For a given $f \in L_\eta^{\mathrm{loc}}$, $\eta > 0$, and $\triangle \in \mathcal{T}_m$, $m \in \mathbb{Z}$ (recall that $\mathcal{T}$ is an LR-triangulation), we define the error of L_η-approximation to f on $\Omega_\triangle$ from $\widetilde{\mathcal{S}}_m$ by

$$\mathbb{S}_\triangle(f)_\eta := \mathbb{S}_\triangle(f, \mathcal{T})_\eta := \inf_{\widetilde{S} \in \widetilde{\mathcal{S}}_m} \| f - \widetilde{S} \|_{L_\eta(\Omega_\triangle)}. \tag{2.20}$$

Similarly as in the polynomial case, we say that $\widetilde{S} \in \widetilde{\mathcal{S}}_m$ is a *near best* L_η-approximation to f on $\Omega_\triangle$ from $\widetilde{\mathcal{S}}_m$ with a constant A if

$$\| f - \widetilde{S} \|_{L_\eta(\Omega_\triangle)} \leqslant A \mathbb{S}_\triangle(f)_\eta.$$

Lemma 2.12. *Suppose $0 < \mu < \eta$ and S is a near best L_μ-approximation to $f \in L_\eta(\Omega_\triangle)$ on $\Omega_\triangle$ from $\widetilde{\mathcal{S}}_m$. Then S is a near best L_η-approximation to f on $\Omega_\triangle$ from $\widetilde{\mathcal{S}}_m$.*

The proof of this lemma is similar to the proof of Lemma 3.2 of [8] (see also Lemma 2.8 above). For completeness, we give it in Appendix B.

The quasi-interpolant (defined above) is a simple and useful tool for constructing projectors into $\widetilde{\mathcal{S}}_m$ with good localization properties. For $\eta > 0$ and $f \in L_\eta^{\mathrm{loc}}$, let $P_{\triangle,\eta} = P_{\triangle,\eta}(f)$ be a near best $L_\eta(\triangle)$-approximation to f from Π_2. Note that if $\eta \geqslant 1$, then $P_{\triangle,\eta}(\cdot)$ can be realized as a linear projector into the space of linear polynomials restricted on $\triangle$. However, $P_{\triangle,\eta}(\cdot)$ is nonlinear if $\eta < 1$. Let

$$S_{m,\eta}(f) := \sum_{\triangle \in \mathcal{T}_m} \mathbb{1}_\triangle \cdot P_{\triangle,\eta}(f) \quad \text{for } m \in \mathbb{Z}.$$

Clearly, $S_{m,\eta}(f) \in \mathcal{S}_m^2$ and $S_{m,\eta}(S) = S$ for every $S \in \mathcal{S}_m^2$. We set

$$T_{m,\eta}(f) = T_{m,\eta}(f, \mathcal{T}) := Q_m\big(S_{m,\eta}(f)\big). \tag{2.21}$$

This construction is well known and is needed when working in L_η with $0 < \eta < 1$. Evidently, $T_{m,\eta}(f) \in \widetilde{\mathcal{S}}_m$ and $T_{m,\eta}(\widetilde{S}) = \widetilde{S}$ for $\widetilde{S} \in \widetilde{\mathcal{S}}_m$.

The next lemma establishes the good local approximation properties of the operators Q_m and T_m.

Lemma 2.13. (a) *If $f \in L_\eta^{\mathrm{loc}}$, $1 \leqslant \eta \leqslant \infty$, and $\triangle \in \mathcal{T}_m$, $m \geqslant 0$, then*

$$\big\|f - Q_m(f)\big\|_{L_\eta(\triangle)} \leqslant c\,\mathbb{S}_\triangle(f)_\eta. \tag{2.22}$$

(b) *If $f \in L_\eta^{\mathrm{loc}}$, $0 < \eta \leqslant \infty$, and $\triangle \in \mathcal{T}_m$, $m \geqslant 0$, then*

$$\big\|f - T_{m,\eta}(f)\big\|_{L_\eta(\triangle)} \leqslant c\,\mathbb{S}_\triangle(f)_\eta. \tag{2.23}$$

The constants above depend only on η and the parameters of $\mathcal{T}$.

Proof. To show that (2.23) holds, we choose $\widetilde{S}_\triangle \in \widetilde{\mathcal{S}}_m$ for which $\mathbb{S}_\triangle(f)_\eta$ is attained, i.e., $\|f - \widetilde{S}_\triangle\|_{L_\eta(\Omega_\triangle)} = \mathbb{S}_\triangle(f)_\eta$. Then

$$\begin{aligned}
\big\|f - T_m(f)\big\|_{L_\eta(\triangle)} &= \big\|f - Q_m\big(S_m(f)\big)\big\|_{L_\eta(\triangle)} \\
&= \big\|f - \widetilde{S}_\triangle + \widetilde{S}_\triangle - Q_m\big(S_m(f)\big)\big\|_{L_\eta(\triangle)} \\
&\leqslant c\|f - \widetilde{S}_\triangle\|_{L_\eta(\triangle)} + c\big\|Q_m\big(\widetilde{S}_\triangle - S_m(f)\big)\big\|_{L_\eta(\triangle)} \\
&\leqslant c\,\mathbb{S}_\triangle(f)_\eta + c\big\|\widetilde{S}_\triangle - S_m(f)\big\|_{L_\eta(\Omega_\triangle)} \\
&\leqslant c\,\mathbb{S}_\triangle(f)_\eta + c\|f - \widetilde{S}_\triangle\|_{L_\eta(\Omega_\triangle)} + c\big\|f - S_m(f)\big\|_{L_\eta(\Omega_\triangle)} \\
&\leqslant c\,\mathbb{S}_\triangle(f)_\eta,
\end{aligned}$$

where we used that $Q_m(\widetilde{S}_\triangle) = \widetilde{S}_\triangle$ on $\triangle$, Lemma 2.11, and the obvious inequality $\|f - S_m(f)\|_{L_\eta(\Omega_\triangle)} \leqslant \|f - \widetilde{S}_\triangle\|_{L_\eta(\Omega_\triangle)}$. Thus (2.23) is proved. The proof of (2.22) is similar and will be omitted. $\square$

Lemma 2.14. (a) *If $f \in L_\eta^{\mathrm{loc}}$, $1 \leqslant \eta \leqslant \infty$, then for $\triangle \in \mathcal{T}$*

$$\big\|f - Q_m(f)\big\|_{L_\eta(\triangle)} \to 0 \quad \textit{as } m \to \infty. \tag{2.24}$$

(b) *If* $f \in L_\eta^{\mathrm{loc}}$, $0 < \eta \leqslant \infty$, *then for* $\triangle \in \mathcal{T}$

$$\big\| f - T_m(f)\big\|_{L_\eta(\triangle)} \to 0 \quad \text{as } m \to \infty. \tag{2.25}$$

Proof. Using (2.1) and simple geometric arguments, one can show that if e is an edge of a descendant of $\triangle$, and e does not emanate from a vertex of $\triangle$, then $|e| \leqslant (1-r)\operatorname{diam}(\triangle)$. By condition (g) on LR-triangulations (Section 2.1), at any given level there can be at most $3N_0$ edges starting from the vertices of $\triangle$. From conditions (c) and (g), it follows that every edge e is subdivided within less than $2N_0$ levels after its first appearance, and by (2.1) each of the pieces of e has length $\leqslant (1-r)\operatorname{diam}(\triangle)$. Combining the above observations, we conclude that after less than $6N_0^2$ levels of refinement all edges of descendants of $\triangle$ will have lengths $\leqslant (1-r)\operatorname{diam}(\triangle)$. From this we derive that

$$\max\big\{\operatorname{diam}(\triangle'):\ \triangle' \in \mathcal{T}_m,\ \triangle' \subset \Omega_\triangle\big\} \to 0 \quad \text{as } m \to \infty.$$

Hence, $\|f - S_m(f)\|_{L_\eta(\Omega_\triangle)} \to 0$ and $\|f - \widetilde{S}_m(f)\|_{L_\eta(\Omega_\triangle)} \to 0$ as $m \to \infty$, where $\widetilde{S}_m(f)$ is a (the) best L_η-approximation to f on $\Omega_\triangle$ from $\widetilde{\mathcal{S}}_m$. Therefore,

$$\begin{aligned}
\big\| f - T_m(f)\big\|_{L_\eta(\triangle)} &\leqslant c\big\| f - \widetilde{S}_m(f)\big\|_{L_\eta(\triangle)} + c\big\| Q_m\big(\widetilde{S}_m(f) - S_m(f)\big)\big\|_{L_\eta(\triangle)} \\
&\leqslant c\big\| f - \widetilde{S}_m(f)\big\|_{L_\eta(\triangle)} + c\big\|\widetilde{S}_m(f) - S_m(f)\big\|_{L_\eta(\Omega_\triangle)} \\
&\leqslant c\big\| f - \widetilde{S}_m(f)\big\|_{L_\eta(\Omega_\triangle)} + c\big\| f - S_m(f)\big\|_{L_\eta(\Omega_\triangle)} \to 0
\end{aligned}$$

as $m \to \infty$, where we used that $Q_m(\widetilde{S}_\triangle) = \widetilde{S}_\triangle$ on $\triangle$ and Lemma 2.11. Thus (2.25) is proved. The proof of (2.24) is similar. □

2.3. Slim B-spaces

In this section, we introduce a collection of smoothness spaces (B-spaces) which we later used for characterization of nonlinear n-term Courant element approximation. Throughout the section, we assume that $\mathcal{T}$ is an arbitrary locally regular triangulation of $\mathbb{R}^2$ (see Section 2.1). The B-spaces will depend on $\mathcal{T}$. This dependence may or may not be indicated explicitly.

Definition of slim B-spaces via local approximation. We define the *slim B-space* $B_{pq}^\alpha(\mathcal{T})$, $\alpha > 0$, $0 < p, q \leqslant \infty$, as the set of all $f \in L_p(\mathbb{R}^2)$ such that

$$\|f\|_{B_{pq}^\alpha(\mathcal{T})} := \|f\|_p + \left(\sum_{m\in\mathbb{Z}}\left[2^{m\alpha}\left(\sum_{\triangle\in\mathcal{T},\ 2^{-m}\leqslant|\triangle|<2^{-m+1}} \mathbb{S}_\triangle(f)_p^p\right)^{1/p}\right]^q\right)^{1/q} < \infty, \tag{2.26}$$

where $\mathbb{S}_\triangle(f)_q := \mathbb{S}_\triangle(f,\mathcal{T})_q$, for $\triangle \in \mathcal{T}_m$, denotes the error of L_q-approximation to f on $\Omega_\triangle$ from $\widetilde{\mathcal{S}}_m$ (see (2.20)), and the ℓ_q-norm is replaced by the sup-norm if $q = \infty$.

We shall further study only a specific class of slim B-spaces which are exactly the smoothness spaces needed for nonlinear Courant L_p-approximation (see Section 3.2). We assume that $0 < p < \infty$ and $\alpha > 0$, and define τ by the identity $1/\tau := \alpha + 1/p$. We shall need the slim B-space $B_\tau^\alpha := B_\tau^\alpha(\mathcal{T})$, which is a slightly modified version of the space $B^\alpha_{\tau\tau}(\mathcal{T})$ from above. We define $B^\alpha_\tau(\mathcal{T})$ as the set of all functions $f \in L_p(\mathbb{R}^2)$ (in place of $f \in L_\tau(\mathbb{R}^2)$) such that

$$\|f\|_{B^\alpha_\tau} = \|f\|_{B^\alpha_\tau(\mathcal{T})} := \left(\sum_{\triangle\in\mathcal{T}} \left(|\triangle|^{-\alpha}\mathbb{S}_\triangle(f)_\tau\right)^\tau\right)^{1/\tau} < \infty. \tag{2.27}$$

Remark. In the above definition, the condition $f \in L_p(\mathbb{R}^2)$ is not restrictive since $B^\alpha_\tau(\mathcal{T})$ is embedded in L_p (see Theorems 2.15 and 2.16 below). Its only role is to eliminate a possible component S_∞ of f, which is a piecewise polynomial on infinite triangles $\triangle_\infty \in \mathcal{T}_{-\infty}$ (see Lemma 2.1). This condition can be replaced, e.g., by the condition: $|\{x\colon |f(x)| > s\}| < \infty$ for each $s > 0$ (see Theorem 2.15 below). It also can be replaced by the condition $f \in L_\tau(\mathbb{R}^2)$ as in the definition of $B^\alpha_{\tau\tau}(\mathcal{T})$ (see (2.26)), which is a little bit restrictive since the spaces $L_p(\mathbb{R}^2)$ and $L_\tau(\mathbb{R}^2)$ ($\tau \neq p$) are not embedded into one another. However, this condition is not too restrictive since our approximation tool in Section 3.2 consists of compactly supported piecewise polynomials and hence all theorems from Section 3.2 would hold if it is used.

Evidently,

$$\|f+g\|^{\tau^*}_{B^\alpha_\tau} \leqslant \|f\|^{\tau^*}_{B^\alpha_\tau} + \|g\|^{\tau^*}_{B^\alpha_\tau}, \quad \tau^* := \min\{\tau, 1\}.$$

Also, if $\|f\|_{B^\alpha_\tau} = 0$, then $\mathbb{S}_\triangle(f)_\tau = 0$ for each $\triangle \in \mathcal{T}$. From this, it readily follows that f coincides with a linear polynomial on each $\triangle_\infty \in \mathcal{T}_{-\infty}$. Therefore, using that $f \in L_p$, we infer that $f = 0$ a.e. Thus, for a fixed LR-triangulation $\mathcal{T}$, $\|\cdot\|_{B^\alpha_\tau(\mathcal{T})}$ is a norm if $\tau \geqslant 1$ and a quasi-norm if $\tau < 1$. In the following "norm" will stand for "norm" or "quasi-norm".

We next introduce other equivalent norms in $B^\alpha_\tau(\mathcal{T})$ which will enable us to operate more freely with B-spaces. For $f \in L^{\mathrm{loc}}_\eta(\mathbb{R}^2)$, $\eta > 0$, we define

$$\begin{aligned} N_{\mathbb{S},\eta}(f) := N_{\mathbb{S},\eta}(f,\mathcal{T}) &:= \left(\sum_{\triangle\in\mathcal{T}} \left(|\triangle|^{-\alpha+1/\tau-1/\eta}\mathbb{S}_\triangle(f)_\eta\right)^\tau\right)^{1/\tau} \\ &= \left(\sum_{\triangle\in\mathcal{T}} \left(|\triangle|^{1/p-1/\eta}\mathbb{S}_\triangle(f)_\eta\right)^\tau\right)^{1/\tau}, \end{aligned} \tag{2.28}$$

where we used that $1/\tau = \alpha + 1/p$. Clearly, $N_{\mathbb{S},\tau}(f) = \|f\|_{B^\alpha_\tau}$.

Atomic decomposition of $B^\alpha_\tau(\mathcal{T})$. For $f \in L_p(\mathbb{R}^2)$, we define

$$N_\Phi(f) = N_\Phi(f,\mathcal{T}) := \inf_{f=\sum_{\theta\in\Theta} c_\theta\varphi_\theta} \left(\sum_{\theta\in\Theta} \left(|\theta|^{-\alpha}\|c_\theta\varphi_\theta\|_\tau\right)^\tau\right)^{1/\tau}, \tag{2.29}$$

where the infimum is taken over all representations $f = \sum_{\theta\in\Theta} c_\theta\varphi_\theta$ with convergence in $L_p(\triangle)$ for each $\triangle \in \mathcal{T}$. (The existence of such representations of f follows by Lemma 2.14.) As will be seen in the proof of Theorem 2.15 below

$$\sum_{\theta\in\Theta} \left(|\theta|^{-\alpha}\|c_\theta\varphi_\theta\|_\tau\right)^\tau < \infty \quad \text{implies} \quad \left\|\sum_{\theta\in\Theta} \left|c_\theta\varphi_\theta(\cdot)\right|\right\|_p < \infty$$

and hence $\sum_{\theta\in\Theta} |c_\theta\varphi_\theta(\cdot)|$ converges a.e. and unconditionally in $L_p(\mathbb{R}^2)$. Therefore, the order of the terms in the series above is not essential. By Lemma 2.7, it follows that

$$N_\Phi(f) \approx \inf_{f=\sum_{\theta\in\Theta} c_\theta\varphi_\theta} \left(\sum_{\theta\in\Theta} \|c_\theta\varphi_\theta\|_p^\tau\right)^{1/\tau}. \tag{2.30}$$

Definition of norms in $B_\tau^\alpha(\mathcal{T})$ via projectors. We define, for $m\in\mathbb{Z}$,

$$q_m := Q_m - Q_{m-1} \quad \text{and} \quad t_{m,\eta} := T_{m,\eta} - T_{m-1,\eta}. \tag{2.31}$$

For a given function $f\in L_\eta^{\mathrm{loc}}(\mathbb{R}^2)$, $1\leqslant\eta\leqslant\infty$, clearly $q_m(f)\in\widetilde{\mathcal{S}}_m$ and we define uniquely the sequence $\{b_\theta(f)\}_{\theta\in\Theta_m}$ ($m\in\mathbb{Z}$) from the expression

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

Also, if $f\in L_\eta^{\mathrm{loc}}(\mathbb{R}^2)$, $0<\eta\leqslant\infty$, then $t_{m,\eta}(f)\in\widetilde{\mathcal{S}}_m$. In this case, we define $\{b_{\theta,\eta}(f)\}_{\theta\in\Theta_m}$ by

$$t_{m,\eta}(f) =: \sum_{\theta\in\Theta_m} b_{\theta,\eta}(f)\varphi_\theta. \tag{2.33}$$

Evidently, $\{b_\theta(\cdot)\}$ and $\{b_{\theta,\eta}(\cdot)\}$ with $\eta\geqslant 1$ are linear functionals, while $\{b_{\theta,\eta}(\cdot)\}$ are nonlinear if $0<\eta<1$.

We define

$$N_{Q,\tau}(f) = N_{Q,\tau}(f,\mathcal{T}) := \left(\sum_{\theta\in\Theta}\left(|\theta|^{-\alpha}\left\|b_{\theta,\tau}(f)\varphi_\theta\right\|_\tau\right)^\tau\right)^{1/\tau}, \tag{2.34}$$

where $b_{\theta,\tau}(f) := b_\theta(f)$ are from (2.32) (or from (2.33)) if $\tau\geqslant 1$ and $b_{\theta,\tau}(f)$ are from (2.33) if $\tau<1$. More generally, we define

$$N_{Q,\eta}(f) = N_{Q,\eta}(f,\mathcal{T}) := \left(\sum_{\theta\in\Theta}\left(|\theta|^{1/p-1/\eta}\left\|b_{\theta,\eta}(f)\varphi_\theta\right\|_\eta\right)^\tau\right)^{1/\tau}. \tag{2.35}$$

By Lemma 2.9, we have

$$N_{Q,\eta}(f) \approx \left(\sum_{\triangle\in\mathcal{T}}\left(|\triangle|^{1/p-1/\eta}\left\|q_m(f)\right\|_{L_\eta(\triangle)}\right)^\tau\right)^{1/\tau} \quad \text{if } \eta\geqslant 1, \tag{2.36}$$

$$N_{Q,\eta}(f) \approx \left(\sum_{\triangle\in\mathcal{T}}\left(|\triangle|^{1/p-1/\eta}\left\|t_{m,\eta}(f)\right\|_{L_\eta(\triangle)}\right)^\tau\right)^{1/\tau} \quad \text{if } 0<\eta<1, \tag{2.37}$$

and, in both cases,

$$N_{Q,\eta}(f) \approx \left(\sum_{\theta\in\Theta}\left\|b_{\theta,\eta}(f)\varphi_\theta\right\|_p^\tau\right)^{1/\tau}. \tag{2.38}$$

Our next step is to show that the slim B-space $B_\tau^\alpha(\mathcal{T})$ is embedded in $L_p(\mathbb{R}^2)$. To do this, we invoke Theorem 3.3, proved later in Section 3.1, which is however completely independent of this section, and can therefore safely be used.

Theorem 2.15. *If* $|\{x\colon\ |f(x)| > s\}| < \infty$ *for each* $s > 0$ *and* $N_{Q,\eta}(f, \mathcal{T}) < \infty$ *for some* $0 < \eta \leqslant \infty$, *then* $f \in L_p(\mathbb{R}^2)$,

$$f = \sum_{\theta\in\Theta} b_{\theta,\eta}(f)\varphi_\theta \quad \textit{absolutely a.e. on } \mathbb{R}^2 \tag{2.39}$$

and unconditionally in $L_p(\mathbb{R}^2)$, *and*

$$\|f\|_p \leqslant \Big\|\sum_{\theta\in\Theta}\big|b_{\theta,\eta}(f)\varphi_\theta(\cdot)\big|\Big\|_p \leqslant cN_{Q,\eta}(f) \tag{2.40}$$

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

Remark. Observe that the condition: $|\{x\colon\ |f(x)| > s\}| < \infty$ for each $s > 0$ is satisfied if $f \in L_q(\mathbb{R}^2)$ for an arbitrary $q < \infty$.

Proof. Let us consider the case when $N_{Q,\eta}(f)$ is defined via the coefficients $b_{\theta,\eta}(f)$ from (2.33). We introduce the following abbreviated notation: $T_m := T_{m,\eta}(f)$, $t_m := t_{m,\eta}(f)$, $b_\theta := b_{\theta,\eta}(f)$, and $N(f) := (\sum_{\theta\in\Theta}\|b_\theta\varphi_\theta\|_p^\tau)^{1/\tau}$. Note that $N_{Q,\eta}(f) \approx N(f)$, by (2.38). Since $\mathcal{T}$ is an LR-triangulation, the sequence $\{\Phi_m\} := \{b_\theta\varphi_\theta\}_{\theta\in\Theta}$ satisfies requirements (i) and (ii) of the general embedding Theorem 3.3 below. Therefore, $\sum_{\theta\in\Theta}|b_\theta\varphi_\theta(\cdot)| < \infty$ a.e. on $\mathbb{R}^2$ and

$$\Big\|\sum_{\theta\in\Theta}\big|b_\theta\varphi_\theta(\cdot)\big|\Big\|_p \leqslant cN(f). \tag{2.41}$$

Hence

$$\sum_{j\in\mathbb{Z}}\big|t_j(\cdot)\big| < \infty \quad \text{a.e. on } \mathbb{R}^2 \tag{2.42}$$

and

$$\Big\|\sum_{j\in\mathbb{Z}}\big|t_j(\cdot)\big|\Big\|_p \leqslant cN(f) < \infty. \tag{2.43}$$

Evidently, (2.39) and (2.41) imply (2.40). Therefore, it suffices to prove that (2.39) holds. To this end, we first show that

$$f = T_0 + \sum_{j=1}^{\infty} t_j \quad \text{absolutely a.e. on } \mathbb{R}^2. \tag{2.44}$$

Set $g := T_0 + \sum_{j=1}^\infty t_j$ pointwise. By (2.42), it follows that g is well defined. Clearly, $g = T_m + \sum_{j=m+1}^\infty t_j$ a.e. for $m \in \mathbb{Z}$. Hence, by (2.43),

$$\|g - T_m\|_p \leqslant \Big\|\sum_{j=m+1}^{\infty}\big|t_j(\cdot)\big|\Big\|_p \to 0 \quad \text{as } m \to \infty. \tag{2.45}$$

On the other hand, $f \in L_\eta^{\mathrm{loc}}(\mathbb{R}^2)$ and by Lemma 2.14 we have, for $\triangle \in \mathcal{T}$,

$$\|f - T_m\|_{L_\eta(\triangle)} \to 0 \quad \text{as } m \to \infty.$$

From this and (2.45), it follows that $g = f$ a.e. and hence (2.44) holds.

We shall next prove that for every $\triangle_\infty \in \mathcal{T}_{-\infty}$ (see Lemma 2.1) there exists a unique linear polynomial $P_{\triangle_\infty}$ such that

$$T_0 - P_{\triangle_\infty} = \sum_{j=-\infty}^{0} t_j \quad \text{absolutely a.e. on } \triangle_\infty. \tag{2.46}$$

Using Lemma 2.7, we have that for any $\triangle \in \mathcal{T}_j$ $(j \in \mathbb{Z})$

$$\|t_j\|_{L_\infty(\triangle)} \leqslant c|\triangle|^{-1/p}\|t_j\|_{L_p(\triangle)} \leqslant c|\triangle|^{-1/p} \sum_{\theta\in\Theta_j:\,\triangle\subset\theta} \|b_\theta\varphi_\theta\|_p \leqslant c|\triangle|^{-1/p}N(f). \tag{2.47}$$

Since $\mathcal{T}$ is an LR-regular triangulation, if $\triangle \subset \triangle'$, $\triangle \in \mathcal{T}_k$, and $\triangle' \in \mathcal{T}_j$, then $|\triangle| \leqslant \rho^{k-j}|\triangle'|$, where $0 < \rho < 1$ is the parameter of $\mathcal{T}$ from (2.1). Using this and (2.47), we obtain, for $\triangle \in \mathcal{T}_k$, $k \in \mathbb{Z}$,

$$\sum_{j=-\infty}^{k} \|t_j\|_{L_\infty(\triangle)} \leqslant cN(f)|\triangle|^{-1/p} \sum_{j=-\infty}^{k} \rho^{(k-j)/p} \leqslant c|\triangle|^{-1/p}N(f) < \infty. \tag{2.48}$$

For $\triangle \in \mathcal{T}_k$, we set $P_\triangle := T_k - \sum_{j=-\infty}^{k} t_j$ pointwise. By (2.42), the series converges absolutely a.e. and, therefore, $P_\triangle$ is well defined. Clearly, $P_\triangle = T_m - \sum_{j=-\infty}^{m} t_j$ for $m \leqslant k$ and, hence, by (2.48),

$$\|T_m - P_\triangle\|_{L_\infty(\triangle)} \leqslant \Bigg\|\sum_{j=-\infty}^{m} \big|t_j(\cdot)\big|\Bigg\|_{L_\infty(\triangle)} \leqslant \sum_{j=-\infty}^{m} \|t_j\|_{L_\infty(\triangle)} \to 0 \quad \text{as } m \to -\infty. \tag{2.49}$$

Since all t_j's, $j < k$, are linear polynomials on $\triangle \in \mathcal{T}_k$, so is $P_\triangle$. Moreover, $P_\triangle$ is the same polynomial for all $\triangle \in \mathcal{T}$ contained in a fixed $\triangle_\infty \in \mathcal{T}_{-\infty}$. Indeed, let $\triangle', \triangle'' \in \mathcal{T}$, $\triangle', \triangle'' \subset \triangle_\infty$ ($\triangle'$ and $\triangle''$ are possibly from different levels). Since $\triangle_\infty$ is an infinite union of nested triangles, there exists $\triangle \in \mathcal{T}$ such that $\triangle', \triangle'' \subset \triangle \subset \triangle_\infty$. By (2.49),

$$\|T_m - P_{\triangle'}\|_{L_\infty(\triangle')} \to 0 \quad \text{and} \quad \|T_m - P_\triangle\|_{L_\infty(\triangle')} \to 0 \quad \text{as } m \to -\infty.$$

Hence $P_{\triangle'} \equiv P_\triangle$. Similarly, $P_{\triangle''} \equiv P_\triangle$. Therefore, there exists a unique linear polynomial $P_{\triangle_\infty}$ such that (2.46) holds.

Combining (2.44) with (2.46), we obtain

$$f - P_{\triangle_\infty} = \sum_{j\in\mathbb{Z}} t_j \quad \text{absolutely a.e. on } \triangle_\infty,\ \triangle_\infty \in \mathcal{T}_\infty. \tag{2.50}$$

Using that $\sum_{j\in\mathbb{Z}} t_j \in L_p(\mathbb{R}^2)$ and the hypothesis of the theorem, we obtain

$$\begin{aligned}\big|\big\{x \in \triangle_\infty\colon \big|P_{\triangle_\infty}(x)\big| > s\big\}\big| &\leqslant \left|\left\{x\colon \big|f(x)\big| > \frac{s}{2}\right\}\right| + \left|\left\{x\colon \left|\sum_{j\in\mathbb{Z}} t_j(x)\right| > \frac{s}{2}\right\}\right| \\ &\leqslant \left|\left\{x\colon \big|f(x)\big| > \frac{s}{2}\right\}\right| + \left(\frac{s}{2}\right)^{-p} \left\|\sum_{j\in\mathbb{Z}} t_j\right\|_p^p < \infty,\end{aligned}$$

for each $s > 0$. Since $\triangle_\infty$ is an infinite triangle or a half plane or $\mathbb{R}^2$ and $P_{\triangle_\infty}$ is a polynomial, this is only possible whenever $P_{\triangle_\infty} \equiv 0$. Thus (2.39) is established.

The proof of the theorem when $N_{Q,\eta}(f)$ is defined via the coefficients $b_{\theta,\eta}(f) := b_\theta(f)$ from (2.32) is the same and will be omitted. □

Theorem 2.16. *For $f \in B^\alpha_\tau(\mathcal{T})$, the norms $\|f\|_{B^\alpha_\tau(\mathcal{T})}$, $N_{\mathbb{S},\eta}(f)$ $(0 < \eta < p)$, $N_\Phi(f)$, and $N_{Q,\eta}(f)$ $(0 < \eta < p)$, defined in* (2.27)–(2.29) *and* (2.35) *are equivalent with constants of equivalence depending only on p, α, η, and the parameters of $\mathcal{T}$.*

Proof. By (2.30), (2.38), and Theorem 2.15, it follows that:

$$N_\Phi(f) \leqslant cN_{Q,\eta}(f). \tag{2.51}$$

Clearly, if $\triangle \in \mathcal{T}_m$ and $\triangle'$ is the (unique) parent of $\triangle$ in $\mathcal{T}_{m-1}$, we have

$$\begin{aligned}\left\|t_{m,\eta}(f)\right\|_{L_\eta(\triangle)} &\leqslant c\left\|f - T_{m,\eta}(f)\right\|_{L_\eta(\triangle)} + c\left\|f - T_{m-1,\eta}(f)\right\|_{L_\eta(\triangle')}\\ &\leqslant c\mathbb{S}_\triangle(f)_\eta + c\mathbb{S}_{\triangle'}(f)_\eta,\end{aligned}$$

where we used (2.23). A similar estimate holds for $\|q_m(f)\|_{L_\eta(\triangle)}$, using (2.22). These imply

$$N_{Q,\eta}(f) \leqslant cN_{\mathbb{S},\eta}(f). \tag{2.52}$$

We next prove that if $N_\Phi(f) < \infty$, then

$$N_{\mathbb{S},\mu}(f) \leqslant cN_\Phi(f) \quad \text{for } 0 < \mu < p. \tag{2.53}$$

By Hölder's inequality, it follows that:

$$N_{\mathbb{S},\mu}(f) \leqslant N_{\mathbb{S},\tau}(f), \quad 0 < \mu \leqslant \tau.$$

Thus it suffices to prove (2.53) only for $\tau < \mu < p$.

Suppose $f \in L_p$ and $N_\Phi(f) < \infty$. Let $f = \sum_{\theta\in\Theta} c_\theta\varphi_\theta$ be an arbitrary representation of f, where the convergence is in $L_p(\triangle)$ for every $\triangle$. Recall that

$$N_{\mathbb{S},\mu}(f) := \left(\sum_{\triangle\in\mathcal{T}} \left(|\triangle|^{\frac{1}{p}-\frac{1}{\mu}}\mathbb{S}_\triangle(f)_\mu\right)^\tau\right)^{\frac{1}{\tau}}, \tag{2.54}$$

where $\mathbb{S}_\triangle(f)_\mu$ is defined in (2.20). Evidently, $\mathbb{S}_\triangle(g)_\mu = 0$ for $\triangle \in \mathcal{T}_m$ if $g \in \widetilde{\mathcal{S}}_m$, and $\mathbb{S}_\triangle(g)_\mu \leqslant \|g\|_{L_\mu(\Omega_\triangle)}$. Now, fix $\triangle \in \mathcal{T}_n$, $n \in \mathbb{Z}$. Using the above properties of $\mathbb{S}_\triangle(g)_\mu$ and Theorem 3.3 with $\{\Phi_m\} := \{c_\theta\varphi_\theta\colon \theta \in \Theta, \theta \subset \Omega^2_\triangle\}$ (for the definition of $\Omega^2_\triangle$, see (2.10)), we obtain

$$\begin{aligned}\mathbb{S}_\triangle(f)^\tau_\mu &= \mathbb{S}_\triangle\left(\sum_{j=n+1}^\infty\sum_{\theta\in\Theta_j} c_\theta\varphi_\theta\right)^\tau_\mu \leqslant \left\|\sum_{j=n+1}^\infty\sum_{\theta\in\Theta_j} c_\theta\varphi_\theta\right\|^\tau_{L_\mu(\Omega_\triangle)}\\ &\leqslant \left\|\sum_{j=n+1}^\infty \sum_{\theta\in\Theta_j,\,\theta\subset\Omega^2_\triangle} c_\theta\varphi_\theta\right\|^\tau_\mu \leqslant c\sum_{\theta\in\Theta,\,\theta\subset\Omega^2_\triangle} \|c_\theta\varphi_\theta\|^\tau_\mu\\ &\leqslant c\sum_{\theta\in\Theta,\,\theta\subset\Omega^2_\triangle} |\theta|^{\tau(\frac{1}{\mu}-\frac{1}{\tau})}\|c_\theta\varphi_\theta\|^\tau_\tau,\end{aligned}$$

where for the last inequality we used that $\|\varphi_\theta\|_q \approx |\theta|^{1/q}$, $0 < q \leqslant \infty$. Substituting the above estimate in (2.54), we get

$$
\begin{aligned}
N_{\mathbb{S},\mu}(f)^\tau &\leqslant c \sum_{\triangle\in\mathcal{T}} |\triangle|^{\tau(\frac{1}{p}-\frac{1}{\mu})} \sum_{\theta\in\Theta,\,\theta\subset\Omega_\triangle^2} |\theta|^{\tau(\frac{1}{\mu}-\frac{1}{\tau})} \|c_\theta\varphi_\theta\|_\tau^\tau \\
&\leqslant c \sum_{\triangle\in\mathcal{T}} \sum_{\theta\in\Theta,\,\theta\subset\Omega_\triangle^2} \left(\frac{|\theta|}{|\triangle|}\right)^{\tau(\frac{1}{\mu}-\frac{1}{p})} \left(|\theta|^{\frac{1}{p}-\frac{1}{\tau}} \|c_\theta\varphi_\theta\|_\tau\right)^\tau \\
&\leqslant c \sum_{\theta\in\Theta} \left(|\theta|^{\frac{1}{p}-\frac{1}{\tau}} \|c_\theta\varphi_\theta\|_\tau\right)^\tau \sum_{\triangle\in\mathcal{T}:\ \theta\subset\Omega_\triangle^2} \left(\frac{|\theta|}{|\triangle|}\right)^{\tau(\frac{1}{\mu}-\frac{1}{p})},
\end{aligned}
\tag{2.55}
$$

where we once switched the order of summation. By condition (g) on LR-triangulations (Section 2.1), we have, for $\theta\in\Theta_j$,

$$
\#\{\triangle\in\mathcal{T}_j:\ \theta\subset\Omega_\triangle^2\} \leqslant c(N_0),
$$

and by (2.1) and (2.2), $|\theta| \leqslant c(N_0,\delta)\rho^j|\triangle|$, if $\theta\subset\Omega_\triangle^2$ with $\triangle\in\mathcal{T}_{m-j}$ and $\theta\in\Theta_m$. Hence, for $\theta\in\Theta$,

$$
\sum_{\triangle\in\mathcal{T}:\ \theta\subset\Omega_\triangle^2} \left(\frac{|\theta|}{|\triangle|}\right)^{\tau(\frac{1}{\mu}-\frac{1}{p})} \leqslant c \sum_{j=0}^{\infty} \rho^{j\tau(\frac{1}{\mu}-\frac{1}{p})} \leqslant c < \infty, \tag{2.56}
$$

where we used that $\rho<1$ and $\mu<p$. Finally, combining (2.56) with (2.55), we obtain

$$
N_{\mathbb{S},\mu}(f)^\tau \leqslant c \sum_{\theta\in\Theta} \left(|\theta|^{-\alpha} \|c_\theta\varphi_\theta\|_\tau\right)^\tau,
$$

which implies (2.53). Evidently, (2.51) and (2.53) imply the theorem. □

Remark. The following simple example shows that, in general, Theorem 2.16 is not valid for $\eta\geqslant p$. Let $f:=\varphi_\theta$ for some $\theta\in\Theta$. It is not hard to see that $\|f\|_{B_\tau^\alpha(\mathcal{T})} \approx |\theta|^{1/p} \approx \|\varphi_\theta\|_p$, while $N_{\mathbb{S},\eta}(f,\mathcal{T})=\infty$, if $\eta\geqslant p$. Therefore, $N_{\mathbb{S},\eta}(f,\mathcal{T})$ is not equivalent to $\|f\|_{B_\tau^\alpha(\mathcal{T})}$ if $\eta\geqslant p$.

2.4. Skinny B-spaces

In this section, we define a second family of B-spaces which we shall use in Section 3.3 for the characterization of nonlinear (discontinuous) piecewise polynomial approximation generated by nested triangulations.

Throughout this section, we assume that $\mathcal{T}$ is an arbitrary weak locally regular triangulation of $\mathbb{R}^2$ (see Section 2.1). We define the *skinny B-space* $\mathcal{B}_{pq}^{\alpha k}(\mathcal{T})$, $\alpha>0$, $0<p,q\leqslant\infty$, $k\geqslant1$, as the set of all $f\in L_p(\mathbb{R}^2)$ such that

$$
\|f\|_{\mathcal{B}_{pq}^{\alpha k}(\mathcal{T})} := \|f\|_p + \left(\sum_{m\in\mathbb{Z}} \left[2^{m\alpha} \left(\sum_{\triangle\in\mathcal{T},\,2^{-m}\leqslant|\triangle|<2^{-m+1}} \omega_k(f,\triangle)_p^p\right)^{1/p}\right]^q\right)^{1/q} < \infty, \tag{2.57}
$$

where $\omega_k(f,\triangle)_p$ is the local modulus of smoothness of f, defined in (2.8).

As for the slim B-spaces, we shall explore in more details only the skinny B-spaces that are needed in nonlinear piecewise polynomial L_p-approximation. Suppose $0<p<\infty$, $\alpha>0$, $k\geqslant1$, and let

$1/\tau := \alpha + 1/p$. We shall need the skinny B-space $\mathcal{B}^{\alpha k}_{\tau}(\mathcal{T})$, which is a slight modification of $\mathcal{B}^{\alpha k}_{\tau\tau}(\mathcal{T})$ from above, and is defined as the set of all $f \in L_p(\mathbb{R}^2)$ (in place of $f \in L_\tau(\mathbb{R}^2)$) such that

$$\|f\|_{\mathcal{B}^{\alpha k}_{\tau}} = \|f\|_{\mathcal{B}^{\alpha k}_{\tau}(\mathcal{T})} := \left(\sum_{\triangle \in \mathcal{T}} \left(|\triangle|^{-\alpha} \omega_k(f, \triangle)_\tau \right)^\tau \right)^{1/\tau} < \infty. \tag{2.58}$$

Whitney's estimate (Lemma 2.6) implies

$$\|f\|_{\mathcal{B}^{\alpha k}_{\tau}(\mathcal{T})} \approx \left(\sum_{\triangle \in \mathcal{T}} \left(|\triangle|^{-\alpha} E_k(f, \triangle)_\tau \right)^\tau \right)^{1/\tau}, \tag{2.59}$$

where $E_k(f, \triangle)_q$ is the error of L_τ-approximation to f on $\triangle$ from Π_k (see (2.7)).

If $\|f\|_{\mathcal{B}^{\alpha k}} = 0$, then $E_k(f, \triangle)_\tau = 0$ for each $\triangle \in \mathcal{T}$. From this, it readily follows that $f = \mathbb{1}_{\triangle_\infty} \cdot P_{\triangle_\infty}$ ($P_{\triangle_\infty} \in \Pi_k$) on each $\triangle_\infty \in \mathcal{T}_{-\infty}$. Therefore, using that $f \in L_p$, we infer that $f = 0$ a.e. Thus, $\|\cdot\|_{\mathcal{B}^{\alpha k}_{\tau}(\mathcal{T})}$ is a norm if $\tau \geqslant 1$ and a quasi-norm if $\tau < 1$.

Remark. The only difference between skinny B-spaces and slim B-spaces is that the local approximation from continuous piecewise linear functions on sets $\Omega_\triangle$, $\triangle \in \mathcal{T}$, is replaced by local polynomial approximation on triangles from $\mathcal{T}$. The key is that the triangles from $\mathcal{T}$ form a tree with respect to the inclusion relation, while the sets $\Omega_\triangle$, $\triangle \in \mathcal{T}$ do not form a tree; they overlap more significantly. This fact allows for developing the theory of the skinny B-spaces and their application to nonlinear (discontinuous) piecewise approximation (see Section 3.3) under less restrictive conditions on the triangulations, namely, for weak locally regular triangulations.

Next, we introduce two other equivalent "norms" in $\mathcal{B}^{\alpha k}_{\tau}(\mathcal{T})$. For $f \in L^{\mathrm{loc}}_\eta(\mathbb{R}^2)$, $\eta > 0$, we define

$$\begin{aligned} \mathcal{N}_{\omega,\eta}(f, \mathcal{T}) &:= \left(\sum_{\triangle \in \mathcal{T}} \left(|\triangle|^{-\alpha + \frac{1}{\tau} - \frac{1}{\eta}} \omega_k(f, \triangle)_\eta \right)^\tau \right)^{\frac{1}{\tau}} \\ &= \left(\sum_{\triangle \in \mathcal{T}} \left(|\triangle|^{\frac{1}{p} - \frac{1}{\eta}} \omega_k(f, \triangle)_\eta \right)^\tau \right)^{\frac{1}{\tau}} \approx \left(\sum_{\triangle \in \mathcal{T}} \left(|\triangle|^{\frac{1}{p} - \frac{1}{\eta}} E_k(f, \triangle)_\eta \right)^\tau \right)^{\frac{1}{\tau}}, \end{aligned} \tag{2.60}$$

where we used that $1/\tau = \alpha + 1/p$. Clearly, $\mathcal{N}_{\omega,\tau}(f, \mathcal{T}) = \|f\|_{\mathcal{B}^{\alpha k}_{\tau}(\mathcal{T})}$.

For each $\triangle \in \mathcal{T}$ and $\eta > 0$, we let $P_{\triangle,\eta}(f)$ be a near best $L_\eta(\triangle)$-approximation to f from Π_k with a constant A which is the same for all $\triangle \in \mathcal{T}$ (see (2.16)). Note that if $\eta \geqslant 1$, then $P_{\triangle,\eta}(f)$ can be realized as a linear projector into the space of polynomials of degree $< k$ restricted on $\triangle$. Let $P_{m,\eta}(f) := \sum_{\triangle \in \mathcal{T}_m} \mathbb{1}_\triangle \cdot P_{\triangle,\eta}(f)$. Clearly, $P_{m,\eta}(f)$ is a near best L_η-approximation to f from $\mathcal{S}^k_m(\mathcal{T})$ and a projector into $\mathcal{S}^k_m(\mathcal{T})$. We define

$$p_{m,\eta}(f) := p_{m,\eta}(f, \mathcal{T}) := P_{m,\eta}(f) - P_{m-1,\eta}(f) \in \mathcal{S}^k_m(\mathcal{T}), \tag{2.61}$$

and set $p_{\triangle,\eta}(f) := \mathbb{1}_\triangle \cdot p_{m,\eta}(f)$ for $\triangle \in \mathcal{T}_m$. We define

$$\mathcal{N}_{P,\eta}(f, \mathcal{T}) := \left(\sum_{\triangle \in \mathcal{T}} \left(|\triangle|^{1/p - 1/\eta} \|p_{\triangle,\eta}(f)\|_\eta \right)^\tau \right)^{1/\tau}. \tag{2.62}$$

Using Lemma 2.7, we obtain

$$\mathcal{N}_{P,\eta}(f,\mathcal{T}) \approx \left(\sum_{\triangle\in\mathcal{T}} \left(|\triangle|^{-\alpha}\left\|p_{\triangle,\eta}(f)\right\|_\tau\right)^\tau\right)^{1/\tau} \approx \left(\sum_{\triangle\in\mathcal{T}} \left\|p_{\triangle,\eta}(f)\right\|_p^\tau\right)^{1/\tau}. \tag{2.63}$$

The following embedding theorem is pivotal for our theory of nonlinear piecewise polynomial approximation.

Theorem 2.17. *If $|\{x\colon |f(x)| > s\}| < \infty$ for each $s > 0$ and $\mathcal{N}_{P,\eta}(f,\mathcal{T}) < \infty$ $(0 < \eta \leqslant \infty)$, then $f \in L_p(\mathbb{R}^2)$,*

$$f = \sum_{m\in\mathbb{Z}} p_{m,\eta}(f) \quad \textit{absolutely a.e. on } \mathbb{R}^2 \tag{2.64}$$

and unconditionally in L_p, and

$$\|f\|_p \leqslant \left\|\sum_{m\in\mathbb{Z}} \left|p_{m,\eta}(f)\right|\right\|_p \leqslant c\,\mathcal{N}_{P,\eta}(f,\mathcal{T}) \tag{2.65}$$

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

Proof. Since $\mathcal{T}$ is a WLR-triangulation, the sequence $\{\Phi_m\} := \{p_{\triangle,\eta}(f)\}_{\triangle\in\mathcal{T}}$ satisfies requirements (i) and (ii) of Theorem 3.3 below. Therefore,

$$\left\|\sum_{\triangle\in\mathcal{T}} \left|p_{\triangle,\eta}(f)\right|\right\|_p \leqslant c\left(\sum_{\triangle\in\mathcal{T}} \left\|p_{\triangle,\eta}(f)\right\|_p^\tau\right)^{1/\tau} \approx c\,\mathcal{N}_{P,\eta}(f,\mathcal{T}) < \infty. \tag{2.66}$$

From this, similarly as in the proof of Theorem 2.15, it follows that for every $\triangle_\infty \in \mathcal{T}_{-\infty}$ (see Lemma 2.1) there exists a polynomial $P_{\triangle_\infty} \in \Pi_k$ such that

$$f - P_{\triangle_\infty} = \sum_{m\in\mathbb{Z}} p_{m,\eta}(f) \quad \text{absolutely a.e. on } \triangle_\infty.$$

Using that $|\{x\colon |f(x)| > s\}| < \infty$ for $s > 0$ and (2.66), we infer $P_{\triangle_\infty} \equiv 0$ and the theorem follows. □

We next give the equivalence of the skinny B-norms introduced above.

Theorem 2.18. *For each $f \in \mathcal{B}_\tau^{\alpha k}(\mathcal{T})$, the norms $\|f\|_{\mathcal{B}_\tau^{\alpha k}(\mathcal{T})}$, $\mathcal{N}_{\omega,\eta}(f,\mathcal{T})$ $(0 < \eta < p)$, and $\mathcal{N}_{P,\eta}(f,\mathcal{T})$ $(0 < \eta < p)$ are equivalent with constants of equivalence depending only on α, k, p, η, and the parameters of $\mathcal{T}$.*

Proof. The proof of this theorem is similar to (but easier than) the one of Theorem 2.16 and will be omitted. The difference is that the role of $\mathbb{S}_\triangle(f)_\mu$ is now played by $\omega_k(f,\triangle)_\mu$. See also the proof of Theorem 2.20 below. □

Remark. The following simple example shows that, in general, $\mathcal{N}_{\omega,\eta}(f,\mathcal{T})$ is not equivalent to $\|f\|_{\mathcal{B}_\tau^{\alpha k}(\mathcal{T})}$ if $\eta \geqslant p$. Let $f := \mathbb{1}_\triangle$ for some $\triangle \in \mathcal{T}$. It is easily seen that $\|f\|_{\mathcal{B}_\tau^{\alpha k}(\mathcal{T})} \approx |\triangle|^{1/p} = \|f\|_p$, while $\mathcal{N}_{\omega,\eta}(f,\mathcal{T}) = \infty$ if $\eta \geqslant p$.

2.5. Fat B-spaces: The link to Besov spaces

Throughout this section, we assume that $\mathcal{T}$ is an arbitrary strong locally regular triangulation of $\mathbb{R}^2$ (Section 2.1). We define the *fat B-space* $\mathbb{B}^{\alpha k}_{pq}(\mathcal{T})$, $\alpha > 0$, $0 < p, q \leqslant \infty$, $k \geqslant 1$, as the set of all $f \in L_p(\mathbb{R}^2)$ such that

$$\|f\|_{\mathbb{B}^{\alpha k}_{pq}(\mathcal{T})} := \|f\|_p + \left(\sum_{m\in\mathbb{Z}}\left[2^{m\alpha}\left(\sum_{\triangle\in\mathcal{T},\,2^{-m}\leqslant|\triangle|<2^{-m+1}}\omega_k(f,\Omega_\triangle)_p^p\right)^{1/p}\right]^q\right)^{1/q} < \infty,$$

where $\Omega_\triangle$ is defined in (2.9).

As in the previous sections, we shall focus our attention only on the scale of fat B-spaces which naturally occur in nonlinear approximation, namely, the spaces $\mathbb{B}^{\alpha k}_{\tau}(\mathcal{T})$, where $\alpha > 0$, $k \geqslant 1$, $0 < p < \infty$, and $1/\tau := \alpha + 1/p$. We define the space $\mathbb{B}^{\alpha k}_{\tau}(\mathcal{T})$ as the set of all functions $f \in L_p(\mathbb{R}^2)$ such that

$$\|f\|_{\mathbb{B}^{\alpha k}_{\tau}(\mathcal{T})} := \left(\sum_{\triangle\in\mathcal{T}}\left(|\triangle|^{-\alpha}\omega_k(f,\Omega_\triangle)_\tau\right)^\tau\right)^{1/\tau} < \infty, \tag{2.67}$$

which is a modification of the space $\mathbb{B}^{\alpha k}_{\tau\tau}(\mathcal{T})$ from above. By Whitney's inequality (Lemma 2.6), we have

$$\|f\|_{\mathbb{B}^{\alpha k}_{\tau}(\mathcal{T})} \approx \left(\sum_{\triangle\in\mathcal{T}}\left(|\triangle|^{-\alpha}E_k(f,\Omega_\triangle)_\tau\right)^\tau\right)^{1/\tau},$$

where $E_k(f,\Omega_\triangle)_\tau$ is the error of L_τ-approximation to f on $\Omega_\triangle$ from Π_k (see (2.7)).

Note that the use of $\Omega_\triangle$ in the definition of $\|f\|_{\mathbb{B}^{\alpha k}_{\tau}(\mathcal{T})}$ is not crucial. It is almost obvious that, for instance,

$$\|f\|_{\mathbb{B}^{\alpha k}_{\tau}(\mathcal{T})} \approx \left(\sum_{\theta\in\Theta}\left(|\theta|^{-\alpha}\omega_k(f,\theta)_\tau\right)^\tau\right)^{1/\tau}.$$

It is critical, however, that the neighboring sets in the collections $\{\Omega_\triangle\}_{\triangle\in\mathcal{T}}$ or $\{\theta\}_{\theta\in\Theta}$ overlap significantly. This makes the difference between the fat and skinny B-norms.

Clearly, for $f \in L_\tau(\mathbb{R}^2)$ and $\triangle \in \mathcal{T}$, we have the inequalities $E_2(f,\triangle)_\tau \leqslant \mathbb{S}_\triangle(f,\mathcal{T})_\tau \leqslant E_2(f,\Omega_\triangle)_\tau$, which yield the following comparison theorem.

Theorem 2.19. *We have*

$$\|f\|_{\mathcal{B}^{\alpha k}_{\tau}(\mathcal{T})} \leqslant \|f\|_{\mathbb{B}^{\alpha k}_{\tau}(\mathcal{T})}$$

and

$$\|f\|_{\mathcal{B}^{\alpha 2}_{\tau}(\mathcal{T})} \leqslant c\|f\|_{B^{\alpha}_{\tau}(\mathcal{T})} \leqslant c\|f\|_{\mathbb{B}^{\alpha 2}_{\tau}(\mathcal{T})}.$$

We next introduce another norm in $\mathbb{B}^{\alpha k}_{\tau}(\mathcal{T})$. For $f \in L^{\mathrm{loc}}_\eta(\mathbb{R}^2)$, $\eta > 0$, we define

$$\mathbb{N}_{\omega,\eta}(f,\mathcal{T}) := \left(\sum_{\triangle\in\mathcal{T}}\left(|\triangle|^{\frac{1}{p}-\frac{1}{\eta}}\omega_k(f,\Omega_\triangle)_\eta\right)^\tau\right)^{\frac{1}{\tau}} \approx \left(\sum_{\triangle\in\mathcal{T}}\left(|\triangle|^{\frac{1}{p}-\frac{1}{\eta}}E_k(f,\Omega_\triangle)_\eta\right)^\tau\right)^{\frac{1}{\tau}}. \tag{2.68}$$

Evidently, $\mathbb{N}_{\omega,\tau}(f,\mathcal{T}) = \|f\|_{\mathbb{B}^{\alpha k}_{\tau}(\mathcal{T})}$.

To prove the equivalence of $\|f\|_{\mathbb{B}^{\alpha k}_\tau(\mathcal{T})}$ and $\mathbb{N}_{\omega,\eta}(f,\mathcal{T})$ for $0<\eta<p$, we need to introduce one more norm in $\mathbb{B}^{\alpha k}_\tau(\mathcal{T})$. For every $\triangle\in\mathcal{T}$, we let $\mathbb{P}_{\triangle,\eta}(f)$ be a near best L_η-approximation to f on $\Omega_\triangle$ from Π_k with a constant A which is the same for all $\Omega_\triangle$, $\triangle\in\mathcal{T}$ (see (2.16)). We define

$$\mathbb{P}_{m,\eta}(f):=\mathbb{P}_{m,\eta}(f,\mathcal{T}):=\sum_{\triangle\in\mathcal{T}_m}\mathbb{1}_\triangle\cdot\mathbb{P}_{\triangle,\eta}(f)$$

and

$$\pi_{\triangle,\eta}(f):=\mathbb{1}_{\Omega_\triangle}\cdot\big(\mathbb{P}_{m+1,\eta}(f)-\mathbb{P}_{\triangle,\eta}(f)\big)\quad\text{if }\triangle\in\mathcal{T}_m.$$

The new norm is defined by

$$\mathbb{N}_{\pi,\eta}(f,\mathcal{T}):=\bigg(\sum_{\triangle\in\mathcal{T}}\big(|\triangle|^{1/p-1/\eta}\big\|\pi_{\triangle,\eta}(f)\big\|_\eta\big)^\tau\bigg)^{1/\tau}. \tag{2.69}$$

Clearly, since $\mathcal{T}$ is an SLR-triangulation,

$$\mathbb{N}_{\pi,\eta}(f,\mathcal{T})\approx\bigg(\sum_{\triangle\in\mathcal{T}}\big(|\triangle|^{-\alpha}\big\|\pi_{\triangle,\eta}(f)\big\|_\tau\big)^\tau\bigg)^{1/\tau}\approx\bigg(\sum_{\triangle\in\mathcal{T}}\big\|\pi_{\triangle,\eta}(f)\big\|_p^\tau\bigg)^{1/\tau}.$$

Theorem 2.20. *For $f\in\mathbb{B}^{\alpha k}_\tau(\mathcal{T})$, the norms $\|f\|_{\mathbb{B}^{\alpha k}_\tau(\mathcal{T})}$, $\mathbb{N}_{\omega,\eta}(f,\mathcal{T})$ $(0<\eta<p)$, and $\mathbb{N}_{\pi,\eta}(f,\mathcal{T})$ $(0<\eta<p)$, defined in* (2.67)–(2.69) *are equivalent with constants of equivalence depending only on α, p, k, η, and the parameters of $\mathcal{T}$.*

Proof. Using Hölder's inequality and the properties of the SLR-triangulations, we readily obtain

$$\mathbb{N}_{\omega,\eta}(f,\mathcal{T})\leqslant c\,\mathbb{N}_{\omega,\mu}(f,\mathcal{T}),\quad 0<\eta<\mu. \tag{2.70}$$

As we pointed out earlier, $\mathbb{N}_{\omega,\tau}(f,\mathcal{T})=\|f\|_{\mathbb{B}^{\alpha k}_\tau(\mathcal{T})}$. Therefore, it suffices to show that

$$\mathbb{N}_{\omega,\mu}(f,\mathcal{T})\approx\mathbb{N}_{\pi,\eta}(f,\mathcal{T})\quad\text{for all }0<\mu,\eta<p.$$

From the definition of $\mathbb{P}_{\triangle,\eta}(f)$ and $\pi_{\triangle,\eta}(f)$, it follows that for any $\triangle'\in\mathcal{T}_m$

$$\begin{aligned}\big\|\pi_{\triangle',\eta}(f)\big\|_\eta&\leqslant c\big\|f-\mathbb{P}_{m+1,\eta}(f)\big\|_{L_\eta(\Omega_{\triangle'})}+c\big\|f-\mathbb{P}_{\triangle',\eta}(f)\big\|_{L_\eta(\Omega_{\triangle'})}\\&\leqslant c\sum_{\triangle\in\mathcal{T}_{m+1},\,\triangle\subset\Omega_{\triangle'}}\big\|f-\mathbb{P}_{\triangle,\eta}(f)\big\|_{L_\eta(\triangle)}+cE_k(f,\Omega_{\triangle'})_\eta\\&\leqslant c\sum_{\triangle\in\mathcal{T}_{m+1},\,\triangle\subset\Omega_{\triangle'}}E_k(f,\Omega_\triangle)_\eta+cE_k(f,\Omega_{\triangle'})_\eta.\end{aligned}$$

Substituting this estimate in the definition of $\mathbb{N}_{\pi,\eta}(f,\mathcal{T})$ in (2.69), we easily obtain

$$\mathbb{N}_{\pi,\eta}(f,\mathcal{T})\leqslant c\,\mathbb{N}_{\omega,\eta}(f,\mathcal{T}),\quad\eta>0. \tag{2.71}$$

We next prove that if $\mathbb{N}_{\pi,\eta}(f,\mathcal{T})<\infty$, $\eta>0$, then

$$\mathbb{N}_{\omega,\mu}(f,\mathcal{T})\leqslant c\,\mathbb{N}_{\pi,\eta}(f,\mathcal{T}),\quad\tau<\mu<p. \tag{2.72}$$

Evidently, (2.70)–(2.72) yield the theorem.

We introduce the following abbreviated notation: $\mathbb{P}_\triangle := \mathbb{P}_{\triangle,\eta}(f)$, $\mathbb{P}_m := \mathbb{P}_{m,\eta}(f)$, and $\pi_\triangle := \pi_{\triangle,\eta}(f)$. We also set $\rho_m := \mathbb{P}_{m+1} - \mathbb{P}_m$ and $\rho_\triangle := \mathbb{1}_\triangle \cdot \rho_m = \mathbb{1}_\triangle \cdot \pi_\triangle$ for $\triangle \in \mathcal{T}_m$. Evidently, $\|\rho_\triangle\|_p \leqslant \|\pi_\triangle\|_p$, and, hence,

$$\left(\sum_{\triangle\in\mathcal{T}} \|\rho_\triangle\|_p^\tau\right)^{1/\tau} \leqslant \left(\sum_{\triangle\in\mathcal{T}} \|\pi_\triangle\|_p^\tau\right)^{1/\tau} \approx \mathbb{N}_{\pi,\eta}(f,\mathcal{T}) < \infty.$$

It is readily seen that the sequence $\{\Phi_m\} := \{\rho_\triangle\}_{\triangle\in\mathcal{T}}$ satisfies conditions (i) and (ii) of Theorem 3.3. Therefore, $\sum_{\triangle\in\mathcal{T}} |\rho_\triangle(\cdot)| < \infty$ a.e. on $\mathbb{R}^2$, and

$$\left\|\sum_{\triangle\in\mathcal{T}} \left|\rho_\triangle(\cdot)\right|\right\|_p \leqslant c\,\mathbb{N}_{\pi,\eta}(f,\mathcal{T}). \tag{2.73}$$

On the other hand, since $f \in L_\eta^{\mathrm{loc}}(\mathbb{R}^2)$, $\|f - \mathbb{P}_m\|_{L_\eta(\triangle)} \to 0$ as $m \to \infty$ for every $\triangle \in \mathcal{T}$. Exactly as in the proof of Theorem 2.15, it follows that $f - \mathbb{P}_{m+1} \in L_p(\mathbb{R}^2)$ and

$$f - \mathbb{P}_{m+1} = \sum_{j=m+1}^{\infty} \rho_j \quad \text{absolutely a.e. on } \mathbb{R}^2 \tag{2.74}$$

and unconditionally in $L_p(\mathbb{R}^2)$. Now, fix $\triangle' \in \mathcal{T}_n$, $n \in \mathbb{Z}$. Since $\mathbb{P}_{\triangle'}$ is a polynomial of degree $< k$ on $\Omega_{\triangle'}$, we have

$$\omega_k(f, \Omega_{\triangle'})_\mu = \omega_k(f - \mathbb{P}_{\triangle'}, \Omega_{\triangle'})_\mu \leqslant c\|f - \mathbb{P}_{\triangle'}\|_{L_\mu(\Omega_{\triangle'})}. \tag{2.75}$$

Using (2.75), (2.74), and Theorem 3.3 with $\{\Phi_m\} := \{\rho_\triangle\colon \triangle \in \mathcal{T}, \triangle \subset \Omega_{\triangle'}\}$, we obtain

$$\begin{aligned}
\omega_k(f, \Omega_{\triangle'})_\mu^\tau &\leqslant c\|\mathbb{P}_{n+1} - \mathbb{P}_{\triangle'}\|_{L_\mu(\Omega_{\triangle'})}^\tau + c\sum_{j=n+1}^{\infty} \|\rho_j\|_{L_\mu(\Omega_{\triangle'})}^\tau \\
&\leqslant c\|\pi_{\triangle'}\|_\mu^\tau + c\sum_{j=n+1}^{\infty} \left\|\sum_{\triangle\in\mathcal{T}_j,\, \triangle\subset\Omega_{\triangle'}} \rho_\triangle\right\|_\mu^\tau \leqslant c\|\pi_{\triangle'}\|_\mu^\tau + c\sum_{\triangle\in\mathcal{T},\, \triangle\subset\Omega_{\triangle'}} \|\rho_\triangle\|_\mu^\tau \\
&\leqslant c\sum_{\triangle\in\mathcal{T},\, \triangle\subset\Omega_{\triangle'}} \|\pi_\triangle\|_\mu^\tau \leqslant c\sum_{\triangle\in\mathcal{T},\, \triangle\subset\Omega_{\triangle'}} |\triangle|^{\tau(\frac{1}{\mu}-\frac{1}{\eta})}\|\pi_\triangle\|_\eta^\tau,
\end{aligned}$$

where we used Lemma 2.7 and the properties of the SLR-triangulations. Substituting the above estimate in the definition of $\mathbb{N}_{\omega,\mu}(f,\mathcal{T})$, we proceed as in the proof of Theorem 2.16, to obtain (2.72). □

Comparison of regular B-spaces with Besov spaces. The *Besov space* $B_q^s(L_p) = B_q^s(L_p(\mathbb{R}^2))$, $s > 0$, $1 \leqslant p, q \leqslant \infty$, is usually defined as the set of all functions $f \in L_p(\mathbb{R}^2)$ such that

$$|f|_{B_q^s(L_p)} := \left(\int_0^\infty \left(t^{-s}\omega_k(f,t)_p\right)^q \frac{\mathrm{d}t}{t}\right)^{1/q} < \infty \tag{2.76}$$

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 kth modulus of smoothness of f in $L_p(\mathbb{R}^2)$, i.e., $\omega_k(f,t)_p := \sup_{|h|\leqslant t} \|\Delta_h^k(f,\cdot)\|_p$. The norm in $B_q^s(L_p)$ is defined by

$$\|f\|_{B_q^s(L_p)} := \|f\|_p + |f|_{B_q^s(L_p)}.$$

It is well known that if k in (2.76) is replaced by any other integer $> s$, then the resulting space would be the same with an equivalent norm. However, the situation is different when $p < 1$ (see [11]). For this reason we introduce k as an independent parameter of the Besov spaces in the next definition.

In this article, we are interested in nonlinear piecewise polynomial (spline) approximation in $L_p(\mathbb{R}^2)$ $(0 < p < \infty)$. The Besov spaces $B^{2\alpha}_\tau(L_\tau)$ with $\alpha > 0$ and $1/\tau := \alpha + 1/p$ play a distinctive role in this theory. Taking into account that $B^{2\alpha}_\tau(L_\tau)$ is embedded in L_p and the above observation regarding the independence of k and the smoothness parameter, we naturally arrive at the following slightly modified version of the Besov space $B^{2\alpha}_\tau(L_\tau)$.

Assuming that $0 < p < \infty$, $\alpha > 0$, $k \geqslant 1$, and $1/\tau := \alpha + 1/p$, we define the Besov space $B^{2\alpha,k}_\tau(L_\tau)$ as the set of all functions $f \in L_p(\mathbb{R}^2)$ (in place of $f \in L_\tau$) such that

$$\|f\|_{B^{2\alpha,k}_\tau(L_\tau)} := \left(\int_0^\infty \bigl(t^{-2\alpha}\omega_k(f,t)_\tau\bigr)^\tau \frac{\mathrm{d}t}{t}\right)^{1/\tau} < \infty. \tag{2.77}$$

Notice that the B-spaces and Besov spaces are normalized differently with respect to the smoothness parameter. Thus, e.g., the fat B-space $\mathbb{B}^{\alpha k}_\tau(\mathcal{T})$ corresponds to the Besov space $B^{2\alpha,k}_\tau(L_\tau)$.

From the properties of $\omega_k(f,t)_\tau$, it readily follows that:

$$\|f\|_{B^{2\alpha,k}_\tau(L_\tau)} \approx \left(\sum_{m\in\mathbb{Z}} \bigl(2^{2\alpha m}\omega_k\bigl(f,2^{-m}\bigr)_\tau\bigr)^\tau\right)^{1/\tau}. \tag{2.78}$$

Next, we give an equivalent norm for the Besov space $B^{2\alpha,k}_\tau(L_\tau)$ in terms of local polynomial approximation. We let D'_m denote the set of all dyadic squares I of the form

$$I = \left[\frac{\nu-1}{2^m}, \frac{\nu}{2^m}\right) \times \left[\frac{\mu-1}{2^m}, \frac{\mu}{2^m}\right), \qquad \nu,\mu \in \mathbb{Z},$$

and let D''_m be the set of all shifts of $I \in D'_m$ by the vector $e := (2^{-m-1}, 2^{-m-1})$, i.e., $D''_m := \{I + e\colon I \in D'_m\}$. We denote $D_m := D'_m \cup D''_m$ and $D := \bigcup_{m\in\mathbb{Z}} D_m$. We now introduce the following norm:

$$\mathbf{N}(f) := \left(\sum_{I\in D} \bigl(|I|^{-\alpha}\omega_k(f,I)_\tau\bigr)^\tau\right)^{1/\tau} \approx \left(\sum_{I\in D} \bigl(|I|^{-\alpha}E_k(f,I)_\tau\bigr)^\tau\right)^{1/\tau}, \tag{2.79}$$

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

Lemma 2.21. *If $f \in B^{2\alpha,k}_\tau(L_\tau)$, then*

$$\mathbf{N}(f) \approx \|f\|_{B^{2\alpha,k}_\tau(L_\tau)}$$

with constants of equivalence depending only on p, α, and k.

Proof. This lemma is well known and fairly easy to prove. Its proof hinges on the following equivalence:

$$\omega_k(f,I)^\tau_\tau \approx \frac{1}{|I|} \int_{[0,\ell(I)]^2} \int_I \bigl|\Delta^k_h(f,x,I)\bigr|^\tau \,\mathrm{d}x\,\mathrm{d}h, \tag{2.80}$$

where $\ell(I) := |I|^{1/2}$ and $\Delta^k_h(f,x,I) := \sum_{j=0}^k (-1)^{k+j}\binom{k}{j} f(x+jh)$ if $[x,x+kh] \subset I$ and $\Delta^k_h(f,x,I) := 0$ otherwise (see [13] for the proof of (2.80) in the univariate case; the same proof applies to the multivariate case as well). (See also [15].) $\square$

We next consider B-spaces over regular triangulations (see Section 2.1).

Theorem 2.22. *If $\mathcal{T}^*$ is a regular triangulation then $\mathbb{B}_\tau^{\alpha k}(\mathcal{T}^*) = B_\tau^{2\alpha,k}(L_\tau)$ with equivalent norms.*

Proof. This theorem is an immediate consequence of Lemma 2.21 and the following lemma. $\square$

Lemma 2.23. *Suppose $\mathcal{T}^*$ is a regular triangulation with minimal angle $\beta > 0$. Then there exists $i_0 = i_0(\beta)$ such that the following hold:*

(a) *If $I \in D_m$ ($m \in \mathbb{Z}$), then there exists $\triangle \in \mathcal{T}^*$ such that $I \subset \Omega_\triangle$ and $|\triangle| \leqslant 2^{-2m+i_0}$.*
(b) *If $\triangle \in \mathcal{T}^*$ and $2^{-2m} \leqslant |\triangle| < 2^{-2m+2}$, then there exists $I \in D_{m-i_0}$ such that $\Omega_\triangle \subset I$.*

Proof. The proof of this obvious lemma will be omitted. $\square$

Exactly as in the case of B-spaces, we introduce the following norm in the Besov space $B_\tau^{2\alpha,k}(L_\tau)$:

$$\mathbf{N}_\eta(f) := \left(\sum_{I\in D}\left(|I|^{\frac{1}{p}-\frac{1}{\eta}}\omega_k(f,I)_\eta\right)^\tau\right)^{\frac{1}{\tau}} \approx \left(\sum_{I\in D}\left(|I|^{\frac{1}{p}-\frac{1}{\eta}}E_k(f,I)_\eta\right)^\tau\right)^{\frac{1}{\tau}} \tag{2.81}$$

which in integral form gives

$$\mathbf{N}_\eta(f) \approx \left(\int_0^\infty\int_{\mathbb{R}^2}\left[t^{2(\frac{1}{p}-\frac{1}{\eta})}\omega_k\big(f,B_t(x)\big)_\eta\right]^\tau t^{-3}\,\mathrm{d}x\,\mathrm{d}t\right)^{\frac{1}{\tau}}, \tag{2.82}$$

where $B_t(x) := \{y \in \mathbb{R}^2\colon \|y-x\|_2 \leqslant t\}$ or $B_t(x) := \{y \in \mathbb{R}^2\colon \|y-x\|_\infty \leqslant t\}$.

Proposition 2.24. *The norms $\mathbf{N}_\eta(\cdot)$ with $0 < \eta < p$ and $\|\cdot\|_{B_\tau^{2\alpha,k}(L_\tau)}$ are equivalent.*

Proof. Using Lemma 2.23 as in the proof of Theorem 2.22, one can show that $\mathbf{N}_\eta(\cdot) \approx \mathbb{N}_{\omega,\eta}(\cdot,\mathcal{T}^*)$ if $\mathcal{T}^*$ is a regular triangulation. From Theorems 2.20 and 2.22, we obtain

$$\mathbb{N}_{\omega,\eta}(\cdot,\mathcal{T}^*) \approx \|\cdot\|_{\mathbb{B}_\tau^{\alpha k}(\mathcal{T}^*)} \approx \|\cdot\|_{B_\tau^{2\alpha,k}(L_\tau)}. \quad \square$$

Remark. This result is (in essence) well known, see [15] and the references therein. The equivalence of $\mathbf{N}_\eta(\cdot)$ and $\|\cdot\|_{B_\tau^{2\alpha,k}(L_\tau)}$ clearly shows the intimate relation of B-spaces with Besov spaces.

Our last goal in this section is to find the range for the smoothness parameter α, where the Besov $B_\tau^{2\alpha}$-spaces coincide with the corresponding slim or skinny B-spaces over regular triangulations.

Theorem 2.25. *Suppose $\mathcal{T}^*$ is a regular triangulation of $\mathbb{R}^2$, $0 < p < \infty$, and $k \geqslant 1$.*

(a) *If $0 < \alpha < 1 + 1/p$ and $1/\tau := \alpha + 1/p$, then $f \in B_\tau^\alpha(\mathcal{T}^*)$ if and only if $f \in B_\tau^{2\alpha,2}(L_\tau)$, and*

$$\|f\|_{B_\tau^{2\alpha,2}(L_\tau)} \approx \|f\|_{B_\tau^\alpha(\mathcal{T}^*)} \tag{2.83}$$

with constants of equivalence depending only on p, α and $\beta = \beta(\mathcal{T}^*)$. This equivalence is no longer true if $\alpha \geqslant 1 + 1/p$. Moreover, for every $\theta \in \Theta(\mathcal{T}^*)$ and $\alpha \geqslant 1 + 1/p$, we have $\|\varphi_\theta\|_{B_\tau^{2\alpha,2}(L_\tau)} = \infty$, while $\|\varphi_\theta\|_{B_\tau^\alpha(\mathcal{T}^*)} \approx \|\varphi_\theta\|_p$.

(b) *If $0 < \alpha < 1/p$ and $1/\tau := \alpha + 1/p$, then $f \in \mathcal{B}_\tau^{\alpha,k}(\mathcal{T}^*)$ if and only if $f \in B_\tau^{2\alpha,k}(L_\tau)$, and*

$$\|f\|_{B_\tau^{2\alpha,k}(L_\tau)} \approx \|f\|_{\mathcal{B}_\tau^{\alpha k}(\mathcal{T}^*)} \tag{2.84}$$

with constants of equivalence depending only on k, p, α, and $\beta = \beta(\mathcal{T}^*)$. This equivalence is no longer true if $\alpha \geqslant 1/p$. Moreover, for every $\triangle \in \mathcal{T}^*$ and $\alpha \geqslant 1/p$, we have $\|\mathbb{1}_\triangle\|_{B_\tau^{2\alpha,k}(L_\tau)} = \infty$, while $\|\mathbb{1}_\triangle\|_{\mathcal{B}_\tau^{\alpha k}(\mathcal{T}^*)} \approx \|\mathbb{1}_\triangle\|_p$.

Proof. (a) From Theorems 2.19 and 2.22, we have $\|f\|_{\mathcal{B}_\tau^{\alpha}(\mathcal{T}^*)} \leqslant c\|f\|_{B_\tau^{2\alpha,2}(L_\tau)}$ for $\alpha > 0$. We next show that

$$\|f\|_{B_\tau^{2\alpha,2}(L_\tau)} \leqslant c\,\|f\|_{\mathcal{B}_\tau^{\alpha}(\mathcal{T}^*)}, \quad \text{if } 0 < \alpha < 1 + 1/p. \tag{2.85}$$

Let $f \in \mathcal{B}_\tau^{\alpha}(\mathcal{T}^*)$. Then by Theorems 2.15 and 2.16, and (2.38), it follows that f can be represented in the form

$$f = \sum_{\theta \in \Theta} b_\theta \varphi_\theta \quad \text{absolutely a.e. on } \mathbb{R}^2 \tag{2.86}$$

and

$$\|f\|_{\mathcal{B}_\tau^{\alpha}(\mathcal{T}^*)} \approx \left(\sum_{\theta \in \Theta} \|b_\theta \varphi_\theta\|_p^\tau\right)^{1/\tau}, \tag{2.87}$$

where $\Theta := \Theta(\mathcal{T}^*)$.

Denote $\Xi_j := \{\theta \in \Theta\colon\ 2^{-2j} \leqslant |\theta| < 2^{-2(j-1)}\}$. Since $\mathcal{T}^*$ is regular, straightforward calculations show that, for each $\theta \in \Theta$,

$$\omega_2(\varphi_\theta, t)_\tau^\tau \approx \begin{cases} |\theta|^{(1-\tau)/2} \cdot t^{1+\tau}, & \text{if } 0 < t < |\theta|^{1/2}, \\ |\theta|, & \text{if } t \geqslant |\theta|^{1/2}, \end{cases}$$

and hence, for $\theta \in \Xi_j$ and $t > 0$,

$$\omega_2(b_\theta \varphi_\theta, t)_\tau^\tau \approx \min\bigl\{\|b_\theta \varphi_\theta\|_p^\tau \cdot 2^{-j(\alpha-1-1/p)\tau} t^{1+\tau},\ \|b_\theta \varphi_\theta\|_p^\tau \cdot 2^{-j2\alpha\tau}\bigr\}, \tag{2.88}$$

where we used that $1/\tau = \alpha + 1/p$.

Denote $f_j := \sum_{\theta \in \Xi_j} b_\theta \varphi_\theta$. Since $\mathcal{T}^*$ is regular, $\#\{\theta \in \Xi_j\colon\ x \in \theta\} \leqslant c(\beta)$ for $x \in \mathbb{R}^2$ and $j \in \mathbb{Z}$. Therefore,

$$\omega_2(f_j, t)_\tau^\tau \leqslant c \sum_{\theta \in \Xi_j} \omega_2(b_\theta \varphi_\theta, t)_\tau^\tau, \quad j \in \mathbb{Z}. \tag{2.89}$$

From (2.88) and (2.89), we derive that for any fixed $m \in \mathbb{Z}$

$$\omega_2\bigl(f_j, 2^{-m}\bigr)_\tau^\tau \leqslant c \sum_{\theta \in \Xi_j} 2^{-j(\alpha-1-1/p)\tau} \cdot 2^{-m(1+\tau)} \|b_\theta \varphi_\theta\|_p^\tau, \quad \text{if } j < m, \tag{2.90}$$

and

$$\omega_2\bigl(f_j, 2^{-m}\bigr)_\tau^\tau \leqslant c \sum_{\theta \in \Xi_j} 2^{-j2\alpha\tau} \|b_\theta \varphi_\theta\|_p^\tau, \quad \text{if } j \geqslant m. \tag{2.91}$$

Let $\lambda := \min\{\tau, 1\}$. Then, using (2.90) and (2.91), we have

$$\omega_2\big(f,2^{-m}\big)_\tau^\lambda \leqslant \sum_{j\in\mathbb{Z}}\omega_2\big(f_j,2^{-m}\big)_\tau^\lambda \leqslant c\sum_{j=-\infty}^{m-1}\bigg(\sum_{\theta\in\Xi_j}2^{-j(\alpha-1-\frac{1}{p})\tau-m(1+\tau)}\|b_\theta\varphi_\theta\|_p^\tau\bigg)^{\frac{\lambda}{\tau}} + c\sum_{j=m}^{\infty}\bigg(\sum_{\theta\in\Xi_j}2^{-j2\alpha\tau}\|b_\theta\varphi_\theta\|_p^\tau\bigg)^{\frac{\lambda}{\tau}}.$$

Substituting this in (2.78), we obtain

$$\|f\|_{B_\tau^{2\alpha,2}(L_\tau)}^\tau \leqslant c\sum_{m\in\mathbb{Z}}2^{m(\alpha-1-\frac{1}{p})\tau}\Bigg(\sum_{j=-\infty}^{m}\bigg(2^{-j(\alpha-1-\frac{1}{p})\tau}\sum_{\theta\in\Xi_j}\|b_\theta\varphi_\theta\|_p^\tau\bigg)^{\frac{\lambda}{\tau}}\Bigg)^{\frac{\tau}{\lambda}} + c\sum_{m\in\mathbb{Z}}2^{m2\alpha\tau}\Bigg(\sum_{j=m}^{\infty}2^{-j2\alpha\tau}\bigg(\sum_{\theta\in\Xi_j}\|b_\theta\varphi_\theta\|_p^\tau\bigg)^{\frac{\lambda}{\tau}}\Bigg)^{\frac{\tau}{\lambda}},$$

where we used that $2\alpha\tau-\tau-1=\tau(\alpha-1-1/p)$ since $1/\tau=\alpha+1/p$. To estimate the above sums, we use the well known discrete Hardy inequalities. Namely, we apply, e.g., the inequality from Lemma 3.10 of [13] to estimate the first sum and Lemma 3.4 from [7] to the second sum. We obtain

$$\|f\|_{B_\tau^{2\alpha,2}(L_\tau)}^\tau \leqslant c\sum_{j\in\mathbb{Z}}\sum_{\theta\in\Xi_j}\|b_\theta\varphi_\theta\|_p^\tau \leqslant c\|f\|_{B_\tau^\alpha(\mathcal{T}^*)}^\tau,$$

which completes the proof of (2.85).

Using (2.88), we obtain

$$\begin{aligned}\|\varphi_\theta\|_{B_\tau^{2\alpha,2}(L_\tau)}^\tau &:= \int_0^\infty \big(t^{-2\alpha}\omega_2(\varphi_\theta,t)_\tau\big)^\tau\frac{\mathrm{d}t}{t}\\ &\approx |\theta|^{(1-\tau)/2}\int_0^{|\theta|^{1/2}} t^{(-2\alpha+1)\tau}\,\mathrm{d}t + |\theta|\int_{|\theta|^{1/2}}^\infty t^{-2\alpha\tau-1}\,\mathrm{d}t\\ &\approx |\theta|^{(1-\tau)/2}\int_0^{|\theta|^{1/2}} t^{(-2\alpha+1)\tau}\,\mathrm{d}t + |\theta|^{\tau/p}.\end{aligned}$$

Therefore, $\|\varphi_\theta\|_{B_\tau^{2\alpha,2}(L_\tau)}=\infty$ if $(-2\alpha+1)\tau\leqslant -1$ which is equivalent to $\alpha\geqslant 1+1/p$, using that $1/\tau=\alpha+1/p$. It is easily seen that $\|\varphi_\theta\|_{B_\tau^\alpha(\mathcal{T}^*)}\approx\|\varphi_\theta\|_p$, which follows from the Bernstein inequality in Theorem 3.7 as well.

(b) Simple calculations show that $\omega_k(\mathbb{1}_\triangle,t)_\tau^\tau\approx\min\{|\triangle|^{1/2}t,|\triangle|\}$ for $\triangle\in\mathcal{T}^*$ and $t>0$. The rest of the proof is similar to the proof of part (a) and will be omitted. □

Comparison between B-spaces over different triangulations and Besov spaces. Suppose $\mathcal{T}$ is an arbitrary strong locally regular triangulation of $\mathbb{R}^2$ (Section 2.1) and $0<p<\infty$. It can be proved that there exists $\alpha_0=\alpha_0(p,\beta,M_0)>0$ such that if $0<\alpha<\alpha_0$ and $f\in B_\tau^\alpha(\mathcal{T})$ with $1/\tau:=\alpha+1/p$, then

$$\|f\|_{\mathbb{B}_\tau^{\alpha 2}(\mathcal{T})}\leqslant c\|f\|_{B_\tau^\alpha(\mathcal{T})}.$$

We leave the proof of this result for elsewhere since it is much longer and more involved than the proof of Theorem 2.25. Thus the fat B-norm $\|f\|_{\mathbb{B}^{\alpha 2}_\tau(\mathcal{T})}$ is equivalent to the slim B-norm $\|f\|_{B^\alpha_\tau(\mathcal{T})}$ for some relatively small range $0<\alpha<\alpha_0$ and becomes much larger when $\alpha \geqslant \alpha_0$. The relationship between fat and skinny B-spaces is quite similar. We skip the details.

It is essential for our theory that the Courant elements φ_θ, $\theta \in \Theta(\mathcal{T})$, have infinite smoothness (smoothness of order $\alpha>0$ for every α) in the slim B-space scale $B^\alpha_\tau(\mathcal{T})$. At the same time each φ_θ has limited smoothness $\alpha<\alpha_0$ in the corresponding fat B-space scale.

If one compares a B^α_τ-space over an arbitrary triangulation with the corresponding Besov space $B^{2\alpha,k}_\tau(L_\tau)$ (or two B-spaces over different triangulations with each other), then everything changes dramatically. As was shown in Section 2.1, there exist strong locally regular triangulations with extremely skinny Courant elements which cause problems to Besov spaces. More precisely, let φ_θ be the Courant element associated with a cell $\theta \in \Theta$ which is convex, has length $l>0$ and width εl with $0<\varepsilon<1$. Simple calculations show that $\omega_2(\varphi_\theta,t)^\tau_\tau \approx \min\{\varepsilon^{-\tau} l^{1-\tau} t^{1+\tau}, \varepsilon l^2\}$. Furthermore, we have $\|\varphi_\theta\|_{B^{2\alpha,2}_\tau(L_\tau)} \approx \varepsilon^{-\alpha}\|\varphi_\theta\|_p$ if $0<\alpha<1+1/p$ and $\|\varphi_\theta\|_{B^{2\alpha,2}_\tau(L_\tau)} = \infty$ if $\alpha \geqslant 1+1/p$. At the same time, $\|\varphi_\theta\|_{B^\alpha_\tau(\mathcal{T})} \approx \|\varphi_\theta\|_p$ for each $\alpha>0$. 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 Courant element approximation in the case of nonregular triangulations.

B-spaces in dimensions $d \neq 2$. Slim, skinny, and fat B-spaces in d dimensions ($d>2$) can be defined and utilized similarly as in the two-dimensional case. We do not consider them in the present article simply to avoid some complications that are unnecessary at this point. Of course, the B-spaces can be defined in the univariate case as well. However, it can be shown that the univariate slim, skinny, and fat B-spaces do not give anything better than the corresponding Besov spaces if $0<p<\infty$ and, therefore, are useless. The point is that in the univariate case the Bernstein inequality holds with no restrictions on $\alpha>0$ (see [11]). In the case of $p=\infty$, however, the B-spaces are different from the corresponding Besov spaces.

3. Nonlinear piecewise polynomial approximation

In this section, we give our main results for nonlinear n-term approximation in $L_p(\mathbb{R}^2)$ ($0<p<\infty$) from: (a) Courant elements generated by LR-triangulations and (b) discontinuous piecewise polynomials over WLR-triangulations.

3.1. Nonlinear n-term approximation: General principles

We begin with a brief description of the general principles that will be guiding us in developing the theory of nonlinear n-term approximation by piecewise polynomials.

Let X be a normed or quasi-normed function space, where the approximation will take place (in this article, $X=L_p(\mathbb{R}^2)$, $0<p<\infty$). 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. We 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 $\Lambda_n \subset \Theta$, $\#\Lambda_n \leqslant n$, and Λ_n varies 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 main objective in this article is to describe the spaces of functions of given rates of n-term approximation. More precisely, we want to characterize the approximation space $A_q^\gamma := A_q^\gamma(\Phi)$, $\gamma > 0$, $0 < q \leqslant \infty$, consisting of all functions $f \in X$ such that

$$\|f\|_{A_q^\gamma} := \|f\|_X + \left(\sum_{n=1}^{\infty} \bigl(n^\gamma \sigma_n(f)\bigr)^q \frac{1}{n}\right)^{1/q} < \infty \tag{3.1}$$

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 $\sigma_n(f) \leqslant cn^{-\gamma}$.

To achieve our goals, we shall use the machinery of Jackson and Bernstein estimates plus interpolation spaces. Suppose $B \subset X$ is a smoothness space with a (quasi-)norm $\|\cdot\|_B$, satisfying the λ-triangle inequality: $\|f + g\|_B^\lambda \leqslant \|f\|_B^\lambda + \|g\|_B^\lambda$ with $0 < \lambda \leqslant 1$ (in our case, B will be some B-space), and let $\Phi \subset B$. The K-functional is defined by

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

The interpolation space $(X, B)_{\mu,q}$ (real method of interpolation) is defined as the set of all $f \in X$ such that

$$\|f\|_{(X,B)_{\mu,q}} := \|f\|_X + \left(\sum_{m=0}^{\infty} \bigl[2^{m\mu} K\bigl(f, 2^{-m}\bigr)\bigr]^q\right)^{1/q} < \infty, \quad 0 \leqslant \mu \leqslant 1,$$

where the ℓ_q-norm is replaced by the sup-norm if $q = \infty$ (see, e.g., [3,4]).

The well known machinery of Jackson and Bernstein estimates allows to characterize the rates of n-term approximation from Φ:

Theorem 3.1. (a) *Suppose the following Jackson estimate holds*: *There is* $\alpha > 0$ *such that for* $f \in B$

$$\sigma_n(f) \leqslant cn^{-\alpha} \|f\|_B, \quad n \geqslant 1. \tag{3.2}$$

Then, for $f \in X$,

$$\sigma_n(f) \leqslant cK\bigl(f, n^{-\alpha}\bigr), \quad n \geqslant 1. \tag{3.3}$$

(b) *Suppose the following Bernstein inequality holds*: *There is* $\alpha > 0$ *such that*

$$\|S\|_B \leqslant cn^\alpha \|S\|_X, \quad \text{for } S \in \Sigma_n,\ n \geqslant 1. \tag{3.4}$$

Then, for $f \in X$,

$$K\bigl(f, n^{-\alpha}\bigr) \leqslant cn^{-\alpha}\left(\left[\sum_{\nu=1}^{n} \frac{1}{\nu}\bigl(\nu^\alpha \sigma_\nu(f)\bigr)^\lambda\right]^{1/\lambda} + \|f\|_X\right), \quad n \geqslant 1. \tag{3.5}$$

Proof. For the proof of this theorem see, e.g., [13]. □

An immediate consequence of Theorem 3.1 is that if the Jackson and Bernstein inequalities (3.2) and (3.4) hold, then $\sigma_n(f) = \mathrm{O}(n^{-\gamma})$, $0 < \gamma < \alpha$, if and only if $K(f, n^{-\alpha}) = \mathrm{O}(n^{-\gamma})$. More generally, Theorem 3.1 readily yields the following characterization of the approximation spaces $A_q^\gamma(\Phi)$:

Theorem 3.2. *Suppose the Jackson and Bernstein inequalities* (3.2) *and* (3.4) *from Theorem* 3.1 *hold. Then*

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

with equivalent norms.

General embedding theorem and Jackson estimate for nonlinear n-term approximation

Theorem 3.3. *Suppose $\{\Phi_m\}$ is a sequence of functions in $L_p(\mathbb{R}^d)$, $d \geqslant 1$, $0 < p < \infty$, which satisfies the following additional properties when $1 < p < \infty$:*

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

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

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

$$\sum_{E_j \ni x,\ |E_j| \geqslant |E_m|} \left(\frac{|E_m|}{|E_j|}\right)^{1/p} \leqslant c_1,$$

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

$$N(f) := \left(\sum_m \|\Phi_m\|_p^\tau\right)^{1/\tau} < \infty. \tag{3.6}$$

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

$$\|f\|_p \leqslant \left\| \sum_m \left|\Phi_m(\cdot)\right| \right\|_p \leqslant cN(f), \tag{3.7}$$

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

Furthermore, if $1 \leqslant p < \infty$, condition (3.6) *can be replaced by the weaker condition*

$$N(f) := \left\| \left\{ \|\Phi_m\|_p \right\} \right\|_{w\ell_\tau} < \infty, \tag{3.8}$$

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

$$\left\|\{x_m\}\right\|_{w\ell_\tau} := \inf\left\{M:\ \#\left\{m:\ |x_m| > Mn^{-1/\tau}\right\} \leqslant n \text{ for } n = 1, 2, \ldots\right\}. \tag{3.9}$$

Theorem 3.4. *Under the hypothesis of Theorem* 3.3, *suppose $\{\Phi_m^*\}_{j=1}^\infty$ is a rearrangement of the sequence $\{\Phi_m\}$ such that $\|\Phi_1^*\|_p \geqslant \|\Phi_2^*\|_p \geqslant \cdots$. Denote $S_n := \sum_{j=1}^n \Phi_j^*$. Then*

$$\|f - S_n\|_p \leqslant cn^{-\alpha} N(f) \quad \text{with } \alpha = 1/\tau - 1/p, \tag{3.10}$$

where $c = 1$ *if* $0 < p \leqslant 1$ *and* $c = c(\alpha, p, c_1)$ *if* $1 < p < \infty$.

Furthermore, the estimate remains valid if condition (3.6) *is replaced by* (3.8) *when* $1 \leqslant p < \infty$.

Proof of Theorems 3.3 and 3.4. *Case* I: $0 < p < 1$. Since $\tau < p$, we have

$$\Bigl\|\sum_m \bigl|\Phi_m(\cdot)\bigr|\Bigr\|_p \leqslant \Bigl(\sum_m \|\Phi_m\|_p^p\Bigr)^{1/p} \leqslant \Bigl(\sum_m \|\Phi_m\|_p^\tau\Bigr)^{1/\tau} = N(f)$$

which proves Theorem 3.3 in this case. To estimate $\|f - S_n\|_p$ we shall use the following simple inequality: If $x_1 \geqslant x_2 \geqslant \cdots \geqslant 0$ and $0 < \tau < p$, then

$$\Biggl(\sum_{j=n+1}^{\infty} x_j^p\Biggr)^{1/p} \leqslant n^{1/p-1/\tau}\Biggl(\sum_{j=1}^{\infty} x_j^\tau\Biggr)^{1/\tau}. \tag{3.11}$$

The proof of this inequality is given in Appendix B. Applying (3.11) with $x_j := \|\Phi_j^*\|_p$, we obtain

$$\begin{aligned}\|f - S_n\|_p &\leqslant \Biggl\|\sum_{j=n+1}^{\infty} \bigl|\Phi_j^*(\cdot)\bigr|\Biggr\|_p \leqslant \Biggl(\sum_{j=n+1}^{\infty} \bigl\|\Phi_j^*\bigr\|_p^p\Biggr)^{1/p} \\ &\leqslant n^{1/p-1/\tau}\Biggl(\sum_{j=1}^{\infty} \bigl\|\Phi_j^*\bigr\|_p^\tau\Biggr)^{1/\tau} = n^{-\alpha} N(f),\end{aligned}$$

which proves Theorem 3.4 in Case I.

Case II: $1 \leqslant p < \infty$. We need the following lemma.

Lemma 3.5. *Let* $F := \sum_{j \in \mathcal{J}_n} |\Phi_j|$, *where* $\#\mathcal{J}_n \leqslant n$, *and* $\|\Phi_j\|_p \leqslant L$ *for* $j \in \mathcal{J}_n$. *Then*

$$\|F\|_p \leqslant cLn^{1/p}$$

with $c = c(p, c_1)$.

Proof. Let $1 < p < \infty$ (the case $p = 1$ is trivial). Using property (i) of the sequence $\{\Phi_m\}$, we have

$$\|F\|_p \leqslant \Biggl\|\sum_{j \in \mathcal{J}_n} \|\Phi_j\|_\infty \cdot \mathbb{1}_{E_j}(\cdot)\Biggr\|_p \leqslant cL\Biggl\|\sum_{j \in \mathcal{J}_n} |E_j|^{-1/p} \cdot \mathbb{1}_{E_j}(\cdot)\Biggr\|_p.$$

We define $E := \bigcup_{j \in \mathcal{J}_n} E_j$ and $\mathcal{E}(x) := \min\{|E_j| \text{ and } j \in \mathcal{J}_n \text{ and } E_j \ni x\}$ for $x \in E$. Property (ii) yields $\sum_{j \in \mathcal{J}_n} |E_j|^{-1/p} \cdot \mathbb{1}_{E_j}(x) \leqslant c_1 \mathcal{E}(x)^{-1/p}$ for $x \in \mathbb{R}^2$. Therefore,

$$\begin{aligned}\|F\|_p &\leqslant cL\bigl\|\mathcal{E}(\cdot)^{-1/p}\bigr\|_{L_p} = cL\Biggl(\int_E \mathcal{E}(x)^{-1}\,\mathrm{d}x\Biggr)^{1/p} \\ &\leqslant cL\Biggl(\sum_{j \in \mathcal{J}_n} |E_j|^{-1}\int_{\mathbb{R}^2} \mathbb{1}_{E_j}(x)\,\mathrm{d}x\Biggr)^{1/p} = cL(\#\mathcal{J}_n)^{1/p} \leqslant cLn^{1/p}. \quad \square\end{aligned}$$

We define $\Xi_\mu := \{j$ and $2^{-\mu}N(f) \leqslant \|\Phi_j\|_p < 2^{-\mu+1}N(f)\}$. Then $\bigcup_{\nu\leqslant\mu}\Xi_\nu = \{j$ and $\|\Phi_j\|_p \geqslant 2^{-\mu}N(f)\}$ and hence, using (3.6) or (3.8), we derive

$$\sum_{\nu\leqslant\mu}\#\Xi_\nu = \#\left(\bigcup_{\nu\leqslant\mu}\Xi_\nu\right) \leqslant 2^{\mu\tau}. \tag{3.12}$$

Therefore,

$$\#\Xi_\mu \leqslant \sum_{\nu\leqslant\mu}\#\Xi_\nu \leqslant 2^{\mu\tau}. \tag{3.13}$$

We denote $M := \sum_{\mu\leqslant m}\#\Xi_\mu$. By (3.12), $M \leqslant 2^{m\tau}$. Let $F_\mu := \sum_{j\in\Xi_\mu}|\Phi_j|$. Using Lemma 3.5 and (3.13), we obtain

$$\begin{aligned}\|f - S_M\|_p &\leqslant \left\|\sum_{\mu=m+1}^{\infty}F_\mu\right\|_p \leqslant \sum_{\mu=m+1}^{\infty}\|F_\mu\|_p\\ &\leqslant c\sum_{\mu=m+1}^{\infty}2^{-\mu}N(f)(\#\Xi_\mu)^{1/p} \leqslant cN(f)\sum_{\mu=m+1}^{\infty}2^{-\mu(1-\tau/p)}\\ &= cN(f)2^{-m(1-\tau/p)} \leqslant cM^{-1/\tau+1/p}N(f) = cM^{-\alpha}N(f).\end{aligned}$$

This estimate readily implies (3.10). Evidently, (3.7) is also contained in the above result (take $S_M := 0$). This completes the proofs of Theorems 3.3 and 3.4. □

As will be seen in Sections 3.2 and 3.3, Theorem 3.4 easily gives the needed Jackson estimates for piecewise polynomial approximation (see Theorems 3.6 and 3.10). However, there is no simple recipe for proving Bernstein estimates (see Appendix A).

3.2. Nonlinear n-term Courant element approximation

In this section, we assume that $\mathcal{T}$ is a locally regular triangulation of $\mathbb{R}^2$. We denote by $\Phi_{\mathcal{T}}$ the collection of all Courant elements φ_θ generated by $\mathcal{T}$ (see Section 2.1). Notice that $\Phi_{\mathcal{T}}$ is not a basis; $\Phi_{\mathcal{T}}$ is redundant. We consider the nonlinear n-term approximation in $L_p(\mathbb{R}^2)$ $(0 < p < \infty)$ from $\Phi_{\mathcal{T}}$. Our main goal is to characterize the approximation spaces generated by this approximation. We let $\widetilde{\Sigma}_n(\mathcal{T})$ denote the nonlinear set consisting of all continuous piecewise linear functions S of the form

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

where $\Lambda_n \subset \Theta(\mathcal{T})$, $\#\Lambda_n \leqslant n$, and Λ_n may vary with S. We denote by $\tilde{\sigma}_n(f,\mathcal{T})_p$ the error of L_p-approximation to $f \in L_p(\mathbb{R}^2)$ from $\widetilde{\Sigma}_n(\mathcal{T})$:

$$\tilde{\sigma}_n(f,\mathcal{T})_p := \inf_{S\in\widetilde{\Sigma}_n(\mathcal{T})}\|f - S\|_p.$$

Throughout this section, we assume that $0 < p < \infty$, $\alpha > 0$, and $1/\tau := \alpha + 1/p$, and denote by $B_\tau^\alpha(\mathcal{T})$ the slim B-space introduced in Section 2.3. We next prove a pair of companion Jackson and Bernstein estimates.

Theorem 3.6 (Jackson estimate). *If $f \in B^\alpha_\tau(\mathcal{T})$, then*

$$\tilde{\sigma}_n(f, \mathcal{T})_p \leqslant cn^{-\alpha}\|f\|_{B^\alpha_\tau(\mathcal{T})} \tag{3.14}$$

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

Remark. Estimate (3.14) remains valid if $\|f\|_{B^\alpha_\tau(\mathcal{T})}$ is replaced by $\|\{\|b_\theta\varphi_\theta\|_p\}\|_{w\ell_\tau}$ with $\{b_\theta\}$ from (2.32) or (2.33) as in the definition of $N_{Q,\tau}(f)$ (see (2.34)), where $\|\cdot\|_{w\ell_\tau}$ is the weak ℓ_τ-norm defined in (3.9).

Proof. By Theorem 2.15, it follows that:

$$f = \sum_{\theta\in\Theta} b_\theta(f)\varphi_\theta \quad \text{absolutely a.e. on } \mathbb{R}^2,$$

where $\{b_\theta\}$ are from (2.32) or (2.33). We use Theorem 3.4, (2.38), and Theorem 2.16 to obtain

$$\tilde{\sigma}_n(f, \mathcal{T})_p \leqslant cn^{-\alpha}\left(\sum_{\theta\in\Theta}\bigl\|b_\theta(f)\varphi_\theta\bigr\|_p^\tau\right)^{1/\tau} \approx cn^{-\alpha}N_{Q,\tau}(f) \approx cn^{-\alpha}\|f\|_{B^\alpha_\tau(\mathcal{T})}. \quad \square$$

Theorem 3.7 (Bernstein estimate). *If $S \in \widetilde{\Sigma}_n(\mathcal{T})$, then*

$$\|S\|_{B^\alpha_\tau(\mathcal{T})} \leqslant cn^\alpha\|S\|_p \tag{3.15}$$

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

The proof of this theorem is more involved than the one of Theorem 3.6. We shall give it in Appendix A.

We denote by $\tilde{A}^\gamma_q := \tilde{A}^\gamma_q(L_p, \mathcal{T})$ the approximation space generated by n-term Courant element approximation (see (3.1)). The Jackson and Bernstein estimates from Theorems 3.6 and 3.7 yield the following characterization of the approximation spaces $\tilde{A}^\gamma_q(L_p, \mathcal{T})$ (see Theorem 3.2):

Theorem 3.8. *If $0 < \gamma < \alpha$ and $0 < q \leqslant \infty$, then*

$$\tilde{A}^\gamma_q(L_p, \mathcal{T}) = \bigl(L_p, B^\alpha_\tau(\mathcal{T})\bigr)_{\frac{\gamma}{\alpha},q}$$

with equivalent norms.

"Algorithm" for nonlinear n-term Courant element approximation. One of our primary motivations for this work was the development of methods for n-term Courant element approximation which capture the rates of the best approximation. The proofs of Theorems 3.3 and 3.6 suggest the following approximation scheme, where we assume that $f \in L_p(\mathbb{R}^2)$, $1 < p < \infty$, and $\mathcal{T}$ is a fixed LR-triangulation of $\mathbb{R}^2$:

Step 1. We use the operators $q_m(f) := q_m(f, \mathcal{T})$ induced by the quasi-interpolant (see (2.31)) to find the following decomposition of f:

$$f = \sum_{m\in\mathbb{Z}} q_m(f) = \sum_{m\in\mathbb{Z}}\sum_{\theta\in\Theta_m} b_\theta(f)\varphi_\theta,$$

where $\{b_\theta(f)\}$ are defined by (2.32) and the identity was established by Theorem 2.15.

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

$$\left\|b_{\theta_1}(f)\varphi_{\theta_1}\right\|_p \geqslant \left\|b_{\theta_2}(f)\varphi_{\theta_2}\right\|_p \geqslant \cdots.$$

Then we define the n-term approximant by

$$\tilde{A}_n(f)_p = \tilde{A}_n(f,\mathcal{T})_p := \sum_{j=1}^{n} b_{\theta_j}(f)\varphi_{\theta_j}.$$

This procedure becomes practically feasible in the setting of approximation of functions defined on compact polygonal domains.

By Theorem 3.4, it follows that:

$$\left\|f - \tilde{A}_n(f)_p\right\|_p \leqslant cn^{-\alpha}\|f\|_{B^\alpha_\tau(\mathcal{T})}.$$

If $0 < p \leqslant 1$, we use the more complicated nonlinear operators $t_{m,\eta}(f)$ ($\eta < p$) from (2.31) instead of $q_m(f)$ and the coefficients $b_\theta(f) := b_{\theta,\eta}$ defined in (2.33). The same estimate for the error holds again by Theorem 3.4.

These results imply that the above algorithm achieves the rates of the best n-term Courant element approximation. We shall further elaborate on this in a forthcoming article.

n-term approximation from the library $\{\Phi_{\mathcal{T}}\}$. We denote by $\tilde{\sigma}_n(f)_p$ the error of n-term approximation to $f \in L_p(\mathbb{R}^2)$ from the best Courant element collection, i.e.,

$$\tilde{\sigma}_n(f)_p := \inf_{\mathcal{T}} \tilde{\sigma}_n(f,\mathcal{T})_p,$$

where the infimum is taken over all LR-triangulations $\mathcal{T}$ with some fixed parameters M_0, N_0, r, ρ, and δ. The following result is immediate from Theorem 3.6.

Theorem 3.9. *Suppose* $\inf_{\mathcal{T}}\|f\|_{B^\alpha_\tau(\mathcal{T})} < \infty$, *where the infimum is taken over all LR-triangulations with some fixed parameters* M_0, N_0, r, ρ, *and* δ, *and let* $f \in L_p(\mathbb{R}^2)$. *Then*

$$\tilde{\sigma}_n(f)_p \leqslant cn^{-\alpha}\inf_{\mathcal{T}}\|f\|_{B^\alpha_\tau(\mathcal{T})},$$

where c depends on α, p, and the parameters $M_0, N_0, r, \rho, \delta$.

It is an *open problem* to characterize the rates of approximation generated by $\{\tilde{\sigma}_n(f)_p\}$. The difficulty stems from the highly nonlinear structure of approximation from the library $\{\Phi_{\mathcal{T}}\}_{\mathcal{T}}$.

3.3. Nonlinear approximation from (discontinuous) piecewise polynomials

In this section, we assume that $\mathcal{T}$ is a weak locally regular triangulation of $\mathbb{R}^2$ (Section 2.1). We denote by $\Sigma_n^k(\mathcal{T})$, $k \geqslant 1$, the nonlinear set of all n-term piecewise polynomial function of the form

$$S = \sum_{\triangle\in\Lambda_n} \mathbb{1}_\triangle \cdot P_\triangle,$$

where $P_\triangle \in \Pi_k$, $\Lambda_n \subset \mathcal{T}$, $\#\Lambda_n \leqslant n$, and Λ_n may vary with S. We denote by $\sigma_n(f,\mathcal{T})_p$ the error of L_p-approximation to $f \in L_p(\mathbb{R}^2)$ from $\Sigma_n^k(\mathcal{T})$:

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

We want to characterize the approximation spaces generated by $\sigma_n(f,\mathcal{T})_p$. To this end we shall proceed according to the recipe from Section 3.1. We shall first prove Jackson and Bernstein estimates. Throughout the rest of the section, we assume that $0 < p < \infty$, $k \geqslant 1$, $\alpha > 0$, and $1/\tau = \alpha + 1/p$. Recall that $\mathcal{B}_\tau^{\alpha,k}(\mathcal{T})$ denotes for the skinny B-space introduced in Section 2.4.

Theorem 3.10 (Jackson estimate). *If $f \in \mathcal{B}_\tau^{\alpha k}(\mathcal{T})$, then*

$$\sigma_n(f,\mathcal{T})_p \leqslant cn^{-\alpha}\|f\|_{\mathcal{B}_\tau^{\alpha k}(\mathcal{T})}$$

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

Remark. The conclusion of Theorem 3.10 remains valid if $\|f\|_{\mathcal{B}_\tau^{\alpha k}(\mathcal{T})}$ is replaced by the weak ℓ_τ-norm $\|\{p_{\triangle,\eta}(f)\}_{\triangle\in\mathcal{T}}\|_{w\ell_\tau}$ of the sequence $\{p_{\triangle,\eta}(f)\}_{\triangle\in\mathcal{T}}$, $0 < \eta < p$, defined in (2.61) (see also (3.9) for the definition of $\|\cdot\|_{w\ell_\tau}$).

Proof. By Theorem 2.17, we have $f = \sum_{\triangle\in\mathcal{T}} p_\triangle$ absolutely a.e. on $\mathbb{R}^2$ and $\|f\|_{\mathcal{B}_\tau^{\alpha k}(\mathcal{T})} \approx (\sum_{\triangle\in\mathcal{T}}\|p_\triangle\|_p^\tau)^{1/\tau}$, where $p_\triangle := p_{\triangle,\eta}(f)$ $(0 < \eta < p)$ are from (2.61). Evidently, the sequence $\{\Phi_j\} := \{p_\triangle\}_{\triangle\in\mathcal{T}}$ satisfies the requirements of Theorem 3.3 and, therefore,

$$\sigma_n(f,\mathcal{T})_p \leqslant cn^{-\alpha}\left(\sum_{\triangle\in\mathcal{T}}\|p_\triangle\|_p^\tau\right)^{1/\tau} \leqslant cn^{-\alpha}\|f\|_{\mathcal{B}_\tau^{\alpha k}(\mathcal{T})}. \quad \square$$

Theorem 3.11 (Bernstein estimate). *If $S \in \Sigma_n^k(\mathcal{T})$, then*

$$\|S\|_{\mathcal{B}_\tau^{\alpha k}(\mathcal{T})} \leqslant cn^\alpha\|S\|_p \tag{3.16}$$

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

We shall give the proof of this theorem together with the proof of Theorem 3.7 in Appendix A.

Now, we denote by $A_q^\gamma := A_q^\gamma(L_p,\mathcal{T})$ the approximation space generated by $\{\sigma_n(f,\mathcal{T})_p\}$ (see (3.1)). The following characterization of the approximation spaces A_q^γ follows by Theorems 3.10 and 3.11 (see Theorems 3.1 and 3.2):

Theorem 3.12. *If $0 < \gamma < \alpha$ and $0 < q \leqslant \infty$, then*

$$A_q^\gamma(L_p,\mathcal{T}) = \big(L_p, \mathcal{B}_\tau^{\alpha k}(\mathcal{T})\big)_{\frac{\gamma}{\alpha},q}$$

with equivalent norms.

Similarly as in the previous section, we set

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

where the infimum is taken over all WLR-triangulations $\mathcal{T}$ with some fixed parameters r and ρ. The following result is immediate from Theorem 3.10.

Theorem 3.13. *Suppose $\inf_{\mathcal{T}}\|f\|_{\mathcal{B}_\tau^{\alpha k}(\mathcal{T})} < \infty$, where the infimum is taken over all WLR-triangulations with some fixed parameters r and ρ, and let $f \in L_p(\mathbb{R}^2)$. Then*

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

It is an *open problem* to characterize the rates of approximation generated by $\{\sigma_n(f)_p\}$.

"Algorithm" for nonlinear n-term piecewise polynomial approximation. We assume that $f \in L_p(\mathbb{R}^2)$, $0 < p < \infty$, and $\mathcal{T}$ is an arbitrary WLR-triangulation of $\mathbb{R}^2$. The proofs of Theorems 3.3 and 3.10 suggest the following approximation scheme:

Step 1. We use the local polynomial approximation to obtain the following decomposition of f:

$$f = \sum_{m\in\mathbb{Z}} p_{m,\eta}(f) = \sum_{\triangle\in\mathcal{T}} p_{\triangle,\eta}(f),$$

where $p_{\triangle,\eta}(f) = \mathbb{1}_\triangle \cdot p_{m,\eta}(f)$ if $\triangle \in \mathcal{T}_m$, and $\eta < p$ (see Theorem 2.17).

Step 2. We order the terms $\{p_{\triangle,\eta}(f)\}_{\triangle\in\mathcal{T}}$ in a sequence $\{p_{\triangle_j,\eta}(f)\}_{j=1}^\infty$ such that

$$\|p_{\triangle_1,\eta}(f)\|_p \geqslant \|p_{\triangle_2,\eta}(f)\|_p \geqslant \cdots.$$

Then we define the n-term approximant by

$$A_n(f)_p = A_n(f,\mathcal{T})_p := \sum_{j=1}^n p_{\triangle_j,\eta}(f).$$

By Theorem 3.10 and its proof, it follows that, for $f \in \mathcal{B}_\tau^{\alpha k}(\mathcal{T})$,

$$\|f - A_n(f)_p\|_p \leqslant cn^{-\alpha} \|f\|_{\mathcal{B}_\tau^{\alpha k}(\mathcal{T})}.$$

Haar bases generated by general triangulations. An important point in this article is that we carry out here nonlinear n-term approximation without using bases. In the exceptional case of nonlinear approximation from piecewise constants, however, Haar bases can be constructed and utilized for nonlinear n-term approximation in L_p, $1 < p < \infty$. To make it simple, suppose that $\mathcal{T}$ is a weak locally regular triangulation of $\mathbb{R}^2$ which is obtained by the standard refinement scheme described in Section 2.1: Every triangle $\triangle \in \mathcal{T}$ has four children obtained by choosing a point on each edge of $\triangle$ and joining these points by line segments. Denote by $\triangle_1, \dots, \triangle_4$ the children of $\triangle$ so that $\triangle_4$ is the triangle in the middle (with its vertices on the three edges of $\triangle$). We associate with $\triangle$ the following three Haar functions: $H_{\triangle,1} := |\triangle_1|^{-1}\mathbb{1}_{\triangle_1} - |\triangle \setminus \triangle_1|^{-1}\mathbb{1}_{\triangle\setminus\triangle_1}$, $H_{\triangle,2} := |\triangle_2|^{-1}\mathbb{1}_{\triangle_2} - |\triangle_3 \cup \triangle_4|^{-1}\mathbb{1}_{\triangle_3\cup\triangle_4}$, and $H_{\triangle,3} := |\triangle_3|^{-1}\mathbb{1}_{\triangle_3} - |\triangle_4|^{-1}\mathbb{1}_{\triangle_4}$. The way we order the children of $\triangle$ is not important. Clearly, $\mathbb{1}_\triangle$, $H_{\triangle,1}$, $H_{\triangle,2}$, and $H_{\triangle,3}$ form an orthogonal system which spans the set of all piecewise constants over $\{\triangle_j\}_{j=1}^4$. Then

$$\mathcal{H}_\mathcal{T} := \{H_{\triangle,1}, H_{\triangle,2}, H_{\triangle,3}\}_{\triangle\in\mathcal{T}}$$

is a Haar basis associated with $\mathcal{T}$. It is easily seen that $\mathcal{H}_\mathcal{T}$ is an orthogonal basis in $L_2(\mathbb{R}^2)$. It can be proved by a standard technique that $\mathcal{H}_\mathcal{T}$ is an unconditional basis for $L_p(\mathbb{R}^2)$, $1 < p < \infty$, and that $\mathcal{H}_\mathcal{T}$ characterizes the skinny $\mathcal{B}_\tau^{\alpha,1}(\mathcal{T})$-norm, $\alpha > 0$, $1/\tau = \alpha + 1/p$. As a consequence, the nonlinear n-term L_p-approximation from $\mathcal{H}_\mathcal{T}$ can be characterized as above (compare with [12]). We skip the details of these results.

3.4. Conclusions and open problems

We bring forward again the fundamental question of how to measure the smoothness of the functions. There is a close connection between sparsity of representation and smoothness of functions that we also wish to discuss here. As we mentioned in Section 1, we believe that in highly nonlinear approximation as well as in some other nonlinear problems the smoothness of the functions should not be measured using a single space scale (like Besov spaces) but by a family (library) of suitable space scales. To explain this concept more precisely we return to n-term Courant element approximation considered in Section 3.2. For this type of approximation, a function f should naturally be considered of smoothness order $\alpha > 0$ if $\inf_{\mathcal{T}} \|f\|_{B^\alpha_\tau(\mathcal{T})} < \infty$, which means that there exists an LR-triangulation $\mathcal{T}_f$ such that $\|f\|_{B^\alpha_\tau(\mathcal{T}_f)} < \infty$. Then the rate of the n-term L_p-approximation of f from the library $\{\Phi_{\mathcal{T}}\}$ is at least $\mathrm{O}(n^{-\alpha})$. It is an *open problem* to develop effective procedures that: (a) determine (or estimate) the maximal smoothness α of a given function f and (b) for a given function f, find an LR-triangulation $\mathcal{T}_f$ such that $\|f\|_{B^\alpha_\tau(\mathcal{T}_f)} \approx \inf_{\mathcal{T}} \|f\|_{B^\alpha_\tau(\mathcal{T})}$. Another related *open problem* is to determine whether for each function $f \in L_p$ there exists a single LR-triangulation $\mathcal{T}_f$ such that the n-term L_p-approximation of f from the library $\{\Phi_{\mathcal{T}}\}$ can be characterized using the B-spaces $B^\alpha_\tau(\mathcal{T}_f)$.

An important issue for discussion is the smoothness of the approximating tool $\Phi_{\mathcal{T}} := \{\varphi_\theta\}_{\theta\in\Theta(\mathcal{T})}$. Clearly, in nonlinear approximation, there is no saturation, which means that the corresponding approximation spaces A^γ_q are nontrivial for all $0 < \gamma < \infty$. Therefore, the smoothness spaces to be used should naturally be designed so that the basis functions $\{\varphi_\theta\}$ are infinitely smooth. This was one of the guiding principles to us in constructing the B-spaces. For instance, the Courant elements $\{\varphi_\theta\}_{\theta\in\Theta(\mathcal{T})}$ are infinitely smooth with respect to the $B^\alpha_\tau(\mathcal{T})$ space scale, namely, $\|\varphi_\theta\|_{B^\alpha_\tau(\mathcal{T})} \leqslant c\|\varphi_\theta\|_p$ for $0 < \alpha < \infty$ (see Section 2.3). This makes it possible that our direct, inverse, and characterization theorems impose no restrictions on the rate of approximation $0 < \alpha < \infty$ (see Sections 3.2 and 3.3). Also, this explains the complete success of Besov spaces in the univariate nonlinear piecewise polynomial (spline) approximation in L_p ($p < \infty$). The important fact is that, any univariate piecewise polynomial (with finitely many pieces) is infinitely smooth with respect to the corresponding Besov spaces. More precisely, for univariate discontinuous piecewise polynomials, the Bernstein inequality holds without any restriction on the smoothness parameter α ($0 < \alpha < \infty$) if $p < \infty$ (see Theorem 2.2 from [11]). In dimensions $d > 1$, however, the situation is totally different. Even for nonlinear approximation from regular piecewise polynomials (piecewise polynomials generated by regular triangulations, in our terms), the Besov spaces are not exactly the right smoothness spaces. Namely, the Besov spaces coincide with the right smoothness spaces only for some range of the smoothness parameter α. For instance, for nonlinear n-term L_p-approximation from Courant elements generated by a regular triangulation of $\mathbb{R}^2$, the Besov spaces $B^{2\alpha,2}_\tau(L_\tau)$, $1/\tau := \alpha + 1/p$, $0 < p < \infty$, are the right spaces only for $0 < \alpha < 1 + 1/p$. In the case of discontinuous piecewise polynomial approximation, the range is $0 < \alpha < 1/p$ (see Section 2.5). For the same reason, the fat B-spaces (Section 2.5) are not exactly the right spaces for characterization of n-term Courant element approximation over general triangulations.

In nonlinear n-term approximation, it is natural to work with bases. Except for the simplest case of n-term piecewise constant approximation (see the end of Section 3.3), we are not aware of good (unconditional) bases for $L_p(\mathbb{R}^2)$ ($1 < p < \infty$) and the B-spaces over general triangulations. However, as was shown in the previous sections there are equally powerful means to tackle the problems. Namely, using simple projectors into subspaces of piecewise polynomials, one can get sufficiently sparse representations of the functions, which allow to capture the rates of the best nonlinear n-term spline approximation. It is an *open problem* to construct good bases consisting of continuous or smooth compactly supported piecewise polynomials (or other functions) over general triangulations.

Methods and algorithms for piecewise polynomial approximation are in demand. This was one of the primary motivations for this work.

Appendix A

A.1. Proof of the Bernstein estimates

In this appendix, we prove Theorems 3.7 and 3.11. We recall our assumptions: $0 < p < \infty$, $\alpha > 0$, and $\tau := (\alpha + 1/p)^{-1}$.

Tree structure in $\mathcal{T}$ generated by $\Lambda \subset \mathcal{T}$. Suppose $\mathcal{T}$ is a multilevel triangulation (WLR or better), and let $\Lambda \subset \mathcal{T}$ and $\#\Lambda < \infty$. The set Λ induces a tree structure in $\mathcal{T}$ that we want to bring forward here and utilize in the proof later on. We shall use the parent–child relation in $\mathcal{T}$ induced by the inclusion relation: Each triangle $\triangle \in \mathcal{T}_m$ has (contains) $\leqslant M_0$ children in $\mathcal{T}_{m+1}$ and has a single parent in $\mathcal{T}_{m-1}$.

Let Γ_0 be the set of all $\triangle \in \mathcal{T}$ such that $\triangle \supset \triangle'$ for some $\triangle' \in \Lambda$. We denote by Γ_{b} the set of all *branching triangles* in Γ_0 (triangles with more than one child in Γ_0) and by Γ_{b}' the set of all *children in $\mathcal{T}$ of branching triangles* (each of them may or may not belong to Γ_0). Now, we extend Γ_0 to $\Gamma := \Gamma_0 \cup \Gamma_{\mathrm{b}}'$. We also extend Λ to $\tilde{\Lambda} := \Lambda \cup \Gamma_{\mathrm{b}} \cup \Gamma_{\mathrm{b}}'$. In addition, we introduce the following subsets of Γ: Γ_{f} the set of all *final triangles* in Γ (triangles in Γ containing no other triangles in Γ) and $\Gamma_{\mathrm{ch}} := \Gamma \setminus \tilde{\Lambda}$ the set of all *chain triangles*. Note that each triangle $\triangle \in \Gamma_{\mathrm{ch}}$ has exactly one child in Γ. Since the final triangles in Γ_0 belong to Λ, then $\#\Gamma_{\mathrm{b}} \leqslant \#\Lambda$ and hence $\#\Gamma_{\mathrm{b}}' \leqslant M_0 \#\Gamma_{\mathrm{b}} \leqslant c\#\Lambda$, $\#\Gamma_{\mathrm{f}} \leqslant \#\Lambda + \#\Gamma_{\mathrm{b}}' \leqslant c\#\Lambda$, and $\#\tilde{\Lambda} \leqslant \#\Lambda + \#\Gamma_{\mathrm{b}} + \#\Gamma_{\mathrm{b}}' \leqslant c\#\Lambda$. Note that $\#\Gamma_{\mathrm{ch}}$ can be uncontrolably larger than $\#\Lambda$.

We next introduce chains in Γ_{ch}. By definition $\lambda = \{\triangle_1, \dots, \triangle_\ell\} \subset \Gamma_{\mathrm{ch}}$ ($\ell \geqslant 1$) is a *finite chain* in Γ_{ch} if $\triangle_\lambda'' \supset \triangle_1 \supset \cdots \supset \triangle_\ell \supset \triangle_\lambda'$ for some $\triangle_\lambda', \triangle_\lambda'' \in \tilde{\Lambda}$, $\triangle_1$ is a child of $\triangle_\lambda''$, $\triangle_j$ is a child of $\triangle_{j-1}$, $j = 2, \dots, \ell$, and $\triangle_\lambda'$ is a child of $\triangle_\ell$. Notice that $\triangle_\lambda'' \notin \Gamma_{\mathrm{b}}$ and hence $\triangle_1$ is the only child of $\triangle_\lambda''$ in Γ. We let $\mathcal{L}$ denote the set of all finite chains in Γ_{ch}. Also, by definition $\lambda = \{\dots, \triangle_{-2}, \triangle_{-1}\} \subset \Gamma_{\mathrm{ch}}$ is an *infinite chain* in Γ_{ch} if we have $\cdots \supset \triangle_{-2} \supset \triangle_{-1} \supset \triangle_\lambda'$ for some $\triangle_\lambda' \in \tilde{\Lambda}$, $\triangle_j$ is a child of $\triangle_{j-1}$, $j = -1, -2, \dots$, and $\triangle_\lambda'$ is a child of $\triangle_{-1}$. We let $\mathcal{L}^\infty$ denote the set of all infinite chains in Γ_{ch}. Clearly, $\mathcal{L} \cup \mathcal{L}^\infty$ consists of disjoint chains of triangles, $\Gamma_{\mathrm{ch}} = \bigcup_{\lambda \in \mathcal{L} \cup \mathcal{L}^\infty} \lambda$, and $\#(\mathcal{L} \cup \mathcal{L}^\infty) \leqslant \#\tilde{\Lambda}$.

Finally, we use the above sets to introduce *rings* generated by $\tilde{\Lambda}$. First, for each $\triangle \in \Gamma \setminus (\Gamma_{\mathrm{b}} \cup \Gamma_{\mathrm{f}})$, we denote by $\tilde{\triangle}$ ($\tilde{\triangle} \neq \triangle$) the unique largest triangle from $\tilde{\Lambda}$ contained in $\triangle$. We associate with each $\triangle \in \Gamma \setminus (\Gamma_{\mathrm{b}} \cup \Gamma_{\mathrm{f}})$ a ring $K_\triangle$ defined by $K_\triangle := \triangle \setminus \tilde{\triangle}$. Also, we define $K_\triangle := \triangle$ if $\triangle \in \Gamma_{\mathrm{f}}$ and $K_\triangle := \emptyset$ if $\triangle \in \Gamma_{\mathrm{b}} \cup (\mathcal{T} \setminus \Gamma)$. Notice that if $\triangle \in \lambda$ for some $\lambda \in \mathcal{L} \cup \mathcal{L}^\infty$, then $\tilde{\triangle} = \triangle_\lambda'$. It is readily seen that $K_{\triangle'}^\circ \cap K_{\triangle''}^\circ = \emptyset$ if $\triangle', \triangle'' \in \tilde{\Lambda}$ and $\triangle' \neq \triangle''$,

$$\triangle = \bigcup_{\triangle' \in \tilde{\Lambda},\, \triangle' \subset \triangle} K_{\triangle'} \quad \text{for } \triangle \in \tilde{\Lambda}, \tag{A.1}$$

and hence

$$\bigcup_{\triangle \in \tilde{\Lambda}} \triangle = \bigcup_{\triangle' \in \tilde{\Lambda}} K_{\triangle'}. \tag{A.2}$$

For the proof of both theorems, we need the following lemma.

Lemma A.1. *Suppose* $S = \sum_{\triangle\in\Lambda} \mathbb{1}_\triangle \cdot P_\triangle$, *where* $P_\triangle \in \Pi_k$ $(k \geqslant 1)$, $\Lambda \subset \mathcal{T}$ *with* $\mathcal{T}$ *a WLR-triangulation, and* $\#\Lambda < \infty$. *Then*

$$\left(\sum_{\triangle\in\Lambda} |\triangle|^{-\alpha\tau} \|S\|_{L_\tau(\triangle)}^\tau\right)^{1/\tau} \leqslant c(\#\Lambda)^\alpha \|S\|_p$$

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

Proof. We adopt all necessary notation from "Tree structure in $\mathcal{T}$ generated by $\Lambda \subset \mathcal{T}$" developed above with $\mathcal{T}$ and Λ from the hypotheses of the lemma. We may assume that

$$S = \sum_{\triangle\in\tilde{\Lambda}} \mathbb{1}_\triangle \cdot P_\triangle.$$

It is an important observation that S is a polynomial of degree $< k$ on each ring $K_\triangle = \triangle \setminus \tilde{\triangle}$. Hence, using Lemma 2.7,

$$\|S\|_{L_\tau(K_\triangle)} \approx |K_\triangle|^{1/\tau - 1/p} \|S\|_{L_p(K_\triangle)} \approx |\triangle|^\alpha \|S\|_{L_p(K_\triangle)}. \tag{A.3}$$

We shall also need the obvious estimate (see (2.1)):

$$\sum_{\triangle\in\Gamma,\, \triangle\supset\triangle'} \left(\frac{|\triangle'|}{|\triangle|}\right)^\gamma \leqslant c(\rho,\gamma) < \infty, \quad \gamma > 0. \tag{A.4}$$

We use (A.1)–(A.4) to obtain

$$\begin{aligned}
\sum_{\triangle\in\tilde{\Lambda}} |\triangle|^{-\alpha\tau} \|S\|_{L_\tau(\triangle)}^\tau &= \sum_{\triangle\in\tilde{\Lambda}} |\triangle|^{-\alpha\tau} \sum_{\triangle'\in\tilde{\Lambda},\, \triangle'\subset\triangle} \|S\|_{L_\tau(K_{\triangle'})}^\tau \\
&= \sum_{\triangle'\in\tilde{\Lambda}} \|S\|_{L_\tau(K_{\triangle'})}^\tau \sum_{\triangle\in\tilde{\Lambda},\, \triangle\supset\triangle'} |\triangle|^{-\alpha\tau} \\
&\leqslant \sum_{\triangle'\in\tilde{\Lambda}} \|S\|_{L_\tau(K_{\triangle'})}^\tau |\triangle'|^{-\alpha\tau} \sum_{\triangle\in\Gamma,\, \triangle\supset\triangle'} \left(\frac{|\triangle'|}{|\triangle|}\right)^{\alpha\tau} \\
&\leqslant c \sum_{\triangle'\in\tilde{\Lambda}} \|S\|_{L_p(K_{\triangle'})}^\tau \leqslant c\left(\sum_{\triangle'\in\tilde{\Lambda}} \|S\|_{L_p(K_{\triangle'})}^p\right)^{\tau/p} (\#\tilde{\Lambda})^{1-\tau/p} \leqslant c(\#\Lambda)^{\alpha\tau} \|S\|_p^\tau,
\end{aligned}$$

where we once switched the order of summation and applied Hölder's inequality. □

Proof of Theorem 3.7. Let $S \in \widetilde{\Sigma}_n(\mathcal{T})$ with $\mathcal{T}$ an LR-triangulation and suppose that $S =: \sum_{\theta\in\mathcal{M}} c_\theta\varphi_\theta$, where $\mathcal{M} \subset \Theta(\mathcal{T})$ and $\#\mathcal{M} \leqslant n$. Let Λ be the set of all triangles $\triangle \in \mathcal{T}$ which are involved in all $\theta \in \mathcal{M}$. Then $S = \sum_{\triangle\in\Lambda} S_\triangle$, where $S_\triangle =: \mathbb{1}_\triangle \cdot P_\triangle$, $P_\triangle \in \Pi_2$. Evidently, $\#\Lambda \leqslant N_0\#\mathcal{M} \leqslant cn$. For the rest of the proof, we adopt all the notation from "Tree structure in $\mathcal{T}$ generated by $\Lambda \subset \mathcal{T}$", given in the beginning of this appendix, with $\mathcal{T}$ and Λ from the above. In addition, we denote

$$\mathcal{X}_m^* := \{\triangle \in \mathcal{T}_m\colon\ \triangle \subset \Omega_{\triangle'} \text{ for some } \triangle' \in \tilde{\Lambda} \cap \mathcal{T}_m\},$$
$$\mathcal{X}_m^{**} := \{\triangle \in \mathcal{T}_m\colon\ \triangle \subset \Omega^2_{\triangle'} \text{ for some } \triangle' \in \tilde{\Lambda} \cap \mathcal{T}_m\}, \quad \text{where for } \triangle \in \mathcal{T}_m,$$
$$\Omega_\triangle := \bigcup \{\triangle' \in \mathcal{T}_m\colon\ \triangle' \cap \triangle \neq \emptyset\} \quad \text{and} \quad \Omega^2_\triangle := \bigcup \{\triangle' \in \mathcal{T}_m\colon\ \triangle' \cap \Omega_\triangle \neq \emptyset\}.$$

Also, we denote $\mathcal{X}^* := \bigcup_{m\in\mathbb{Z}} \mathcal{X}_m^*$ and $\mathcal{X}^{**} := \bigcup_{m\in\mathbb{Z}} \mathcal{X}_m^{**}$. Evidently, we have $\#\mathcal{X}^* \leqslant 3N_0\#\tilde{\Lambda} \leqslant cn$ and $\#\mathcal{X}^{**} \leqslant 3N_0^2\#\tilde{\Lambda} \leqslant cn$.

For $m \in \mathbb{Z}$, we denote $S_m := \sum_{\theta\in\mathcal{M},\, \mathrm{level}(\theta)\leqslant m} c_\theta\varphi_\theta$. Clearly, $S_m \in \widetilde{\mathcal{S}}_m$ and, therefore, for $\triangle \in \mathcal{T}_m$,

$$\mathbb{S}_\triangle(S)_\tau = \mathbb{S}_\triangle(S - S_m)_\tau \leqslant \|S - S_m\|_{L_\tau(\Omega_\triangle)}. \tag{A.5}$$

We shall also use the obvious inequality $\mathbb{S}_\triangle(S)_\tau \leqslant \|S\|_{L_\tau(\Omega_\triangle)}$.

Next, we estimate $\|S\|^\tau_{B^\alpha_\tau(\mathcal{T})} := \sum_{\triangle\in\mathcal{T}} |\triangle|^{-\alpha\tau}\mathbb{S}_\triangle(S)^\tau_\tau$ by splitting up $\mathcal{T}$ into two subsets, namely, $\mathcal{X}^*$ and $\mathcal{T} \setminus \mathcal{X}^*$.

(i) If $\triangle \in \mathcal{X}_m^*$, then $\triangle \subset \Omega_{\triangle'}$ for some $\triangle' \in \tilde{\Lambda} \cap \mathcal{T}_m$ and hence $\Omega_\triangle \subset \Omega^2_{\triangle'}$. From this, we find

$$\mathbb{S}_\triangle(S)^\tau_\tau \leqslant \|S\|^\tau_{L_\tau(\Omega_\triangle)} = \sum_{\triangle^\star\in\mathcal{T}_m,\, \triangle^\star\subset\Omega_\triangle} \|S\|^\tau_{L_\tau(\triangle^\star)} \leqslant \sum_{\triangle^\star\in\mathcal{T}_m,\, \triangle^\star\subset\Omega^2_{\triangle'}} \|S\|^\tau_{L_\tau(\triangle^\star)}$$

and hence, using (2.2),

$$|\triangle|^{-\alpha\tau}\mathbb{S}_\triangle(S)^\tau_\tau \leqslant c \sum_{\triangle^\star\in\mathcal{T}_m,\, \triangle^\star\subset\Omega^2_{\triangle'}} |\triangle^\star|^{-\alpha\tau}\|S\|^\tau_{L_\tau(\triangle^\star)}.$$

Therefore,

$$\sum_{\triangle\in\mathcal{X}_m^*} |\triangle|^{-\alpha\tau}\mathbb{S}_\triangle(S)^\tau_\tau \leqslant c \sum_{\triangle\in\mathcal{X}_m^{**}} |\triangle|^{-\alpha\tau}\|S\|^\tau_{L_\tau(\triangle)}$$

and, summing over $m \in \mathbb{Z}$, we find

$$\sum_{\triangle\in\mathcal{X}^*} |\triangle|^{-\alpha\tau}\mathbb{S}_\triangle(S)^\tau_\tau \leqslant c \sum_{\triangle\in\mathcal{X}^{**}} |\triangle|^{-\alpha\tau}\|S\|^\tau_{L_\tau(\triangle)} \leqslant c(\#\mathcal{X}^{**})^{\alpha\tau}\|S\|^\tau_p \leqslant cn^{\alpha\tau}\|S\|^\tau_p, \tag{A.6}$$

where we applied Lemma A.1 to S with Λ replaced by $\mathcal{X}^{**}$ which is legitimate since $\mathcal{X}^{**} \supset \Lambda$ and hence S has the required representation.

(ii) Let $\triangle \in \mathcal{T}_m \setminus \mathcal{X}_m^*$. Then $\Omega_\triangle =: \bigcup_{j=1}^{n_\triangle} \triangle_j$ for some $\triangle_j \in (\Gamma_{\mathrm{ch}} \cap \mathcal{T}_m) \cup (\mathcal{T}_m \setminus \Gamma)$, $j = 1, \dots, n_\triangle$, with $n_\triangle \leqslant 3N_0$. We have, using (A.5),

$$\mathbb{S}_\triangle(S)^\tau_\tau = \mathbb{S}_\triangle(S - S_m)^\tau_\tau \leqslant \sum_{j=1}^{n_\triangle} \|S - S_m\|^\tau_{L_\tau(\triangle_j)}. \tag{A.7}$$

Note that if $\triangle_j \in \mathcal{T}_m \setminus \Gamma$, then $S|_{\triangle_j} = S_m|_{\triangle_j}$ and hence $\|S - S_m\|_{L_\tau(\triangle_j)} = 0$.

Suppose $\triangle_j \in \Gamma_{\mathrm{ch}} \cap \mathcal{T}_m$. It is an important observation that, in this case, $S|_{K_{\triangle_j}} = S_m|_{K_{\triangle_j}} = \mathbb{1}_{K_{\triangle_j}} \cdot P_{\triangle_j}$ and $S_m|_{\triangle_j} = \mathbb{1}_{\triangle_j} \cdot P_{\triangle_j}$, for some $P_{\triangle_j} \in \Pi_2$, where $K_{\triangle_j} := \triangle_j \setminus \tilde{\triangle}_j$ ($\tilde{\triangle}_j \in \tilde{\Lambda}$) is the ring associated with $\triangle_j$. Using this, we find

$$\begin{aligned}\|S - S_m\|^\tau_{L_\tau(\triangle_j)} &= \|S - S_m\|^\tau_{L_\tau(\tilde{\triangle}_j)} \leqslant c\|S\|^\tau_{L_\tau(\tilde{\triangle}_j)} + c\|P_{\triangle_j}\|^\tau_{L_\tau(\tilde{\triangle}_j)}\\ &\leqslant c\|S\|^\tau_{L_\tau(\tilde{\triangle}_j)} + c|\tilde{\triangle}_j||\triangle_j|^{\alpha\tau-1}\|S\|^\tau_{L_p(K_{\triangle_j})}.\end{aligned} \tag{A.8}$$

For the last inequality in (A.8) we used that

$$\begin{aligned}\|P_{\triangle_j}\|^\tau_{L_\tau(\tilde\triangle_j)} &\leqslant |\tilde\triangle_j|\|P_{\triangle_j}\|^\tau_{L_\infty(\triangle_j)} \leqslant c|\tilde\triangle_j|\|P_{\triangle_j}\|^\tau_{L_\infty(K_{\triangle_j})}\\ &\leqslant c|\tilde\triangle_j||\triangle_j|^{-\tau/p}\|P_{\triangle_j}\|^\tau_{L_p(K_{\triangle_j})} \leqslant c|\tilde\triangle_j||\triangle_j|^{\alpha\tau-1}\|S\|^\tau_{L_p(K_{\triangle_j})},\end{aligned}\tag{A.9}$$

where we applied Lemma 2.7 and used that $S|_{K_{\triangle_j}} = P_{\triangle_j}|_{K_{\triangle_j}}$. From (A.7) and (A.8), we infer

$$\begin{aligned}\sum_{\triangle\in\mathcal{T}\setminus\mathcal{X}^*}|\triangle|^{-\alpha\tau}\mathbb{S}_\triangle(S)^\tau_\tau &\leqslant c\sum_{m\in\mathbb{Z}}\sum_{\triangle\in\Gamma_{\text{ch}}\cap\mathcal{T}_m}\left(|\triangle|^{-\alpha\tau}\|S\|^\tau_{L_\tau(\tilde\triangle)} + \frac{|\tilde\triangle|}{|\triangle|}\|S\|^\tau_{L_p(K_\triangle)}\right)\\ &\leqslant c\sum_{\triangle\in\Gamma_{\text{ch}}}|\triangle|^{-\alpha\tau}\|S\|^\tau_{L_\tau(\tilde\triangle)} + c\sum_{\triangle\in\Gamma_{\text{ch}}}\frac{|\tilde\triangle|}{|\triangle|}\|S\|^\tau_{L_p(K_\triangle)}\\ &=: \Sigma_1 + \Sigma_2.\end{aligned}$$

Switching the order of summation and applying (A.4), we obtain

$$\begin{aligned}\Sigma_1 &= c\sum_{\triangle'\in\tilde\Lambda}\|S\|^\tau_{L_\tau(\triangle')}\sum_{\triangle\in\Gamma_{\text{ch}},\,\triangle\supset\triangle'}|\triangle|^{-\alpha\tau}\\ &\leqslant c\sum_{\triangle'\in\tilde\Lambda}\|S\|^\tau_{L_\tau(\triangle')}|\triangle'|^{-\alpha\tau}\sum_{\triangle\in\Gamma,\,\triangle\supset\triangle'}\left(\frac{|\triangle'|}{|\triangle|}\right)^{\alpha\tau}\\ &\leqslant c\sum_{\triangle'\in\tilde\Lambda}|\triangle'|^{-\alpha\tau}\|S\|^\tau_{L_\tau(\triangle')} \leqslant c(\#\tilde\Lambda)^{\alpha\tau}\|S\|^\tau_p,\end{aligned}\tag{A.10}$$

where for the latter estimate we applied Lemma A.1 to S with $\tilde\Lambda$ in place of Λ.

To estimate Σ_2, we shall use the representation of Γ_{ch} as a disjoint union of chains: $\Gamma_{\text{ch}} = \bigcup_{\lambda\in\mathcal{L}\cup\mathcal{L}^\infty}\lambda$. Let $\lambda\in\mathcal{L}$ and suppose $\lambda = \{\triangle_1,\dots,\triangle_\ell\}$, where $\triangle''_\lambda\supset\triangle_1\supset\cdots\supset\triangle_\ell\supset\triangle'_\lambda$ with $\triangle'_\lambda, \triangle''_\lambda\in\tilde\Lambda$ ($\triangle''_\lambda\notin\Gamma_{\text{b}}$). Then

$$\begin{aligned}\sum_{\triangle\in\lambda}|\tilde\triangle||\triangle|^{-1}\|S\|^\tau_{L_p(K_\triangle)} &\leqslant \|S\|^\tau_{L_p(\triangle''_\lambda\setminus\triangle'_\lambda)}\sum_{j=1}^{\ell}|\triangle'_\lambda||\triangle_j|^{-1}\\ &\leqslant \|S\|^\tau_{L_p(K_{\triangle''_\lambda})}\sum_{j=1}^{\ell}\rho^{\ell-j+1} \leqslant c\|S\|^\tau_{L_p(K_{\triangle''_\lambda})}.\end{aligned}$$

If $\lambda\in\mathcal{L}^\infty$ and $\triangle\in\lambda$, then $S|_{K_\triangle} = 0$ and hence $\|S\|_{L_p(K_\triangle)} = 0$.

Summing the above inequalities over all $\lambda\in\mathcal{L}$, we obtain

$$\Sigma_2 \leqslant c\sum_{\triangle^\star\in\tilde\Lambda}\|S\|^\tau_{L_p(K_{\triangle^\star})} \leqslant c\left(\sum_{\triangle^\star\in\tilde\Lambda}\|S\|^p_{L_p(K_{\triangle^\star})}\right)^{\tau/p}(\#\tilde\Lambda)^{1-\tau/p} \leqslant c(\#\tilde\Lambda)^{\alpha\tau}\|S\|^\tau_p,\tag{A.11}$$

where we used Hölder's inequality and (A.2). Estimates (A.10) and (A.11) yield

$$\sum_{\triangle\in\mathcal{T}\setminus\mathcal{X}^*}|\triangle|^{-\alpha\tau}\mathbb{S}_\triangle(S)^\tau_\tau \leqslant c(\#\tilde\Lambda)^{\alpha\tau}\|S\|^\tau_p \leqslant cn^{\alpha\tau}\|S\|^\tau_p.$$

This and (A.6) imply $\|S\|^\tau_{B^\alpha_\tau(\mathcal{T})} \leqslant cn^{\alpha\tau}\|S\|^\tau_p$. $\square$

Proof of Theorem 3.11. Let $\mathcal{T}$ be a WLR-triangulation and $S \in \Sigma^k_n(\mathcal{T})$. Then S can be written in the form $S = \sum_{\triangle\in\Lambda} \mathbb{1}_\triangle \cdot P_\triangle$, where $P_\triangle \in \Pi_k$, $\Lambda \subset \mathcal{T}$, and $\#\Lambda \leqslant n$. As in the previous proof, we adopt all the notation from "Tree structure in $\mathcal{T}$ generated by $\Lambda \subset \mathcal{T}$" with $\mathcal{T}$ and Λ from the above.

To estimate $\|S\|^\tau_{\mathcal{B}^\alpha_\tau(\mathcal{T})} := \sum_{\triangle\in\mathcal{T}} |\triangle|^{-\alpha\tau}\omega_k(S,\triangle)^\tau_\tau$, we shall split $\mathcal{T}$ into three subsets:

(i) If $\triangle \in \mathcal{T}\setminus\Gamma$, then S is a polynomial of degree $< k$ on $\triangle$ and hence $\omega_k(S,\triangle)_\tau = 0$.

(ii) If $\triangle \in \tilde\Lambda$, then evidently $\omega_k(S,\triangle)_\tau \leqslant c\|S\|_{L_\tau(\triangle)}$ and hence

$$\sum_{\triangle\in\tilde\Lambda} |\triangle|^{-\alpha\tau}\omega_k(S,\triangle)^\tau_\tau \leqslant c\sum_{\triangle\in\tilde\Lambda} |\triangle|^{-\alpha\tau}\|S\|^\tau_{L_\tau(\triangle)} \leqslant c(\#\tilde\Lambda)^{\alpha\tau}\|S\|^\tau_p, \tag{A.12}$$

where for the last inequality we used Lemma A.1 (with Λ replaced by $\tilde\Lambda$).

(iii) Let $\triangle \in \Gamma_{\mathrm{ch}}$ (recall that $\Gamma_{\mathrm{ch}} := \Gamma\setminus\tilde\Lambda$). Clearly, $S|_{K_\triangle} = \mathbb{1}_{K_\triangle}\cdot P_\triangle$ for some $P_\triangle \in \Pi_k$, where $K_\triangle := \triangle\setminus\tilde\triangle$ is the ring associated with $\triangle$. Therefore,

$$\begin{aligned}\omega_k(S,\triangle)^\tau_\tau = \omega_k(S-P_\triangle,\triangle)^\tau_\tau &\leqslant c\|S\|^\tau_{L_\tau(\tilde\triangle)} + c\|P_\triangle\|^\tau_{L_\tau(\tilde\triangle)}\\ &\leqslant c\|S\|^\tau_{L_\tau(\tilde\triangle)} + c|\tilde\triangle||\triangle|^{\alpha\tau-1}\|S\|^\tau_{L_p(K_\triangle)},\end{aligned} \tag{A.13}$$

where we used that $\|P_\triangle\|^\tau_{L_\tau(\tilde\triangle)} \leqslant c|\tilde\triangle||\triangle|^{\alpha\tau-1}\|P_\triangle\|^\tau_{L_p(K_\triangle)}$ which follows by Lemma 2.7 exactly as in (A.9). From (A.13), we infer

$$\begin{aligned}\sum_{\triangle\in\Gamma_{\mathrm{ch}}} |\triangle|^{-\alpha\tau}\omega_k(S,\triangle)^\tau_\tau &\leqslant c\sum_{\triangle\in\Gamma_{\mathrm{ch}}} |\triangle|^{-\alpha\tau}\|S\|^\tau_{L_\tau(\tilde\triangle)} + c\sum_{\triangle\in\Gamma_{\mathrm{ch}}} |\tilde\triangle||\triangle|^{-1}\|S\|^\tau_{L_p(K_\triangle)}\\ &=: \Sigma^\star_1 + \Sigma^\star_2.\end{aligned}$$

We estimate $\Sigma^\star_1$ and $\Sigma^\star_2$ exactly as the sums Σ_1 and Σ_2 were estimated in (A.10) and (A.11), respectively. We obtain

$$\sum_{\triangle\in\Gamma_{\mathrm{ch}}} |\triangle|^{-\alpha\tau}\omega_k(S,\triangle)^\tau_\tau \leqslant c(\#\tilde\Lambda)^{\alpha\tau}\|S\|^\tau_p \leqslant cn^{\alpha\tau}\|S\|^\tau_p.$$

Combining this estimate with (A.12), we find $\|S\|_{\mathcal{B}^\alpha_\tau(\mathcal{T})} \leqslant cn^{\alpha\tau}\|S\|_p$ and the proof of Theorem 3.11 is complete. $\square$

Appendix B

Proof of Lemma 2.6 (*Whitney*). Suppose $\mathcal{P} \subset \mathbb{R}^2$ is a parallelogram and $f \in L_q(\mathcal{P})$. Evidently, there exists an affine transform $\mathbf{A}$ which maps $\mathcal{P}$ one-to-one onto $[0,1]^2$. Whitney's estimate

$$E_k(f,\mathcal{P})_q \leqslant c\,\omega_k(f,\mathcal{P})_q \tag{B.1}$$

is invariant under affine transforms and, hence, follows from the case $\mathcal{P} := [0,1]^2$. For the proof of Whitney's inequality on $[0,1]^2$, we refer the reader to [1] (for the case of $q \geqslant 1$) and [14] (for the case of $0 < q < 1$).

Now, having (B.1), we can prove Whitney's estimate for a triangle as well. Fix an arbitrary triangle $\triangle = [x_1, x_2, x_3]$. Let $y_1 := (x_2 + x_3)/2$, $y_2 := (x_1 + x_3)/2$, and $y_3 := (x_1 + x_2)/2$ be the midpoints of its edges, and let $\triangle' := [y_1, y_2, y_3]$. Consider now the three parallelograms $\mathcal{P}_1 := [x_1, y_3, y_1, y_2]$, $\mathcal{P}_2 := [x_2, y_1, y_2, y_3]$, and $\mathcal{P}_3 := [x_3, y_2, y_3, y_1]$. Clearly, $\triangle = \bigcup_{j=1}^{3} \mathcal{P}_j$ and $\triangle' = \bigcap_{j=1}^{3} \mathcal{P}_j$. We select polynomials $P_{\triangle'}, P_1, P_2, P_3 \in \Pi_k$ such that $\|f - P_{\triangle'}\|_{L_q(\triangle')} = E_k(f, \triangle')_q$ and $\|f - P_j\|_{L_q(\mathcal{P}_j)} = E_k(f, \mathcal{P}_j)_q$ for $j = 1, 2, 3$. Evidently, since $\triangle' \subset \mathcal{P}_j$ and $|\mathcal{P}_j| = 2|\triangle'|$, using Lemma 2.7 and (B.1), we have

$$\begin{aligned}
\|P_j - P_{\triangle'}\|_{L_q(\mathcal{P}_j)} &\leqslant c\|P_j - P_{\triangle'}\|_{L_q(\triangle')} \leqslant c\|f - P_j\|_{L_q(\triangle')} + c\|f - P_{\triangle'}\|_{L_q(\triangle')} \\
&\leqslant c\|f - P_j\|_{L_q(\mathcal{P}_j)} + cE_k(f, \triangle')_q \leqslant cE_k(f, \mathcal{P}_j)_q \\
&\leqslant c\,\omega_k(f, \mathcal{P}_j)_q \leqslant c\,\omega_k(f, \triangle)_q
\end{aligned}$$

with $c = c(q, k)$. From this, we obtain

$$\begin{aligned}
E_k(f, \triangle)_q &\leqslant \|f - P_{\triangle'}\|_{L_q(\triangle)} \leqslant c\sum_{j=1}^{3} \|f - P_{\triangle'}\|_{L_q(\mathcal{P}_j)} \\
&\leqslant c\sum_{j=1}^{3} \|f - P_j\|_{L_q(\mathcal{P}_j)} + c\sum_{j=1}^{3} \|P_{\triangle'} - P_j\|_{L_q(\mathcal{P}_j)} \leqslant c\,\omega_k(f, \triangle)_q,
\end{aligned}$$

where we again used (B.1). Thus (2.11) is proved for a triangle.

To prove (2.11) in the second case one can proceed similarly, using that the estimate is invariant under affine transforms and most importantly that $\mathcal{T}$ is an SLR-triangulation (see Section 2.1). We omit the details. □

Proof of Lemma 2.12. Let $\widetilde{S} \in \widetilde{\mathcal{S}}_m$ be an element of best L_η-approximation to f on $\Omega_\triangle$ from $\widetilde{\mathcal{S}}_m$. Using Lemma 2.7(c) and Hölder's inequality, we obtain

$$\begin{aligned}
\|f - S\|_{L_\eta(\Omega_\triangle)} &\leqslant c\|f - \widetilde{S}\|_{L_\eta(\Omega_\triangle)} + c\|\widetilde{S} - S\|_{L_\eta(\Omega_\triangle)} \\
&\leqslant c\,\mathbb{S}_\triangle(f)_\eta + c|\Omega_\triangle|^{1/\eta - 1/\mu}\|\widetilde{S} - S\|_{L_\mu(\Omega_\triangle)} \\
&\leqslant c\,\mathbb{S}_\triangle(f)_\eta + c|\Omega_\triangle|^{1/\eta - 1/\mu}\big(\|f - \widetilde{S}\|_{L_\mu(\Omega_\triangle)} + \|f - S\|_{L_\mu(\Omega_\triangle)}\big) \\
&\leqslant c\,\mathbb{S}_\triangle(f)_\eta + c|\Omega_\triangle|^{1/\eta - 1/\mu}\|f - \widetilde{S}\|_{L_\mu(\Omega_\triangle)} \\
&\leqslant c\,\mathbb{S}_\triangle(f)_\eta + c\|f - \widetilde{S}\|_{L_\eta(\Omega_\triangle)} \leqslant c\,\mathbb{S}_\triangle(f)_\eta. \quad \square
\end{aligned}$$

Proof of inequality (3.11). We shall use the obvious inequality

$$a^\alpha b^{s-\alpha} \leqslant (a + b)^s, \quad \text{if } 0 < \alpha \leqslant s \text{ and } a, b > 0, \tag{B.2}$$

which is immediate from $(a/b)^\alpha \leqslant (a/b + 1)^\alpha \leqslant (a/b + 1)^s$. Now, set $\alpha := 1/\tau - 1/p$, $s := 1/\tau > \alpha$, $a := nx_n^\tau$, and $b := \sum_{j=n+1}^{\infty} x_j^\tau$. Applying inequality (B.2), we find

$$\begin{aligned}
\left(\sum_{j=n+1}^{\infty} x_j^p\right)^{1/p} &\leqslant \left(x_n^{p-\tau} \sum_{j=n+1}^{\infty} x_j^\tau\right)^{1/p} = x_n^{1-\tau/p}\left(\sum_{j=n+1}^{\infty} x_j^\tau\right)^{1/p} \\
&= n^{-\alpha} a^\alpha b^{1/\tau - \alpha} \leqslant n^{-\alpha}(a + b)^{1/\tau} \leqslant n^{-\alpha}\left(\sum_{j=1}^{\infty} x_j^\tau\right)^{1/\tau}. \quad \square
\end{aligned}$$

References

[1] Yu. Brudnyi, Approximation of functions of n-variables by quasi-polynomials, Math. USSR Izv. 4 (1970) 568–586.
[2] Yu. Brudnyi, 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.
[3] J. Bergh, J. Löfström, Interpolation Spaces. An introduction, in: Grundlehren der Mathematischen Wissenschaften, Vol. 223, Springer-Verlag, Berlin/New York, 1976.
[4] C. Bennett, R. Sharpley, Interpolation of Operators, in: Pure and Applied Mathematics, Vol. 129, Academic Press, Boston, MA, 1988.
[5] R. DeVore, Nonlinear approximation, Acta Numer. (1998) 51–150.
[6] R. DeVore, B. Jawerth, V. Popov, Compression of wavelet decompositions, Amer. J. Math. 114 (1992) 737–785.
[7] R. DeVore, G.G. Lorentz, Constructive Approximation, in: Grundlehren, Vol. 303, Springer, Heidelberg, 1993.
[8] R. DeVore, V. Popov, Interpolation of Besov spaces, Trans. Amer. Math. Soc. 305 (1998) 297–314.
[9] R. DeVore, P. Petrushev, X. Yu, Nonlinear wavelet approximation in the space $C(R^d)$, in: A.A. Gonchar, E.B. Saff (Eds.), Progress in Approximation Theory, New York, Springer-Verlag, 1992, pp. 261–283.
[10] P. Oswald, Multilevel Finite Element Approximation, Teubner, Stuttgart, 1994.
[11] P. Petrushev, Direct and converse theorems for rational and spline approximation and Besov spaces, in: Lecture Notes in Mathematics, Vol. 1302, Proceedings of the Conference “Function Spaces and Applications”, Lund 1986, 1988, pp. 363–377.
[12] P. Petrushev, Multivariate n-term rational and piecewise polynomial approximation, J. Approx. Theory 121 (2003) 158–197.
[13] P. Petrushev, V. Popov, Rational Approximation of Real Functions, Cambridge Univ. Press, 1987.
[14] E. Storozhenko, P. Oswald, Jackson’s theorem in the space $L_p(\mathbb{R}^k)$, $0 < p < 1$, Siberian Math. J. 19 (1978) 630–639.
[15] H. Triebel, Theory of Function Spaces II, in: Monographs in Mathematics, Vol. 84, Birkhäuser, Basel, 1992.