Anisotropic Spaces and Nonlinear n-term Spline Approximation

Pencho Petrushev *

Abstract. This article is a survey of some new (and old) results in nonlinear n-term spline approximation and related topics. A substantial part of the paper is devoted to the anisotropic spaces generated by sequences of nested triangulations of compact polygonal domains in $\mathbb{R}^2$ and their characterization via corresponding Franklin systems. The emphasis is placed on practical algorithms for nonlinear n-term approximation from the scaling functions of a multiresolution analysis in the uniform norm, which are capable of achieving the rates of the best n-term approximation. Results on the relation between bivariate n-term rational and spline approximation are also given.

§1. Introduction

A fundamental idea in Harmonic analysis is to represent functions or distributions as linear combinations of functions of a particularly simple nature. Familiar examples include the Fourier series representation on the circle or on the torus and atomic decompositions of Hardy spaces and more general Triebel-Lizorkin and Besov spaces. In both examples, however, there are problems with the representation. The problem with the Fourier series is that only a few spaces (related to L_2) have simple characterization in terms of the Fourier coefficients. The difficulty with the atomic decompositions of spaces is that the representing atoms vary with the function or distribution being represented. In the case of classical spaces on $\mathbb{R}^d$ or on the torus, both limitations can be avoided by using wavelets.

The situation in Approximation theory is quite similar. For instance, linear approximation from trigonometric or algebraic polynomials and

xxx
xxx and xxx (eds.), pp. 1–4.

ISBN 0-9728482-x-x

*Supported by NSF Grant DMS-0200665.

nonlinear n-term approximation from large redundant systems (dictionaries) are ineffective. The problem with polynomials is that they have poor localization properties, while the problem with nonlinear n-term approximation from large dictionaries is that it is hard to find a good (optimal or near optimal) representation for a given function and hence to approximate. Not long ago the wavelets came into play with a great success.

Wavelets, however, cannot solve all our problems. Many applications involve more complex geometries for which wavelet bases with the desired properties are hard to construct or even are not available at all. Secondly, most of the functions have anisotropic structure e.g. they may have singularities along curves or surfaces, which causes problems to wavelets. The situation is again quite different when approximating in the uniform norm which is a major concern and guiding issue in this article. Wavelet representations of functions are not well aligned with the uniform norm. Thus in a wide range of problems wavelets have now run their course. The challenge is to find new tools and methods for tasks that the existing tools are not able to handle efficiently.

This survey is centered around the idea of utilizing to nonlinear n-term approximation (anisotropic) redundant systems with the structure of a multiresolution analysis (MRA). Such hierarchies allow for a great deal of flexibility and simultaneously provide enough structure for the development of a coherent theory and effective algorithms for nonlinear n-term approximation.

We next outline the main topics covered in this article. The principles of nonlinear n-term approximation are given in §2.

In §3 we focus our attention on spline MRAs generated by sequences of nested triangulations. More precisely, consider a sequence $\mathcal{T}_0, \mathcal{T}_1, \dots$ of nested triangulations of a polygonal domain E in $\mathbb{R}^2$ (for simplicity). Mild natural conditions are imposed on the triangulation to prevent them from possible deterioration and at the same time to allow the triangles to change in shape, size, and orientation quickly when moving around at a given level and through the levels. In particular, skinny triangles with arbitrarily sharp angles may occur in any location. A sequence of triangulations like these immediately creates a geometric structure on E, in particular, it determines a quasi-distance $d_{\mathcal{T}}(\cdot,\cdot)$, defined by means of the areas of the triangles (§3.1). In turn $d_{\mathcal{T}}(\cdot,\cdot)$ along with the Lebesgue measure on E induces spaces of homogeneous type, which generate anisotropic spaces on E such as anisotropic Hardy spaces, BMO, and B-spaces (anisotropic Besov spaces). Section 3 is devoted to the description of the anisotropic structure generated by a sequence of nested triangulations of E.

Further, any sequence $\mathcal{T}_0, \mathcal{T}_1, \dots$ (set $\mathcal{T} := \cup_{m\ge 0}\mathcal{T}_m$) allows to introduce a Franklin system $\mathcal{F}_{\mathcal{T}}$ by applying the Gram-Schmidt orthogonalization process to the corresponding sequence of Courant elements (linear continuous elements). As will be shown in §4 (see [19]) the anisotropic

Franklin system $\mathcal{F}_{\mathcal{T}} = \{f_\theta\}_{\theta\in\Theta^*}$ is a Schauder basis for $L_p(E), 1 \le p \le \infty$, with $L_\infty(E) := C(E)$ and a unconditional basis for $L_p(E), 1 < p < \infty$ and the corresponding Hardy space $H_1(E, \mathcal{T})$. Also, the Franklin bases characterize the corresponding anisotropic BMO spaces and B-spaces. Thus all basic results for classical Franklin systems have analogues in the anisotropic case.

Denote by $\mathcal{S}^{k,r}(\mathcal{T}_m)$ the set of all r-times differentiable piecewise polynomials of degree $< k$ over the triangles of $\mathcal{T}_m$. Assume that there exists a spline multiresolution analysis $\mathcal{M}$ consisting of a ladder of spaces

$$\mathcal{S}_0 \subset \mathcal{S}_1 \subset \cdots, \qquad \mathcal{S}_m \subset \mathcal{S}^{k,r}(\mathcal{T}_m),$$

with bases (Φ_m) of $(\mathcal{S}_m)$. Set $\Phi_{\mathcal{T}} := \cup_{m\ge 0}\Phi_m$. Consider now the problem for nonlinear n-term approximation from the set $\Sigma_n(\Phi_{\mathcal{T}})$ of all piecewise polynomials of the form $s = \sum_{j=1}^n c_j\phi_j$, where $\phi_j \in \Phi_{\mathcal{T}}$ may come from arbitrary levels and locations. The theory of nonlinear n-term approximation from such collection of elements is developed in [7, 16]. We describe the B-spaces (generalized Besov spaces) needed for this theory in §5 and the theory itself in §6.

Our primary interest in this article is focused on algorithms for nonlinear n-term approximation in L_p $(0 < p < \infty)$ and most importantly in the uniform norm. The "Push-the-Error" algorithm for nonlinear n-term approximation in the uniform norm was first developed in [17]. In [6] this algorithm is further refined so that it is capable of achieving all rates of the best n-term approximation. The results on algorithms from [6, 17] are discussed in §7.

The last Section 8 is concerned with nonlinear n-term rational approximation, which is closely related to nonlinear n-term spline approximation.

Most of the presented results are brand new and exist only in preprints, available on line. This article should be regarded as complimentary to the survey paper [24], where the basic results on nonlinear n-term spline approximation are presented in detail.

Notation. Throughout this article for a given set $G \subset \mathbb{R}^2$, $|G|$ denotes the Lebesgue measure of G, while G° means the interior of G; $\mathbb{1}_G$ denotes the characteristic function of G, and $\tilde{\mathbb{1}}_G := |G|^{-1/2}\mathbb{1}_G$. For a finite set G, $\#G$ denotes the cardinality of G. Positive constants are denoted by $c, c_1, \ldots$ (if not specified, they may vary at every occurrence), $A \approx B$ means $c_1 A \le B \le c_2 B$, and $A := B$ or $B =: A$ stands for "A is by definition equal to B".

§2. The Principles of Nonlinear n-term Approximation

In this section we give a brief description of the general guiding principles of the theory of nonlinear n-term approximation as well as of the algorithms for n-term approximation. (See e.g. [9, 11, 25].)

Let X be a normed or quasi-normed function space, where the approximation will take place (in this article, $X = L_p(E)$, $0 < p \le \infty$). Suppose $\Phi = \{\phi_\theta\}_{\theta\in\Theta}$ 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

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

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

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

Approximation spaces. The primary objective of nonlinear n-term approximation is to describe the spaces of functions of given rates of n-term approximation such as the set of all $f \in X$ for which $\sigma_n(f) \le cn^{-\gamma}$ ($\gamma > 0$). More precisely, the goal is to characterize the approximation space $A_q^\gamma := A_q^\gamma(\Phi)$, $\gamma > 0$, $0 < q \le \infty$, consisting of all functions $f \in X$ such that

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

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

The machinery of Jackson-Bernstein estimates. A standard technique for description of approximation spaces is to prove Jackson and Bernstein estimates, and then apply interpolation.

Suppose $B \subset X$ is a smoothness space with a (quasi-)norm $\|\cdot\|_B$, satisfying the τ-triangle inequality: $\|f+g\|_B^\tau \le \|f\|_B^\tau + \|g\|_B^\tau$ with $0 < \tau \le 1$, and let $\Phi \subset B$. The K-functional is defined by

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

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

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

where the ℓ_q-norm is replaced by the sup-norm if $q = \infty$ (see e.g. [1]).

Theorem 1. (a) *Suppose the following Jackson estimate holds: There is* $\alpha > 0$ *such that for* $f \in B$

$$\sigma_n(f) \leq cn^{-\alpha}\|f\|_B, \quad n \geq 1. \tag{2}$$

Then, for $f \in X$,

$$\sigma_n(f) \leq cK(f, n^{-\alpha}), \quad n \geq 1.$$

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

$$\|S\|_B \leq cn^\alpha \|S\|_X, \quad \textit{for } S \in \Sigma_n,\ n \geq 1. \tag{3}$$

Then, for $f \in X$,

$$K(f, n^{-\alpha}) \leq cn^{-\alpha}\left(\left[\sum_{\nu=1}^{n} \frac{1}{\nu}(\nu^\alpha \sigma_\nu(f))^\tau\right]^{1/\tau} + \|f\|_X\right), \quad n \geq 1.$$

For the proof of this theorem, see e.g. [25].

An immediate consequence of Theorem 1 is that if the Jackson and Bernstein inequalities (2) and (3) hold, then $\sigma_n(f) = O(n^{-\gamma})$, $0 < \gamma < \alpha$, if and only if $K(f, n^{-\alpha}) = O(n^{-\gamma})$. More generally, Theorem 1 readily yields the following characterization of the approximation spaces $A_q^\gamma(\Phi)$:

Theorem 2. *Suppose the Jackson and Bernstein inequalities* (2) *and* (3) *from Theorem 1 hold. Then*

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

with equivalent norms.

It is an important observation that in nonlinear n-term approximation there is no saturation. This means that no matter how large γ is the approximation space $A_q^\gamma(\Phi)$ is nontrivial. This is the motivation of our desire to characterize the approximation spaces $A_q^\gamma(\Phi)$ for all $\gamma > 0$. Thus if the Jackson and Bernstein inequalities (2) and (3) from Theorem 1 hold for all $\alpha > 0$, then Theorem 2 characterizes $A_q^\gamma(\Phi)$ for all $\gamma > 0$.

Algorithms for nonlinear n-term approximation. It is desirable to construct practically feasible algorithms for nonlinear n-nonlinear approximation from specific sets Φ, which achieve the rates of the best approximation. More precisely, for a given Φ, we are interested in an algorithm which associates with each $f \in X$ a function $\mathcal{A}_n(f)$ of the form

$$\mathcal{A}_n(f) := \sum_{\theta \in \mathcal{M}_n} d_\theta \phi_\theta, \quad \mathcal{M}_n \subset \Phi, \text{ and } \#\mathcal{M}_n \leq n,$$

which is optimal in the following sense. Denote $A_n(f) := \|f - \mathcal{A}_n(f)\|_X$ and let

$$\|f\|_{A_q^\gamma(\mathcal{A})} := \|f\|_X + \Big(\sum_{n=1}^{\infty}(n^\gamma A_n(f))^q \frac{1}{n}\Big)^{1/q} < \infty. \tag{4}$$

We are interested in algorithms such that $A_q^\gamma(\mathcal{A}) = A_q^\gamma(\Phi)$ (with equivalent norms) for all $\gamma > 0$. Of course, for large values of γ the space $A_q^\gamma(\Phi)$ is tiny and not quite important but still an algorithm should be deemed *optimal* if it can capture all rates of best n-term approximation.

Direct theorem for n-term approximation in L_p ($p < \infty$). It is a general truth that in nonlinear approximation it is much easier to prove Jackson estimates than Bernstein estimates. We next give a general embedding theorem and a Jackson estimate which provide the needed estimates in many cases of nonlinear n-term approximation in L_p, $p < \infty$, in the case of compactly supported approximating elements. The proof of these estimates is not hard and can be found in [16].

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

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

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

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

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

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 $0 < \tau < p$

$$N(f) := \Big(\sum_m \|\Phi_m\|_p^\tau\Big)^{1/\tau} < \infty. \tag{5}$$

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

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

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

Furthermore, if $1 \le p < \infty$, condition (5) can be replaced by the weaker condition

$$N(f) := \|\{\|\Phi_m\|_p\}\|_{w\ell_\tau} < \infty, \tag{6}$$

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

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

Theorem 4. *Under the hypothesis of Theorem 3, suppose* $\{\Phi_m^*\}_{j=1}^\infty$ *is a rearrangement of the sequence* $\{\Phi_m\}$ *such that* $\|\Phi_1^*\|_p \ge \|\Phi_2^*\|_p \ge \dots$. *Denote* $S_n := \sum_{j=1}^n \Phi_j^*$. *Then*

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

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

Furthermore, the estimate remains valid if condition (5) *is replaced by* (6) *when* $1 \le p < \infty$.

§3. Anisotropic Structures on Compact Polygonal Domains

In this section we collect all prerequisites regarding triangulations, maximal operators, hierarchies of spline bases, spaces of homogeneous type, etc. which are needed in the development nonlinear n-term spline approximation and anisotropic Franklin bases.

3.1. Multilevel Triangulations

A set $E \subset \mathbb{R}^2$ is said to be a *bounded polygonal domain* if its interior E° is connected and E is the union of a finite set $\mathcal{T}_0$ of closed triangles with disjoint interiors: $E = \cup_{\triangle \in \mathcal{T}_0} \triangle$.

Locally Regular (LR) Triangulations. We call $\mathcal{T} = \bigcup_{m=0}^\infty \mathcal{T}_m$ a locally regular triangulation of E or briefly an *LR-triangulation* with levels $(\mathcal{T}_m)_{m \ge 0}$ if the following conditions are fulfilled:

(a) Every level $\mathcal{T}_m$ is a partition of E, that is, $E = \cup_{\triangle \in \mathcal{T}_m} \triangle$ and $\mathcal{T}_m$ consists of closed triangles with disjoint interiors.

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

(c) Each triangle $\triangle \in \mathcal{T}_m$ has at least two and at most M_0 children (subtriangles) in $\mathcal{T}_{m+1}$, where $M_0 \ge 2$ is a constant.

(d) 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.

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

(f) There exist constants $0 < r < \rho < 1$ $(r \le \frac{1}{2})$ such that for each $\triangle \in \mathcal{T}_m$ $(m \ge 0)$ and any child $\triangle' \in \mathcal{T}_{m+1}$ of $\triangle$,

$$r|\triangle| \le |\triangle'| \le \rho|\triangle|. \tag{7}$$

(g) There exists a constant $0 < \delta_1 \le 1$ such that for $\triangle', \triangle'' \in \mathcal{T}_m (m \ge 0)$ with a common vertex,

$$\delta_1 \le |\triangle'|/|\triangle| \le \delta_1^{-1}. \tag{8}$$

Strong Locally Regular (SLR) Triangulations. We call $\mathcal{T} = \bigcup_{m=0}^\infty \mathcal{T}_m$ a strong locally regular (SLR) triangulation of E if $\mathcal{T}$ is an LR-triangulation of E and satisfies the following additional condition:

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

$$|\mathrm{conv}\,(\triangle' \cup \triangle'')|/|\triangle'| \le \delta_2^{-1}, \tag{9}$$

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

Regular Triangulations. We call $\mathcal{T} = \bigcup_{m=0}^\infty \mathcal{T}_m$ a regular triangulation of a bounded polygonal domain $E \subset \mathbb{R}^2$ if $\mathcal{T}$ satisfies conditions (a)-(e) of LR-triangulations and also the *minimal angle condition*, that is, $\min\mathrm{angle}\,(\triangle) \ge \beta$ for every triangle $\triangle \in \mathcal{T}$, where $\beta > 0$ is a constant.

Evidently, every SLR-triangulation is an LR-triangulation but not the other way around and every regular triangulation is an SLR-triangulation. For other types of triangulations, see [16].

We denote by $\mathcal{V}_m$ the set of all vertices of triangles from $\mathcal{T}_m$ and by $\mathcal{E}_m$ the set of all edges of triangles in $\mathcal{T}_m$. We also set $\mathcal{V} := \cup_{m \ge 0} \mathcal{V}_m$ and $\mathcal{E} := \cup_{m \ge 0} \mathcal{E}_m$.

We next give some basic facts about LR-triangulations which is our main type of triangulations. For more details we refer the reader to [16] and [19].

The constants M_0, N_0, r, ρ, δ, and $\#\mathcal{T}_0$ associated with an LR-triangulation $\mathcal{T}$ are assumed fixed. We refer to them as *parameters* of $\mathcal{T}$.

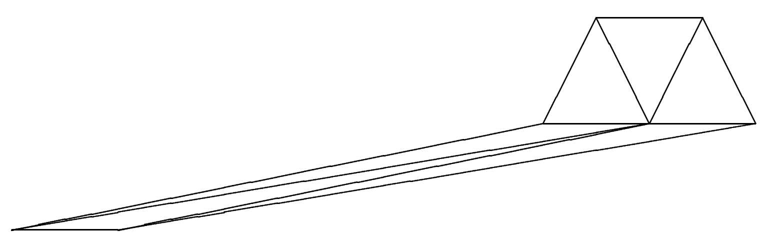

Figure 1: A skewed cell in an LR-triangulation

The most important conditions (f)-(g) on LR-triangulations involve only areas of triangles but not angles. Consequently, if $\mathcal{T}$ is an LR-triangulation and $\triangle', \triangle'' \in \mathcal{T}_m$ have a common edge, then it may happen that $\triangle'$ is an equilateral triangle (or close to an equilateral triangle) but $\triangle''$ has an uncontrollably sharp angle (see Figure 1).

In an LR-triangulation $\mathcal{T}$ there can be an equilateral (or close to such) triangle $\triangle^\diamond$ at any level $\mathcal{T}_m$ with descendants $\triangle_1 \supset \triangle_2 \supset \ldots$ such that $\min\text{angle}\,(\triangle_j) \to 0$ as $j \to \infty$.

It is important to know how fast the area $|\triangle|$ of a triangle $\triangle \in \mathcal{T}_m$ may change when $\triangle$ moves away from a fixed triangle within the same level. Condition (f) suggests a geometric rate of change but in fact it is polynomial: If $\triangle$, $\triangle' \in \mathcal{T}_m$ can be connected by n intermediate edges from $\mathcal{E}_m$, then

$$c_1^{-1}(n+1)^{-s} \le |\triangle'|/|\triangle''| \le c_1(n+1)^s, \qquad (10)$$

where $s, c_1 > 0$ depend only on the parameters of $\mathcal{T}$.

Further, for any sequence of triangles $\triangle_1 \supset \triangle_2 \supset \ldots$ from an LR-triangulation, $\operatorname{diam}(\triangle_j) \to 0$ as $j \to \infty$.

We want to mention only one property of SLR-triangulations, namely, triangles of an SLR-triangulation may have arbitrarily sharp angles, but the configuration of Figure 1 is impossible (for more details see [16]).

Graph Distance. We now introduce the mth level graph distance between vertices: For any two vertices $v', v'' \in \mathcal{V}_m$, $m \ge 0$, we define the *graph distance* $\rho_m(v', v'')$ as the minimum number of edges from $\mathcal{E}_m$ needed to connect v' and v''.

Cells. For any vertex $v \in \mathcal{V}_m$ $(m \ge 0)$, we denote by ω_v the union of all triangles from $\mathcal{T}_m$ which have v as a common vertex. We denote by $\mathcal{O}_m$ the set of all such cells ω_v with $v \in \mathcal{V}_m$ and set $\mathcal{O} = \cup_{m\ge 0}\mathcal{O}_m$. For a given cell $\omega \in \mathcal{O}$, we shall denote by v_ω the "central" vertex of ω.

For given $\omega', \omega'' \in \mathcal{O}_m$, we define the *graph distance* $\rho_m(\omega', \omega'')$ between ω' and ω'' by $\rho_m(\omega', \omega'') := \rho_m(v_{\omega'}, v_{\omega''})$, where $v_{\omega'}, v_{\omega''} \in \mathcal{V}_m$ are the "central vertices" of ω', ω''.

Definition of ω_x^m. We want to associate with each $x \in E$ a cell $\omega_x^m \in \mathcal{O}_m$, $m \ge 0$, which contains x. To this end we first associate with each triangle $\triangle \in \mathcal{T}_m$ a cell $\omega_\triangle^m \in \mathcal{O}_m$ such that $\triangle \subset \omega_\triangle^m$. Such a cell can be selected in three different ways. We choose one of them for each $\triangle \in \mathcal{T}_m$. Then for each $x \in E$ such that $x \in \triangle^\circ$ with $\triangle \in \mathcal{T}_m$, we define $\omega_x^m := \omega_\triangle^m$. If x lies on the edge of a triangle from $\mathcal{T}_m$, we define ω_x^m as any cell from ω_m such that x belongs to its interior, but if $x = v_\omega$ for some $\omega \in \mathcal{O}_m$, we set $\omega_x^m := \omega$.

Stars. In order to deal with graph distances and neighborhood relations it is convenient to employ the notion of the mth *level star of a set*: For any set $G \subset E$ and $m \ge 0$, we define the first mth level star of G by

$$\operatorname{Star}_m(G) := \operatorname{Star}_m^1(G) := \cup\{\omega \in \mathcal{O}_m : \mathcal{O}^\circ \cap G \ne \emptyset\} \qquad (11)$$

and inductively, $\mathrm{Star}\,^k_m(G) := \mathrm{Star}\,^1_m(\mathrm{Star}\,^{k-1}_m(G))$, $k > 1$.

Quasi-distance and Maximal Operator. Every LR-triangulation $\mathcal{T}$ of E naturally generates a quasi-distance and a maximal operator.

We begin by recalling the definition of a quasi-distance on a set X: The mapping $d : X \times X \to [0, \infty)$ is called a *quasi-distance* on X if for $x, y, z \in X$,

(a) $d(x, y) = 0 \Longleftrightarrow x = y$,

(b) $d(y, x) = d(x, y)$,

(c) $d(x, z) \le K(d(x, y) + d(y, z))$ with $K \ge 1$.

Assuming that $\mathcal{T}$ is an LR-triangulation of a polygonal domain $E \subset \mathbb{R}^2$, we define the quasi-distance $d_{\mathcal{T}} : E \times E \to [0, \infty)$ by

$$d_{\mathcal{T}}(x, y) := \min\{|\omega| : \omega \in \mathcal{O} \ \text{ and } \ x, y \in \omega\}, \tag{12}$$

if x, y belong to at least one cell from $\mathcal{O}$, and by $d_{\mathcal{T}}(x, y) := |E|$ otherwise. It is not hard to see that $d_{\mathcal{T}}(\cdot, \cdot)$ satisfies the axioms of a quasi-distance.

The following inequality relates the quasi-distance $d_{\mathcal{T}}(\cdot, \cdot)$ with the mth level graph distance introduced above: There exist constants $\beta > 0$ and $c > 0$ such that for $\omega \in \mathcal{O}_m$ $(m \ge 0)$ and $x \in E$,

$$d_{\mathcal{T}}(v_\omega, x) \le c|\omega|\rho_m(\omega, \omega_x^m)^\beta \quad \text{if} \quad \rho_m(\omega, \omega_x^m) \ge 2.$$

The quasi-distance $d_{\mathcal{T}}(\cdot, \cdot)$ generates a maximal operator. Denote by $B(y, a)$ the "ball" centered at y of radius $a > 0$ with respect to this quasi-distance, i.e. $B(y, a) := \{x : d_{\mathcal{T}}(x, y) < a\}$. Then for any $s > 0$ the maximal operator $\mathcal{M}^s_{d_{\mathcal{T}}}$ is defined by

$$(\mathcal{M}^s_{d_{\mathcal{T}}} f)(x) := \sup_{B:\, x \in B} \Big(\frac{1}{|B|} \int_B |f(y)|^s \, dy\Big)^{1/s}, \quad x \in E,$$

where the supremum is over all balls B containing x.

For our purposes it is more convenient to use the equivalent maximal operator $\mathcal{M}^s_{\mathcal{T}}$ defined by

$$(\mathcal{M}^s_{\mathcal{T}} f)(x) := \sup_{\omega:\, x \in \omega} \left(\frac{1}{|\omega|} \int_\omega |f(y)|^s dy\right)^{1/s},$$

where the supremum is over all cells $\omega \in \mathcal{O}$ containing x or $\omega = E$.

It is important that the Fefferman-Stein [14] vector valued maximal inequality holds for the maximal operator $\mathcal{M}^s_{\mathcal{T}}$ (for more details, see [19]):

Theorem 5. *Let $\mathcal{T}$ be an LR-triangulation of $E \subset \mathbb{R}^2$. If $0 < p < \infty$, $0 < q \le \infty$, and $0 < s < \min\{p, q\}$, then for any sequence of functions $(f_j)_{j=1}^{\infty}$ on E,*

$$\Big\| \Big(\sum_{j=1}^{\infty} |\mathcal{M}_{\mathcal{T}}^{s} f_j|^q \Big)^{1/q} \Big\|_p \le c \Big\| \Big(\sum_{j=1}^{\infty} |f_j|^q \Big)^{1/q} \Big\|_p, \tag{13}$$

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

3.2. Hierarchical Families of Spline Bases

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

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

- $\mathcal{S}_m \subset \mathcal{S}_{m+1}$.
- $\Pi_{\tilde{k}} \subset \mathcal{S}_m$, for some $1 \le \tilde{k} \le k$ ($\tilde{k}$ independent of m).
- For any $s \in \mathcal{S}_m$ there exists a unique sequence of real coefficients $\{a_\theta(s)\}_{\theta \in \Theta_m}$ such that

  $$s = \sum_{\theta \in \Theta_m} a_\theta(s) \phi_\theta.$$

  (Thus Φ_m is a basis for $\mathcal{S}_m$ and $\{a_\theta(\cdot)\}_{\theta \in \Theta_m}$ are the dual functionals.)
- For each $\theta \in \Theta_m$ there is a vertex $v = v_\theta \in \mathcal{V}_m$ such that

  $$\operatorname{supp} \phi_\theta \subset \operatorname{Star}_m^{\ell}(v) =: \mathcal{E}_\theta, \ \text{(see (11))} \tag{14}$$

  $$\|\phi_\theta\|_{L_\infty(E)} = \|\phi_\theta\|_{L_\infty(\mathcal{E}_\theta)} = 1, \tag{15}$$

  $$|a_\theta(s)| \le \beta \|s\|_{L_\infty(\mathcal{E}_\theta)}, \quad s \in \mathcal{S}_m, \tag{16}$$

  where $\ell \ge 1$ and $\beta > 0$ are constants, independent of θ and m.

Since $\mathcal{S}_m \subset \mathcal{S}_{m+1}$, we have

$$\phi_\theta = \sum_{\eta \in \Theta_m,\, \eta \subset \theta} a_{\theta,\eta} \phi_\eta, \quad \theta \in \Theta_{m-1}. \tag{17}$$

Moreover, by (15)-(16) it follows that $|a_{\theta,\eta}| = |a_\eta(\phi_\theta)| \le \beta$.

We denote $\mathcal{M} := (\mathcal{S}_m)_{m\ge0}$, $\Phi := \cup_{m\ge0}\Phi_m$ and $\Theta := \cup_{m\ge0}\Theta_m$. We shall call $\mathcal{M}$ a *spline multiresolution analysis* over $\mathcal{T}$ with a family of basis functions Φ.

Courant elements. A simple example of a spline MRA is the sequence $(\mathcal{S}_m)_{m\ge0}$ of all continuous piecewise linear functions ($r = 0$, $k = 2$) over the levels $(\mathcal{T}_m)_{m\ge0}$ of a given LR-triangulation $\mathcal{T}$ of a compact polygonal domain E in $\mathbb{R}^2$. A basis for each space $\mathcal{S}_m$ is given by the set Φ_m of the Courant elements ϕ_θ, supported on the cells θ of $\mathcal{T}_m$ (θ is the union of all triangles of $\mathcal{T}_m$ attached to a vertex, say, v_θ). Thus $\Theta = \mathcal{O}$. The function ϕ_θ takes the value 1 at v_θ and the value 0 at all other vertices.

A concrete construction of differentiable spline basis functions (from C^r, $r \ge 1$) associated with a spline multiresolution analysis over general triangulations is given in [7], see also the discussion in [6, 7] of examples of spline MRAs in regular set-ups.

Note that Θ and Θ_m ($m \ge 0$) above are simply index sets, which in the case of Courant elements can be identified as sets of cells (supports of basis functions). In general, several basis functions in Φ_m may have the same support. However, the supports of only $\le$ *constant* of them may overlap.

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

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

with constants of equivalence independent of $\{b_\theta\}$ and m.

Depending on the domain E in some settings one can even construct wavelet or prewavelet bases. For simplicity, whenever we assume in this article the existence of wavelets we assume the existence of a biorthogonal wavelet basis $\Psi = \{\psi_\lambda : \lambda \in \mathcal{L}\}$ on E with a dual $\widetilde{\Psi} = \{\widetilde{\psi}_\lambda : \lambda \in \mathcal{L}\}$, where $\mathcal{L} = \cup_{m\ge0}\mathcal{L}_m$ is the index set of the "true" wavelets. Then each $f \in L_p(E)$ ($1 \le p \le \infty$) has the representation

$$f = \sum_{\theta\in\Theta_0} c_\theta(f)\phi_\theta + \sum_{m\ge0}\sum_{\lambda\in\mathcal{L}_m} c_\lambda(f)\psi_\lambda, \qquad c_\lambda(f) := \langle f, \widetilde{\psi}_\lambda\rangle, \tag{18}$$

which is assumed to be unconditional if $1 < p < \infty$. In addition, we assume that $\psi_\lambda, \widetilde{\psi}_\lambda$ ($\lambda \in \mathcal{L}_m$) are compactly supported with $\operatorname{supp}\psi_\lambda, \operatorname{supp}\widetilde{\psi}_\lambda \subset \lambda$, and $\lambda = \mathrm{Star}_m^{\ell'}(v_\lambda)$, where $v_\lambda \in \mathcal{V}_m$ and $\ell' = constant$. Also, we assume that for $\lambda \in \mathcal{L}_m$, $\psi_\lambda \in \mathcal{S}_{m+1}$, i.e. $\psi_\lambda = \sum_{\theta\in\Theta_{m+1}} a_{\lambda,\theta}\phi_\theta$, and $|a_{\lambda,\theta}| \le \beta'$ with β' a uniform constant.

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

$$\|f - P_{\triangle,q}(f)\|_{L_q(\triangle)} \le cE_k(f,\triangle)_q, \quad \text{for } f \in L_q(\triangle), \tag{19}$$

where Π_k denotes the set of all algebraic polynomials of total degree $< k$. We define a linear operator $Q_m : \mathcal{S}^{k,-1}(\mathcal{T}_m) \to \mathcal{S}_m$ as follows. For each $\theta \in \Theta_m$, let $\lambda_\theta : \mathcal{S}^{k,-1}(\mathcal{T}_m)|_{\mathcal{E}_\theta} \to \mathbb{R}$ be a linear functional such that

$$\lambda_\theta(s|_{\mathcal{E}_\theta}) = a_\theta(s), \text{ if } s \in \mathcal{S}_m, \text{ and}$$

$$|\lambda_\theta(f)| \le \beta \|f\|_{L_\infty(\mathcal{E}_\theta)}, \quad f \in \mathcal{S}^{k,-1}(\mathcal{T}_m)|_{\mathcal{E}_\theta}.$$

Such linear functionals always exist due to the Hahn-Banach theorem. Set

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

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

$$\|Q_m(s)\|_{L_q(\triangle)} \le c\|s\|_{L_q(\Omega^\ell_\triangle)}$$

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

We now extend Q_m to $L_q(E)$, $0 < q \le \infty$. Let $P_{\triangle,q} : L_q(\triangle) \to \Pi_k$ be a projector satisfying (19). We define

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

and the quasi-interpolant that we need is defined by

$$Q_{m,q}(f) := Q_m(p_{m,q}(f)), \quad \text{for } f \in L_q(E), \tag{20}$$

which is a projector of $L_q(E)$ onto $\mathcal{S}_m$.

The next estimate shows that $Q_{m,q}$ provides a good local L_q-approximation from $\mathcal{S}_m$. Denote first, for $\triangle \in \mathcal{T}_m$,

$$\Omega^\ell_\triangle := \mathrm{Star}^{\,\ell}_{\,m}(\triangle) = \cup\{\mathrm{Star}^{\,\ell}_{\,m}(v) : v \in \mathcal{V}_m,\ v \in \triangle\} \tag{21}$$

and let $\mathbb{S}_\triangle(f)_q$ denote the error of L_q-approximation from $\mathcal{S}_m$ on $\Omega^\ell_\triangle$, i.e.

$$\mathbb{S}_\triangle(f)_q := \inf_{s\in\mathcal{S}_m} \|f - s\|_{L_q(\Omega^\ell_\triangle)}, \quad \triangle \in \mathcal{T}_m. \tag{22}$$

The good local approximation properties of $Q_{m,q}$ can be described as follows: If $f \in L_q(E)$, $0 < q \le \infty$ ($f \in C$ if $q = \infty$), then

$$\|f - Q_{m,q}(f)\|_{L_q(\triangle)} \le c\mathbb{S}_\triangle(f)_q, \quad \triangle \in \mathcal{T}_m \ (m \ge 0),$$

with c independent of f, m, and $\triangle$. Further, it is not hard to see that if $f \in L_q(E)$, $0 < q \leq \infty$, then

$$\|f - Q_{m,q}(f)\|_{L_q(E)} \to 0 \quad \text{as} \quad m \to \infty. \tag{23}$$

Decomposition via quasi-interpolants. We use the projectors $Q_{m,q}$ from (20) for decomposing a given function into multilevel components. For $f \in L_q(E)$, we denote by

$$q_m(f) = q_{m,q}(f) := Q_{m,q}(f) - Q_{m-1,q}(f), \quad Q_{-1,q}(f) := 0, \tag{24}$$

the "detail" of f between levels m and $m-1$. Evidently, $q_m(f) \in \mathcal{S}_m$. Let $\{b_{\theta,q}(f)\}_{\theta\in\Theta_m}$ be defined by the identity

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

Now, (23) yields the following representation of $f \in L_q(E)$, $0 < q \leq \infty$:

$$f = \sum_{m\geq 0} q_m(f) = \sum_{m\geq 0} \sum_{\theta\in\Theta_m} b_{\theta,q}(f)\phi_\theta \quad \text{in } L_q. \tag{25}$$

See [7] for more details.

Decomposition via wavelets. Whenever a wavelet basis is available, q_m can be defined as the associated canonical projectors, i.e.

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

Then one can rewrite $q_m(f)$ at the mth level: $q_m(f) =: \sum_{\theta\in\Theta_m} b_\theta(f)\phi_\theta$ and use these to replace the corresponding terms in (25).

3.3. Spaces of Homogeneous Type on Polygonal Domains

Spaces of homogeneous type were first introduced in [4] as a means to extend the Calderon-Zygmund theory of singular integral operators to more general settings.

Let X be a topological space endowed with a Borel measure μ and a quasi-distance $d(\cdot,\cdot)$. Assume that the balls $B(x,r) := \{y \in X : d(x,y) < r\}$, $x \in X$, $r > 0$, form a basis for the topology T in X, and $\mu(B(x,r)) > 0$ if $r > 0$. The space (X, d, μ) is said to be of *homogeneous type* if there exists a constant A such that for all $x \in X$ and $r > 0$,

$$0 < \mu(B(x,2r)) \leq A\mu(B(x,r)). \tag{26}$$

The homogeneous type space $(E, d_\mathcal{T}, m)$. Suppose that E is a bounded polygonal domain and let $\mathcal{T}$ be a LR-triangulation on E. Also, let $d_\mathcal{T}(\cdot,\cdot)$

be the quasi-distance on E, defined in (12). Finally, denote by m the Lebesgue measure on E. It is easy to see that $(E, d_{\mathcal{T}}, m)$ is a space of homogeneous type, so one can utilize the machinery developed in [5].

The Hardy space $H_1(E, \mathcal{T})$. We next define the Hardy space $H_1 := H_1(E, \mathcal{T})$ associated with the space $(E, d_{\mathcal{T}}, m)$ by means of atomic representations (see [5]).

According to Coifmann and Weiss [5], a function $a(x)$ is said to be a q-atom $(1 < q \le \infty)$ if there exist $x_0 \in E$ and $r > 0$ such that

(i) $\operatorname{supp} a \subset B(x_0, r)$, (ii) $\|a\|_q \le |B(x_0, r)|^{1/q-1}$, (iii) $\int a(x)\, dx = 0$.
In addition, $|E|^{-1} \mathbb{1}_E$ is by definition a q-atom as well.

We adopt the following slightly different but equivalent definition for a q-atom which better suits our purposes.

Definition 1. A function $a(x)$ is said to be a q-atom $(1 < q \le \infty)$ for $H_1(E, \mathcal{T})$ if there is $\omega \in \mathcal{O}$ or $\omega = E$ such that

(a) supp $a \subset \omega$,
(b) $\|a\|_q \le |\omega|^{1/q-1}$,
(c) $\int_E a(x)\, dx = 0$.
We also postulate $|E|^{-1} \mathbb{1}_E$ to be a q-atom.

Definition 2. The space $H_1^q := H_1^q(E, \mathcal{T})$ $(1 < q \le \infty)$ is defined as the set of all functions $f \in L_1(E)$ admitting an atomic decomposition

$$f = \sum_{j=0}^{\infty} \lambda_j a_j,$$

where the a_j's are q-atoms and $\sum_{j=0}^{\infty} |\lambda_j| < \infty$. Moreover, the norm of $f \in H_1^q$ is given by

$$\|f\|_{H_1^q} := \inf \Big\{ \sum_{j=0}^{\infty} |\lambda_j| : f = \sum_{j=0}^{\infty} \lambda_j a_j,\ a_j\ q\text{-atoms} \Big\}.$$

A fundamental fact in the theory of Hardy spaces is that $H_1^q = H_1^\infty$ whenever $1 < q \le \infty$ with equivalent norms (see [5], Theorem A). Thus all spaces H_1^q are the same and we shall drop the index q.

It is an important observation that the spaces $H_1(E, \mathcal{T})$ essentially depend on the triangulations $\mathcal{T}$. We call $H_1(E, \mathcal{T}^*)$ a *regular* H_1*-space* if $\mathcal{T}^*$ is a regular multilevel triangulation of E (see §3.1). It is readily seen that if $H_1(E, \mathcal{T}^*)$ is regular, then it is the same (with equivalent norms) as the space $H_1(E)$ defined using atoms generated by the Euclidean distance on E. Thus all regular spaces $H_1(E, \mathcal{T})$ are the same. As is shown in [19] there exists an LR-triangulation $\mathcal{T}$ such that $H_1(E, \mathcal{T}) \ne H_1(E)$. The reason for this is that there exist LR-triangulations on E containing triangles with uncontrollably sharp angles (see §3.1).

It is not hard to prove that $H_1(E,\mathcal{T})$ is a Banach space and $\|f\|_{L_1(E)} \le c\|f\|_{H_1(E,\mathcal{T})}$ for $f \in H_1(E,\mathcal{T})$.

Another fundamental result is that the dual of $H_1(E,\mathcal{T})$ is the space $BMO := BMO(E,\mathcal{T})$ which can be defined in our case as the set of all functions f on E such that

$$\|f\|_{BMO} := \Big|\int_E f(x)dx\Big| + \sup_\omega \Big(\frac{1}{|\omega|}\int_\omega |f(x)-f_\omega|^2dx\Big)^{1/2} < \infty, \quad (27)$$

where $f_\omega := \frac{1}{|\omega|}\int_\omega f(x)dx$ and the supremum is taken over all $\omega \in \mathcal{O}$ or $\omega = E$. More precisely, for $g \in BMO(E,\mathcal{T})$ and $f \in H_1(E,\mathcal{T})$ with an atomic decomposition $f = \sum_{j=1}^\infty \lambda_j a_j$,

$$\lim_{m\to\infty}\sum_{j=1}^m \lambda_j \int_E g(x)a_j(x)dx \quad (28)$$

defines a continuous linear functional on H_1 whose norm is equivalent to $\|g\|_{BMO}$ and vice versa each continuous linear functional on H_1 is of this form.

Note that an equivalent norm in $BMO(E,\mathcal{T})$ can be defined by replacing in (27) $\Big(\frac{1}{|\omega|}\int_\omega |f(x)-f_\omega|^2dx\Big)^{1/2}$ by $\frac{1}{|\omega|}\int_\omega |f(x)-f_\omega|dx$. For more details, see [5].

Finally, we observe that since $H_1(E,\mathcal{T}) \ne H_1(E)$ for some LR-triangu lations $\mathcal{T}$, then by a duality argument it follows that for the same triangulations $BMO(E,\mathcal{T}) \ne BMO(E)$, where $BMO(E)$ stands for the regular BMO space on E. Thus in general $BMO(E,\mathcal{T})$ depends on the triangulation $\mathcal{T}$.

§4. Anisotropic Franklin Bases

It is desirable to work with spline bases over general nested triangulations e.g. LR- or SLR-triangulations. However, to the best of our knowledge there are no constructions of compactly supported spline wavelet or prewavelet bases over such triangulations up to now.

Franklin systems, however, are easier to explore and provide a tool for decomposition of various anisotropic spaces generated by multilevel nested triangulations. We next define the Franklin system $\mathcal{F}_\mathcal{T}$ generated by Courant elements and present the main results on Franklin bases obtained in [19].

Let $\mathcal{T} := \bigcup_{m\ge0}\mathcal{T}_m$ be an LR-triangulation of E and recall that $\mathcal{V}_m$ denotes the set of all vertices of triangles from $\mathcal{T}_m$. Denote by $\Phi = \{\phi_\theta\}_{\theta\in\Theta}$ the set of all Courant elements supported on the cells of $\mathcal{T}$. Therefore, Θ now is the set $\mathcal{O}$ of all cells over $\mathcal{T}$ (see §3.1).

We set $\mathcal{V}_0^* = \mathcal{V}_0$ and $\mathcal{V}_m^* = \mathcal{V}_m \setminus \mathcal{V}_{m-1}$ for $m \geq 1$ and write $\mathcal{V}^* = \bigcup_{m=0}^{\infty} \mathcal{V}_m^*$.

Let $\theta_0 := E$. Choose $\theta_{\max} \in \Theta_0$ to be of maximum area and denote $\Theta_0^* := \{\theta_0\} \cup \Theta_0 \setminus \{\theta_{\max}\}$, i.e. we replace $\theta_{\max}$ by $\theta_0 = E$. Moreover, we associate θ_0 with $v_{\theta_{\max}}$ and set $\phi_{\theta_0} := \mathbb{1}_{\theta_0}$. For $m \geq 1$ denote by Θ_m^* the set of all cells $\theta \in \Theta_m$ with "central" vertices $v_\theta \in \mathcal{V}_m^*$ and set $\Theta^* := \bigcup_{m=0}^{\infty} \Theta_m^*$.

Note that for each m, the set $\{\phi_\theta : \theta \in \bigcup_{i=0}^{m} \Theta_i^*\}$ is linearly independent. Also, $\mathcal{S}_m = \operatorname{span}\{\phi_\theta : \theta \in \Theta_m\} = \operatorname{span}\{\phi_\theta : \theta \in \bigcup_{i=0}^{m} \Theta_i^*\}$.

We consider an arbitrary (but fixed) *linear order* $\preceq$ on Θ^* satisfying the following conditions:

(i) If $\theta \in \Theta_m^*$ and $\theta' \in \Theta_n^*$ with $m < n$, then $\theta \preceq \theta'$ and (ii) $\theta_0 \preceq \theta$, $\forall \theta \in \Theta^*$.

The *Franklin system* $\mathcal{F}_\mathcal{T}$ now is defined by applying the Gram-Schmidt orthogonalization process to $\{\phi_\theta\}_{\theta \in \Theta^*}$ in $L_2(E)$ with respect to the order $\preceq$. We obtain an orthonormal system $\mathcal{F}_\mathcal{T} := \{f_\theta\}_{\theta \in \Theta^*}$ in $L_2(E)$ consisting of continuous piecewise linear functions. Each Franklin function f_θ is uniquely determined (up to a multiple ± 1) by the conditions:

(a) $f_\theta \in \operatorname{span}\{\phi_{\theta'} : \theta' \preceq \theta\}$.

(b) $\langle f_\theta, \phi_{\theta'} \rangle = 0$ for all $\theta' \prec \theta$,

(c) $\|f_\theta\|_2 = 1$.

Note that $f_{\theta_0} = \pm \tilde{\mathbb{1}}_{\theta_0} := \pm |E|^{-1/2} \mathbb{1}_E$.

As is well known the localization properties of the Franklin functions play crucial role in the study of the Franklin systems.

Theorem 6. *There exist constants $0 < q_1 < 1$ and $c > 0$ depending only on the parameters of $\mathcal{T}$ such that for any $\theta \in \Theta_m^*$ $(m \geq 0)$,*

$$|f_\theta(x)| \leq c|\theta|^{-1/2} q_1^{\rho_m(\theta, \theta_x^m)}, \qquad x \in E, \tag{29}$$

where $\rho_m(\cdot,\cdot)$ and θ_x^m are defined in §3.1.

The main results on anisotropic Franklin systems from [19] read as follows:

Theorem 7. *The Franklin system $\mathcal{F}_\mathcal{T} := \{f_\theta\}_{\theta \in \Theta^*}$ is a Schauder basis for $L_p(E)$, $1 \leq p \leq \infty$, with $L_\infty(E) := C(E)$.*

Theorem 8. *The Franklin system $\mathcal{F}_\mathcal{T} := \{f_\theta\}_{\theta \in \Theta^*}$ is an unconditional basis for $H_1(E, \mathcal{T})$ and $L_p(E), 1 < p < \infty$.*

Theorem 9. *The following conditions are equivalent:*

(a) $f \in H_1(E, \mathcal{T})$;

(*b*) *The series* $\sum_{\theta\in\Theta^*}\langle f,f_\theta\rangle f_\theta$ *converges unconditionally in* L_1*;*

(*c*) $S_f(x):=\Big(\sum_{\theta\in\Theta^*}|\langle f,f_\theta\rangle|^2|f_\theta(x)|^2\Big)^{1/2}\in L_1$*;*

(*d*) $F_f(x):=\Big(\sum_{\theta\in\Theta^*}|\langle f,f_\theta\rangle|^2|\tilde{\mathbb{1}}_\theta(x)|^2\Big)^{1/2}\in L_1$.

Furthermore, if $f\in H_1(E,\mathcal{T})$*, then*

$$\|f\|_{H_1}\approx\|S_f\|_{L_1}\approx\|F_f\|_{L_1}. \tag{30}$$

Theorem 10. *A function* $f\in BMO(E,\mathcal{T})$ *if and only if*

$$\sup_\theta\Big(\frac{1}{|\theta|}\sum_{\eta\in\Theta^*:\,\eta\subset\theta}|\langle f,f_\eta\rangle|^2\Big)^{1/2}<\infty, \tag{31}$$

where the supremum is taken over all $\theta\in\Theta$ *or* $\theta=E$*. Furthermore,* $\|f\|_{BMO(E,\mathcal{T})}$ *is equivalent to the quantity in* (31).

These results show that the basic and well-known results on Franklin bases in the regular case (see e.g. [18], [3], and [15]) have analogues in the anisotropic case. Finally, we note that Franklin systems generated by hierarchical sequences of spline bases $\Phi=\{\phi_\theta\}$ other than Courant elements can be developed with an equal success and would have similar properties.

§5. B-spaces

We begin by introducing the B-space $B^\alpha_{pq}:=B^\alpha_{pq}(\mathcal{M})$ induced by a spline MRA $\mathcal{M}$ generated by a hierarchical sequence of spline bases over an LR or better triangulation $\mathcal{T}$ of a compact polygonal domain E in $\mathbb{R}^2$. We consider only B-spaces which are imbedded in $L_1(E)$. We say that the indices α, p, and q are *admissible* if one of the following holds:

(a) $0<p,q\le\infty$ and $\alpha>(1/p-1)_+$ or

(b) $0<p<1$, $0<q\le 1$, and $\alpha=1/p-1$.

It is easy to see [20] that these conditions guarantee the desired embedding.

We define $B^\alpha_{pq}(\mathcal{M})$ as the set of all functions $f\in L_p(E)$ such that

$$|f|_{B^\alpha_{pq}(\mathcal{M})}:=\Big(\sum_{m=0}^\infty\Big(\sum_{\triangle\in\mathcal{T}_m}(|\triangle|^{-\alpha}\,\mathbb{S}_\triangle(f)_p)^p\Big)^{q/p}\Big)^{1/q}<\infty, \tag{32}$$

where $\mathbb{S}_\triangle(f)_p$ is the error of L_p-approximation to f from $\mathcal{S}_m$ on $\Omega^\ell_\triangle$ (see (22)). We set

$$\|f\|_{B^\alpha_{pq}(\mathcal{M})}:=|E|^{-\alpha}\|f\|_p+|f|_{B^\alpha_{pq}(\mathcal{M})}. \tag{33}$$

Evidently, $\|\cdot\|_{B^\alpha_{pq}(\mathcal{M})}$ is a norm if $p, q \geq 1$ and quasi-norm otherwise.

The B-space B^α_{pq} has an atomic decomposition. We define

$$\|f\|^A_{B^\alpha_{pq}(\mathcal{M})} := \inf_{f=\sum_{\theta\in\Theta} a_\theta\phi_\theta} \Big(\sum_{m=0}^{\infty}\Big(\sum_{\theta\in\Theta_m}(|\theta|^{-\alpha}\|a_\theta\phi_\theta\|_p)^p\Big)^{q/p}\Big)^{1/q}, \quad (34)$$

where the infimum is taken over all representations $f = \sum_{\theta\in\Theta} a_\theta\phi_\theta$ in $L_p(E)$.

A third approach to the B-spaces B^α_{pq} is by using the decomposition via quasi-interpolants from (25). We define

$$\|f\|^Q_{B^\alpha_{pq}(\mathcal{M})} := \Big(\sum_{m=0}^{\infty}\Big(\sum_{\theta\in\Theta_m}(|\theta|^{-\alpha}\|b_{\theta,p}(f)\phi_\theta\|_p)^p\Big)^{q/p}\Big)^{1/q}. \quad (35)$$

Theorem 11. *For a given spline MRA $\mathcal{M}$ and admissible indices α, p, and q the the norms $\|\cdot\|_{B^\alpha_{pq}(\mathcal{M})}$, $\|\cdot\|^A_{B^\alpha_{pq}(\mathcal{M})}$, and $\|\cdot\|^Q_{B^\alpha_{pq}(\mathcal{M})}$, defined in* (33)-(35)*, are equivalent with constants of equivalence depending only on α, p, q, and the parameters of $\mathcal{T}$.*

The proof of this theorem is fairly simple and can be carried out as the proofs of the corresponding results in [6, 7, 16]; see also the more complicated proof of Theorem 2 in [20].

Remark. As was shown above the B-spaces B^α_{pq} are in essence sequence spaces and hence they can be interpolated by utilizing standard techniques. For some interpolation results on B-spaces, see [7].

In general the B-spaces are different from Besov spaces. However, if $\mathcal{T}$ is a regular triangulation of a compact polygonal domain E in $\mathbb{R}^2$, then the B-space $B^\alpha_{pq}(\mathcal{T})$ coincides with the Besov space $B^{2\alpha}_q(L_p(E))$ for $0 < \alpha < \alpha_0$ with $\alpha_0 > 0$ sufficiently small. For more details, see [7, 16].

Franklin basis decomposition of B-spaces. Suppose that the spline MRA $\mathcal{M}$ is generated by spaces $(\mathcal{S}_m)_{m\geq 0}$ consisting of piecewise linear and continuous functions induced by an LR-triangulation of E. In this case the B-spaces $B^\alpha_{pq}(\mathcal{M})$ can be characterized via representations using the corresponding Franklin basis [20]. Define

$$\|f\|^F_{B^\alpha_{pq}(\mathcal{M})} := \Big(\sum_{m=0}^{\infty}\Big(\sum_{\theta\in\Theta^*_m}(|\theta|^{-\alpha}\|c_\theta(f)f_\theta\|_p)^p\Big)^{q/p}\Big)^{1/q},$$

where $c_\theta(f) := \langle f, f_\theta\rangle$.

Theorem 12. *Suppose α, p, and q are admissible indices and let $\mathcal{T}$ be an LR-triangulation of a bounded polygonal domain $E \subset \mathbb{R}^2$. Then $f \in B^\alpha_{pq}(\mathcal{M})$ if and only if $\|f\|^F_{B^\alpha_{pq}(\mathcal{M})} < \infty$ and*

$$\|f\|^F_{B^\alpha_{pq}(\mathcal{M})} \approx \|f\|_{B^\alpha_{pq}(\mathcal{M})}.$$

The B-spaces of nonlinear n-term approximation. The primary application of B-spaces is to nonlinear n-term approximation in $L_p(E)$ $(0 < p \le \infty)$ from hierarchical sequence of spline bases $(\Phi_m)_{m\ge 0}$ associated with a spline MRA $\mathcal{M}$ on E (see §6 below). Denote briefly

$$B^\alpha_\tau(\mathcal{M}) := B^\alpha_{\tau\tau}(\mathcal{M})$$

with τ determined from $1/\tau = \alpha + 1/p$ according to two specific choices of p and α: (a) $0 < p < \infty$ and $\alpha > 0$; or (b) $p = \infty$ and $\alpha \ge 1$. In fact, the indices p, α, and τ are selected so that $B^\alpha_\tau(\mathcal{M})$ lies on the Sobolev embedding line; $B^\alpha_\tau(\mathcal{M})$ is embedded in $L_p(E)$. There are several other useful norms in $B^\alpha_\tau(\mathcal{M})$ apart from the norms coming from (33)-(35). For instance, for any $0 < \eta < p$,

$$\begin{aligned} \|f\|_{B^\alpha_\tau(\mathcal{M})} &:= |E|^{-\alpha}\|f\|_{L_\tau} + \Big(\sum_{\triangle\in\mathcal{T}}(|\triangle|^{-\alpha}\,\mathbb{S}_\triangle(f)_\tau)^\tau\Big)^{1/\tau} \\ &\approx |E|^{1/p-1/\eta}\|f\|_{L_\eta} + \Big(\sum_{\triangle\in\mathcal{T}}(|\triangle|^{1/p-1/\eta}\,\mathbb{S}_\triangle(f)_\eta)^\tau\Big)^{1/\tau} \end{aligned}$$

and also

$$\begin{aligned} \|f\|_{B^\alpha_\tau(\mathcal{M})} &\approx \Big(\sum_{\theta\in\Theta}(|\theta|^{-\alpha}\|b_{\theta,\tau}(f)\phi_\theta\|_\tau)^\tau\Big)^{1/\tau} \\ &\approx \Big(\sum_{\theta\in\Theta}(|\theta|^{1/p-1/\eta}\|b_{\theta,\eta}(f)\phi_\theta\|_\eta)^\tau\Big)^{1/\tau}. \end{aligned} \tag{36}$$

The point here is that normally $\tau < 1$ and the space L_τ is not very friendly in this case, while if $p > 1$ then η can be selected so that $1 < \eta < p$, which allows one to work in L_η instead of L_τ. For more details, see [7, 16]

§6. Best Nonlinear n-term Approximation

We now consider nonlinear n-term approximation from the scaling functions of a spline MRA. Suppose $\mathcal{M} := (\mathcal{S}_m)_{m\ge 0}$ is a spline MRA generated by a locally regular or better triangulation $\mathcal{T}$ of a compact polygonal domain E in $\mathbb{R}^2$. Let $(\Phi_m)_{m\ge 0}$ be the corresponding bases. (See §3.2.)

Let Σ_n denote the nonlinear set consisting of all functions g of the form

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

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

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

The primary goal now is to characterize the approximation spaces generated by $(\sigma_n(f)_p)$. To this end we shall use the machinery of Jackson-Bernstein estimates as explained in §2. As elsewhere, our standing assumption is that $0 < p \leq \infty$ and $\alpha \geq 1$ for $p = \infty$ and $\alpha > 0$ if $p < \infty$; in both cases we set $1/\tau := \alpha + 1/p$.

Theorem 13. **[Jackson estimate]** *If $f \in B_\tau^\alpha(\mathcal{M})$, then*

$$\sigma_n(f)_p \leq cn^{-\alpha}\|f\|_{B_\tau^\alpha(\mathcal{M})} \tag{37}$$

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

Estimate (37) follows from the basic estimates of the error of the "Push-the-Error" algorithm ($p = \infty$) and "Threshold" algorithm ($0 < p < \infty$), stated in Theorems 18 and 22 below.

Theorem 14. **[Bernstein estimate]** *If $g \in \Sigma_n$, then*

$$\|g\|_{B_\tau^\alpha(\mathcal{M})} \leq cn^\alpha\|g\|_p \tag{38}$$

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

For the proof of this theorem, see [7, 16, 17].

We denote the corresponding approximation space by $A_q^\gamma = A_q^\gamma(\Phi, L_p)$ (see (1)). The following characterization of the approximation spaces A_q^γ is immediate from estimates (37)-(38):

Theorem 15. *If $0 < \gamma < \alpha$ and $0 < q \leq \infty$, then*

$$A_q^\gamma(\Phi, L_p) = (L_p(E), B_\tau^\alpha(\mathcal{M}))_{\frac{\gamma}{\alpha},q}$$

with equivalent norms.

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

Theorem 16. *Assuming that $\alpha > 0$ if $p < \infty$ and $\alpha > 1$ if $p = \infty$, and $1/\tau := \alpha + 1/p$ in both cases, we have*

$$A_\tau^\alpha(\Phi, L_p) = B_\tau^\alpha(\mathcal{M}) \tag{39}$$

with equivalent norms.

For the proofs and more details, see [7, 16].

We next turn to a constructive realization of best n-term approximation.

§7. Algorithms for nonlinear n-term approximation

In this section we present practical algorithms for nonlinear n-term approximation which capture all rates of the best n-term approximation in L_p with $0 < p \leq \infty$. As elsewhere in this survey we assume that $\mathcal{M} := (\mathcal{S}_m)_{m \geq 0}$ is a spline MRA generated by a locally regular or better triangulation $\mathcal{T}$ of a compact polygonal domain E in $\mathbb{R}^2$ (see §3.1-2).

7.1. "Push-the-Error" Algorithm in L_∞

As will be shown in the next section simple threshold strategy allows for achieving the rates of best n-term approximation in L_p when $p < \infty$. The situation is quite different when approximating in the uniform norm. The "piling up" effect of multilevel structures is not well aligned with the L_∞-norm. Nevertheless, an efficient way of realizing optimal L_∞-approximation rates for approximation spaces induced by best n-term approximation is offered by another algorithmic paradigm, called "Push-the-Error" algorithm. It should be stressed that the "Push-the-Error" algorithm is, in principle, very flexible in that it essentially requires only refinability of single scale basis functions, i.e. has a potential to work under fairly general circumstances. The algorithm is presented here in the set-up of a spline MRA. The main idea is to complement thresholding strategies, i.e. keeping terms with large coefficients, with transferring small terms (pushing the error) to higher levels with the aid of refinement equations. This accounts for the fact that small terms may add up over different levels to form after all a significant contribution in the uniform norm because even the best multilevel bases are no longer able to properly separate the contributions from different length scales. The "Push-the-Error" algorithm has been developed in [17] and [6]; the essence of this algorithm originates from [10]. We present here the refined version of the algorithm developed in [6].

In [17] there is another algorithm (named "Trim & Cut") developed for nonlinear n-term approximation in L_p, $0 < p \leq \infty$. The idea of this algorithm originates in the proof of the Jackson estimate in [12]. A similar algorithm has been suggested by Yu. Brudnyi and I. Kozlov as well (see [2] and the references therein). The execution of the "Trim & Cut" algorithm relies heavily on a coloring procedure used to represent the set of all supports of basis functions as a disjoint union of trees with respect to the inclusion relation. This renders the scheme practically infeasible. Consequently, it is less valuable compared to the "Push-the-Error" algorithm.

7.1.1 Description of the Algorithm

For a given function $f \in C(E)$, we use the decomposition scheme from (25) with $1 < q < \infty$ to represent f in the form

$$f = \sum_{\theta \in \Theta} b_\theta(f)\phi_\theta = \sum_{m \geq 0} \sum_{\theta \in \Theta_m} b_\theta(f)\phi_\theta, \tag{40}$$

where the coefficients $b_\theta := b_\theta(f)$ depend linearly on f and the series converges uniformly on E. Whenever f has a wavelet expansion (see (18)), we rewrite the wavelets in terms of scaling functions to obtain (40).

For the purpose of designing an algorithm capable of achieving the rates of the best n-term approximation from $\Phi = \{\phi_\theta\}_{\theta \in \Theta}$ in the uniform norm, the initial decomposition (40) must provide an efficient representation of f. In our case this means that the terms in (40) should characterize the norm in $B^\alpha_\tau(\mathcal{M})$, $\alpha \geq 1$, $\tau := 1/\alpha$, as in (36) which can be achieved by employing simple projectors onto the spaces $(\mathcal{S}_m)$.

To describe the "Push-the-Error" algorithm we need a few preliminaries. For any $\eta, \theta \in \Theta$ with $l(\eta) > l(\theta)$, we say that η is *connected* with θ via sets from Θ if there exists a sequence of elements $\eta =: \eta_0, \eta_1, \ldots, \eta_k := \theta$ with $k := l(\eta) - l(\theta)$ such that

(i) $l(\eta_i) = l(\eta_{i+1}) + 1, \quad i = 0, \ldots, k-1;$

(ii) η_i sits on η_{i+1}, $i = 0, \ldots, k-1$, i.e. $\eta_i^\circ \cap \eta_{i+1}^\circ \neq \emptyset$.

Here $l(\theta)$ denotes the level of θ ($l(\theta) = m$ if $\theta \in \Theta_m$).

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

$$\begin{aligned} \mathcal{U}'_\theta &:= \{\eta \in \Theta : l(\eta) > l(\theta),\ \eta \text{ is connected with } \theta\} \quad \text{and} \\ \mathcal{U}_\theta &:= \mathcal{U}'_\theta \cup \{\theta\}. \end{aligned}$$

Note that $\eta \in \mathcal{U}_\theta$ implies $\mathcal{U}_\eta \subseteq \mathcal{U}_\theta$, and hence there is a constant N_* such that

$$\eta \in \mathcal{U}_\theta \quad \Longrightarrow \quad \eta \subset \operatorname{Star}^{N_*}_m(\theta), \quad m := l(\theta). \tag{41}$$

The following local error terms will play a critical role in the algorithm:

$$E(f,\theta) = E(\theta) := |b_\theta(f)| + \Big\| \sum_{\eta \in \mathcal{U}'_\theta} b_\eta(f)\phi_\eta \Big\|_\infty. \tag{42}$$

By the properties of the bases functions, there exists a constant $\nu_* \geq \ell$ such that for each $\theta \in \Theta_m$, $m \geq 0$,

$$\theta \in \operatorname{Star}^{\nu_*}_m(x) \quad \text{for} \quad x \in \theta.$$

Then for each $\theta \in \Theta_m$, we define its *"concrete"* Ω_θ by

$$\Omega_\theta := \operatorname{Star}^{N_* + 4\nu_*}_m(\theta),$$

where N_* is from (41).

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

$$\mathcal{X}_\theta := \{\eta \in \Theta_m : \eta^\circ \cap \Omega_\theta^\circ \neq \emptyset\} \quad \text{with} \quad m := l(\theta). \tag{43}$$

We call the elements of $\mathcal{X}_\theta$ the *neighbors* of θ.

We now describe conceptually the "Push-the-error" algorithm; corresponding practical ramifications will be discussed later on.

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

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

by the following steps:

Step 1. [*Decomposition*] We represent f in the form (40).

Step 2. [*"Prune the shrubs"*] We discard all terms $b_\theta\phi_\theta$ such that

$$E(f,\eta) \le \varepsilon, \quad \forall\, \eta \in \mathcal{U}_\theta. \tag{44}$$

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

$$f_\Gamma := \sum_{\theta \in \Gamma} b_\theta \phi_\theta.$$

Evidently, there exists some $M \ge 0$ such that

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

i.e. Γ is a finite set.

Step 3. [*"Push the error"*] Let $\tilde{\Lambda}_0$ be the set of all $\theta \in \Theta_0 \cap \Gamma$ such that $|b_\theta(f)| > \varepsilon$ and set $\Lambda_0 := (\cup_{\theta \in \tilde{\Lambda}_0} \mathcal{X}_\theta) \cap \Gamma$. We define

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

Using the refinement equations (17), we represent (rewrite) each of the remaining terms $b_\theta\phi_\theta$, $\theta \in (\Theta_0 \cap \Gamma) \setminus \Lambda_0$, as a linear combination of $\{\phi_\eta\}_{\eta \in \Theta_1}$ and add to the resulting terms the existing terms $b_\theta\phi_\theta$, $\theta \in \Theta_1 \cap \Gamma$. As a result we obtain a representation of f_Γ in the form

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

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

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

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

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

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

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

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

7.1.2 Error Analysis of "Push-the-Error"

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

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

and

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

The main conceptual tool in the error analysis is the following weak quasi-subadditivity of the counting functional $N(\varepsilon)$.

Theorem 17. *There exist constants c_* and $\tilde{c}$ depending only on the parameters of the MRA such that if $f = f_0 + f_1$, $f_j \in C(E)$, and the "Push-the-Error" algorithm is applied to f_j with $\varepsilon_j > 0$ $(j = 0, 1)$ and to f with $\varepsilon := c_*(\varepsilon_0 + \varepsilon_1)$, then*

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

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

Theorem 18. *If $f \in B^\alpha_\tau(\mathcal{M})$, $\alpha \geq 1$, $\tau := 1/\alpha$, then for each $\varepsilon > 0$*

$$A_{N(\varepsilon)}(f) \leq c\varepsilon \quad \text{and} \quad N(\varepsilon) \leq c\varepsilon^{-\tau} \|f\|^\tau_{B^\alpha_\tau(\mathcal{M})}$$

and, therefore,

$$A_n(f) \leq cn^{-\alpha} \|f\|_{B^\alpha_\tau(\mathcal{M})}, \quad n = 1, 2, \ldots.$$

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

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

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

We can now address the program outlined in §2. Let us denote by $K(f,t)_\infty$ the K-functional generated by the spaces $C(E)$ and $B^\alpha_\tau(\mathcal{M})$ with $\tau := 1/\alpha$.

Theorem 19. *Suppose that $f \in C(E)$ and $\alpha \geq 1$. Then one has*

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

and, therefore,

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

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

The following result is an immediate consequence of Theorem 19:

Theorem 20. *For $f \in C(E)$ and $\gamma > 0$, $A_n(f) = O(n^{-\gamma})$ if and only if $\sigma_n(f)_\infty = O(n^{-\gamma})$.*

More generally, let $A^\gamma_q(\sigma) = A^\gamma_q(L_\infty, \sigma)$ be the approximation spaces generated by the best nonlinear n-term approximation from the scaling functions of the spline MRA $\mathcal{M}$. Let $A^\gamma_q(\mathcal{A})$ be the set of all functions $f \in C(E)$ such that

$$\|f\|_{A^\gamma_q(\mathcal{A})} := \|f\|_\infty + \left(\sum_{n=1}^{\infty} (n^\gamma A_n(f))^q \frac{1}{n} \right)^{1/q} < \infty \tag{45}$$

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

Theorem 19 yields the following more general result:

Theorem 21. *For any $\gamma > 0$ and $0 < q \leq \infty$, we have $A^\gamma_q(\mathcal{A}) = A^\gamma_q(\sigma)$ and $\|f\|_{A^\gamma_q(\mathcal{A})} \approx \|f\|_{A^\gamma_q(\sigma)}$ for $f \in A^\gamma_q(\mathcal{A}) = A^\gamma_q(\sigma)$.*

We want to point out here that the results from Theorems 20-21 hold for all $0 < \gamma < \infty$ and this is the meaning of "the algorithm captures all rates of best n-term approximation". We refer the reader to [6] for the proofs of the above theorem and further details.

7.1.3 Practical aspects of "Push-the-Error"

Complexity. Assume now that the function f (a surface or multidimensional data) has an initial representation (approximation) in some "finest" space $\mathcal{S}_M$ of a spline MRA involving $O(N)$ terms. Let us assume that the "Push the error" algorithm (as described in §7.1.1) is applied to this f. The decomposition Step 1 of "Push-the-Error" will run in $O(N)$ flops. Step 2 ["Prune the shrubs"] of the algorithm can evidently be realized in $O(N \log N)$ flops by rewriting all terms of interest at the finest level. Step 3 ["Push the error"] works in $O(N)$ flops. The reconstruction Step 4 runs also in $O(N)$ flops. Therefore, the "Push-the-Error" algorithm appears to be an attractive approximation scheme from practical point of view. We next present a more economical version of the second step of the algorithm.

Scalable second version of Step 2 ["Prune the shrubs"]. We define a new local error term $\widetilde{E}(f,\theta)$ by

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

Now, the condition $E(f,\eta) \le \varepsilon$ in (44) is replaced by the condition $\widetilde{E}(f,\eta) \le \varepsilon$ (see (42)) which is practically easier to be verified. The new version of Step 2 of the algorithm can be realized in $O(N)$ flops by employing a well-know principle of Dynamic Programming. One uses the coefficient $\{b_\theta(f)\}$ obtained in Step 1 to compute

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

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

It is easy to see that for this new version of "Push-the-Error" Theorem 18 remains valid. However, it is impossible to establish Theorem 17 in this case, which makes this version less attractive from a theoretical point of view.

Further observations and practical modifications. As already mentioned in the beginning of §7.1.1, for an optimal performance of the "Push-the-Error" algorithm it is important to have an initial sparse representation of the function f being approximated. To this end the dual functionals $\{a_\theta(\cdot)\}$ should be bounded in L_q for some $q > 1$. In turn, this means that decomposition methods based on interpolatory schemes do not provide efficient representations and should be avoided.

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

Finally, one can run the "Push-the-Error" algorithm without executing Step 2 at all. An algorithm consisting of only Step 1 and Step 2 is also reasonable in some situations. Other modifications are also possible. However, one should be aware of the existence of several traps which may defeat such modifications of the algorithms (see [17]).

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

Here we show that the usual threshold scheme used in nonlinear n-term approximation from wavelets in L_p $(1 < p < \infty)$ can be successfully utilized for n-term approximation from the scaling functions of spline MRA in L_p $(0 < p < \infty)$ (see [6, 17]).

We begin with a description of the algorithm.

Step 1. [*Decomposition*] We represent the function f being approximated by using the decomposition (25) with $0 < q < p$. As a result we have $f = \sum_{\theta \in \Theta} b_\theta(f)\phi_\theta$ in $L_p(E)$.

Step 2. [*"Threshold"*] We first order the terms $\{b_\theta \phi_\theta\}_{\theta\in\Theta}$ in a sequence $(b_{\theta_j}\phi_{\theta_j})_{j\ge 1}$ so that

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

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

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

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

As elsewhere we assume that $\alpha > 0$, $0 < p < \infty$, and $\tau := (\alpha + 1/p)^{-1}$. The following theorem is an immediate consequence of Theorem 4.

Theorem 22. *If $f \in \mathcal{B}_\tau^\alpha(\mathcal{M})$, then*

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

Furthermore,

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

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

We next explain in what sense the "Threshold" algorithm captures the rates of the best nonlinear n-term approximation in L_p, $0 < p < \infty$. Denote by $A_\tau^\alpha(\sigma, L_p) := A_\tau^\alpha(\Phi, L_p)$ the approximation space generated by the best n-term approximation from Φ in L_p and let $A_q^\gamma(\mathcal{A}^T, L_p)$ be the set of all functions $f \in L_p(E)$ such that

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

with the usual modification when $q = \infty$ (see also (45)).

Theorem 23. *For any $\alpha > 0$ and $1/\tau = \alpha + 1/p$, we have $A_\tau^\alpha(\mathcal{A}^T, L_p) = B_\tau^\alpha(\mathcal{M}) = A_\tau^\alpha(\sigma, L_p)$ and for each f in this space*

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

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

As for the "Push-the-Error" algorithm, it is critical to have an efficient initial decomposition of the function f being approximated, i.e. the representation should provide a decomposition of the norm in $B_\tau^\alpha(\mathcal{M})$, $1/\tau = \alpha + 1/p$. For the "Threshold" algorithm this is guaranteed by employing the decompositions from (25) with $q < p$.

The estimate $A_n^T(f)_p \leq c\|f\|_p$ fails to be true in general (even if $1 < p < \infty$) since the convergence in the representation of the function f being approximated that is used (see (25)) is not assumed to be unconditional. (This problem does not arise in the case when wavelets exist.) This is why the result from Theorem 23 is somewhat weaker than the result from Theorem 21.

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

§8. n-term rational approximation

The univariate rational approximation on $\mathbb{R}$ is a relatively well developed area in Approximation theory (see, e.g. [25]) . At the same time, the multivariate rational approximation is virtually not existing yet. A reason for this is that it is extremely hard to deal with rational functions of the form $R = P/Q$, where P and Q are algebraic polynomials in d variables ($d > 1$). Very little is known about this type of rational functions. It seems natural to consider nonlinear n-term approximation from the dictionary $\mathcal{R}$ consisting of all functions on $\mathbb{R}^d$ of the form

$$R = \sum_{j=1}^{n} r_j \tag{46}$$

where r_j are partial fractions. In [23], it is considered the case when the r_j's are of the form $r(x) = \prod_{\mu=1}^{d} \frac{a_\mu x_\mu + b_\mu}{(x_\mu - \alpha)^2 + \beta_\mu^2}$. The main estimate from [23] relates the n-term rational approximation with nonlinear piecewise polynomial approximation over an arbitrary dyadic partition of $\mathbb{R}^d$. As a consequence a direct estimate is obtained for n-term rational approximation in terms of a minimal B-norm (over all dyadic partitions).

K. Park [22] obtained similar results for the more complicated case of n-term rational approximation in $\mathbb{R}^2$, when the r_j's are of the form

$$r_i = \prod_{\mu=1}^{6} \frac{a_\mu x_1 + b_\mu x_2 + c_\mu}{1 + (\alpha_\mu x_1 + \beta_\mu x_2 + \gamma_\mu)^2} \tag{47}$$

with $a_\mu, b_\mu, c_\mu, \alpha_\mu, \beta_\mu, \gamma_\mu \in \mathbb{R}$.

Results of the same character are obtained also by S. Dekel and D. Leviatan [8] with no restrictions on the triangulations when approximating in L_p with $0 < p \le 1$ and under the condition that the piecewise polynomials are over triangulations satisfying the minimal angle condition (regular triangulations, see §3.1) when $1 < p < \infty$.

We next give a brief description of the results in [22]. Let $\mathcal{R}_n$ be the set of all n-term rational functions on $\mathbb{R}^2$ of the form (46) with r_i from (47). Denote by $R_n(f)_p$ the error of L_p-approximation to f from $\mathcal{R}_n$:

$$R_n(f)_p := \inf_{R \in \mathcal{R}_n} \|f - R\|_p.$$

Clearly, each $R \in \mathcal{R}_n$ depends on $\le 36n$ parameters and $\mathcal{R}_n$ is a nonlinear set, however, $c\mathcal{R}_n = \mathcal{R}_n$ $(c \ne 0)$ and $\mathcal{R}_n + \mathcal{R}_m = \mathcal{R}_{n+m}$. A fundamental property of $\mathcal{R}_n$ is that it is invariant under affine transforms, i.e. if $R \in \mathcal{R}_n$, then $R \circ A \in \mathcal{R}_n$ for every affine transform A.

Assume that $\mathcal{T}$ is an SLR-triangulation on $\mathbb{R}^2$ (see [16]). Let $\Sigma_n^k(\mathcal{T})$, $k \ge 1$, denote the set of all n-term piecewise polynomial function of the form

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

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

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

The following theorem contains the main result from [22].

Theorem 24. *Let $f \in L_p(\mathbb{R}^2)$, $0 < p < \infty$, $\alpha > 0$, and $k \geq 1$. Then*

$$R_n(f)_p \leq cn^{-\alpha}\Big(\sum_{m=1}^{n} \frac{1}{m}(m^\alpha \sigma_m(f,\mathcal{T})_p)^{p^*} + \|f\|_p^{p^*}\Big)^{1/p^*}, \quad n = 1, 2, \dots,$$

where $p^* = \min\{1, p\}$ and c depends only on α, p, k, and the parameters of $\mathcal{T}$.

The proof of this theorem is based on Newman's famous result on rational approximation of $|x|$ on $[-1, 1]$ and estimates for certain maximal functions, and in particular, estimate (13).

It is an important observation that in Theorem 24 there is no restriction on $\alpha > 0$ (but c depends on α). The next corollary follows immediately from the above theorem.

Corollary 1. *If $\sigma_n(f, \mathcal{T})_p = O(n^{-\gamma})$ for an arbitrary SLR-triangulation $\mathcal{T}$, $0 < p < \infty$, and $\gamma > 0$, then $R_n(f)_p = O(n^{-\gamma})$.*

Combining the Jackson estimate for n-term piecewise polynomial approximation from Theorem 13 with Theorem 24, gives the following result.

Corollary 2. *If $f \in \bigcap_{\mathcal{T}} B^\alpha_\tau(\mathcal{M}_\mathcal{T})$, where $\alpha > 0$, $1/\tau := \alpha + 1/p$, $0 < p < \infty$, then*

$$R_n(f)_p \leq cn^{-\alpha} \inf_{\mathcal{T}} \|f\|_{B^\alpha_\tau(\mathcal{M}_\mathcal{T})},$$

where the infimum is taken over all SLR-triangulations $\mathcal{T}$ and associated spline MRAs $\mathcal{M}_\mathcal{T}$ with some fixed parameters.

§9. References

1. J. Bergh, J. Löfström, *Interpolation spaces: An introduction*, Grundlehren der Mathematischen Wissenschaften, No. 223. Springer-Verlag, Berlin-New York, 1976.
2. Yu. Brudnyi, Nonlinear approximation by refinable functions and piecewise polynomials, in *Approximation Theory X: Abstract and Classical Analysis*, C. K. Chui, L. L. Schumaker, and J. Stöckler (eds.), Vanderbilt University Press, Nashville, 2002, pp. 91–112.
3. Z. Cieselski, A. Kamont, *Survey on the orthogonal Franklin system*, in Approximation Theory, a volume dedicated to Bl. Sendov, B. Bojanov (ed.), DARPA, Sofia, 2002, pp. 84–132.

4. R. Coifman, G. Weiss, Analyse harmonique non-commutative sur certains espaces homogenes. Etude de certaines integrales singulieres, *Lecture Notes in Mathematics*, Vol. 242. Springer-Verlag, Berlin-New York, 1971.
5. R. Coifman, G. Weiss, Extensions of Hardy spaces and their use in analysis, Bull. Amer. Math. Soc. **83** (1977), 569–645.
6. W. Dahmen and P. Petrushev, "Push-the-Error" algorithm for nonlinear n-term approximation, 2004, preprint. `http://www.math.sc.edu/~pencho/`
7. O. Davydov, P. Petrushev, Nonlinear approximation from differentiable piecewise polynomials, SIAM J. Math. Anal. **35** (2004), 708–758.
8. S. Dekel, D. Leviatan, On the relation between piecewise polynomial and rational approximation, preprint.
9. R. DeVore, Nonlinear approximation, Acta Numerica No 7, Cambridge University Press, 1998, 51–150.
10. R. DeVore, B. Jawerth, and B. Lucier, Surface compression, Comput. Aided Geom. Design **9** (1992), 219–239.
11. R. DeVore, G.G. Lorentz, *Constructive Approximation*, Springer Grundlehren Vol. 303, Heidelberg, 1993.
12. R. DeVore, P. Petrushev, and X. Yu, Nonlinear wavelet approximation in the space $C(R^d)$, *Progress in Approximation Theory* (A. A. Gonchar, E. B. Saff, eds.), New York, Springer-Verlag, 1992, pp. 261–283.
13. R. DeVore, V. Popov, Interpolation of Besov spaces, Trans. Amer. Math. Soc. **305** (1988), 397–414.
14. C. Fefferman and E. Stein, Some maximal inequalities, Amer. J. Math. **93** (1971), 107–115.
15. A. Jonsson, A. Kamont, Piecewise linear bases and Besov spaces on fractal sets, Anal. Math., 27 (2001), 77–117.
16. B. Karaivanov, P. Petrushev, Nonlinear piecewise polynomial approximation beyond Besov spaces, Appl. Comput. Harmonic Anal. **15** (2003), 177–223.
17. B. Karaivanov, P. Petrushev, and R.C. Sharpley, Algorithms for nonlinear piecewise polynomial approximation, Trans. Amer. Math. Soc. **355** (2003), 2585–2631.
18. B. Kashin, A. Saakian, *Orthogonal series*, Nauka, Moscow, 1984 (in Russian), also English translation Amer. Math. Soc., 1989.
19. G. Kyriazis, K. Park, and P. Petrushev, Anisotropic Franklin bases on polygonal domains, 2004, preprint. `http://www.math.sc.edu/~pencho/`

20. G. Kyriazis, K. Park, and P. Petrushev, B-spaces and their characterization via anisotropic Franklin bases, 2004, preprint.
`http://www.math.sc.edu/~pencho/`

21. P. Oswald, *Multilevel Finite Element Approximation*, Teubner, Stuttgart, 1994.

22. K. Park, Multivariate n-term rational approximation, preprint, 2004.

23. P. Petrushev, Multivariate n-term rational and piecewise polynomial approximation, J. Approx. Theory **121** (2003), 158–197.

24. P. Petrushev, Nonlinear n-term approximation from hierarchical spline bases, *Constructive Theory of Functions* , *Varna 2002* (*B. Bojanov, Ed.*), DARBA, Sofia, 2003, pp. 33–85.

25. P. Petrushev, V. Popov, *Rational approximation of real functions*, Cambridge University Press, 1987.

Pencho Petrushev
University of South Carolina
Columbia, USA
`pencho@math.sc.edu`
`http://www.math.sc.edu/~pencho/`