Anisotropic Smoothness Spaces via Level Sets

RONALD DEVORE
University of South Carolina

GUERGANA PETROVA
Texas A&M University

AND

PRZEMYSŁAW WOJTASZCZYK
University of Warsaw

Abstract

It has been understood for sometime that the classical smoothness spaces, such as the Sobolev and Besov classes, are not satisfactory for certain problems in image processing and nonlinear PDEs. Their deficiency lies in their isotropy. Functions in these smoothness spaces must be simultaneously smooth in all directions. The anisotropic generalizations of these spaces also have the deficiency that they are biased in coordinate directions. While they allow different smoothness in certain directions, these directions must be aligned to the coordinate axes. In the application areas mentioned above, it would be desirable to measure smoothness in new ways which would allow one to have more local control over the smoothness directions. We introduce one possible approach to this problem based on defining smoothness via level sets. We present this approach in the case of functions defined on $\mathbb{R}^d$. Our smoothness spaces depend on two smoothness indices (s_1, s_2). The first reflects the smoothness of the level sets of the function, while the second index reflects how smoothly the level sets themselves are changing. As a motivation, we start with $d = 2$ and investigate Besov smooth domains.

1 Introduction

The classical smoothness spaces, such as Sobolev and Besov spaces, have been an important staple in mathematical analysis for more than a half century. They serve to classify the regularity of a function in a way that is useful in describing when a function can be efficiently approximated, or computed, or compressed by certain frequently used methods such as splines or wavelets. They are also used to specify regularity of solutions to partial differential equations. However, it has been recognized for some time that they are not satisfactory in classifying functions in certain application areas. We mention two of these areas.

THIS RESEARCH WAS COMPLETED WHILE R.D. WAS A VISITING PROFESSOR AT THE COURANT INSTITUTE

Communications on Pure and Applied Mathematics, Vol. 000, 0001–0032 (2000)

In image compression, Candes and Donoho [1] have pointed out that certain very simple functions, known as cartoon images, are wrongly classified by the Sobolev-Besov classification. By this we mean that the index of smoothness of a function in these spaces does not accurately reflect its compressibility. A cartoon image is the characteristic function $f = \chi_\Omega$ of a domain $\Omega \subset \mathbb{R}^2$. Suppose that this domain is the interior of a simple closed C^2 curve. Then, the Sobolev-Besov smoothness index of f is at most one. This smoothness index will determine the approximation rate of f by classical linear and nonlinear methods using say splines or piecewise polynomials. For example, if we approximate f in L_2 by such approximation methods using n parameters (or n bits in the case of encoding) then we will only achieve the distortion rate $\mathcal{O}(n^{-1/2})$, where the 2 reflects the curse of dimensionality. On the other hand, we can approximate such a characteristic function by simply approximating the domain Ω by a second domain Ω_0 with distortion measured by the area of the symmetric difference between Ω and Ω_0. Then χ_{Ω_0} is the approximation to χ_Ω. For example, taking Ω_0 as an appropriate polygonal domain with a piecewise linear boundary consisting of n pieces we will achieve L_2 distortion between χ_Ω and χ_{Ω_0} bounded by Cn^{-1}. In other words, the Besov-Sobolev smoothness index predicts the wrong compressibility of such functions f. Stated in another way, the classical approximation methods are not efficient in approximating such functions. This observation has led to the introduction of alternative approximation methods using generalizations of wavelets known as curvelets.

Consider next the numerical approximation of the solution to certain nonlinear evolution equations known as conservation laws. A scalar conservation law in $\mathbb{R}^d$ is given by the initial value problem

$$u_t + \nabla_x \cdot f(u) = 0, \quad u(x,0) = u_0(x), \tag{1.1}$$

for $(x,t) \in \mathbb{R}^d \times \mathbb{R}_+$ and flux $f = (f_1, \dots, f_d)$. Even when the initial condition u_0 is in C^∞ and the flux f has arbitrarily prescribed differentiability, the solution u will develop shock discontinuities in finite time. The question arises as to how efficiently one can compute an approximation to u in the sense of L_1 at a later time $t > 0$. This question is usually tackled by means of regularity theorems. A regularity theorem guarantees that the solution $u(\cdot,t)$ will have a certain smoothness at a later time $t > 0$, provided the initial condition is suitably smooth. The classical result here is that $u(\cdot,t)$ is in BV whenever $u_0 \in BV$. The BV regularity theorem would only guarantee that the solution at later time $t > 0$ could be approximated to accuracy $\mathcal{O}(n^{-1/d})$ in L_1 by classical approximation methods using n parameters such as piecewise polynomials on uniform partitions. However in the case $d = 1$, it was shown in [7] that whenever $u_0 \in C^\infty$, the solution has arbitrary high regularity in a certain class of Besov spaces and more importantly that this implies that the solution could be approximated to accuracy $\mathcal{O}(n^{-s})$ for all $s > 0$ by piecewise polynomials on adaptive partitions. In other words, in the case $d = 1$, shock formation is not detrimental to numerical efficiency. However, no corresponding results are known in higher space dimensions $d > 1$. Indeed, it is known that the solution has limited Besov regularity for the same essential reason as was the case in the image processing example. The question arises whether there are higher smoothness results if smoothness is measured in other ways and perhaps more importantly, can the solution be computed to higher order accuracy by new approximation methods. One such method that is based on this philosophy is front tracking (see for example Chapter 11 of [2]).

These two examples motivate our search for new ways to measure smoothness which would treat cartoon images and shock discontinuities more favorably than the Sobolev-Besov scales. We shall introduce a new family of smoothness spaces based on level sets which measure smoothness in a new way. These spaces will judge a function based on two criteria. The first is how smooth are its level sets. The second criteria will be how smoothly the level sets evolve as the level changes.

The main results of this paper, given in Theorems 6.1 and 7.2, show that functions in the spaces we introduce can be approximated by functions whose level sets have a simple description in terms of piecewise algebraic polynomials.

The first part of this paper will provide motivation for the definitions and theorems that follow. There, we introduce a new class of domains $\Omega \subset \mathbb{R}^2$, which we call Besov domains because of their analogy to Lipschitz domains, that are geometrically very natural, but may not be (as shown in Example 4.1 and Example 4.2) level sets of functions in $\mathbb{R}^2$. We show that these Besov domains can be approximated with a certain accuracy by piecewise polynomial domains. This motivates us to view the level sets of functions in $\mathbb{R}^d$, $d \geq 2$, as sets that can be approximated well by piecewise polynomial domains (as described in §4).

2 Besov spaces and approximation

Our first goal is to define certain *Besov domains* whose boundary can be described by Besov regularity. We begin by recalling properties of univariate Besov classes and then proceed to defining the Besov domains and deriving some of their properties.

Consider a function ϕ defined on an interval $I := [a,b]$. We define the Besov space $B^s_\tau(L_\tau(I))$ as in [6] (see §10 of Chapter 2). These Besov spaces can also be defined by wavelet decompositions. The wavelet approach gives the same spaces with an equivalent norm provided the Besov space is compactly embedded in L_1 which we shall always assume. Analogues of the Sobolev embedding theorem say that the Besov space $B^s_\tau(L_\tau(I))$ is compactly embedded in $L_p(I)$ provided $1/\tau < s + 1/p$. We identify the Besov space $B^s_\tau(L_\tau(I))$ with the point $(1/\tau, s)$ in the upper right quadrant of the plane. The line segment in the upper right quadrant of the plane which consists of all pairs $(1/\tau, s)$ such that $1/\tau = s + 1/p$ is called the Sobolev embedding line or sometimes the critical line. Thus, we see that the Besov space is compactly embedded in $L_p(I)$ if its parameter point $(1/\tau, s)$ lies above the critical line.

We shall use some well-known embedding inequalities. To describe them, we define $\mathbf{P}_r$ as the class of all univariate algebraic polynomials of degree at most $r-1$ (order at most r). We are interested in how well we can approximate ϕ by such polynomials which leads us to defining

$$E_r(\phi, L_p(I)) := \inf_{P \in \mathbf{P}_r} \|\phi - P\|_{L_p(I)}, \tag{2.1}$$

which is the error in approximating ϕ in the $L_p(I)$ norm by polynomials of order r. If $\phi \in B^s_\tau(L_\tau(I))$ with parameters above the critical line and if $s < r$, then (see Chapter 12 of [6])

$$E_r(\phi, L_p(I)) \le C_r |I|^\delta |\phi|_{B^s_\tau(L_\tau(I))}, \quad \delta := s + 1/p - 1/\tau, \tag{2.2}$$

where C_r is a constant depending only on r. The parameter δ measures how far $(1/\tau, s)$ is above the critical line. Actually the inequality (2.2) easily follows from the embedding inequality

$$E_r(\phi, L_p([0,1])) := \inf_{P \in \mathbf{P}_r} \|\phi - P\|_{L_p([0,1])} \le C_r |\phi|_{B^s_\tau(L_\tau([0,1]))}$$

by a change of variables.

The embedding inequalities (2.2) are the basis for proving theorems about non-linear approximation. We recall one class of results related to approximation by piecewise polynomials. We define $\Sigma_n := \Sigma_n^r$ to be the set of all functions S defined on I which are a piecewise polynomial with at most n pieces and whose components are polynomials of order at most r. We define the error in approximating ϕ by the elements of Σ_n^r by

$$\sigma_n(\phi, L_p(I)) := \sigma_{n,r}(\phi, L_p(I)) := \inf_{S \in \Sigma_n^r} \|\phi - S\|_{L_p(I)}, \quad n \ge 1. \tag{2.3}$$

Using (2.2), one has (a proof is given below) that whenever $s < r$,

$$\sigma_{n,r}(\phi, L_p(I)) \le C'(r, \delta) |I|^\delta |\phi|_{B^s_\tau(L_\tau(I))} n^{-s}, \quad n \ge 1. \tag{2.4}$$

Here it is important to note that the constant $C'(r,\delta)$ depends only on r and δ.

Next, notice that there is even a numerically friendly way to generate good approximations from Σ_n^r to a given ϕ through adaptive refinement. For this one uses the local error $E(J) := E_r(\phi, L_p(J))$. We start with the interval I and a given error tolerance $\varepsilon > 0$. If $E(I) > \varepsilon$ then we divide I into its two children by halving I. At a general step, given an interval J, we subdivide J into its two children whenever $E(J) > \varepsilon$; otherwise we leave J alone. We denote by $\mathscr{P}_\varepsilon$ the final partition we obtain in this way. We recall the following lemma which has been essentially proved but not stated in [4]. For completeness, we sketch its proof.

Lemma 2.1. *Fix r and let $s < r$. If $\phi \in B^s_\tau(L_\tau(I))$ with the parameter point $(1/\tau, s)$ above the critical line, then for any $\varepsilon > 0$, the final partition $\mathscr{P}_\varepsilon$ given by the above adaptive procedure satisfies*

$$\#(\mathscr{P}_\varepsilon) \le C(r,\delta) \left[|I|^\delta \varepsilon^{-1} |\phi|_{B^s_\tau(L_\tau(I))} \right]^{\frac{1}{s+1/p}}. \tag{2.5}$$

Proof: For each $k \geq 0$, we count the number N_k of dyadic intervals $J \in \mathscr{P}_\varepsilon$ at level k (this means J is dyadic with respect to I and $|J| = 2^{-k}|I|$). We claim that

$$N_k \leq \min\left\{2^k, \left[C_r 2^{s+1/p} 2^{-k\delta}|I|^\delta \varepsilon^{-1}|\phi|_{B^s_\tau(L_\tau(I))}\right]^\tau\right\}. \tag{2.6}$$

This inequality is obvious if the minimum is 2^k. If the minimum is the second term, then we note that the intervals J at dyadic level k are disjoint. Let $\bar{J}$ denote the parent of J. Then for each J at level k in our final partition, we have from (2.2) that

$$\varepsilon^\tau < E_r(\phi, L_p(\bar{J}))^\tau \leq C_r^\tau |\bar{J}|^{\delta\tau}|\phi|^\tau_{B^s_\tau(L_\tau(\bar{J}))}. \tag{2.7}$$

If we add these inequalities over all $J \in \mathscr{P}_\varepsilon$ with $|J| = 2^{-k}|I|$ and use the fact that $|\phi|^\tau_{B^s_\tau(L_\tau(\bar{J}))}$ is set sub-additive and that each $\bar{J}$ is associated to at most two J, we arrive at the inequality (2.6).

To finish the proof of the lemma, we have

$$\#(\mathscr{P}_\varepsilon) \leq \sum_{k=0}^\infty N_k \leq \sum_{k=0}^K 2^k + 2^{\tau(s+1/p)} C_r^\tau |I|^{\delta\tau}|\phi|^\tau_{B^s_\tau(L_\tau(I))}\varepsilon^{-\tau}\sum_{k=K}^\infty 2^{-k\delta\tau}, \tag{2.8}$$

where K is the largest integer where the minimum is taken by 2^k. It follows that $2^K \approx \left[2^{s+1/p} C_r |I|^\delta |\phi|_{B^s_\tau(L_\tau(I))} 2^{-K\delta}\varepsilon^{-1}\right]^\tau$. Since the sums are geometric we obtain $\#(\mathscr{P}_\varepsilon) \leq C2^K$. Solving for K, we arrive at (2.5). □

Let us note that if we are given an $n \geq 1$, then choosing ε to satisfy the equation

$$n = C(r,\delta)\left[|I|^\delta \varepsilon^{-1}|\phi|_{B^s_\tau(L_\tau(I))}\right]^{\frac{1}{s+1/p}},$$

we get a partition with at most n elements. The $L_p(I)$ error in approximating ϕ by the piecewise polynomial S which consists of the best polynomial approximation to ϕ on each interval of this partition satisfies

$$\|\phi - S\|_{L_p(I)} \leq n^{1/p}\varepsilon \leq C'(r,\delta)|I|^\delta |\phi|_{B^s_\tau(L_\tau(I))} n^{-s}. \tag{2.9}$$

In other words, we obtain the estimate (2.4) through adaptive approximation.

One final note about piecewise polynomial approximation. We can obtain the same estimates as described above with continuous piecewise polynomials. For example, suppose we have a partition and a piecewise polynomial $S \in \Sigma_n^r$ which is generated by the adaptive procedure and achieves the error estimate (2.4) but S is possibly discontinuous at some of its break points. By adding at most $(n-1)$ additional break points and linear segments we can modify the partition and S and obtain the same estimates with a continuous approximant.

3 Besov domains in $\mathbb{R}^2$

In order to motivate this section, we recall that our goal is to define smoothness classes via level sets. If a function f is continuous on $\mathbb{R}^d$, then its level sets are open subsets of $\mathbb{R}^d$. One of the ingredients we shall use in our definition of smoothness spaces is a quantitative measure of the smoothness of a set Ω. Ultimately our definition will be made by how well Ω can be approximated by very simple sets which are open and have piecewise polynomial boundaries. However, it is useful to have an idea of which sets Ω have this property. Therefore, we begin this section by defining a class of bounded open sets in $\mathbb{R}^2$ which have a certain smoothness for their boundary and end by showing that these domains can be approximated by piecewise polynomial domains. Our definition is quite similar to that of Lipschitz graph domain or minimally smooth domains in the sense of Stein [11, CH. VI.3.3] except that the latter uses Lipschitz spaces in its definition and we want to use more general Besov spaces.

By a coordinate system for $\mathbb{R}^2$ we mean a pair of unit vectors $(\mathbf{u},\mathbf{v})$ which are orthogonal. Given a point $x \in \mathbb{R}^2$, its coordinates with respect to these vectors are $x' = \langle x,\mathbf{u}\rangle$ and $y' = \langle x,\mathbf{v}\rangle$. We shall typically work in the following setting. We have an open rectangle R whose sides are parallel to the given coordinate vectors $\mathbf{u}$ and $\mathbf{v}$. The rectangle can be written as $R = I \times J$ in the coordinate system $(\mathbf{u},\mathbf{v})$. We assume that there exist a measurable function ϕ defined on I and an $\varepsilon > 0$ such that for each $x' \in I$, the numbers $\phi(x') - \varepsilon$ and $\phi(x') + \varepsilon$ are in J. This means that each point $(x', \phi(x'))$ is in the rectangle which we get from R by raising the bottom and lowering the top by ε. We call $G(\phi,R) := \{(x',y') \in R : y' < \phi(x')\}$ a *fundamental graph domain* associated to R and ϕ. In the case when ϕ is a polynomial of order r, $G(\phi,R)$ is called a *fundamental r-polynomial graph domain*. Figure 3.1 shows a typical fundamental graph domain.

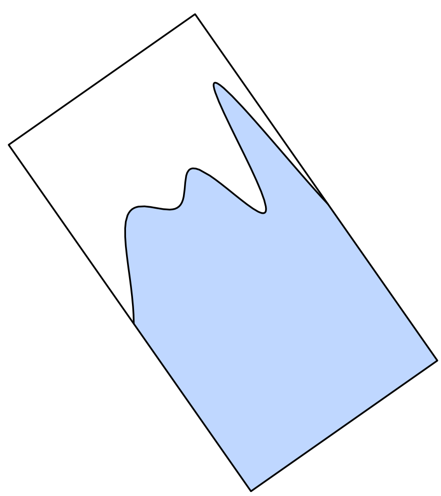

Figure 3.1: The typical fundamental graph domain.

Note that for continuous ϕ's, the set $G(\phi,R)$ is always an open subset of $\mathbb{R}^2$ and it has a simple geometrical interpretation (see Figure 3.1).

It is the rectangle R with its top trimmed off by the graph of the function ϕ. In general, $G(\phi,R)$ is not an open set in $\mathbb{R}^2$ but it is measurable. We abuse language throughout this paper by referring to sets as domains even though they are not necessarily open.

Let $B := B(I) := B^s_\tau(L_\tau(I))$ be a univariate Besov space which compactly embeds into L_1. This means that $1/\tau < s+1$. If the function ϕ in the definition of a fundamental graph domain $G(\phi,R)$ belongs to the class B then we will call such a domain a *fundamental* $B^s_\tau(L_\tau(I))$ *graph domain* associated to R and ϕ. We shall use the shorter notation fundamental B graph domain where it is understood that $B = B^s_\tau(L_\tau(I))$. Note that the elements of a Besov space are equivalence classes and each representative of this equivalence class gives its own fundamental B graph domain. However, the sets are equal almost everywhere. Note also that the boundary of a fundamental B graph domain can be very pathological, in particular, it need not be of finite length.

We say that a bounded open set $\Omega \in \mathbb{R}^2$ is a $B^s_\tau(L_\tau)$ *domain* (or B-domain for short) if there exists a positive integer N and a collection of N open rectangles $\{R_j := I_j \times J_j\}_{j=1}^N$ (each with respect to a coordinate system $(\mathbf{u}_j, \mathbf{v}_j)$), $j = 1,2,\dots,N$, which cover the boundary of Ω, with the following properties (see Figure 3.2): For each of the rectangles R_j, there is a function $\phi_j \in B^s_\tau(L_\tau(I_j))$ such that $\Omega_j := \Omega \cap R_j$ is a fundamental $B^s_\tau(L_\tau(I_j))$ graph domain associated to R_j and ϕ_j, more precisely $\Omega_j := \Omega \cap R_j = G(\phi_j, R_j)$.

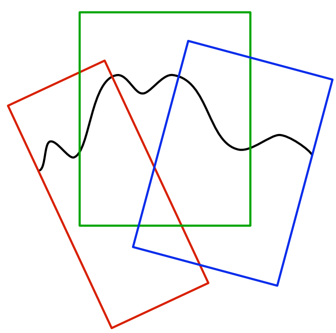

Figure 3.2: A portion of a Besov domain.

Notice that a B-domain will have many coverings of the above form. These coverings may have different values of N. Rather than try to choose a best N we shall simply introduce a measure of the complexity of the best representation for the given value of N. Namely, when $B = B^s_\tau(L_\tau)$ and $N > 0$, we define

$$(3.1) \qquad |\Omega|_{B,N} := \inf\left\{ \left[\sum_{j=1}^N |I_j|\right]^\delta \left[\sum_{j=1}^N |\phi_j|^\tau_{B^s_\tau(L_\tau(I_j))}\right]^{1/\tau}\right\}, \qquad \delta := s+1-1/\tau,$$

where the infimum is taken over all covers $\{R_j\}_{j=1}^N$. Let us make some remarks about this definition. First of all, this may be undefined for certain N since there may be no such covers with N components. Then, we shall define $|\Omega|_{B,N} = \infty$. Secondly, these quantities are non-increasing as $N \to \infty$. Notice also that δ is the discrepancy we introduced earlier which measures how far B is above the Sobolev embedding line for L_1. Lastly, this is not a norm (or semi-norm) in spite the suggestive notation; it simply measures in some way how nice Ω is.

Next, we define the distance between two measurable sets Ω_1, Ω_2 in $\mathbb{R}^2$ as

$$d_1(\Omega_1,\Omega_2) := |\Omega_1 \setminus \Omega_2| + |\Omega_2 \setminus \Omega_1|, \tag{3.2}$$

where for a measurable set $S \subset \mathbb{R}^2$, $|S|$ is its Lebesgue measure. Notice that changing a set S by a set of measure zero will not change its distance to any other set. Our goal is to approximate a general B-domain by simpler domains in such a way, that the distance between them is small. For us, these simpler domains will be ones whose boundary can be described by a finite number of polynomials. The simplest case here would be a polygonal domain.

To give a precise definition of what we mean by simpler domains, we shall return to the fundamental r-polynomial graph domains we have introduced earlier. Suppose that we have a finite collection of open rectangles R_j, $j = 1,\dots N$, each with respect to its own coordinate system $(\mathbf{u}_j, \mathbf{v}_j)$. The set $\Omega_1 := \bigcup_{j=1}^N R_j$ is open and its complement Ω_0 is closed in $\mathbb{R}^2$. We shall need further information about the structure of Ω_0. The following lemma is geometrically quite obvious.

Lemma 3.1. *Let R_j, $j = 1,\dots,N$, be open rectangles (not necessarily parallel to the coordinate axis) in $\mathbb{R}^2$. Then the set $\Omega_0 := \mathbb{R}^2 \setminus \bigcup_{j=1}^N R_j$ is equal almost everywhere to $\Omega_0' \cup \Omega_0''$, where Ω_0'' is the unbounded component* [1] *of Ω_0 and if $\Omega_0' \neq \emptyset$,*

$$\Omega_0' = \bigcup_{j=1}^m \Delta_j, \quad \textit{almost everywhere,} \tag{3.3}$$

where each Δ_j is an open finite triangular region and $m \le CN^2$ with C an absolute constant.

Proof: It is clear that Ω_0' and Ω_0'' are equal almost everywhere to the union of some of the pieces into which $\mathbb{R}^2$ is divided by the $4N$ lines which determine the boundaries of the R_j's, $j = 1,\dots,N$.

We first bound the maximal number $n(K)$ of pieces into which K lines in general position divide $\mathbb{R}^2$. Clearly $n(1) = 2 = 1+1$, $n(2) = 4 = 1+1+2$, and it can be easily seen that $n(K) \le n(K-1)+K$. Thus, inductively we get $n(K) \le 1+1+2+\cdots+K = 1+\frac{K(K+1)}{2}$, and so $4N$ lines in general position divide the plane into $n(4N) = 8N^2+2N+1$ regions (see also [12]). This estimate was further refined

[1] By a component of Ω_0 we mean a maximal open connected subset of Ω_0

(see [13] and the references therein), when it was taken into account that some of the lines may be parallel to each other or that more than two lines may pass through one point.

Let us denote by a the number of pieces (bounded and unbounded) in which the $4N$ lines which determine the boundaries of the R_j's, $j = 1,\dots,N$, divide $\mathbb{R}^2$ and let q be the total number of distinct intersection points of these lines . For our purposes, it is enough to observe that $a < n(4N) < CN^2$. Note that the set Ω_0' can be written as $\Omega_0' = \bigcup_{j=1}^{\ell} U_j$, where U_j are bounded convex regions and $\ell < a$. If v_j is the number of vertices of U_j, then each U_j is a union of $(v_j - 2)$ triangles $\Delta_{i,j}$, and thus $\Omega_0' = \bigcup_{j=1}^{\ell} \bigcup_{i=1}^{v_i-2} \Delta_{i,j}$, which gives a triangulation for Ω_0' with $m < \sum_{j=1}^{\ell} v_j$ triangular regions.

We say that the i-th intersection point is a point with multiplicity λ_i, if exactly λ_i lines pass through this point. We follow the sweep-line argument from [13], (originally introduced in [9]) and take a line (called a sweep line) that initially intersects each of the given lines, and all given intersection points are on one side of this line. It identifies $4N+1$ regions, containing the two rays and $4N-1$ line segments, obtained from the intersection of this line with the original $4N$ lines. Then, we sweep this line through the plane, keeping it always parallel to its initial position. New regions are encountered at the intersection points of the given lines: $(\lambda - 1)$ new regions at each point of multiplicity λ. This way the line sweeps through $\sum_{j=1}^{q}(\lambda_j - 1)$ new regions, and therefore the total number of regions is $a = 4N + 1 + \sum_{j=1}^{q}(\lambda_j - 1)$. Now, $\sum_{j=1}^{q} \lambda_j = a - 4N - 1 + q < CN^2$, since $q \le 2N(4N-1)$ and $a < CN^2$. On the other hand, $m < \sum_{j=1}^{\ell} v_j < \sum_{j=1}^{a} v_j = 2\sum_{j=1}^{q} \lambda_j$, since if we sum the number of vertices of all regions (bounded and unbounded), then we end up counting every intersection point j, $2\lambda_j$ times. This, together with the above estimate proves the lemma. □

We shall also use the next remark which is easy to prove.

Remark 3.2. If the open rectangles R_j, $j = 1,\dots,N$, are a cover of the boundary of a bounded open set Ω, then any triangular region Δ_j from Lemma 3.1 is either completely contained in the interior of Ω or in the interior of the complement of Ω.

We are now prepared to give our definition of simple domains. Given the positive integers n, r, we are now going to define a collection $\mathbf{D}(n,r)$ of certain sets $\hat{\Omega}$ which have a very simple description. To begin with, we suppose that we have m_2 rectangles $R_j = I_j \times J_j$, with respect to coordinate vectors $(\mathbf{u}_j, \mathbf{v}_j)$ and corresponding polynomials P_j, each with respect to the coordinate system $(\mathbf{u}_j, \mathbf{v}_j)$ such that $G(P_j, R_j)$ is a fundamental r-polynomial graph domain (recall that this requires that the graph of P_j on I_j is contained in R_j), $j = 1,\dots,m_2$. We define

$$\hat{\Omega} := \left[\bigcup_{j=1}^{m_1} \Delta_j\right] \bigcup \left[\bigcup_{j=1}^{m_2} G(P_j, R_j)\right], \tag{3.4}$$

where each Δ_j is an open triangular region. We say that such a set $\hat{\Omega}$ is in $\mathbf{D}(n,r)$ if $m_1+m_2 \le n$. Notice that if m_1 or m_2 is equal to 0, then $\hat{\Omega}$ consists of only fundamental r-polynomial graph domains or triangle regions, respectively. In the case $n=0$, we set $\mathbf{D}(0,r)=\emptyset$.

Remark 3.3. We could define our simple domains as the set of $\hat{\Omega}$'s such that,

$$\hat{\Omega} := b\left(\mathbb{R}^2 \setminus \bigcup_{j=1}^{m_2} R_j\right) \bigcup \left[\bigcup_{j=1}^{m_2} G(P_j, R_j)\right], \quad m_2 \le n, \tag{3.5}$$

where $b\left(\mathbb{R}^2 \setminus \bigcup_{j=1}^{m_2} R_j\right)$ is the bounded component of $\mathbb{R}^2 \setminus \bigcup_{j=1}^{m_2} R_j$. Instead, we chose the definition given in (3.4), which has at least two advantages. First, it is more general (see Lemma 3.1). Second, if we use definition (3.5) then when encoding the portion $b\left(\mathbb{R}^2 \setminus \bigcup_{j=1}^{m_2} R_j\right)$ of $\hat{\Omega}$ (see Section 6.2), it would not be enough to only quantize the vertices of the rectangles $\{R_j\}_{j=1}^{m_2}$, since a small error in the quantization may lead to a significant error in the domain description. We circumvent this difficulty by using our more general definition of simple domains involving triangles. Namely, we can now simply quantize the vertices of the triangles Δ_j, constituting $b\left(\mathbb{R}^2 \setminus \bigcup_{j=1}^{m_2} R_j\right)$. This allows us to prove a bit rate theorem (Theorem 6.2) on encoding our function classes in Section 6.2.

Note that in definition (3.4) we allow the regions to possibly overlap. However, in most constructions, such as in the following theorem, the triangular regions are disjoint.

Theorem 3.4. *Let $\Omega \subset \mathbb{R}^2$ be a $B := B^s_\tau(L_\tau)$ domain, where $s>0$ and $1/\tau < s+1$. Then, for any $N>0$, there is an $\hat{\Omega} \in \mathbf{D}(C_1 n, r)$, $r>s$, such that*

$$d_1(\Omega, \hat{\Omega}) \le C_2(r,\delta) n^{-s} |\Omega|_{B,N}, \quad n \ge N^2, \tag{3.6}$$

where the constant C_1 depends only on the order of the polynomials r, while the constant $C_2(r,\delta)$ depends only on r and the offset $\delta := s+1-\frac{1}{\tau}$ of the point $(1/\tau, s)$ above the critical line for Sobolev embedding into L_1.

Proof: We can assume without loss of generality that $|\Omega|_{B,N} < \infty$. From the definition of $B^s_\tau(L_\tau)$ domain, for each $\varepsilon > 0$, there is a collection $\{R_j\}_{j=1}^N$, $j=1,\dots N$, of open rectangles and corresponding coordinate systems $(\mathbf{u}_j, \mathbf{v}_j)$, where $R_j = I_j \times J_j$ in these coordinate systems, and a collection of functions $\phi_j \in B^s_\tau(L_\tau(I_j))$, such that $\Omega_j := \Omega \cap R_j = G(\phi_j, R_j)$ and

$$\left[\sum_{j=1}^N |I_j|\right]^\delta \left[\sum_{j=1}^N |\phi_j|^\tau_{B^s_\tau(L_\tau(I_j))}\right]^{1/\tau} \le |\Omega|_{B,N} + \varepsilon =: M. \tag{3.7}$$

Let us now define

$$n_j := \left\lceil n \frac{L_j}{\sum_{j=1}^N L_j} \right\rceil, \tag{3.8}$$

where

$$L_j = |I_j|^{\delta/(s+1)} |\phi_j|^{1/(s+1)}_{B^s_\tau(L_\tau(I_j))}. \tag{3.9}$$

According to (2.4), for each j, there is a piecewise polynomial S_j defined on I_j (with respect to the coordinate system for R_j) of order r with n_j pieces that approximates ϕ_j in $L_1(I_j)$ to accuracy

$$\|\phi_j - S_j\|_{L_1(I_j)} \leq C'(r,\delta)|I_j|^{\delta} |\phi_j|_{B^s_\tau(L_\tau(I_j))} n_j^{-s}. \tag{3.10}$$

The total number of pieces in these piecewise polynomials does not exceed

$$\sum_{j=1}^{N} n_j \leq \sum_{j=1}^{N} \left(1 + \frac{nL_j}{\sum_{i=1}^{N} L_i}\right) \leq N + n \leq 2n. \tag{3.11}$$

We may assume that the graphs of the piecewise polynomial approximations S_j are within the rectangles R_j. This follows from the fact (see the definition of a fundamental graph domain) that there is an ε_j, such that the graph of ϕ_j lies between the lines $y' = c_j + \varepsilon_j$ and $y' = d_j - \varepsilon_j$ in R_j, where $R_j = I_j \times J_j$, $J_j = [c_j, d_j]$ (with respect to the coordinate system $(\mathbf{u}_j, \mathbf{v}_j)$). Therefore, if the graph of the piecewise polynomial S_j goes above or below these two lines, we trim it by substituting the pieces outside the strip $[c_j + \varepsilon_j, d_j - \varepsilon_j]$ by the corresponding line segments. This way, instead of S_j, we construct a new piecewise polynomial $\widetilde{S}_j$ with at most $(2r-1)n_j$ polynomial pieces of order r, for which the error is even better since we have $\|\phi_j - \widetilde{S}_j\|_{L_1(I_j)} \leq \|\phi_j - S_j\|_{L_1(I_j)}$. In going further, we will use S_j to denote these new approximations. The above arguments show that each of the sets $\widetilde{\Omega}_j := G(S_j, R_j)$, $j = 1, \ldots, N$, is almost everywhere equal to a union of $(2r-1)n_j$ fundamental r-polynomial graph domains.

Next, we denote by $\Omega_0 := \Omega \setminus \bigcup_{j=1}^{N} \Omega_j = \Omega \setminus \bigcup_{j=1}^{N} R_j$. From Lemma 3.1 and Remark 3.2 we know that the interior of $\Omega_0 := \Omega \setminus \bigcup_{j=1}^{N} R_j$ is almost everywhere the union of at most CN^2, triangular regions, i.e. $\Omega_0 = \bigcup_{j=1}^{m} \Delta_j$ with $m \leq CN^2$. Now we define

$$\hat{\Omega} := \Omega_0 \cup \left(\bigcup_{j=1}^{N} \widetilde{\Omega}_j\right) = \left(\bigcup_{j=1}^{m} \Delta_j\right) \bigcup \left(\bigcup_{j=1}^{N} \widetilde{\Omega}_j\right).$$

Thus, $\hat{\Omega}$ is up to a set of measure zero equal to a domain in the set $\mathbf{D}(2(2r-1)n + CN^2, r) \subset \mathbf{D}(C_1 n, r)$ with $C_1 = 2(2r-1) + C$.

It remains to only show that $\hat{\Omega}$ approximates Ω sufficiently well, so we need a bound for $d_1(\Omega, \hat{\Omega})$. We first note that $\hat{\Omega} \setminus \Omega \subset \bigcup_{j=1}^{N} (\widetilde{\Omega}_j \setminus \Omega_j)$ and similarly

$\Omega\setminus\hat\Omega\subset\bigcup_{j=1}^N(\Omega_j\setminus\widetilde\Omega_j)$. Hence, we have from (3.10) that

$$d_1(\Omega,\hat\Omega)\le\sum_{j=1}^N d_1(\Omega_j,\widetilde\Omega_j)\le\sum_{j=1}^N\|\phi_j-S_j\|_{L_1(I_j)}\tag{3.12}$$

$$\le C'(r,\delta)\sum_{j=1}^N|I_j|^\delta|\phi_j|_{B^s_\tau(L_\tau(I_j))}n_j^{-s}.$$

Using (3.8) we can bound the sum on the right side of (3.12) by

$$\sum_{j=1}^N|I_j|^\delta|\phi_j|_{B^s_\tau(L_\tau(I_j))}\left(n\frac{L_j}{\sum_{j=1}^N L_j}\right)^{-s},$$

which, using (3.9) equals

$$n^{-s}\left(\sum_{j=1}^N|I_j|^{\delta/(s+1)}|\phi_j|^{1/(s+1)}_{B^s_\tau(L_\tau(I_j))}\right)^{s+1}.$$

Since $(s+1)/\delta>1$, by Hölder's inequality this is bounded by $n^{-s}M$. Since $\varepsilon>0$ is arbitrary, this completes the proof of Theorem 3.4. $\square$

4 Domains in $\mathbb{R}^d$

In this section, we shall extend our smoothness classification of domains to sets in $\mathbb{R}^d$. Rather than generalize the notion of Besov domain directly, we shall begin with the fact, proved in the previous section, that each such domain can be well approximated by polynomial domains from $\mathbf{D}(n,r)$. Thus, our approach is to first generalize the classes $\mathbf{D}(n,r)$ to $\mathbb{R}^d$ and then classify sets according to how well they can be approximated by sets from $\mathbf{D}(n,r)$.

Let us first explain why this approach seems better than to generalize the definition of B-domains. B-domains have a very appealing geometric structure, but they are not closed under approximation in the metric $d_1(\cdot,\cdot)$. Moreover, even C^∞ functions may have level sets that are not B-domains. There are many examples we could give to illustrate this. Here is one.

Example 4.1. Let f be a C^∞ function on $\mathbb{R}^2$ such that $f(x)\ge0$ and $f(x)>0\iff x\in(0,1)^2$. We choose an infinite collection of disjoint open squares $Q_n\subset[0,1]^2$ whose boundaries do not intersect and $|Q_n|=2^{-2n}$. We denote by f_n the function f dilated and shifted so that it is supported precisely on Q_n. One easily checks that the function $F(x)=\sum_{n=1}^\infty\frac{1}{n!}f_n(x)$ is a well defined C^∞ function on $\mathbb{R}^2$. Clearly the level set $\Omega:=\{x:\ F(x)>0\}$ equals $\bigcup_{n=1}^\infty Q_n$, and Ω is not a B-domain. On the other hand it can be very well approximated by triangular domains. For each N, the set $\Omega_N:=\bigcup_{n=1}^N Q_n$ is a union of $2N$ triangles and obviously

$$d_1(\Omega,\Omega_N)=\sum_{n=N+1}^\infty2^{-2n}\le C2^{-2N}.$$

One could argue that the above example is not that convincing. Namely later in our considerations, we use the averages over z of the level set norms and thus the fact that some of the level sets are bad does not really matter. In fact, from the Sard theorem (see [8] Theorem 3.4.3) and the implicit function theorem it follows that for C^2 functions f in $\mathbb{R}^2$, the set of all $z \in \mathbb{R}$ for which the boundary of the level set $\{x: f(x) > z\}$ is not locally a graph of a function has measure zero. However, as illustrated in the next example, we may have less smooth functions for which all level sets are not B-domains. Thus, a classification of domains in $\mathbb{R}^d$, suitable for treating level sets, requires an approximation process. This leads us to take the approach given in this section.

Example 4.2. Let us fix a sequence of points $\{a_n\}_{n=1}^{\infty}$ in $\mathbb{R}^2$, such that $a_n = (x_n, y_n)$ with $x_n^2 + y_n^2 = 1$, $x_n > 0$, $y_n \searrow 0$, and a sequence of numbers $\lambda_n \searrow 1$. Let Δ_n be the triangle with vertices $a_n, \lambda_n a_n, a_{n+1}$. We define the set $\Omega := \bigcup_{n=1}^{\infty} \Delta_n \cup \{(x,y): x^2+y^2 \leq 1\}$ and the function μ on $\mathbb{R}^2$, $\mu(x,y) := \inf\{\lambda^{-1}: \lambda > 0 \text{ and } \lambda(x,y) \in \Omega\}$. Note, that in the neighbourhood of the point $(1,0)$, the boundary of Ω is not a graph of any function in any coordinate system. Then, for any C^∞ function ϕ on $\mathbb{R}$, decreasing on $\mathbb{R}_+$ with $\phi(0) = 1$ and $\phi(x) = 0$ for $x > 1$, all nontrivial level sets of the function $F(x,y) = \phi(\mu(x,y))$ are homotetic images of Ω, so none of them is a B-domain.

Next, we continue with our definition. We fix a positive integer r and consider polynomials in $\mathbb{R}^d$ or order r (degree $r-1$). We define a fundamental r-polynomial graph domain $G(\phi, R)$, where the d-dimensional parallelepiped $R = I \times J$, with I a $(d-1)$ dimensional parallelepiped, is with respect to some orthogonal coordinate system, and ϕ is a polynomial in $(d-1)$ variables of order r in that coordinate system. As in the case $d = 2$, we require that the graph of ϕ on I is completely contained in R.

We define $\mathbf{D}(n,r)$, $n \geq 0$, as the collection of all sets $\hat{\Omega}$ in $\mathbb{R}^d$ which can be written as

$$\hat{\Omega} = \left(\bigcup_{j=1}^{m_1} \Delta_j \right) \bigcup \left(\bigcup_{j=1}^{m_2} G(P_j, R_j) \right), \tag{4.1}$$

where Δ_j is a d-dimensional open simplex in $\mathbb{R}^d$, $G(P_j, R_j)$ is a fundamental r-polynomial graph domain, and $m_1 + m_2 \leq n$. In the case $n = 0$, we set $\mathbf{D}(0,r) := \emptyset$. If m_1 or m_2 is equal to 0, then $\hat{\Omega}$ is a union of only fundamental r-polynomial graph domains or simplexes, respectively.

For any measurable set $\Omega \subset \mathbb{R}^d$, we define the error in approximating Ω by sets from $\mathbf{D}(n,r)$ by

$$\sigma_{n,r}(\Omega) := \inf_{\hat{\Omega} \in \mathbf{D}(n,r)} d_1(\Omega, \hat{\Omega}), \quad n \geq 0, \tag{4.2}$$

where as in the two dimensional case, $d_1(\Omega, \hat{\Omega}) := |\Omega \setminus \hat{\Omega}| + |\hat{\Omega} \setminus \Omega|$. It is clear that $\sigma_{0,r}(\Omega) = |\Omega|$. We gather these domains into approximation classes. For any

$s > 0$, let $\mathscr{A}^s$ be the class of all domains Ω such that

$$\sigma_{n,r}(\Omega) \leq M(n+1)^{-s}, \quad \text{for all } n \geq 0. \tag{4.3}$$

The smallest M for which (4.3) holds is defined to be $|\Omega|_{\mathscr{A}^s}$. We consider the parameter s as a measure of smoothness of sets. Namely, sets in $\mathscr{A}^s$ have smoothness of order s.

It is clear that any piecewise polynomial domain in $\mathbf{D}(m,r)$ belongs to $\mathscr{A}^s$ for all $s > 0$. Note that in the case $d = 2$ any $B^s_\tau(L_\tau)$ domain with $s > 0$ and $1/\tau < s+1$, is in $\mathscr{A}^s$ by virtue of Theorem 3.4.

5 Level sets

We have mentioned in the introduction, that there are two main ingredients for how we measure the smoothness of a function. One is the smoothness of its level sets and the second is how smoothly these level sets change with a change in height. We have already discussed our way of measuring the smoothness of level sets. We now turn to the question of quantifying how the level sets are changing. We begin in this section by recalling how an L_p-norm of a function f defined on $\mathbb{R}^d$ is computed from these level sets.

We shall consider functions defined on $\mathbb{R}^d$ with finite support. We first consider nonnegative functions f defined on $\mathbb{R}^d$. We denote the level set of f at level $z \geq 0$ by $\Omega(z) := \Omega(f,z) := \{x : f(x) > z\}$. We fix a value of $p \in [1,\infty)$ and recall that for every $f \in L_p$

$$\|f\|^p_{L_p} = p\int_0^\infty A(f,z)z^{p-1}\,dz, \tag{5.1}$$

where $A(f,z) := |\Omega(f,z)|$ is the measure of this level set. Here and later in this paper, $\|\cdot\|_{L_p}$ means the L_p norm on $\mathbb{R}^d$; all other L_p norms will indicate the domain of integration. As seen from this formula, the measure $d\mu_p(z) := pz^{p-1}\,dz$ plays an important role when using level sets.

Intuitively, if two functions $f, g \in L_p$ have level sets that are close then the norm $\|f-g\|_{L_p}$ should be small. Our next goal will be to give a quantitative version of this observation. We can work in more generality than just level sets of functions and this will be useful for us later. We suppose that for each $z \geq 0$, we have a Lebesgue measurable set $\Omega(z)$. From this family of sets we construct the function

$$F(x) := \int_0^\infty \chi_{\Omega(z)}(x)\,dz, \quad x \in \mathbb{R}^d, \tag{5.2}$$

where $\chi_{\Omega(z)}$ is the characteristic function of $\Omega(z)$. We assume that for almost all $x \in \mathbb{R}^d$, the function $\chi_{\Omega(z)}(x)$ is measurable as a function of z. If f is a nonnegative function and $\Omega(z) = \Omega(f,z)$ are the level sets of f, then $F = f$. However, we shall frequently work in cases where the $\Omega(z)$ are not nested and in this case the sets $\Omega(z)$ are generally not the level sets of F.

Now suppose we have two families of sets $\Omega(z)$ and $\bar{\Omega}(z)$. We define

$$\Delta(z) := (\Omega(z) \setminus \bar{\Omega}(z)) \cup (\bar{\Omega}(z) \setminus \Omega(z)), \tag{5.3}$$

which is the symmetric difference of these sets. We also let

$$\delta(z) := |\Delta(z)| = d_1(\Omega(z), \bar{\Omega}(z)), \tag{5.4}$$

which is the Lebesgue measure of this symmetric difference.

Lemma 5.1. *Given two families of sets $\Omega(z)$ and $\bar{\Omega}(z)$, let F and $\bar{F}$ be defined by* (5.2) *and let δ be defined as in* (5.4). *Then, for any $1 \le p < \infty$, we have*

$$\|F - \bar{F}\|_{L_p}^p \le \int_0^\infty \delta(z) d\mu_p(z). \tag{5.5}$$

Proof: From the definitions of F and $\bar{F}$, we have

$$|F(x) - \bar{F}(x)| \le \int_0^\infty \chi_{\Delta(z)}(x)\, dz. \tag{5.6}$$

Fix x and let $\Lambda(x) := \{z : \chi_{\Delta(z)}(x) = 1\}$ and $L(x) := |\Lambda(x)|$. Our main observation is that for each x we have

$$\begin{aligned} \left[\int_0^\infty \chi_{\Delta(z)}(x)\, dz\right]^p &= \left[\int_0^{L(x)} dz\right]^p = \int_0^{L(x)} d\mu_p(z) \\ &\le \int_{\Lambda(x)} d\mu_p(z) = \int_0^\infty \chi_{\Delta(z)}(x)\, d\mu_p(z). \end{aligned} \tag{5.7}$$

Indeed, the inequality follows because the two sets, namely $[0, L(x)]$ and $\Lambda(x)$ on which we are integrating $d\mu_p(z)$ have the same Lebesgue measure and the function pz^{p-1} is increasing. We now integrate both sides of (5.7) with respect to x and use Fubini's theorem to interchange the integrals on the right side to arrive at (5.5). □

Let us single out the special case of this lemma when we are dealing with functions f and g. In this case, we use the notation

$$\delta(f, g; z) := \delta(z), \tag{5.8}$$

where δ is defined by using the level sets of f for $\Omega(z)$ and those of g for $\bar{\Omega}(z)$.

Theorem 5.2. *Given $\Omega \subset \mathbb{R}^d$ and any two nonnegative functions $f, g \in L_p(\Omega)$, we have*

$$\|f - g\|_{L_p(\Omega)}^p \le \int_0^\infty \delta(f, g; z) d\mu_p(z). \tag{5.9}$$

Proof: This is an immediate consequence of Lemma 5.1. □

5.1 Evolution of level sets

In this section, we introduce ways to measure how smoothly the level sets of a given function f evolve with the level z. This will be the second ingredient in our definition of smoothness spaces in $\mathbb{R}^d$. We will introduce a crude measure of smoothness which is the analogue of first order smoothness (or one order of differentiability) for classical smoothness spaces. In Section 7, we shall discuss possible higher order measures of smoothness.

We begin by considering nonnegative functions f on $\mathbb{R}^d$. Recall that we denote the level set of f at level $z \geq 0$ by $\Omega(z) := \Omega(f,z) := \{x : f(x) > z\}$, and its Lebesgue measure by $A(z) := A(f,z) := |\Omega(f,z)|$. We will measure the regularity of the evolution of the level sets of f by the smoothness of the function $A(z)$ as measured by its approximability by piecewise constant functions. Notice that $A(z)$ is a decreasing function on $\mathbb{R}_+$.

We fix a value of $p \in [1,\infty)$ which corresponds to the L_p norm in which we want to measure the approximation of f. We continue to use the measure $d\mu_p(z) := pz^{p-1}\,dz$ as introduced in the previous section. We will consider approximation of $A(z)$ in the $L_1(d\mu_p)$ norm. We first consider the approximation of a general monotone function g by constants. We fix an interval $I = [a,b]$ and define

$$E(g,I)_p := \inf_{c\in\mathbb{R}} \|g-c\|^{1/p}_{L_1(I,d\mu_p)}. \tag{5.10}$$

We will need the following technical lemma.

Lemma 5.3. *Let $g \in L_1(I,d\mu_p)$ be a decreasing (possibly unbounded) positive function defined on an interval $I = (a,b) \subset \mathbb{R}_+$. We put $g(a) = \lim_{x\to a_+} g(x)$ (where we allow this limit to be infinity). Then, there is an interval $J = (a',b') \subset I$, $\mu_p(J) = \mu_p(I)/2$, such that for each $\eta \in (a',b')$, we have*

$$\|g-g(\eta)\|^{1/p}_{L_1(I,d\mu_p)} \leq 3^{1/p}E(g,I)_p. \tag{5.11}$$

Proof: First note that the constant c^* of best approximation to g on I obviously satisfies $c^* \in (g(b),g(a))$. Let $J = (a',b') \subset I$ be an interval with measure $\mu_p(J) = \mu_p(I)/2$ such that $c^* \in (g(b'),g(a'))$. We choose this interval J so that $\nu = |(g(a') - c^*) - (c^* - g(b'))|$ is as small as possible given the other constraints on J. Either ν will be zero or one of the endpoints of I will be an endpoint of J . In either case, since g is decreasing, it follows that for every $z \notin J$ and $\eta \in J$ we have $|g(z)-c^*| \geq |g(\eta)-c^*|$. We integrate this inequality over $J^c = I \setminus J$ and obtain

$$E^p := E(g,I)^p_p \geq \int_{J^c} |g(z)-c^*|d\mu_p(z) \geq \mu_p(J^c)|g(\eta)-c^*|, \quad \eta \in J,$$

which gives

$$|g(\eta)-c^*| \leq \frac{E^p}{\mu_p(J^c)} = \frac{2E^p}{\mu_p(I)}, \quad \text{for every } \eta \in J. \tag{5.12}$$

Now we use the triangle inequality and (5.12) to derive

$$\|g-g(\eta)\|_{L_1(I,d\mu_p)} \le \|g-c^*\|_{L_1(I,d\mu_p)} + \|c^*-g(\eta)\|_{L_1(I,d\mu_p)} \le E^p + 2E^p = 3E^p,$$

and the proof is completed. □

We now turn to approximation of A by piecewise constants. Let Σ_n^1 be the set of univariate piecewise constant functions on $\mathbb{R}_+$ with n pieces, $n>0$, i.e. taking at most n non-zero values, and let Σ_0^1 consist of the zero function. For any univariate function $g \in L_1(\mathbb{R}_+, d\mu_p)$ and any $1 \le p < \infty$ we define the error

$$\sigma_n(g)_p := \inf_{S\in\Sigma_n^1} \|g-S\|_{L_1(\mathbb{R}_+,d\mu_p)}^{1/p}, \quad n \ge 0, \tag{5.13}$$

of approximating g by the elements in Σ_n^1 using the $L_1(\mu_p)$ norm. Clearly we have $\sigma_0(g)_p = \|g\|_{L_1(\mathbb{R}_+,d\mu_p)}^{1/p}$. For any $s>0$, we define the approximation class $\mathscr{B}_p^s$ that consists of all functions g, such that

$$\sigma_n(g)_p \le M(n+1)^{-s}, \quad \text{for all } n \ge 0. \tag{5.14}$$

As usual, we define $|g|_{\mathscr{B}_p^s}$ to be the smallest M for which (5.14) hold.

It follows from results on nonlinear approximation that we have already mentioned that any function g in $B_\tau^s(L_\tau(I))$ is in $\mathscr{B}_1^s$ for $s \in (0,1)$ and $1/\tau < s+1$. There is a way to characterize $\mathscr{B}_p^s$. By a simple change of variables, $g \in \mathscr{B}_p^s$ if and only if the function $g(z^p)$ can be approximated by piecewise constant functions with n pieces in the L_1 norm to accuracy $O(n^{-sp})$. The latter class of functions contains the Besov space $B_\tau^{sp}(L_\tau)$ for $\tau \ge (sp+1)^{-1}$. Also, any piecewise constant function with finite number of pieces clearly belongs to $\mathscr{B}_p^s$ for all $s>0$ and all p.

Note that if f is a finite linear combination of characteristic functions of sets B_j of finite measure, $f=\sum_{j=1}^m c_j \chi_{B_j}$, then $A(f,z)$ is a piecewise constant function with a finite number of pieces and therefore $A(f,z)$ will belong to $\mathscr{B}_p^s$ for all $s>0$ and all p.

6 The smoothness spaces for $\mathbb{R}^d$

We are now in position to give the definition of our new smoothness spaces. Given a function f defined on $\mathbb{R}^d$, we can decompose $f=f_+-f_-$, where $f_\pm$ are its positive and negative parts. We fix any $p \in [1,\infty)$ and any $s_1, s_2 > 0$. We say a function f defined on $\mathbb{R}^d$ is in $\mathscr{F}_p^{(s_1,s_2)}$ if

$$\begin{aligned} M_1(f) := &\left\{\int_0^\infty |\Omega(f_+,z)|_{\mathscr{A}^{s_1 p}} d\mu_p(z)\right\}^{1/p} \\ &+ \left\{\int_0^\infty |\Omega(f_-,z)|_{\mathscr{A}^{s_1 p}} d\mu_p(z)\right\}^{1/p} < \infty, \end{aligned} \tag{6.1}$$

and

$$M_2(f) := |A(f_+,\cdot)|_{\mathscr{B}_p^{s_2}} + |A(f_-,\cdot)|_{\mathscr{B}_p^{s_2}} < \infty. \tag{6.2}$$

Here $\mathscr{A}^s$ and $\mathscr{B}_p^s$ are the approximation classes defined in the previous sections. We define

$$|f|_{\mathscr{F}_p^{(s_1,s_2)}} := M_1(f) + M_2(f). \tag{6.3}$$

The role of $M_1(f)$ is to measure the smoothness of the level sets of f. It measures it in an averaged sense. The role of $M_2(f)$ is to measure how smoothly these level sets evolve with changing height.

Notice that these smoothness spaces depend on the choice of r, a dependence we shall usually not indicate. Clearly every finite linear combination of characteristic functions $\{\chi_{\Lambda_j}\}$ of domains $\Lambda_j \in \mathbf{D}(n_j, r)$, $f = \sum_{j=1}^m c_j \chi_{\Lambda_j}$, is an element of $\mathscr{F}_p^{(s_1,s_2)}$ for all $s_1 > 0$, $s_2 > 0$, and p. Also, the characteristic function χ_Ω of an open bounded $B_\tau^s(L_\tau)$ domain $\Omega \subset \mathbb{R}^2$ with $s > 0$ and $1/\tau < s+1$ belongs to $\mathscr{F}_p^{(s/p,s_2)}$ for all $s_2 > 0$ and all p (see Theorem 3.4). Note also that $M_1(f) < \infty$ implies that $f \in L_p(\mathbb{R}^d)$.

6.1 Approximation of functions in $\mathscr{F}_p^{(s_1,s_2)}$

Our next goal will be to show that the functions in the smoothness space $\mathscr{F}_p^{(s_1,s_2)}$ can be approximated well by linear combinations of characteristic functions χ_Λ for which the set Λ is particularly simple. Given (n,r), we define $\mathscr{C}_{n,r}$ as the collection

$$\mathscr{C}_{n,r} := \{S : \ S = \sum_{j=1}^m c_j \chi_{\Lambda_j},\ \Lambda_j \in \mathbf{D}(n_j, r),\ \sum_{j=1}^m n_j \le n\}, \tag{6.4}$$

of all functions that are linear combinations of characteristic functions of piecewise polynomial domains. The following theorem holds.

Theorem 6.1. *If $f \in \mathscr{F}_p^{(s_1,s_2)}$, with $1 \le p < \infty$, $0 < s_1, s_2$, then there is an $S \in \mathscr{C}_{8n,r}$ such that*

$$\|f - S\|_{L_p} \le 3^{1/p} |f|_{\mathscr{F}_p^{(s_1,s_2)}} n^{-s}, \tag{6.5}$$

where

$$\frac{1}{s} = \frac{1}{s_1} + \frac{1}{s_2}. \tag{6.6}$$

If $f \ge 0$ then we have $S \in \mathscr{C}_{4n,r}$ and $S = \sum_{j=1}^m c_j \chi_{\Lambda_j}$ with $c_j > 0$ and $c_j \le \|f\|_{L_\infty}$ if $f \in L_\infty(\mathbb{R}^d)$.

Proof: We will prove (6.5) for f nonnegative and with $S \in \mathscr{C}_{4n,r}$. The theorem follows from this by decomposing f into its positive and negative parts. Given any integer $n = 1, 2, \ldots$, we define

$$n_1 := \left\lceil n^{\frac{s}{s_1}} \right\rceil, \quad n_2 := \left\lceil n^{\frac{s}{s_2}} \right\rceil.$$

The assumption that $f \in \mathscr{F}_p^{(s_1,s_2)}$, guarantees that there is a decreasing piecewise constant function $\bar{S}_2(z)$ defined on $\mathbb{R}_+$ with at most n_2 pieces and breakpoints $z_1 < \cdots < z_{n_2}$, such that for $A(z) := A(f,z)$, we have

$$\sigma_{n_2}(A)_p = \|A - \bar{S}_2\|_{L_1(\mathbb{R}_+,d\mu_p)}^{1/p} \le M_2(f)(n_2+1)^{-s_2} \le M_2(f)n^{-s}. \tag{6.7}$$

We define $I_j := (z_j, z_{j+1})$, $j = 0,\dots,n_2$, where by definition $z_0 := 0$ and $z_{n_2+1} := \infty$. Since by assumption $f \in L_p(\mathbb{R}^d)$, it follows that $\bar{S}_2$ is zero on (z_{n_2},∞).

We shall now invoke Lemma 5.3. It follows from the lemma that for each interval I_j, there is a subinterval J_j with $\mu_p(J_j) = \mu_p(I_j)/2$ such that for any points $\xi_j \in J_j$, $j = 0,\dots,n_2-1$, we have

$$\|A - A(\xi_j)\|_{L_1(I_j,d\mu_p)} \le 3E(A,I_j)_p^p, \quad j = 0,\dots,n_2-1.$$

Summing the above inequalities over j and denoting by S_2 the piecewise constant function $S_2 := \sum_{j=0}^{n_2-1} A(\xi_j)\chi_{I_j}$, we arrive at

$$\|A - S_2\|_{L_1((0,z_{n_2}),d\mu_p)} \le 3\sum_{j=0}^{n_2-1} E(A,I_j)_p^p.$$

Now we add to both sides $\int_{z_{n_2}}^{\infty} A(z)d\mu_p(z)$ and since $S_2(z) = 0$ for $z > z_{n_2}$, we have

$$\begin{aligned} \|A - S_2\|_{L_1(\mathbb{R}_+,d\mu_p)} &\le 3\sum_{j=0}^{n_2-1} E(A,I_j)_p^p + \|A\|_{L_1((z_{n_2},\infty),d\mu_p)} \\ &\le 3\|A - \bar{S}_2\|_{L_1(\mathbb{R}_+,d\mu_p)}. \end{aligned} \tag{6.8}$$

We now choose the $\xi_j \in J_j$ so that for each $j = 0,\dots,n_2-1$, we have

$$\begin{aligned} |\Omega(f,\xi_j)|_{\mathscr{A}^{s_1 p}} &\le \frac{1}{\mu_p(J_j)}\int_{J_j} |\Omega(f,z)|_{\mathscr{A}^{s_1 p}}\, d\mu_p(z) \\ &\le \frac{2}{\mu_p(I_j)}\int_{I_j} |\Omega(f,z)|_{\mathscr{A}^{s_1 p}}\, d\mu_p(z). \end{aligned} \tag{6.9}$$

In going further, we define $\Omega_j := \Omega(f,\xi_j)$, $j = 0,\dots,n_2-1$. According to the definition of $\mathscr{A}^{s_1 p}$, for each integer $n_1 > 0$ there is a domain $\bar{\Omega}_j \in \mathbf{D}(n_1,r)$ such that

$$d_1(\Omega_j,\bar{\Omega}_j) \le |\Omega_j|_{\mathscr{A}^{s_1 p}} n_1^{-s_1 p} \le |\Omega_j|_{\mathscr{A}^{s_1 p}} n^{-sp}, \quad j = 0,\dots,n_2-1. \tag{6.10}$$

We will take as our approximant to f the function

$$S := \sum_{j=0}^{n_2-1} (z_{j+1} - z_j)\chi_{\bar{\Omega}_j}. \tag{6.11}$$

Note that $S \in \mathscr{C}_{n_1 n_2, r}$, where $n_1 n_2 = \left\lceil n^{\frac{s}{s_1}} \right\rceil \left\lceil n^{\frac{s}{s_2}} \right\rceil \le 4n$.

We shall now estimate $\|f-S\|_{L_p}$. We let $f_c := \sum_{j=0}^{n_2-1}(z_{j+1}-z_j)\chi_{\Omega_j}$. From the triangle inequality, we have

$$\|f-S\|_{L_p} \le \|f-f_c\|_{L_p} + \|f_c-S\|_{L_p}. \tag{6.12}$$

To complete the proof of the theorem, we shall estimate each of the two terms appearing on the right side of (6.12).

To estimate the first term, we apply Lemma 5.1. Notice that the level sets of f are $\Omega(f,z)$ and those of f_c are Ω_j when $z \in (z_j, z_{j+1})$, so $\delta(f,f_c;z) = |A(z)-S_2(z)|$. Thus using (5.9), we obtain from (6.7) and (6.8)

$$\|f-f_c\|_{L_p} \le \|A-S_2\|^{1/p}_{L_1(\mathbb{R}_+,d\mu_p)} \le 3^{1/p}M_2(f)n^{-s}. \tag{6.13}$$

To estimate the second term in (6.12), we again apply Lemma 5.1. The level sets of f_c are $\Omega_j = \Omega(f,\xi_j)$ for $z \in (z_j, z_{j+1})$. Note that the level sets of S are not necessarily the $\bar{\Omega}_j$'s because these sets may not be nested. However, we do have

$$S(x) = \sum_{j=0}^{n_2-1} \int_{z_j}^{z_{j+1}} \chi_{\bar{\Omega}_j}(x)\,dz, \tag{6.14}$$

and therefore, we are in the setting of Lemma 5.1. By (6.10) we have

$$\delta(z) = d_1(\Omega_j, \bar{\Omega}_j) \le |\Omega_j|_{\mathscr{A}^{s_1 p}} n^{-sp}, \quad z \in (z_j, z_{j+1}), \quad j = 0, \dots, n_2-1. \tag{6.15}$$

Note that $\delta(z) = 0$ for $z > z_{n_2}$. It follows from this and (6.9) that

$$\int_{z_j}^{z_{j+1}} \delta(z)d\mu_p(z) \le 2n^{-sp} \int_{z_j}^{z_{j+1}} |\Omega(f,z)|_{\mathscr{A}^{s_1 p}} d\mu_p(z), \quad j = 0, \dots, n_2-1. \tag{6.16}$$

If we sum these estimates over $j = 0, \dots, n_2-1$ and use them in Lemma 5.1, we arrive at

$$\|f_c-S\|^p_{L_p} \le \int_0^{z_{n_2}} \delta(z)d\mu_p(z) \le 2n^{-sp} \int_0^{z_{n_2}} |\Omega(f,z)|_{\mathscr{A}^{s_1 p}} d\mu_p(z) \le 2n^{-sp}M_1(f)^p,$$

and therefore

$$\|f_c-S\|_{L_p} \le 2^{1/p}M_1(f)n^{-s} < 3^{1/p}M_1(f)n^{-s}. \tag{6.17}$$

Inequalities (6.12), (6.13) and (6.17) and the fact that $S \in \mathscr{C}_{4n,r}$ give the proof of the theorem. $\square$

Let us define now the error

$$\Theta_n(f)_p := \inf_{S \in \mathscr{C}_{n,r}} \|f-S\|_{L_p}, \quad n \ge 1,$$

of approximating $f \in L_p$, $1 \le p < \infty$, by elements from $\mathscr{C}_{n,r}$ and set $\Theta_0(f)_p := \|f\|_{L_p}$. Further, for any $s > 0$, we define the approximation class $\mathscr{E}^s_p$ as the set of all functions f for which

$$\Theta_n(f)_p \le M(n+1)^{-s}, \quad \text{for all } n \ge 0.$$

The class $\mathscr{E}_p^s$ is clearly a linear function space, and by Theorem 6.1, we have $\mathscr{F}_p^{(s_1,s_2)} \subset \mathscr{E}_p^s$ for $1/s = 1/s_1 + 1/s_2$. However, at the moment we do not know whether the class $\mathscr{F}_p^{(s_1,s_2)}$ is linear or not.

6.2 Encoding of functions in $\mathscr{F}_p^{(s_1,s_2)}$

We have shown in the previous section (see Theorem 6.1) that functions in the space $\mathscr{F}_p^{(s_1,s_2)}$ can be approximated by functions from $\mathscr{C}_{8n,r}$ to L_p accuracy $\mathscr{O}(n^{-s})$, where $s = (1/s_1 + 1/s_2)^{-1}$. In this section, we want to discuss a theoretical encoder of functions from this class, i.e. how to assign to f a bitstream (a sequence of $0,1$'s) from which f can be approximately recovered. The encoder is based on quantization of the piecewise polynomial domains that approximate the level sets of f and thereby gives an upper bound for the Kolmogorov entropy of the spaces $\mathscr{F}_p^{(s_1,s_2)}$. We shall carry out the details only in the case of functions defined on the unit square $\Omega := [0,1]^2$ in $\mathbb{R}^2$ which is a setting commonly used in image processing and compression of digital terrain maps.

Note that the unit ball $U(\mathscr{F}_p^{(s_1,s_2)}) := \{f \in L_p(\Omega) : |f|_{\mathscr{F}_p^{(s_1,s_2)}} \leq 1\}$ is not a compact subset of $L_p(\Omega)$. For example, in the case $r \geq 1$ and $p = 1$ any function $\beta\chi_R$, with $R \subset \Omega$ a square and $2\beta|R| = 1$, is in this space. By making R small, we can take an arbitrarily large number of these functions so that their supports are disjoint and hence they are an $L_1(\Omega)$ distance one apart. Therefore, we need to add some condition to get a compact set. We consider the class $K := K_p^{(s_1,s_2)} := \{f \in U(\mathscr{F}_p^{(s_1,s_2)}) : \|f\|_{L_\infty(\Omega)} \leq 1\}$. There are various less restrictive possibilities to assuming that the functions f are bounded but the above condition meshes well with typical application areas such as image and surface compression.

To encode $f \in K$, we shall approximate it by a function $S \in \mathscr{C}_{8n,r}$ and then encode S. We fix the value of n which corresponds to the accuracy we will want in the final approximation of f by the encoded function. We define $k := \lceil sp \rceil$ as the smallest integer larger than sp and choose $\alpha := 2^{-m+1}$, where m is the smallest positive integer such that $\alpha \leq n^{-k-1}$. The value of α and m are now fixed for the remainder of this section. We denote by $\alpha\mathbb{Z}^2$ the lattice consisting of all points in $\mathbb{R}^2$ of the form $\alpha(i_1,i_2)$ with $i_1,i_2 \in \mathbb{Z}$.

We begin by describing an encoding of domains in $\mathbf{D}(\bar{n},r)$ whenever $\bar{n} \leq n$. We assume for the moment that the order r of the polynomials is known and fixed. (Later we shall actually encode the value of r.) Given $\Lambda \in \mathbf{D}(\bar{n},r)$, there are integers $\bar{m}_1, \bar{m}_2 \geq 0$, with $\bar{m}_1 + \bar{m}_2 \leq \bar{n}$, such that Λ is the union of $\bar{m}_1$ triangular domains Δ_j, $j = 1,\dots,\bar{m}_1$, together with $\bar{m}_2$ fundamental r-polynomial graph domains $G(P_j,R_j)$, $j = 1,\dots,\bar{m}_2$ (see (4.1)) . Each of the polynomials P_j has order $\leq r$ and each rectangle $R_j = I_j \times J_j$ is with respect to a coordinate system with basis vectors $(\mathbf{u}_j,\mathbf{v}_j)$.

In order to make our arguments simpler, we shall without loss of generality assume that $\Omega = [1/4,3/4]^2$. Under this assumption, we can also require that all of the rectangles and triangular domains are contained in $[0,1]^2$. Otherwise we could replace $G(P_j,R_j)$ and Δ_j with r-polynomial graph domains and triangles with this property, without worsening the approximation, and whose total number is $C(r)\bar{n}$, where $C(r)$ is a constant depending only on r.

We shall associate to each of the components Δ_j and $G(P_j,R_j)$ at most $\le C_0 L(n)$ bits where $L(n) := \lceil \log n \rceil$ and C_0 is an integer which will depend only on r, s, and p.

Encoding triangular domains Δ: Given a triangular domain $\Delta \subset [0,1]^2$, we denote its vertices by $\mathbf{a},\mathbf{b},\mathbf{c} \in \mathbb{R}^2$. We define $\hat{\mathbf{a}},\hat{\mathbf{b}},\hat{\mathbf{c}}$, as points in $\alpha\mathbb{Z}^2$ which are each respectively the closest points from $\alpha\mathbb{Z}^2$ to $\mathbf{a},\mathbf{b},\mathbf{c}$. Here in the case of ties we use the convention of picking the point with smallest coordinates to break the tie. This is easily seen to be unique. We denote by $\hat{\Delta}$ the triangular domain with vertices $\hat{\mathbf{a}},\hat{\mathbf{b}},\hat{\mathbf{c}}$. Each of the points $\hat{\mathbf{a}},\hat{\mathbf{b}},\hat{\mathbf{c}}$ is of the form $\alpha(i_1,i_2)$ where i_1,i_2 are integers. Since the vertices $\mathbf{a},\mathbf{b},\mathbf{c}$ are in $[0,1]^2$, the integers i_1,i_2 are non-negative and $\le 1/\alpha = 2^{m-1}$. We can represent any integer $0 \le i \le 2^{m-1}$ as $i = \sum_{j=0}^{m-1} \beta_j 2^j$, with $\beta_j \in \{0,1\}$, $j = 0,\dots,m-1$. We define $B(\Delta)$ as the bitstream consisting of the $6m$ entries $0,1$ with the first m consisting of the bits of the first coordinate of $\hat{\mathbf{a}}$, the second m consisting of the bits of the second coordinate of $\hat{\mathbf{a}}$ and so on. Upon receiving the bits $B(\Delta)$ and the knowledge of m, the receiver can construct $\hat{\Delta}$. We claim that

$$d_1(\Delta,\hat{\Delta}) \le 12\alpha \le 12n^{-k-1}. \tag{6.18}$$

To prove this, we first observe that

$$|\mathbf{a}-\hat{\mathbf{a}}|, |\mathbf{b}-\hat{\mathbf{b}}|, |\mathbf{c}-\hat{\mathbf{c}}| \le \frac{\alpha}{\sqrt{2}}. \tag{6.19}$$

Since any point on the line segment $[\mathbf{a},\mathbf{b}]$ connecting $\mathbf{a}$ and $\mathbf{b}$ is a convex combination of $\mathbf{a}$ and $\mathbf{b}$ (similarly for $[\hat{\mathbf{a}},\hat{\mathbf{b}}]$), it follows that any point on $[\mathbf{a},\mathbf{b}]$ is at distance at most $\frac{\alpha}{\sqrt{2}}$ from a point on $[\hat{\mathbf{a}},\hat{\mathbf{b}}]$ and vice versa. Now, suppose we wish to bound $d_1(\Delta,\hat{\Delta})$. Every point in Δ is in the set of points $\hat{\Delta}_+$ whose distance is at most $\frac{\alpha}{\sqrt{2}}$ from $\hat{\Delta}$. Similarly, consider the set of points $\hat{\Delta}_-$ which are in the interior of $\hat{\Delta}$ and distance $\ge \frac{\alpha}{\sqrt{2}}$ from the boundary of $\hat{\Delta}$. This set of points is contained in Δ. Hence,

$$d_1(\Delta,\hat{\Delta}) \le |\hat{\Delta}_+ \setminus \hat{\Delta}_-| \le 6(1+\alpha)\alpha. \tag{6.20}$$

Here for the last inequality, we argue as follows. We claim that $\hat{\Delta}_+ \setminus \hat{\Delta}_-$ is contained in a union of three rectangles associated to each of the sides of $\hat{\Delta}$. For example, for the side $[\mathbf{a},\mathbf{b}]$ we symmetrically enlarge its length by $\sqrt{2}\alpha$ resulting in a line segment L. The length of L is $\le \sqrt{2}+\sqrt{2}\alpha$. We take the rectangle whose median is L and whose width is $\sqrt{2}\alpha$. This rectangle has area $\le 2\alpha(1+\alpha)$. We repeat this procedure for the other two sides of Δ. Then the three rectangles cover $\hat{\Delta}_+ \setminus \hat{\Delta}_-$ and so we arrive at (6.20). From (6.20), using the value of α, we derive (6.18).

Encoding fundamental r-polynomial graph domains $G(P,R)$: This encoding will be a little more difficult to analyze. Any rectangle $R \subset [0,1]^2$ has sides of length at most $\sqrt{2}$, so it is the rotation (about the origin) and then translation of a rectangle R_0 of the form $R_0 = [0,a] \times [0,b]$, $0 < a,b \leq \sqrt{2}$, where this representation of R_0 is with respect to the standard coordinate system $\mathbf{e_1} = (0,1), \mathbf{e_2} = (1,0)$. The rotation is given by describing the unit vector $\mathbf{u}$ which is the image of $(0,1)$. The translation is given by a vector $\mathbf{t} = (t_1,t_2)$ with $0 \leq t_1,t_2 < 1$. We shall encode R by giving in order the first m binary bits of a, the first m binary bits of b, the first m binary bits of the first coordinate of $\mathbf{u}$, the first m binary bits of the second coordinate of $\mathbf{u}$, the first m binary bits of the first coordinate of $\mathbf{t}$ and lastly the first m binary bits of the second coordinate of $\mathbf{t}$. This will utilize exactly $6m$ bits. We denote by $\hat{R}$ and $\hat{R}_0$ the open rectangles that will be recovered from these bits (for $\hat{R}_0$ we need just the first $2m$ bits). Notice that if either a or b is smaller than 2^{-m} then these rectangles will be empty.

The domain $G := G(P,R)$ is likewise the rotation and then translation of a fundamental r-polynomial graph domain $G_0 := G(P_0,R_0)$ with R_0 as above and the rotation and translation are the same as in the transformation from R_0 to R . We need to describe how we encode the polynomial P_0 with respect to the standard coordinate system. We assume that a, $b > 2^{-m}$ since otherwise $\hat{R}_0$ is empty. We know that P_0 takes values in $(0,\sqrt{2})$ on the interval $[0,a]$. We represent $P_0(x) = \sum_{j=0}^{r-1} \gamma_j x^j$. We know that $\gamma_0 = P(0)$ is in $(0,\sqrt{2})$. From Markov's inequality for polynomials, we get that

$$|\gamma_j| \leq \frac{r^{2j}}{j!}(a/2)^{-j}\sqrt{2} < B2^{mr}, \quad B := \sqrt{2}r^{2r}2^r. \tag{6.21}$$

We encode the polynomial P_0 by sending the first $2mr$ bits for each of the numbers $B^{-1}2^{-mr}|\gamma_j|$ which are in $(0,1)$ and one bit for the sign of γ_j. Knowing the values of m and r, these bits will allow the receiver to determine B and to construct a $\hat{\gamma}_j$ such that

$$|\gamma_j - \hat{\gamma}_j| \leq B2^{-mr}, \quad j = 0,1,\dots,r-1. \tag{6.22}$$

This means that the polynomial $\hat{P}_0 := \sum_{j=0}^{r-1} \hat{\gamma}_j x^j$ satisfies

$$|P_0(x) - \hat{P}_0(x)| \leq Br(\sqrt{2})^r 2^{-mr}, \quad x \in [0,a]. \tag{6.23}$$

In summary, each fundamental polynomial domain $G(P,R)$ is encoded by a bitstream $B(G(P,R))$ consisting of the $6m$ bits $B(R)$ followed by the $(2mr+1)r$ bits which encode the polynomial P. Hence this bitstream has length $(2mr+1)r+6m$. Upon receiving this bitstream, the receiver can construct the rectangle $\hat{R}_0$ and the polynomial cap $\hat{P}_0$. Let us denote by $\hat{G}_0$ the domain formed from this rectangle by using this polynomial cap. It is clear that this domain approximates G_0 with accuracy

$$d_1(G_0,\hat{G}_0) \leq Br(\sqrt{2})^{r+1}2^{-mr} + \sqrt{2}2^{-m} \leq Br((\sqrt{2})^{r+1} + \sqrt{2})n^{-k-1}, \tag{6.24}$$

where the first term represents the error caused by having the wrong cap (see (6.23)) and the second term represents the error in approximating a by $\hat{a}$.

The receiver can now use the approximate rotation $\hat{\mathbf{u}}$ and the approximate translation $\hat{\mathbf{t}}$ to construct a domain $\hat{G} := G(\hat{P}, \hat{R})$. We claim that

$$d_1(G, \hat{G}) \leq C2^{-m}, \tag{6.25}$$

with C a constant depending only on r and s. Since we already know that $\hat{G}_0$ approximates G_0 to this accuracy, to establish this claim we need only consider the effects of using quantized rotations and translations versus the true rotations and translations. We shall confine ourselves to sketching how to treat the effect of the quantized rotation which is the most difficult case. The difference between the true rotation and the quantized rotation is a small rotation ρ whose angle is $\leq C2^{-m}$. Estimating the effect of this error is equivalent to estimating the distance between $\rho\hat{G}_0$ and G_0. This leads to bounding the measure of the two sets $\rho\hat{G}_0 \setminus G_0$ and $G_0 \setminus \rho\hat{G}_0$. We describe the argument only for the first of these. Here it is enough to bound the measure of the set $G_1 := \rho\hat{G}_0 \setminus \hat{G}_0$ since we already have a good bound for $\hat{G}_0 \setminus G_0$.

The key observation for estimating $|G_1|$ is the fact that for any $x \in \Omega$, $|\rho x - x| \leq C_0 2^{-m}$ with an absolute constant C_0. This means that any point in $\rho\hat{G}_0$ is at distance at most $C_0 2^{-m}$ from $\hat{G}_0$. Therefore, similar to our argument for triangular domains, we can cover G_1 by four sets. Each of these consists of the set of points which are at distance $\leq C_0 2^{-m}$ respectively from the sides, bottom and the polynomial cap. The same argument as for triangular domains shows that the first three of these sets all have measure $\leq C2^{-m}$ with C an absolute constant. The set corresponding to the cap has measure $\leq C2^{-m}$, with C now depending on r, because the length of the graph of P_0 on $[0, \hat{a}]$ is $\leq C$ with C depending only on r. This proves that $|G_1| \leq C2^{-m}$ as desired.

Encoding a domain $\Lambda \in \mathbf{D}(\bar{n}, r)$: We assume for this encoding that the receiver knows n, k, r from which he can deduce m. (Later when discussing the encoding of functions f this information will be sent to the receiver.) To encode a domain $\Lambda \in \mathbf{D}(\bar{n}, r)$, we send the following bitstream $B(\Lambda)$. The first bits are $\bar{m}_1$ ones followed by a zero. This will enable the receiver to know $\bar{m}_1$. Next, we send $\bar{m}_2$ ones followed by a zero. This allows the receiver to know $\bar{m}_2$ and hence $\bar{n} = \bar{m}_1 + \bar{m}_2$. We next send the $\bar{m}_1$ bitstreams $B(\Delta_j)$, $j = 1, \ldots, \bar{m}_1$. Since the receiver knows the length of each of these bitstreams is $6m$ and the total number is $\bar{m}_1$, he will know that he is receiving $6\bar{m}_1 m$ bits that describe the triangles $\hat{\Delta}_j$. Similarly, we next send the bitstreams $B(G(P_j, R_j))$. The receiver knows that there will be $\bar{m}_2$ of these bitstreams and each will be of length $(2mr+1)r + 6m$. It follows that

$$\#(B(\Lambda)) \leq C\bar{n}m \leq C\bar{n}\log n, \tag{6.26}$$

with C a constant depending only r, s, and p. Upon obtaining these bits, the receiver can construct $\hat{\Lambda} := \left[\bigcup_{j=1}^{\bar{m}_1} \hat{\Delta}_j\right] \cup \left[\bigcup_{j=1}^{\bar{m}_2} G(\hat{P}_j, \hat{R}_j)\right]$ which satisfies

$$d_1(\Lambda, \hat{\Lambda}) \leq \sum_{j=1}^{\bar{m}_1} d_1(\Delta_j, \hat{\Delta}_j) + \sum_{j=1}^{\bar{m}_2} d_1(G(P_j, R_j), G(\hat{P}_j, \hat{R}_j)) \leq Cn \cdot n^{-k-1} \leq Cn^{-k}, \tag{6.27}$$

where we have used the estimates (6.18) and (6.25) together with the fact that $\bar{n} \leq n$.

Encoding functions in $K_p^{(s_1,s_2)}$: Now, suppose we are given a function $f \in K_p^{(s_1,s_2)}$. As before, we decompose f into its positive and negative parts and encode each separately. Thus, it is sufficient to describe how we encode a nonnegative function $f \in K_p^{(s_1,s_2)}$. From Theorem 6.1, we know that there is a function $S = \sum_{j=1}^{\ell} c_j \chi_{\Lambda_j} \in \mathscr{C}_{4n,r}$ with $0 < c_j \leq 1$, such that

$$\|f - S\|_{L_p} \leq Cn^{-s}. \tag{6.28}$$

with C an absolute constant, $\Lambda_j \in \mathbf{D}(n_j, r)$ and $\sum_{j=1}^{\ell} n_j \leq 4n$. To encode c_j, we shall take m binary bits (denoted by $B(c_j)$) for each c_j, $j = 1, \ldots, \ell$. Obtaining these bits allows the receiver to construct the coefficients $\hat{c}_j$, $j = 1, \ldots, \ell$, which approximate c_j to accuracy 2^{-m}:

$$|c_j - \hat{c}_j| \leq 2^{-m} \leq n^{-k-1}. \tag{6.29}$$

Our bitstream for f now takes the following form. We first send n ones followed by a zero. The receiver will then know the value of n. We next send k ones followed by a zero and then r ones followed by a zero. The receiver will now be able to determine k, m, r. Next we send ℓ ones followed by a zero. Obtaining these bits allows the receiver to determine ℓ. Then we send the m bits $B(c_1)$ followed by the bitstream $B(\Lambda_1)$, followed by $B(c_2), B(\Lambda_2)$ and so on. We denote the total bitstream by $B(f)$. Let us note that the length of this bitstream is

$$\#(B(f)) \leq (n+1) + (k+1) + (r+1) + (\ell+1) + \ell m + C\sum_{j=1}^{\ell} n_j \log n \leq C_0 n \log n, \tag{6.30}$$

where C_0 depends only on r, s, and p. Here we have used the estimate (6.26) to bound $B(\Lambda_j)$ for each j and the facts that $\ell \leq 4n$, $\sum_{j=1}^{\ell} n_j \leq 4n$, and $m \leq C \log n$. Upon receiving these bits, the receiver can construct the function

$$\hat{S} = \sum_{j=1}^{\ell} \hat{c}_j \chi_{\hat{\Lambda}_j}. \tag{6.31}$$

The following theorem gives the properties of the encoding/decoding.

Theorem 6.2. *If f is a nonnegative function in $K_p^{(s_1,s_2)}$ and n is any integer, then the bitstream $B(f)$ consists of $\le C_0 n \log n$ bits where the constant C_0 depends only on r, s, and p. The bitstream $B(f)$ is decoded into a function $\hat{S} \in \mathscr{C}_{4n,r}$ such that*

$$\|f - \hat{S}\|_{L_p} \le C n^{-s}, \tag{6.32}$$

where

$$\frac{1}{s} = \frac{1}{s_1} + \frac{1}{s_2}, \tag{6.33}$$

and C depends only on r, s, and p.

Proof: We can write

$$f - \hat{S} = f - S + S - S_0 + S_0 - \hat{S}, \tag{6.34}$$

where $S = \sum_{j=1}^{\ell} c_j \chi_{\Lambda_j}$ is the function from Theorem 6.1 and $S_0 := \sum_{j=1}^{\ell} \hat{c}_j \chi_{\Lambda_j}$. By Theorem 6.1, $\|f - S\|_{L_p}$ is bounded by the right side of (6.32). From (6.29), we have

$$\|S - S_0\|_{L_p} \le \sum_{j=1}^{\ell} |c_j - \hat{c}_j| \le \ell n^{-k-1} \le 4n^{-k} \le 4n^{-sp} \le 4n^{-s}, \tag{6.35}$$

where we have used the fact that $\ell \le 4n$. To estimate $S_0 - \hat{S}$, we use Lemma 5.1 which gives

$$\|S_0 - \hat{S}\|_{L_p} \le \left[\int_0^1 \delta(z) d\mu_p(z) \right]^{1/p}, \tag{6.36}$$

where $\delta(z) = d_1(\Lambda_j, \hat{\Lambda}_j)$, $z \in (\hat{z}_{j-1}, \hat{z}_j)$, $j = 1, \dots, \ell$. It follows from (6.27) that $\delta(z) \le C n^{-k}$ for all z, and therefore

$$\|S_0 - \hat{S}\|_{L_p} \le C^{1/p} n^{-k/p} \le C^{1/p} n^{-s}. \tag{6.37}$$

We have therefore proved the error estimate (6.32). □

7 Generalizations of the smoothness spaces $\mathscr{F}_p^{(s_1,s_2)}$

In this section, we shall mention ways in which we can generalize the smoothness spaces $\mathscr{F}_p^{(s_1,s_2)}$. There will be two ingredients to these generalizations. The first is to incorporate other possibilities for measuring the smoothness of the evolution of level sets. The second is to indicate how the notion of fundamental domains can be generalized by using alternatives to polynomial domains. Our development will be given without proofs since the proofs of the results stated in this section would be very similar to what we have already done in the previous sections of the paper.

7.1 Smoother evolution

The spaces that we have defined and analyzed in the previous sections were based on the simplest ideas to quantify how smoothly level sets evolve with z. Since we have used piecewise constant approximation (for $A(f,z)$), we measured first order smoothness of the evolution. We have chosen to emphasize that approach because when speaking of higher order evolution of level sets, there is no canonical way to measure such an evolution. For example, given two level sets $\Omega(f,z_0)$ and $\Omega(f,z_1)$ corresponding to levels z_0 and z_1, there is no clearly preferred way to describe a higher order interpolation between these sets. There are several possible approaches such as constrained minimization, evolution equations, or interpolation of sets. Each of these would lead to a notion of higher order smoothness which could be used to define analogues of the spaces $\mathscr{F}_p^{(s_1,s_2)}$. A good choice for such a definition would depend on the potential application area and what interesting theorems could be proved under such a definition. So rather than presuppose the definition, we shall proceed entirely formally leaving open the particular option to be chosen. What we want to point out in this section are sufficient conditions on such an interpolation method in order that the analogue of Theorem 6.1 holds for the corresponding new smoothness spaces.

Given $0 \le t_0 < \cdots < t_m$ and measurable sets $\Lambda_0, \Lambda_1, \cdots, \Lambda_m \subset \mathbb{R}^d$, we say that $\mathscr{I}$ is a set interpolation operator (of order m) if it assigns to every $z \ge 0$ a set $\Lambda(z) := \mathscr{I}(z; t_0, \Lambda_0, \ldots, t_m, \Lambda_m)$ with $\Lambda(t_j) = \Lambda_j$, $j = 0, 1, \ldots, m$. As a minimal requirement, we will assume that the sets $\{(z,x) : x \in \Lambda(z)\} \subset \mathbb{R}_+ \times \mathbb{R}^d$ are measurable, so in particular the function $\varphi_x(z) = \chi_{\Lambda(z)}(x)$ is measurable for almost all x. We will require the following stability property for the set interpolation operator. Suppose that $\{\Lambda_j\}$ and $\{\bar{\Lambda}_j\}$, $j = 0, \ldots, m$, are two given families of sets (we think of $\bar{\Lambda}_j$ as an approximation of Λ_j). Then, we will require that

$$d_1(\mathscr{I}(z; t_0, \Lambda_0, \ldots, t_m, \Lambda_m), \mathscr{I}(z; t_0, \bar{\Lambda}_0, \ldots, t_m, \bar{\Lambda}_m)) \le C \max_{0 \le j \le m} d_1(\Lambda_j, \bar{\Lambda}_j), \tag{7.1}$$

where $z \ge 0$ and C depends only on $\mathscr{I}$. Starting with such a method of set interpolation, we can define approximation spaces that describe how smoothly the level sets of a function evolve.

Fix $m \ge 0$, which is the order of the set interpolation operator. Given a non-negative function f and any points $t_0 < t_1 < \cdots < t_m$, we shall use the notation $\Omega(z; t_0, \ldots, t_m) := \mathscr{I}(z; t_0, \Omega(f,t_0), \ldots, t_m, \Omega(f,t_m))$, where $z \ge 0$. Given any finite interval $I \subset \mathbb{R}_+$, we can measure the average performance of such interpolation by

$$E(f,I)_p^p := \frac{1}{\nu_p(\Delta_I)} \int_{\Delta_I} \int_I d_1(\Omega(f,z), \Omega(z; t_0, \ldots, t_m))\, d\mu_p(z)\, d\nu_p(\mathbf{t}), \tag{7.2}$$

where ν_p is the $m+1$ dimensional product of μ_p and $\Delta_I := \Delta_I(0,1,\ldots,m)$ is the $(m+1)$-dimensional simplex consisting of all $\mathbf{t} = (t_0, \ldots, t_m)$ with $t_0 \le \cdots \le t_m$,

$t_j \in I$, $j = 0, \dots, m$. Clearly we have that

$$\nu_p(\Delta_I) = \frac{1}{(m+1)!}\mu_p(I)^{m+1}. \tag{7.3}$$

Now, given any $n \geq 1$, we consider any sequence $0 =: z_0 < z_1 < \cdots < z_{n-1} < z_n$, and the intervals $I_j := (z_{j-1}, z_j)$, $j = 1, \dots, n+1$, where $z_{n+1} := \infty$. Given any nonnegative function f and any $n \geq 1$, we define $\sigma_n(f)_p^p$ as

$$\sigma_n(f)_p^p := \inf_{z_0,\dots,z_n} \sum_{j=1}^{n+1} E(f, I_j)_p^p, \tag{7.4}$$

where for an unbounded interval I, $E(f,I)_p^p := \int_I |\Omega(f,z)|\, d\mu_p(z)$, with $|\Omega(f,z)|$ the Lebesgue measure of $\Omega(f,z)$. We set $\sigma_0(f)_p^p := \int_0^\infty |\Omega(f,z)|\, d\mu_p(z) = \|f\|_p^p$ (see (5.1)). We can use this σ_n as a replacement for (5.13) and define the approximation spaces $\mathscr{B}_p^s$ as spaces of nonnegative functions on $\mathbb{R}^d$ such that

$$\sigma_n(f)_p \leq M(n+1)^{-s}, \quad n \geq 0. \tag{7.5}$$

We define $|f|_{\mathscr{B}_p^s}$ as the smallest M for which (7.5) holds.

We refer the reader back to §4 where we introduced the approximation of level sets by piecewise polynomial domains and the approximation classes $\mathscr{A}^s$ for domains.

We can now give our generalization of the spaces $\mathscr{F}_p^{(s_1,s_2)}$. Given $p \in [1,\infty)$ and $s_1, s_2 > 0$, we define $\mathscr{F}_p^{(s_1,s_2)}(\mathscr{I})$ as the space of all functions on $\mathbb{R}^d$ satisfying (6.1) and

$$|f_+|_{\mathscr{B}_p^{s_2}} + |f_-|_{\mathscr{B}_p^{s_2}} := M_2(f) < \infty, \tag{7.6}$$

where the spaces $\mathscr{B}_p^s$ are defined by (7.5). For the remainder of this section we will assume that $\mathscr{I}$ is fixed and we will abbreviate $\mathscr{F}_p^{(s_1,s_2)}(\mathscr{I})$ to $\mathscr{F}_p^{(s_1,s_2)}$. We have the following lemma.

Lemma 7.1. *Given any nonnegative function f, $1 \leq p < \infty$, $s > 0$, and any finite interval $I \subset \mathbb{R}_+$, there exist a constant C, depending only on m, and points $t_0 < t_1 < \cdots < t_m$ from I, such that*

$$\int_I d_1(\Omega(f,z), \Omega(z; t_0, \dots, t_m))\, d\mu_p(z) \leq CE(f,I)_p^p, \tag{7.7}$$

and

$$|\Omega(f,t_j)|_{\mathscr{A}^s} \leq C\frac{1}{\mu_p(I)} \int_I |\Omega(f,z)|_{\mathscr{A}^s}\, d\mu_p(z), \quad j = 0, \dots, m. \tag{7.8}$$

Using this lemma, we can provide a method that uses the set interpolation operator $\mathscr{I}$ to approximate a given function $f \in \mathscr{F}_p^{(s_1,s_2)}$. The construction is analogous to that given in Section 6.1, where we approximate f by linear combination of characteristic functions.

Given $n \geq 1$, s_1 and s_2, we denote by $s := (1/s_1 + 1/s_2)^{-1}$, and define $n_1 := \lceil n^{\frac{s}{s_1}} \rceil$ and $n_2 = \lceil n^{\frac{s}{s_2}} \rceil$. We first choose $0 = z_0 < z_1 < \cdots < z_{n_2} < \infty$, such that for $I_j := (z_{j-1}, z_j)$, $j = 1, \ldots, n_2 + 1$, where $z_{n_2+1} := \infty$, we have (see (7.4)),

$$\sum_{j=1}^{n_2+1} E(f, I_j)_p^p \leq 2\sigma_{n_2}(f)_p^p. \tag{7.9}$$

On each of the intervals I_j, $j = 1, \ldots, n_2$, we use Lemma 7.1 and choose points $t_{0,j} < t_{1,j} < \cdots < t_{m,j}$, such that

$$\int_{I_j} d_1(\Omega(f,z), \Omega(z; t_{0,j}, \ldots, t_{m,j})) d\mu_p(z) \leq CE(f, I_j)_p^p, \tag{7.10}$$

and

$$|\Omega(f, t_{i,j})|_{\mathscr{A}^{s_1 p}} \leq C \frac{1}{\mu_p(I_j)} \int_{I_j} |\Omega(f,z)|_{\mathscr{A}^{s_1 p}} \, d\mu_p(z), \quad i = 0, \ldots, m. \tag{7.11}$$

For each of the level sets $\Omega(f, t_{i,j})$ we choose an approximation $\bar{\Omega}_{i,j} \in \mathbf{D}(n_1, r)$ such that $d_1(\Omega(f, t_{i,j}), \bar{\Omega}_{i,j}) \leq n_1^{-s_1 p} |\Omega(f, t_{i,j})|_{\mathscr{A}^{s_1 p}}$, and therefore from (7.11) we have for $i = 0, \ldots, m$, and $j = 1, \ldots, n_2$,

$$d_1(\Omega(f, t_{i,j}), \bar{\Omega}_{i,j}) \leq C n_1^{-s_1 p} \frac{1}{\mu_p(I_j)} \int_{I_j} |\Omega(f,z)|_{\mathscr{A}^{s_1 p}} \, d\mu_p(z). \tag{7.12}$$

We now define

$$\bar{\Omega}_n(z) := \mathscr{I}(z; t_{0,j}, \bar{\Omega}_{0,j}, \ldots, t_{m,j}, \bar{\Omega}_{m,j}), \quad z \in I_j, \quad j = 1, \ldots, n_2. \tag{7.13}$$

From the sets $\bar{\Omega}_n(z)$, we construct the approximation

$$S_n(x) := \int_0^\infty \chi_{\bar{\Omega}_n(z)}(x) \, dz. \tag{7.14}$$

Notice that S_n is completely determined by the parameters $\{t_{i,j}\}$, $i = 0, \ldots, m$, $j = 1, \ldots n_2$, and the domains $\bar{\Omega}_{i,j} \in \mathbf{D}(n_1, r)$. To describe a simplex in $\mathbb{R}^d$ we need $c_1(d)$ parameters, to describe a parallelepiped no more than $c_2(d)$ parameters and to describe a polynomial of order r no more than $c_3(r,d)$ parameters, where $c_3(r,d)$ is a constant that depends on d and r. So for each $t_{i,j}$, there are at most $C(r,d)n_1$ parameters that determine $\bar{\Omega}_{i,j}$. Thus, S_n is determined by at most $C(r,d)mn_1n_2 \leq C(r,d,m)n$ parameters in all and the following theorem holds.

Theorem 7.2. *If $f \in \mathscr{F}_p^{(s_1,s_2)}$, $1 \leq p < \infty$, then*

$$\|f - S_n\|_{L_p} \leq C|f|_{\mathscr{F}_p^{(s_1,s_2)}} n^{-s}, \quad n = 1, 2, \ldots, \tag{7.15}$$

where $\frac{1}{s} = \frac{1}{s_1} + \frac{1}{s_2}$ and the constant C depends on m, the interpolation process $\mathscr{I}$ and p.

We omit the proof which is almost identical to that of Theorem 6.1.

Notice that if $m=0$, the interpolation operator $\mathscr{I}$, defined as $\mathscr{I}(z;t_0,\Lambda_0)=\Lambda_0$ for $z\geq 0$, satisfies condition (7.1). Now, using (7.2), we get

$$E(f,I)_p^p=\frac{1}{\mu_p(I)}\int_I\int_I|A(f,z)-A(f,t)|\,d\mu_p(z)d\mu_p(t). \tag{7.16}$$

So we have

$$\begin{aligned} E(f,I)_p^p &\geq \int_I\left|A(f,z)-\frac{1}{\mu_p(I)}\int_I A(f,t)\,d\mu_p(t)\right|d\mu_p(z)\\ &\geq \inf_{c\in\mathbb{R}}\|A(f,\cdot)-c\|_{L_1(I,d\mu_p)}. \end{aligned} \tag{7.17}$$

Adding and subtracting an arbitrary constant c under the absolute value in (7.16), we get

$$E(f,I)_p^p\leq 2\inf_{c\in\mathbb{R}}\|A(f,\cdot)-c\|_{L_1(I,d\mu_p)}. \tag{7.18}$$

Comparing (7.17) and (7.18) with (5.10), we see that for such interpolation process $\mathscr{I}$ we obtain the spaces $\mathscr{F}_p^{(s_1,s_2)}$ defined in §6 and Theorem 7.2 becomes Theorem 6.1.

7.2 General approximating sets

The second generalization of our approach to generating smoothness spaces is to use other families $\mathbf{D}(n)$ for approximating level sets, instead of the families $\mathbf{D}(n,r)$. Let $\mathbf{D}(n)$, $n=0,1,2,\dots$, $\mathbf{D}(0):=\emptyset$, be an increasing sequence of families of measurable subsets of $\mathbb{R}^d$. For practical purposes, we assume that $\hat{\Omega}\in\mathbf{D}(n)$ can be described by n parameters in an efficient way. For a set $\Omega\subset\mathbb{R}^d$, we define $\sigma_n(\Omega)=\inf_{\hat{\Omega}\in\mathbf{D}(n)}d_1(\Omega,\hat{\Omega})$. For each $s>0$, we define the class $\mathscr{A}^s$ of subsets $\Omega\subset\mathbb{R}^d$, such that $\sigma_n(\Omega)\leq M(n+1)^{-s}$, $n\geq 0$. The smallest M for which this inequality holds for all n is defined to be $|\Omega|_{\mathscr{A}^s}$. Next, we associate with these new approximation families the following linear combinations of more general characteristic functions,

$$\mathscr{C}(n,\mathbf{D})=\left\{S\,:\,S=\sum_{j=1}^m c_j\chi_{\Lambda_j}\quad \Lambda_j\in\mathbf{D}(n_j)\quad \sum_{j=1}^m n_j\leq n\right\}. \tag{7.19}$$

Using these new approximation spaces $\mathscr{A}^s$, we define for $p\in[1,\infty)$ and $s_1,s_2>0$, smoothness spaces $\mathscr{F}_p^{(s_1,s_2)}$. We say that a function $f=f_+-f_-$, (f_+ and f_- are its positive and negative parts) on $\mathbb{R}^d$ is in $\mathscr{F}_p^{(s_1,s_2)}$ if

$$M_1(f):=\left\{\int_0^\infty|\Omega(f_+,z)|_{\mathscr{A}^{s_1p}}d\mu_p(z)\right\}^{1/p}+\left\{\int_0^\infty|\Omega(f_-,z)|_{\mathscr{A}^{s_1p}}d\mu_p(z)\right\}^{1/p}<\infty,$$

and $M_2(f):=|A(f_+,\cdot)|_{\mathscr{B}_p^{s_2}}+|A(f_-,\cdot)|_{\mathscr{B}_p^{s_2}}<\infty$, where $\mathscr{B}_p^s$ are the spaces introduced in Section 5.1.

We can repeat the arguments from Section 6.1 and obtain the following generalization of Theorem 6.1.

Theorem 7.3. *If $f \in \mathscr{F}_p^{(s_1,s_2)}$, with $1 \le p < \infty$, and $s_1, s_2 > 0$, then there exists an $S \in \mathscr{C}(8n, \mathbf{D})$, such that*

$$\|f - S\|_{L_p} \le 3^{1/p} |f|_{\mathscr{F}_p^{(s_1,s_2)}} n^{-s},$$

where $\frac{1}{s} := \frac{1}{s_1} + \frac{1}{s_2}$.

Note that for each sequence of sets $\mathbf{D}(n)$ we get (possibly) different smoothness spaces $\mathscr{F}_p^{(s_1,s_2)}$.

In previous sections, we have used fundamental r-polynomial graph domains and simplexes as the basic building blocks in the definition of the spaces $\mathbf{D}(n,r)$. We started with these building blocks because they are simple to encode. However there are many other ways to generate interesting classes $\mathbf{D}(n)$, and we list a few.

- In $\mathbb{R}^2$, we define $\mathbf{D}(n)$ as unions of polygons whose total number of vertices is n.
- In $\mathbb{R}^d$, we define $\mathbf{D}(n)$ as unions of n convex bodies. Those classes are rather impossible to describe by n parameters.
- In the definition of r-polynomial graph domains, for example, we could replace graphs of polynomials of order r, with graphs of rational functions of order r, graphs of trigonometric polynomials of order r, or algebraic curves of order r. Actually, we can replace polynomials of order r by any set of functions. Note that the empty set of functions would result with $\mathbf{D}(n)$ being a union of n simplexes.
- We could construct classes $\mathbf{D}(n) := \mathbf{D}^d(n)$ of subsets of $\mathbb{R}^d$ by induction on the dimension d. For example, if for $d = 2$ we set $\mathbf{D}^2(n) = \mathbf{D}(n,r)$, we construct a set of functions $\mathscr{C}(n, \mathbf{D}^2)$, as defined in (7.19) (see also (6.4)). Using this class of functions we define $\mathbf{D}^3(n)$ as fundamental graph domains whose caps are functions from $\mathscr{C}(n, \mathbf{D}^2)$. Continuing in this way we define the classes $\mathbf{D}^d(n)$ inductively.

Acknowledgment. This research was supported by the Office of Naval Research Contracts ONR-N00014-03-1-0051, ONR/DEPSCoR N00014-03-1-0675 and ONR/DEPSCoR N00014-00-1-0470; the ARO Contract DAAD 19-02-1-0028; DARPA Grant HM1582-05-20001; the ARO/MURI Grant W911NF-07-1-0185; the NSF contracts DMS-0221642 and DMS 0505501; the NRC New Investigators Twinning Program 2003-2004; and the Polish Grant KBN 1 PO03A 038 27 and the Foundation for the Polish Science.

Bibliography

[1] Candes E. , and Donoho D. New tight frames of curvelets and optimal representation of objects with piecewise C^2 singularities. *Comm. Pure Appl. Math.* **57** (2004), 219–266.

[2] Dafermos C. *Hyperbolic Conservation Laws in Continuum Physics*. Grundlehren, vol. 325, Springer, Berlin, 2005.

[3] DeVore R. Nonlinear approximation. *Acta Numer.* **7** (1998), 51–150.

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

[5] DeVore R., and Sharpley R. Besov spaces on domains in $\mathbb{R}^d$. *Trans. Amer. Math. Soc.* **335** (1993), no1, 843–864.

[6] DeVore R., and Lorentz G.G. *Constructive Approximation*. Grundlehren, vol. 283, Springer, Berlin, 1993.

[7] DeVore R., and Lucier B. High order regularity for conservation laws *Indiana Math. J.* **39** (1990), 413–430.

[8] Federer H. *Geometric measure theory*. Springer Verlag, Springer, Berlin 1969 (reprinted 1996).

[9] Hadwiger H. Eulers charakteristik und kombinatorische geometrie. *J. Reine Angew. Math.* **194** (1955), 101–110.

[10] Osher S., and Fedkiw R. *Level Set Methods and Dynamic Implicit Surfaces*. Applied Math. Sci., vol 153, Springer, Berlin, 2003.

[11] Stein E. *Singular Integrals and Differentiability Properties of Functions*. Princeton University Press, Princeton, 1970.

[12] Steiner J. Einige gesetze über die theilung der ebene und des raumes *J. Reine Angew. Math* **1** (1826), 349–364.

[13] Wetzel J. On the division of the plane by lines. *Amer. Math. Monthly* **85** (1978), no 8, 647–656.

Received Month 200X.