Constr. Approx. (2004) 20: 625–646
DOI: 10.1007/s00365-003-0549-y

CONSTRUCTIVE APPROXIMATION

On Bivariate Smoothness Spaces Associated with Nonlinear Approximation

S. Dekel, D. Leviatan, and M. Sharir

Abstract. In recent years there have been various attempts at the representations of multivariate signals such as images, which outperform wavelets. As is well known, wavelets are not optimal in that they do not take full advantage of the geometrical regularities and singularities of the images. Thus these approaches have been based on tracing curves of singularities and applying bandlets, curvelets, ridgelets, etc. (e.g., [3], [4], [8], [15], [18], [26], [27], [29]), or allocating some weights to curves of singularities like the Mumford–Shah functional [25] and its modifications. In the latter approach a function is approximated on subdomains where it is smoother but there is a penalty in the form of the total length (or other measurement) of the partitioning curves. We introduce a combined measure of smoothness of the function in several dimensions by augmenting its smoothness on subdomains by the smoothness of the partitioning curves. Also, it is known that classical smoothness spaces fail to characterize approximation spaces corresponding to multivariate piecewise polynomial nonlinear approximation. We show how the proposed notion of smoothness can almost characterize these spaces. The question whether the characterization proposed in this work can be further "simplified" remains open.

1. Introduction

Let $\Omega \subset \mathbb{R}^2$ be a bounded connected domain whose boundary is piecewise Lipschitz smooth. It is constructive to think of $\Omega = [0,1]^2$. For $f : \Omega \subseteq \mathbb{R}^m \to \mathbb{R}^n$, $h \in \mathbb{R}^m$, and $r \in \mathbb{N}$ we recall the rth-order difference operator $\Delta_h^r(f) : \Omega \subseteq \mathbb{R}^m \to \mathbb{R}^n$,

$$\Delta_h^r(f,x) := \Delta_h^r(f,\Omega,x) := \begin{cases} \sum_{k=0}^{r} (-1)^{r+k} \binom{r}{k} f(x+kh), & [x, x+rh] \subset \Omega, \\ 0, & \text{otherwise}, \end{cases}$$

where $[x, y]$ denotes the line segment connecting any two points $x, y \in \mathbb{R}^m$. The *modulus of smoothness* (see [13] for the univariate case) is defined for $0 < p \le \infty$ by

$$\omega_r(f,t)_{L_p(\Omega)} := \sup_{|h| \le t} \|\Delta_h^r(f,\Omega,x)\|_{L_p(\Omega)}, \qquad t > 0, \tag{1.1}$$

Date received: August 26, 2002. Date revised: March 20, 2003. Date accepted: April 1, 2003. Communicated by Ronald A. DeVore. Online publication: October 10, 2003.
AMS classification: 41A15, 41A17, 41A63, 65T60, 68U05, 68U10.
Key words and phrases: Multivariate nonlinear approximation, Smoothness spaces, Besov spaces, Modulus of smoothness, K-Functional, Mumford–Shah functional, Piecewise polynomials approximation, Wavelets.

where for $x \in \mathbb{R}^n$, $|x|$ denotes the norm of x. For a bounded domain $\Omega \subset \mathbb{R}^2$ and $f : \Omega \to \mathbb{R}$ we define

$$\omega_r(f, \Omega)_p := \sup_{h \in \mathbb{R}^2} \|\Delta_h^r(f, \Omega, x)\|_{L_p(\Omega)}.$$

Another notion of smoothness is the *K-functional* which employs the use of the Sobolev spaces. In the bivariate case, the Sobolev space, $W_p^r(\Omega)$, is the space of functions $g : \Omega \subseteq \mathbb{R}^2 \to \mathbb{R}$, $g \in L_p(\Omega)$, which have all their distributional derivatives of order r, $D^\gamma g := \partial^r g / \partial x_1^{\gamma_1} \partial x_2^{\gamma_2}$, $\gamma = (\gamma_1, \gamma_2)$, $\gamma_i \geq 0$, $|\gamma| := \gamma_1 + \gamma_2 = r$, in $L_p(\Omega)$. The semi-norm of this space is given by $|g|_{W_p^r(\Omega)} := \sum_{|\gamma|=r} \|D^\gamma g\|_{L_p(\Omega)} < \infty$. For $f : \Omega \subseteq \mathbb{R}^2$ we define

$$K_r(f, t)_p := K(f, t, L_p(\Omega), W_p^r(\Omega)\big) := \inf_{g \in W_p^r(\Omega)} \{\|f - g\|_{L_p(\Omega)} + t|g|_{W_p^r(\Omega)}\}. \tag{1.2}$$

We also denote for a bounded domain $\widetilde{\Omega} \subseteq \mathbb{R}^2$,

$$K_r(f, \widetilde{\Omega})_p := K_r(f, \operatorname{diam}(\widetilde{\Omega})^r)_p. \tag{1.3}$$

It is known that the above two notions of smoothness, (1.1) and (1.2), are equivalent (see [13, Chapter 6] for the univariate case, [2] for the case $\Omega = \mathbb{R}^d$, and [20] for the case of Lipschitz-graph multivariate domains). That is, for $1 \leq p \leq \infty$, there exist $C_1, C_2 > 0$ such that, for any $t > 0$,

$$C_1 K_r(f, t^r)_{L_p(\Omega)} \leq \omega_r(f, t)_{L_p(\Omega)} \leq C_2 K_r(f, t^r)_{L_p(\Omega)}. \tag{1.4}$$

It is easy to show that C_2 depends only on r. But, whenever Ω is not a univariate domain or a "simple" multivariate domain such as a cube, the constant C_1 also depends on the geometry of Ω.

In our setup, the choice of the K-functional as a measure of smoothness seems to be more appropriate. Specifically, we will measure the smoothness of a "surface" piece given by $f : \widetilde{\Omega} \subseteq \mathbb{R}^2 \to \mathbb{R}$ in the p-norm by (1.3). Observe that by (1.4) we always have

$$C_1 K_r(f, \widetilde{\Omega})_p \leq \omega_r(f, \widetilde{\Omega})_p \leq C_2 K_r(f, \widetilde{\Omega})_p, \tag{1.5}$$

with C_1 depending on the geometry of $\widetilde{\Omega}$.

The smoothness of a continuous planar curve $b : [0, 1] \to \mathbb{R}^2$ will be measured by the K-functional

$$\begin{aligned} K_r(b, t)_{\infty,1} &:= K(b, t, C[0, 1], W^{r-1}(BV[0, 1])) \\ &:= \inf_{g \in W^{r-1}(BV[0,1])} \{\|b - g\|_\infty + t|g^{(r-1)}|_{BV}\}, \end{aligned}$$

where, for a planar curve $\varphi : [0, 1] \to \mathbb{R}^2$,

$$|\varphi|_{BV} := \sup_{0=t_0<\cdots<t_n=1} \sum_{i=0}^{n-1} |\varphi(t_{i+1}) - \varphi(t_i)|.$$

Using the appropriate variant of (1.4) one shows that, for $t > 0$,

$$K_r(b, t^r)_{\infty,1} \le C K_r(b, t^r)_\infty \le C\omega_r(b, t)_\infty. \tag{1.6}$$

Using the equivalence (1.4) one can define the *Besov space* $B_q^\alpha(L_p(\Omega))$ (see [11]) as the set of functions $f \in L_p(\Omega)$ for which

$$|f|_{B_q^\alpha(L_p(\Omega))} := \begin{cases} \left(\int_0^1 (t^{-\alpha} K_r(f, t^r)_p)^q \dfrac{dt}{t}\right)^{1/q}, & 0 < q < \infty, \\ \sup\limits_{0<t\le 1} t^{-\alpha} K_r(f, t^r)_p, & q = \infty, \end{cases} \tag{1.7}$$

is finite for some $r \ge \lfloor \alpha \rfloor + 1$. The integration in (1.7) is over [0, 1] due to the fact that Ω is bounded. By the monotonicity of the modulus of smoothness or the K-functional, an equivalent discrete form is

$$|f|_{B_q^\alpha(L_p(\Omega))} \sim \begin{cases} \left(\sum_{n=0}^\infty (2^{n\alpha} K_r(f, 2^{-nr})_p)^q\right)^{1/q}, & 0 < q < \infty, \\ \sup\limits_{n\ge 0} 2^{n\alpha} K_r(f, 2^{-nr})_p, & q = \infty. \end{cases} \tag{1.8}$$

Observe that one usually finds in the literature the discrete sum (1.8) with $K_r(f, 2^{-nr})_p$ replaced by $\omega_r(f, 2^{-n})_p$. The equivalence of these two representations again follows from (1.4).

The Besov spaces play an important role in approximation theory since they characterize the approximation spaces corresponding to some important approximation methods: Linear algorithms such as polynomial and spline approximation and nonlinear algorithms such as univariate free-knot splines [13, Chapter 12], univariate rational approximation [23, Section 10.6], and multivariate wavelets [11, Chapter 7]. In the multivariate setting, the Besov spaces have perhaps the following disadvantage. From (1.1) it is clear that once a direction $h \in \mathbb{R}^2$ is chosen, the function is "differentiated" in that direction over the whole of the domain. One may then argue that in the multivariate setting the measure of smoothness should be more adaptive to piecewise smoothness over certain disjoint subdomains.

Another point is this: Besov spaces and their multivariate anisotropic variants are linear spaces. At the same time approximation spaces associated with multivariate piecewise polynomial approximation are not linear (see Subsection 2.1). Therefore it is not possible that Besov spaces or other "classical" smoothness spaces can characterize these approximation spaces.

We propose that in the multivariate setting, a measure of smoothness that incorporates measures of smoothness in several dimensions is more appropriate. Our efforts to proceed in this direction rely on recent attempts (e.g., [3], [4], [8], [15], [18], [26], [27], [29]) to find compact representations for multivariate signals by combining traditional coding methods such as wavelet compression with computation of lower-dimensional structure, for example, segmentation in the case of images. Indeed, one of the goals of this work is to try and understand on which "class" of functions these approaches outperform wavelets. Thus, we try to quantify, in an approximation theoretical sense, the amount of "structure" present in the signal.

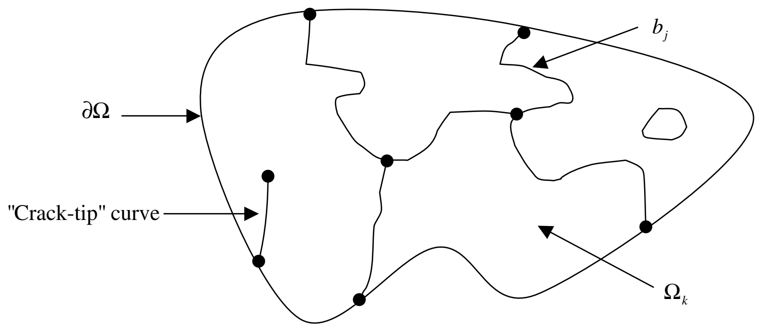


Fig. 1.1. A partition Λ of the domain Ω.

Definition 1.1. For $t > 0$ we define $\Lambda(t)$ as the set of partitions Λ of a bounded domain Ω with the following properties:

(i) The partition Λ is defined by nonintersecting curves $b_j : [0,1] \to \Omega$, $j = 1, \ldots, n_E(\Lambda)$, each of finite length, denoted by $\text{len}(b_j)$. The curves may intersect only at endpoints and a subset of the curves should compose the boundary of Ω. Observe that we allow the curves to have "crack-tips" (see, e.g., [25]), that is, an end of a curve possibly does not touch the end of any of the other curves.
(ii) To each curve b_j we associate a parameter $0 < t_j \le 1$ such that $\sum_{j=1}^{n_E(\Lambda)} t_j^{-1} \le t^{-1}$ (in particular, this implies that $n_E(\Lambda) \le t^{-1}$).
(iii) The curves partition Ω into open connected subdomains Ω_k, $k = 1, \ldots, n_F(\Lambda)$.

The following notion of smoothness combines measures of smoothness at several dimensions.

Definition 1.2. For $f \in L_p(\Omega)$, $1 \le p < \infty$, $t > 0$, and $r_1, r_2 \in \mathbb{N}$, we define

$$\text{(1.9)} \quad \widetilde{K}_{r_1,r_2}(f,t)_p := \inf_{\Lambda\in\Lambda(t)} \left(\sum_{j=1}^{n_E(\Lambda)} \text{len}(b_j) K_{r_1}(b_j, t_j^{r_1})_{\infty,1} + \sum_{k=1}^{n_F(\Lambda)} K_{r_2}(f,\Omega_k)_p^p \right)^{1/p}.$$

One can see from (1.9) that the $\widetilde{K}$*-functional* is defined using sums of "curve" and "surface" smoothness terms. The $\widetilde{K}$-functional is highly nonlinear in the following sense:

(i) $\widetilde{K}$ is nondecreasing as a function of t but, in general, is not continuous.
(ii) $\widetilde{K}$ in general is not sublinear, that is, there does not exist any constant C such that for any $f, g \in L_p(\Omega)$ we have $\widetilde{K}_{r_1,r_2}(f+g,t)_p \le C(\widetilde{K}_{r_1,r_2}(f,t)_p + \widetilde{K}_{r_1,r_2}(g,t)_p)$.

In Section 4 we discuss the relationships between the $\widetilde{K}$-functional and the Mumford–Shah-type functionals of [25].

Using (1.9) we define the following bivariate smoothness spaces:

Definition 1.3. A function $f \in L_p(\mathbb{R}^2)$, $1 \le p < \infty$, is said to be in the $\widetilde{B}$-space $\widetilde{B}_q^{\alpha,r_1,r_2}(L_p(\Omega))$ if

$$(1.10)\quad (f)_{\widetilde{B}_q^{\alpha,r_1,r_2}(L_p(\Omega))} := \begin{cases} \left(\int_0^1 (t^{-\alpha} \widetilde{K}_{r_1,r_2}(f,t)_p)^q \frac{dt}{t} \right)^{1/q}, & 0 < q < \infty, \\ \sup_{0<t\le 1} t^{-\alpha} \widetilde{K}_{r_1,r_2}(f,t)_p, & q = \infty, \end{cases}$$

is finite. In similar manner to (1.8) we have

$$(1.11)\quad (f)_{\widetilde{B}_q^{\alpha,r_1,r_2}(L_p(\Omega))} \sim \begin{cases} \left(\sum_{n=0}^{\infty} (2^{n\alpha} \widetilde{K}_{r_1,r_2}(f,2^{-n})_p)^q\right)^{1/q}, & 0 < q < \infty, \\ \sup_{n\ge 0} 2^{n\alpha} \widetilde{K}_{r_1,r_2}(f,2^{-n})_p, & q = \infty. \end{cases}$$

Observe that $(\cdot)_{\widetilde{B}_q^{\alpha,r_1,r_2}(L_p(\Omega))}$ serves as a measure of smoothness, but it is not a seminorm because in general the triangle inequality is not fulfilled. From (1.10) or (1.11) it is clear that $\widetilde{B}_q^{\alpha,r_1,r_2}(L_p(\Omega)) \subseteq \widetilde{B}_q^{\beta,r_1,r_2}(L_p(\Omega))$, moreover, $(\cdot)_{\widetilde{B}_q^{\beta,r_1,r_2}(L_p(\Omega))} \le (\cdot)_{\widetilde{B}_q^{\alpha,r_1,r_2}(L_p(\Omega))}$, whenever $\beta \le \alpha$. We find it useful to denote $\widetilde{B}_q^{\alpha,r}(L_p(\Omega)) := \widetilde{B}_q^{\alpha,2,r}(L_p(\Omega))$ and $\widetilde{B}_q^{\alpha}(L_p(\Omega)) := \widetilde{B}_q^{\alpha,2,r}(L_p(\Omega))$ with $r = \lfloor \alpha \rfloor + 1$.

The following two results show the relations between the classical Besov smoothness spaces and the $\widetilde{B}$-spaces.

Theorem 1.4. *For $\Omega = [0,1]^2$, $1 \le p < \infty$, $r \in \mathbb{N}$, and $0 < t \le 1$,*

$$(1.12)\qquad \widetilde{K}_{2,r}(f,t^2)_p \le C\omega_r(f,t)_p.$$

In particular, for any $\alpha > 0$ and $0 < q \le \infty$, we have that the space $B_q^\alpha(L_p([0,1]^2))$ is contained in $\widetilde{B}_q^{\alpha/2}(L_p([0,1]^2))$, moreover, $(\cdot)_{\widetilde{B}_q^{\alpha/2}(L_p([0,1]^2))} \le C|\cdot|_{B_q^\alpha(L_p([0,1]^2))}$.

One can improve the above by using the characterization of Besov spaces by wavelet approximation spaces (see [11, Section 7.6]).

Theorem 1.5. *For $\alpha > 0$, $1 < p < \infty$, $q = (\alpha/2 + 1/p)^{-1}$, and $r > \alpha + 1 - 1/p$, the space $B_q^\alpha(L_q([0,1]^2))$ is contained in $\widetilde{B}_q^{\alpha/2,r}(L_p([0,1]^2))$, moreover, we have $(\cdot)_{\widetilde{B}_q^{\alpha/2,r}(L_p([0,1]^2))} \le C|\cdot|_{B_q^\alpha(L_p([0,1]^2))}$.*

It seems that these results are sharp in the sense that the Besov space on the left is not contained in $\widetilde{B}_q^{\beta/2,r}(L_p([0,1]^2))$ for any $\beta > \alpha$.

The difference between the Besov spaces and the $\widetilde{B}$-spaces is that the $\widetilde{B}$-spaces exhibit the most significant singularities along curves penalized by a measure of the lower-dimensional smoothness of those curves. If a function in $\widetilde{B}$ represents an image, then the singularities along curves are the edges in the images and the other parts are smooth in the two-dimensional gauge. These curves of singularities we call the "structure" present in a given function.

In general, we cannot expect multivariate functions of weak-type smoothness to have any lower-dimensional geometric structure and, in fact, in general they are of oscillatory type. This was demonstrated by Donoho [16] by manipulating the wavelet coefficients of real-life images such that, on the one hand, their Besov semi-norm remains unchanged and thus also the performance of nonlinear wavelet approximation (see [16]) and, on the other hand, they turn into visually incoherent texture.

The next simple result verifies that, unlike the Besov spaces, $\widetilde{B}$-spaces contain functions that do have lower-dimensional structure or smoothness (see also Example 1.7 in [9]).

Example 1.6. Let $\widetilde{\Omega} \subset \Omega$ and assume that $\partial\widetilde{\Omega}$, $\partial\Omega$ are piecewise $\mathrm{Lip}^*(\alpha)$ curves. Then, $\mathbb{1}_{\widetilde{\Omega}} \in \widetilde{B}_q^{\beta,r_1,r_2}(L_p(\Omega))$ for all $r_1 \geq \lfloor\alpha\rfloor + 1$, $r_2 \geq 1$, $\beta < \alpha$, $1 \leq p < \infty$, $0 < q \leq \infty$. For example, if $\partial\widetilde{\Omega} \in C^\infty$ then $\mathbb{1}_{\widetilde{\Omega}} \in \widetilde{B}_q^{\alpha,r_1,r_2}(L_p(\Omega))$, whenever $r_1 \geq \lfloor\alpha\rfloor + 1$. On the other hand, we have that $\mathbb{1}_{\widetilde{\Omega}} \notin B_q^\alpha(L_p(\Omega))$ if $\alpha > 1/p$.

Let $S_m^{r_1,r_2}(\Omega)$ denote the collection of piecewise polynomials of type $\sum_{k=1}^m \mathbb{1}_{\Omega_k} P_k$, where $\Omega_k \subset \Omega$ are domains with disjoint interiors whose boundary is composed a fixed number of nonintersecting piecewise polynomial segments of degree $r_1 - 1$ and P_k are bivariate polynomials of degree $r_2 - 1$. In the special case where $r_1 = 2$, the approximation takes the form of piecewise polynomials over polygonal domains. By triangulating these polygonal domains, we may consider $S_m^{2,r}(\Omega)$ to be the collection of functions of type $\sum_{k=1}^m \mathbb{1}_{\Delta_k} P_k$, where Δ_k are triangles with disjoint interiors and P_k are bivariate polynomials of degree $r - 1$.

The parameters r_1, r_2 allow us to "tune" the approximation method to the lower- or higher-dimensional smoothness of the approximated functions. For $f \in L_p(\Omega)$ we define the degree of approximation

$$\sigma_{m,r_1,r_2}(f)_p := \inf_{\varphi \in S_m^{r_1,r_2}} \|f - \varphi\|_{L_p(\Omega)}.$$

Denoting $\sigma_{m,r}(f)_p := \sigma_{m,2,r}(f)_p$, we have the following Jackson-type inequality for approximation by piecewise polynomials over triangles.

Theorem 1.7. *Let Ω be a bounded domain with a piecewise* $\mathrm{Lip}^*(2)$ *boundary and let $f \in L_\infty(\Omega)$. Then, for $1 \leq p < \infty$ and each $m \geq 1$, we have that*

$$\sigma_{m,r}(f)_p \leq C_1(p,r)\max(\|f\|_{L_\infty(\Omega)}, 1)\widetilde{K}_{2,r}(f, C_2 m^{-1})_p. \tag{1.13}$$

Remark. It seems like a drawback of Theorem 1.7, that we have to assume that $f \in L_\infty(\Omega)$, while we would like estimates in the L_p-norm, and (1.13) may look a bit awkward in view of the involvement of the quantity $\|f\|_{L_\infty(\Omega)}$ in the estimate. However, we wish to point out that images are always in L_∞ so that in normal applications we usually obtain L_2 estimates of functions in $L_2(\Omega) \cap L_\infty(\Omega)$. It is pretty clear that our $\tilde{K}$-functional does not distinguish between the characteristic function of the unit disk and a 10^6 blowup of that function.

The main ingredients in the proof of the Jackson inequality (1.13) are the "local" polynomial approximation result, Theorems 2.2 and 3.1, the geometric result concerning polygonal approximation of partitions of planar domains.

Definition 1.8 (Approximation Spaces). For $\alpha > 0$ and $0 < q \le \infty$, let $A_q^{\alpha,r_1,r_2}(L_p(\Omega), \Sigma)$ denote the set of functions $f \in L_p(\Omega)$ for which

$$(1.14) \qquad (f)_{A_q^{\alpha,r_1,r_2}(L_p,\Sigma)} := \begin{cases} \left(\sum_{m=0}^{\infty}(2^{m\alpha}\sigma_{2^m,r_1,r_2}(f)_p)^q\right)^{1/q}, & 0 < q < \infty, \\ \sup_{m\ge 0} 2^{m\alpha}\sigma_{2^m,r_1,r_2}(f)_p, & q = \infty, \end{cases}$$

is finite. In the special case where $r_1 = 2$ and the approximation takes the form of piecewise polynomials over triangles we denote $A_q^{\alpha,r}(L_p(\Omega), \Delta) := A_q^{\alpha,2,r}(L_p(\Omega), \Sigma)$.

The $\widetilde{B}$-spaces can "almost" characterize nonlinear approximation algorithms corresponding to piecewise polynomial approximation in the following way:

Theorem 1.9. *Let Ω be a bounded domain with a piecewise linear boundary. Then, for any $\alpha > 0$, $1 \le p < \infty$, $0 < q \le \infty$, and $r \in \mathbb{N}$, the set $A_q^{\alpha,r}(L_p(\Omega), \Delta)$ is contained in $\widetilde{B}_q^{\alpha,r}\big(L_p(\Omega)\big)$, moreover $(f)_{\widetilde{B}_q^{\alpha,r}(L_p(\Omega))} \le C(f)_{A_q^{\alpha,r}(L_p(\Omega),\Delta)}$. On the other hand,*

$$(1.15) \qquad f \in \widetilde{B}_q^{\alpha,r}(L_p(\Omega)) \cap L_\infty(\Omega) \quad \Longrightarrow \quad f \in A_q^{\alpha,r}(L_p(\Omega), \Delta).$$

Perhaps our discussion so far quantifies the following "intuition." Besov spaces cannot capture singularities well along curves while nonlinear piecewise polynomial approximation does. Thus piecewise polynomials over triangles outperform wavelets significantly if the approximated function represents, for instance, an image which normally has edge singularities, what we have referred to as "structure." On the other hand, if a function is a typical Besov-type function that is smooth only in a weak sense with oscillations "randomly" distributed over the time and frequency domains, then we should not expect m-term piecewise polynomials approximation to perform any better than the m-term wavelet approximation.

Another form of nonlinear approximation we consider is approximation by rational functions. Denote by $\mathcal{R}_n$ the set of all bivariate rational functions of degree n, i.e.,

$$\mathcal{R}_n := \{R = P_1/P_2 : P_1, P_2 \in \Pi_n(\mathbb{R}^2),\ P_2 > 0\}.$$

We restrict ourselves to elements of $\mathcal{R}_n$ that are taken from a collection of functions that can be described by n parameters and denote this collection by $\widetilde{\mathcal{R}}_n$ (see [9] for an exact description of these parameters). Thus, for $f \in L_p(\mathbb{R}^2)$, $0 < p \le \infty$, we denote

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

The corresponding rational approximation spaces $A_q^\alpha(L_p, \widetilde{\mathcal{R}})$ are defined by replacing in (1.14) the terms $\sigma_{2^m,r_1,r_2}(f)_p$ by $\widetilde{\rho}_{2^m}(f)_p$. Applying (1.15) and [9, Theorem 1.6], we may conclude that bivariate rational approximation also performs well in the presence of lower-dimensional structure.

Corollary 1.10. *For any $\gamma < \alpha$, $0 < q \le \infty$, and $r \in \mathbb{N}$ we have*

$$f \in \widetilde{B}_q^{\alpha,r}(L_1(\Omega)) \cap L_\infty(\Omega) \quad \Longrightarrow \quad f \in A_q^\gamma(L_1(\Omega), \widetilde{\mathcal{R}}).$$

2. Approximation by Piecewise Polynomials

2.1. *Nonlinearity of the Piecewise Polynomial Approximation Spaces*

As mentioned in the Introduction, all classical smoothness spaces are linear spaces. Here, following the discussion in [11, Section 6.5], we give a simple proof that the multivariate approximation spaces we are interested in are nonlinear and therefore cannot be characterized by the classical smoothness spaces.

Theorem 2.1. *For $\alpha > 0$, $0 < p < \infty$, and $0 < q \le \infty$ there exist functions $f, g \in L_p[(0,1]^2)$ such that $f, g \in A_q^{\alpha,1}(L_p([0,1]^2), \Delta)$ but $f + g \notin A_q^{\alpha,1}(L_p([0,1]^2), \Delta)$.*

Proof. Assume first that $1/p < \alpha < 2/p$. We construct recursively two sequences of functions $\{f_n\}_{n\ge1}$, $\{g_n\}_{n\ge1}$. The function f_1 is described in Figure 2.1 and f_2 is described in Figure 2.2.

Assume that f_n, $n \ge 2$, is already defined as 0 on the upper right corner, a square I_n of side length 2^{-n}. We set $f_{n+1} = f_n$ on $[0,1]^2 \backslash I_n$ and proceed to define it on I_n. We divide I_n by a vertical line into two rectangles $I_{n,1}$ and $I_{n,2}$ each of vertical side length 2^{-n} and horizontal side length of $2^{-(n+1)}$. We now divide $I_{n,1}$ into 2^{n+1} equal vertical strips on which f_{n+1} assumes the alternating values of $(-1)^m$, $m = 0, \dots, 2^{n+1} - 1$. We divide $I_{n,2}$ by a horizontal line into two equal size squares of side length $2^{-(n+1)}$, and ascribe to f_{n+1} the value $(-1)^{n+1}$ in the lower square and 0 in the upper square, i.e., in I_{n+1}.

The function g_n is obtained by reflection of f_n through the main diagonal of $I_0 := [0,1]^2$. The function g_2 is described in Figure 2.3.

We can summarize the properties of the functions f_n and g_n:

- The functions f_n and g_n are piecewise constant over $2^{n+1} + n - 2$ rectangles.
- The functions f_n, g_n take the values 0 and ±1.
- On $[0,1]^2 \backslash I_n$ we have that $f_{n+1} = f_n$, $g_{n+1} = g_n$, and therefore also $f_{n+1} + g_{n+1} = f_n + g_n$.
- The function $f_n + g_n$ takes the values 0 and ±2 in rectangles, the total of which is $\ge 4^{n-1}$. The function $f_n + g_n$ is zero in about half of these rectangles, so we have more than 4^{n-2} rectangles with nonzero values.

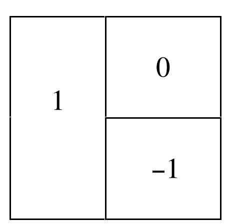


Fig. 2.1. The function f_1.

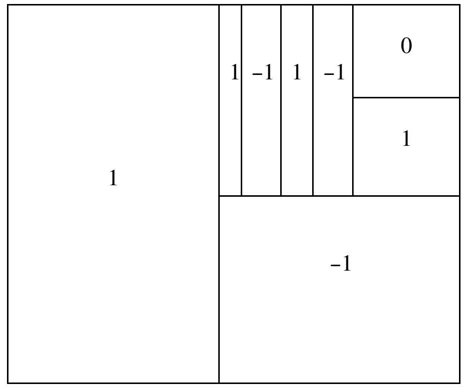


Fig. 2.2. The function f_2.

- The sequences $\{f_n\}_{n\geq 1}$, $\{g_n\}_{n\geq 1}$ converge in the p-metric for all $0 < p < \infty$, to (measurable) limits $f := \lim_{n\to\infty} f_n$, $g := \lim_{n\to\infty} g_n$.

Since on $[0,1]^2\backslash I_n$ we have $f = f_n$, we can triangulate it into $2(2^{n+1}+n-1)$ triangles to obtain

$$\sigma_{2^{n+3}}(f)_p \leq \sigma_{2(2^{n+1}+n-1)}(f)_p \leq |I_{2^n}|^{1/p} = 2^{-2n/p},$$

which implies that $f \in A_q^{\alpha,1}(L_p, \Delta)$ for $\alpha < 2/p$. The same is true for g.

When we approximate $f + g$ with piecewise constants on 2^n triangles, we obviously have to assign the same nonzero values taken by $f_n + g_n$ on the biggest rectangles. Observe that the number of rectangles on which $f_{[(n+4)/2]} + g_{[(n+4)/2]}$ takes nonzero values is larger than 2^{n-1} (so that the number of such triangles is larger than 2^n). Note that by our construction the rectangles we get for the function $f_{[(n+4)/2]+1} + g_{[(n+4)/2]+1}$, in the square in the lower left corner of $I_{[(n+4)/2]}$, are smaller than any of the above, thus on that area the approximating function is 0. On about half of it $f_n + g_n = \pm 2$ and this

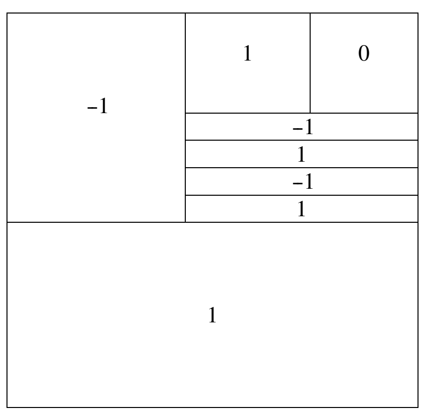


Fig. 2.3. The function g_2.

area is one-eighth of the area of $I_{[(n+4)/2]}$, namely, no less than 2^{-n-4}. Hence

$$\sigma_{2^n}(f+g)_p \geq 2^{-4}2^{-n/p},$$

and, consequently, $f+g \notin A_q^{\alpha,1}(L_p,\Delta)$ for $\alpha > 1/p$.

To treat the case $\alpha \geq 2/p$, we modify the above construction by assigning to the functions f_n, g_n the values 0 and $\pm 2^{n(2/p-\alpha-\varepsilon)}$ over I_n at the nth step of the construction, with sufficiently small $\varepsilon > 0$. ∎

Remark. Observe that the approximation spaces corresponding to piecewise polynomials of type

$$\sum_{k=1}^{m} \mathbb{1}_{\Delta_k} P_k,$$

where the triangles Δ_k are allowed to intersect, are linear spaces and therefore strictly contain the (nonlinear) spaces $A_q^{\alpha,r}(L_p,\Delta)$.

2.2. *On Polynomial Approximation Over Triangles*

For a bounded domain $\Omega \subset \mathbb{R}^2$ and $r \in \mathbb{N}$ we denote the degree of polynomial approximation

$$E_{r-1}(f,\Omega)_p := \min_{P\in\Pi_{r-1}} \|f-P\|_{L_p(\Omega)}.$$

The following is the main result of this section.

Theorem 2.2. *Let* $\bigcup_{n=1}^{N} \Delta_n \subseteq \widetilde{\Omega}$, *where* Δ_n *are triangles with disjoint interiors. Then, for any* $f \in L_p(\widetilde{\Omega})$, $1 \leq p < \infty$,

$$\sum_{n=1}^{N} E_{r-1}(f,\Delta_n)_p^p \leq C(p,r) K_r(f,\widetilde{\Omega})_p^p, \tag{2.1}$$

where $K_r(f,\widetilde{\Omega})_p$ *is defined by* (1.3).

Whitney-type estimates for polynomial approximation are estimates of the type

$$E_{r-1}(f,\Omega)_p \leq C\omega_r(f,\Omega)_p\ ,$$

where the constant C usually depends on r and p but in the multivariate case may also depend on the geometry of the domain. To characterize approximation of piecewise polynomials over triangles we require Whitney-type results where the constant does not depend on the "thinness" of the triangles. Indeed, it is proved in [21] that, for any triangle Δ and $f \in L_p(\Delta)$, $0 < p \leq \infty$, we have

$$E_{r-1}(f,\Delta)_p \leq C(r,p)\omega_r(f,\Delta)_p\ . \tag{2.2}$$

In our case, we require a variant of (2.2) that uses the appropriate K-functional because the modulus of smoothness (1.1) is not suitable when we want to add up estimates over several triangles.

Remark. In the univariate case one can add up smoothness terms over disjoint intervals using the averaged modulus of smoothness (see [13, Section 6.5]). This technique also works for disjoint multivariate cubes (see [14]).

The Whitney estimate (2.2) and the right-hand side of (1.4) yield

$$E_{r-1}(f,\Delta)_p \le C(r,p)K_r(f,\Delta)_p, \tag{2.3}$$

for functions in $f \in L_p(\Delta)$, $1 \le p \le \infty$, and for functions $g \in W_p^r(\Delta)$, they give

$$E_{r-1}(g,\Delta)_p \le C(r,p)(\mathrm{diam}(\Delta))^r |g|_{W_p^r(\Omega)}. \tag{2.4}$$

In [10] we generalize (2.3) and show that for any bounded convex domain $\Omega \subset \mathbb{R}^d$ and $f \in L_p(\Omega)$, $1 \le p \le \infty$,

$$E_{r-1}(f,\Omega)_p \le C(r,d)K_r(f,\Omega)_p.$$

Proof of Theorem 2.2. Let $g \in W_p^r(\widetilde{\Omega})$ such that

$$\|f-g\|_{L_p(\widetilde{\Omega})} + \mathrm{diam}(\widetilde{\Omega})^r |g|_{W_p^r(\widetilde{\Omega})} \le 2K_r(f,\widetilde{\Omega})_p.$$

Let $P_n \in \Pi_{r-1}$, $n = 1,\ldots,N$, such that $E_{r-1}(g,\Delta_n)_p = \|g-P_n\|_{L_p(\Delta_n)}$ and define the piecewise polynomial function

$$\varphi := \sum_{n=1}^N \mathbb{1}_{\Delta_n} P_n.$$

Then, for $U := \bigcup_{n=1}^N \Delta_n$ and $1 \le p < \infty$,

$$\begin{aligned}\sum_{n=1}^N E_{r-1}(f,\Delta_n)_p^p &\le \|f-\varphi\|_{L_p(U)}^p \\ &\le (\|f-g\|_{L_p(U)} + \|g-\varphi\|_{L_p(U)})^p.\end{aligned}$$

Application of (2.4) yields

$$\begin{aligned}\|g-\varphi\|_{L_p(U)}^p &\le C(r,p)\sum_{n=1}^N (\mathrm{diam}(\Delta_n))^{rp}|g|_{W_p^r(\Delta_n)}^p \\ &\le C(r,p)(\mathrm{diam}(\widetilde{\Omega}))^{rp}|g|_{W_p^r(\widetilde{\Omega})}^p.\end{aligned}$$

Consequently,

$$\begin{aligned}\sum_{n=1}^N E_{r-1}(f,\Delta_n)_p^p &\le (\|f-g\|_{L_p(U)} + \|g-\varphi\|_{L_p(U)})^p \\ &\le (\|f-g\|_{L_p(\widetilde{\Omega})} + C(\mathrm{diam}(\widetilde{\Omega}))^r|g|_{W_p^r(\widetilde{\Omega})})^p \\ &\le CK_r(f,\widetilde{\Omega})_p^p.\end{aligned}$$

∎

3. Polygonal Approximation of a Partition of Planar Domain

The following main result of this section is required for the proof of the Jackson inequality (1.13).

Theorem 3.1. *Let* Λ *be a partition of a bounded domain* Ω (*see Definition* 1.1). *Assume that each curve* b_j *of the partition* Λ *is approximated by an interpolating polygon* s_j *such that* $\max_{0\le u\le 1}|b_j(u)-s_j(u)|\le \varepsilon_j$, $j=1,\dots,n_E(\Lambda)$, *and that the total number of linear segments of all the polygons is* m. *Then there exist pairwise interior disjoint polygonal connected domains* $\widetilde{\Omega}_{k,l}$, $k=1,\dots,n_F(\Lambda)$, $l=1,\dots,n_F(k)$, *with* $\widetilde{\Omega}_{k,l}\subseteq\Omega_k$, *such that the total complexity of* $\{\widetilde{\Omega}_{k,l}\}$ *is* $O(m)$ *and*

$$|\Omega| - \sum_{k=1}^{n_F(\Lambda)}\sum_{l=1}^{n_F(k)}|\widetilde{\Omega}_{k,l}| \le 4\sum_{j=1}^{n_E(\Lambda)}\operatorname{len}(b_j)\varepsilon_j. \tag{3.1}$$

Let γ be a closed Jordan curve in the plane and let $P=\{p_1,\dots,p_n\}$ be a set of n points on γ which appear in this counterclockwise order along the curve. For each $i=1,\dots,n$, let γ_i denote the arc between p_i and p_{i+1} (where we put $p_{n+1}:=p_1$). Let C_i denote the convex hull of γ_i, and let R_i be a circumscribed rectangle of C_i (or of γ_i), one of whose sides is parallel to the straight segment $[p_i,p_{i+1}]$. Put $U:=\bigcup_{i=1}^n C_i$ and $W:=\bigcup_{i=1}^n R_i$. Clearly, $U\subseteq W$.

We first observe that the union $W:=\bigcup_{i=1}^n R_i$ may have quadratic complexity. A construction that illustrates the lower bound is shown in Figure 3.1.

(In Figure 3.1 not all the rectangles R_i are shown but the presence of the missing ones would not have affected the quadratic complexity of W.)

Our goal is to first find an intermediate polygonal region U^* that contains U, is contained in W, and has complexity $O(n)$. In what follows we show how to construct such a region.

Lemma 3.2. *The number of intersection points of the boundaries of the sets* C_i *that lie on* ∂U *is at most* $6n-12$ *for* $n\ge 3$.

Proof. We claim that $\{C_i\}$ is a collection of *pseudo-disks*, i.e., simply connected planar regions, each pair of whose boundaries intersect at most twice. This is a well-known

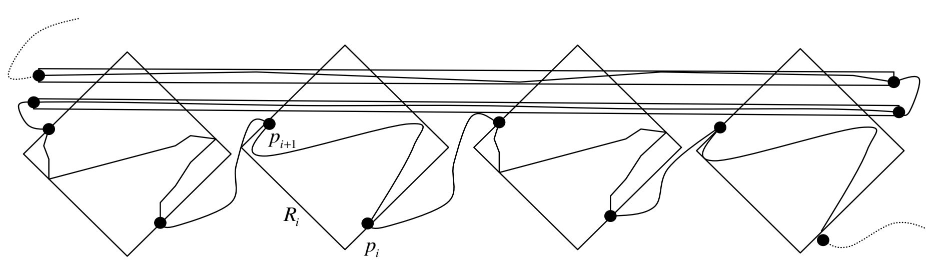


Fig. 3.1. The union W of the rectangles R_i may have quadratic complexity.

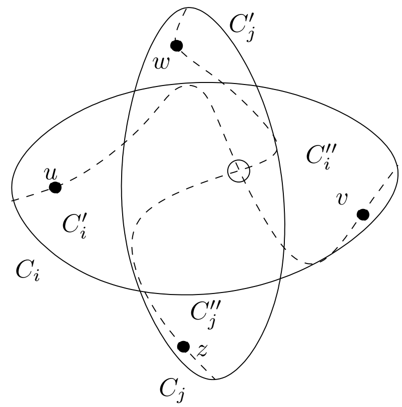


Fig. 3.2. Two hull boundaries ∂C_i, ∂C_j cannot intersect at four points.

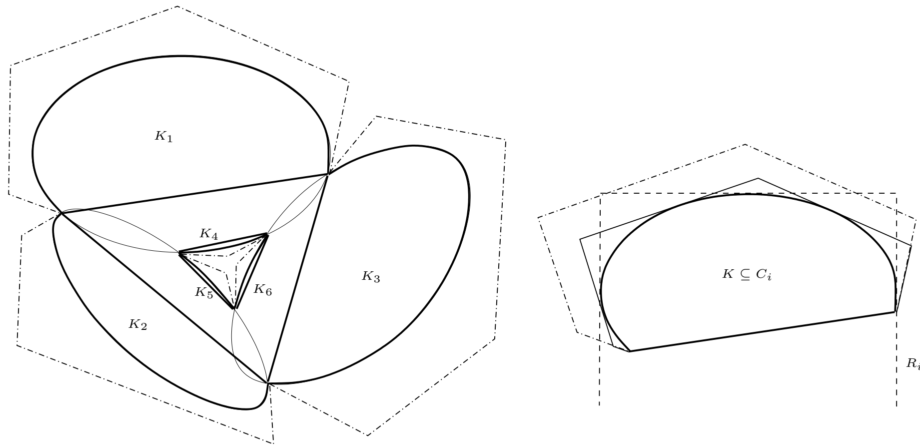
property (see, e.g., [6]) but we include its proof for the sake of completeness. Let C_i, C_j be a fixed pair of these sets, and suppose to the contrary that ∂C_i and ∂C_j intersect each other in at least four points. Since these sets are convex, $C_i \cup C_j \setminus (C_i \cap C_j)$ consists of at least four nonempty connected components, at least two of which, denoted C_i', C_i'', are contained in $C_i \setminus C_j$, and at least two others, denoted C_j', C_j'', are contained in $C_j \setminus C_i$; see Figure 3.2.

Note that each of the components C_i', C_i'' (resp., C_j', C_j'') must intersect γ_i (resp.,γ_j), for otherwise, if say, $\gamma_i \cap C_i' = \emptyset$, then we can replace C_i by $C_i \setminus C_i'$, which is a convex set that contains γ_i, contradicting the fact that C_i is the convex hull of γ_i. Choose four points $u \in C_i' \cap \gamma_i$, $v \in C_i'' \cap \gamma_i$, $w \in C_j' \cap \gamma_j$, and $z \in C_j'' \cap \gamma_j$, and observe that the portion of γ_i between u and v must cross the portion of γ_j between w and z; see Figure 3.2. This contradiction implies that $\{C_i\}$ is a family of pseudo-disks. The claim is now an immediate consequence of the linear bound on the complexity of the union of pseudo-disks, given in [22]. ∎

For each pair of consecutive intersection points u, v along (some connected component of) ∂U, connect u and v by a straight segment. This chord and the portion of ∂U between u and v bound a convex subregion of U. Let $\mathcal{K}$ denote the set of resulting subregions. The regions in $\mathcal{K}$ are pairwise openly disjoint, as is easily verified. See Figure 3.3(a) for an illustration. We now use the following result, due to Edelsbrunner et al. [17].

Lemma 3.3. *Let $\mathcal{K}$ be a collection of m pairwise openly disjoint convex regions in the plane. One can cover each region in $\mathcal{K}$ by a convex polygon, so that the resulting polygons are also pairwise openly disjoint, and the total number of their edges is at most* $3m - 6$.

Let K be a region in $\mathcal{K}$, and let V be the convex polygon that covers K. We shrink V by translating each of its edges so that it becomes tangent to K. The resulting polygon V' is clearly contained in V and contains K. Finally, let C_i be the (unique) convex hull



(a) (b)

Fig. 3.3. (a) The subregions in K and their containing polygons. (b) Shrinking a covering polygon.

that contains K and let R_i be the rectangle containing C_i. We replace V' by $V' \cap R_i$. This increases the number of edges of V' by at most four. See Figure 3.3(b).

In summary, we have obtained a collection $\mathcal{V}$ of at most $6n - 12$ pairwise openly disjoint convex polygons with a total of at most $3(6n - 12) - 6 = 18n - 42$ edges. Let U^* denote the union of U with the union of $\mathcal{V}$. Then $U \subseteq U^* \subseteq W$. Moreover, ∂U^* consists exclusively of edges of the polygons in $\mathcal{V}$ (the inclusion of U just fills holes in $\mathcal{V}$), so U^* is a polygonal region with at most $18n - 42$ edges. We have thus shown

Theorem 3.4. *Let γ, P, and W be as above. There exists a polygonal region U^* with at most $18|P| - 42$ edges that contains γ and is contained in W.*

Proof of Theorem 3.1. Let Ω_k be a subdomain of the partition defined by the subset of curves $b_{k,j}$, $j = 1, \dots, n_E(k)$. Without loss of generality we may assume that Ω_k is of genus 1, which implies that $\partial\Omega_k$ is a Jordan curve. Otherwise, we may subdivide Ω_k into regions of genus 1 by adding at most $n_E(k)$ line segments and associating with them the "error" $\varepsilon = 0$. We denote by $b_{k,j,i}$, $i = 1, \dots, n_s(k, j)$, the ith portion of the curve $b_{k,j}$ and by $s_{k,j,i}$ the corresponding approximating linear segment of $s_{k,j}$. We associate with each such portion the rectangle $R_{k,j,i}$ one of whose sides is parallel to $s_{k,j,i}$ as described in Figure 3.4. Assuming the segment $s_{k,j,i}$ is placed on the x-axis, then $R_{k,j,i}$ is defined by min/max points of $b_{k,j,i}$ in the x and y directions. The horizontal length of the rectangle cannot exceed the length of the curve segment $b_{k,j,i}$, while the vertical length is bounded by $2\varepsilon_j$ because the distance between each point on $b_{k,j,i}$ and the segment $s_{k,j,i}$ does not exceed ε_j.

Therefore, denoting

$$W_k := \bigcup_{\substack{i=1,\dots,n_s(k,j)\\ j=1,\dots,n_E(k)}} R_{k,j,i}$$

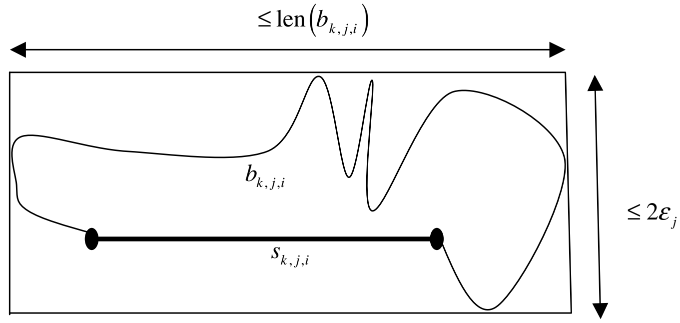


Fig. 3.4. The rectangle $R_{k,j,i}$.

we have

$$|W_k| \le 2 \sum_{j=1}^{n_E(k)} \mathrm{len}(b_{k,j})\varepsilon_{k,j}.$$

By Theorem 3.4 there exists a polygonal domain U_k^* of total complexity $\le C \sum_{j=1}^{n_E(k)} n_s(k, j)$ such that $\partial\Omega_k \subset U_k^* \subseteq W_k$. This implies that $\Omega_k \backslash U_k^* = \bigcup_{l=1}^{n_F(k)} \widetilde{\Omega}_{k,l}$, where $\widetilde{\Omega}_{k,l}$, $l = 1, \ldots, n_F(k)$, are pairwise interior disjoint polygonal connected domains whose total complexity is smaller than the complexity of U_k^* and for which

$$\left|\Omega_k \backslash \bigcup_{l=1}^{n_F(k)} \widetilde{\Omega}_{k,l}\right| \le |U_k^*| \le |W_k| \le 2 \sum_{j=1}^{n_E(k)} \mathrm{len}(b_{k,j})\varepsilon_{k,j}. \tag{3.2}$$

The result is obtained by summing up the total number of edges of $\{\widetilde{\Omega}_{k,l}\}$ and the area estimate (3.2) over all Ω_k, $k = 1, \ldots, n_F(\Lambda)$. ∎

4. Relations Between the $\widetilde{K}$ and the Mumford–Shah Functionals

We would like to draw attention to a connection between the $\widetilde{K}$ and the Mumford–Shah functionals. In their seminal paper [24], Mumford and Shah introduced a technique for segmenting a bivariate function, thereby obtaining a "compact" representation of functions that have some lower-dimensional structure. Let Λ be a collection of continuous curves b_j, $j = 1, \ldots, n_E(\Lambda)$, that partition the domain Ω to open subdomains Ω_k, $k = 1, \ldots, n_F(\Lambda)$. In the case of the Mumford–Shah functional, there are no "smoothness" parameters t_j associated with the curves b_j (see Definition 1.1). In particular, there is no limit on their number $n_E(\Lambda)$. Let D be a differential operator of degree r, such as the gradient of degree 1 used in [25]. Let $g \in L_2(\Omega)$ such that $g \in W_2^r(\Omega_k)$, $k = 1, \ldots, n_F(\Lambda)$. Then for weights $\mu_1, \mu_2 > 0$ and any $f \in L_p(\Omega)$ we define an *energy gauge*

$$\mathbf{E}(f, \Lambda) = \sum_{k=1}^{N_F(\Lambda)} \|f - g\|_{L_2(\Omega_k)}^2 + \mu_1 \sum_{k=1}^{N_F(\Lambda)} \|Dg\|_{L_2(\Omega_k)}^2 + \mu_2 \sum_{j=1}^{n_E(\Lambda)} \mathrm{len}(b_j), \tag{4.1}$$

which is the error in the approximation of f by g, combined with penalty terms of various types. The Mumford–Shah functional seeks to minimize (4.1) over all partitions Λ and piecewise smooth functions g. The first term in (4.1) measures the approximation of f by g, the second the (piecewise) smoothness of g, and the third asks that the curves that determine the partition be as short as possible. The weight μ_1 controls the balance between approximation and smoothness and the weight μ_2 the "amount" of segmentation one expects in the solution. Indeed, these parameters depend on the specific application where the technique is used and are sometimes implicitly controlled by the user.

As noted in [25], the choice of the L_2-norm implies that the Mumford–Shah functional depends only on the partition Λ. Indeed, once the partition is fixed, standard calculus of variations shows that **E** is a positive definite quadratic function with a unique minimum.

Now, a modified version of the $\widetilde{K}$-functional in the case of $p = r_1 = 2$ can be expressed as the infimum of

$$\text{(4.2)} \quad \widetilde{\mathbf{E}}(g, \Lambda) := \|f - g\|^2_{L_2(\Omega)} + \mu_1 \sum_{k=1}^{n_F(\Lambda)} \|Dg\|^2_{L_2(\Omega_k)} + \sum_{j=1}^{n_E(\Lambda)} \operatorname{len}(b_j) t_j^2 \|b_j''\|_2^2,$$

with

$$\sum_{j=1}^{n_E(\Lambda)} t_j^{-1} \le \mu_2^{-1},$$

where again μ_1, μ_2 are weights that play the same role as in (4.1). Comparing (4.1) with (4.2) we see that the main difference between **E** and $\widetilde{\mathbf{E}}$ lies in the different notions of lower-dimensional "structure." The energy gauge **E** uses only the length as a measure of lower-dimensional structure and does not distinguish, for example, between a straight line and a circle, both of the same length. Obviously, from an approximation theoretical point of view the circle is more complex. Also, note that $\widetilde{\mathbf{E}}$ counts the number of curves in the partition, $n_E(\Lambda)$, and ensures that it does not exceed μ_2^{-1}. This implies that $\widetilde{\mathbf{E}}$ implicitly considers the end points of the curves as vertices of the partition where breakpoints in curves are allowed.

One of our future goals is to investigate if functionals of the type (4.2) have any advantage over known variants of the Mumford–Shah in applications such as segmentations of images.

Another possible application of the $\widetilde{K}$-functional is the following. In [5] it is shown that wavelet shrinkage methods can provide a near-minimizer for the K-functional

$$\text{(4.3)} \qquad \|f - g\|_{L_2([0,1]^2)} + t|g|_{B_1^1(L_1([0,1]^2))}.$$

Thus, the wavelet shrinkage algorithm which is both fast and robust can be used to solve a variational problem that traditionally was considered computationally intensive and nonstable. In [7] it was shown that wavelet shrinkage methods also produce near-minimizers for a variational problem similar to (4.3) where the "smoothness" measure $|g|_{B_1^1(L_1)}$ is replaced by $|g|_{BV}$. This is the Total-Variation functional introduced in [28].

Now, consider the following problem. For a given $m \ge 2$, $r \ge 1$, and a function $f \in L_p([0,1]^2)$ find a "near-best" piecewise polynomial $\varphi = \sum_{k=1}^m \mathbb{1}_{\Delta_k} P_k$ over m triangles such that $\|f-\varphi\|_p \le C\sigma_{m,r}(f)_p$. While for wavelet approximation finding a "near-best" m-term approximation is a relatively simple task using the "greedy algorithm" (see [11]), finding a "near-best" piecewise polynomial approximation might be computationally impossible. Indeed, in [1] it is shown that the discrete version of this problem is NP-hard. Roughly speaking, this means the following. Assume that for given samples $f_{i,j} = f(i/N, j/N)$, $0 \le i, j \le N$, tolerance $\varepsilon > 0$ and $1 \le p < \infty$, there exists an (optimal) piecewise polynomial $\varphi \in S_m^{2,r}(\mathbb{R}^2)$ such that

$$\left(\sum_{i,j=1}^{N} |f_{i,j} - \varphi(i/N, j/N)|^p \right)^{1/p} \le \varepsilon, \tag{4.4}$$

where m is the smallest possible number of triangles for which (4.4) can be satisfied. It seems that there is no algorithm that runs in polynomial time (in the number of samples) and finds a "near-best" piecewise polynomial, over, say Cm triangles and satisfies (4.4). However, it is shown in [1] that there exists an algorithm that runs in $O(N^{16})$ and finds a piecewise polynomial over $Cm \log m$ triangles which satisfies (4.4).

So the following approach could be considered. Since, in this work we show that in some sense (ignoring for a moment the constants)

$$\sigma_{m,r}(f)_p \approx \widetilde{K}_{2,r}(f, m^{-1})_p,$$

perhaps one can "reverse" the approach of [5] and find a near-minimizer of the nonlinear approximation problem by applying Mumford–Shah techniques to minimize the $\widetilde{K}$-functional? This suggests the following algorithm for computing a "good" piecewise polynomial approximation over triangles:

1. Find a (local) minimum of $\widetilde{\mathbf{E}}$, given in (4.2) with $\mu_2 \approx m^{-1}$. The solution is determined by a partition $\Lambda \in \Lambda(\mu_2)$.
2. Approximate the curves of Λ by near-best "free-knot" polygons. The expected total of segments of the polygons is $O(m)$.
3. From the polygons of Step 2 compute $O(m)$ disjoint triangles that "almost cover" the subdomains of Λ. This can be done using geometric algorithms that correspond to the constructive techniques of Section 3.
4. For each triangle Δ computed in Step 3, calculate a near-best polynomial $P \in \Pi_{r-1}$ so that $\|f - P\|_{L_p(\Delta)} \le C E_{r-1}(f)_{L_p(\Delta)}$.

5. Proofs of the Main Results

Proof of Theorem 1.4. Let $f \in L_p([0,1]^2)$. We partition the square $[0,1]^2$ into smaller squares with side lengths $\lceil t^{-1} \rceil^{-1}$. There are $C\lceil t^{-2} \rceil$ of those. They define a partition where the curves b_j are simply the edges of the squares with attached parameters $t_j = 1$ and the domains Ω_k are the squares themselves. The partition is in $\Lambda(Ct^2)$ and we clearly have, by (1.3) and (1.4),

$$\begin{aligned}\widetilde{K}_{2,r}(f, Ct^2)_p^p &\leq \sum_{k=1}^{n_F(\Lambda)} K_r(f, \Omega_k)_p^p \\ &\leq CK_r(f, t^r, [0,1]^2)_p^p \\ &\leq C(p,r)\omega_r(f,t)_{L_p([0,1]^2)}^p.\end{aligned}$$

By virtue of (1.8) and (1.11), it is easy to see that $(f)_{\widetilde{B}_q^{\alpha/2}(L_p)} \leq C|f|_{B_q^\alpha(L_p)}$. ■

Lemma 5.1. *Let Ω be a bounded domain with a piecewise linear boundary and let $f \in L_p(\Omega)$. Then*

$$\widetilde{K}_{2,r}(f, C(\Omega)m^{-1})_p \leq \sigma_{m,r}(f)_p.$$

Proof. Let $\varphi \in S_m^r(\mathbb{R}^2)$ with $\varphi = \sum_{k=1}^m \mathbb{1}_{\Delta_k} P_k$ such that $\|f - \varphi\|_{L_p(\Omega)} \leq (1 + \varepsilon)\sigma_{m,r}(f)_p$ for some $\varepsilon > 0$. Then, the triangles $\{\Delta_k\}_{k=1}^m$ determine a partition $\Lambda_\varphi \in \Lambda(C(\Omega)m^{-1})$ where the first curves b_j, $j = 1, \ldots, n_E(\Omega)$, are the edges of the domains boundary and the remaining curves b_j, $j = n_E(\Omega)+1, \ldots, n_E(\Omega)+3m$, are the edges of the triangles with parameters $t_j = 1$. The domains of the partition are the triangles Δ_k and $\widetilde{\Omega} := \Omega \backslash \bigcup_{k=1}^m \Delta_k$. Thus,

$$\begin{aligned}\widetilde{K}_{2,r}(f, Cm^{-1})_p^p &\leq \sum_{k=1}^m K_r(f, \Delta_k)_p^p + K_r(f, \widetilde{\Omega})_p^p \\ &\leq \sum_{k=1}^m \|f - P_k\|_{L_p(\Delta_k)}^p + \|f\|_{L_p(\widetilde{\Omega})}^p \\ &\leq \|f - \varphi\|_{L_p(\Omega)}^p \\ &\leq (1+\varepsilon)^p \sigma_{m,r}(f)_p^p.\end{aligned}$$ ■

Proof of Theorem 1.5. It is well known that the Besov space $B_q^\alpha(L_q([0,1]^2))$ is characterized by its n-term approximation by B-spline wavelets of order $r > \alpha + 1 - 1/p$, on the dyadic cubes (see [12] and [14]). Note that these are piecewise polynomials on dyadic rings. By triangulating these dyadic rings we obtain a partition of $[0,1]^2$ where the curves b_j are the edges so that $K_2(b_j, 1)_{\infty,1} = 0$. The proof now follows by Lemma 5.1 when we observe that $\sigma_{C2^n,r}(f)_p$ is smaller than the degree of wavelet approximation on 2^n dyadic cubes. ■

In what follows we use the notation $\sigma_{m,r}(\varphi)_\infty$ to also denote the degree of approximation of a univariate continuous function $\varphi \in C([0,1])$ by C^{r-2}-"free-knot" splines of degree $r-1$ over m pieces (see [13, Section 12.4]).

Lemma 5.2. *Let $\varphi : [0,1] \to \mathbb{R}$ be continuous. Then, for $r \geq 1$,*

$$\sigma_{m,r}(\varphi)_\infty \leq C(r)K_r(\varphi, m^{-r})_{\infty,1}.$$

Proof. Let $g \in W^{r-1}(BV)$ such that

$$\|\varphi - g\|_\infty + m^{-r}|g^{(r-1)}|_{BV} \ \le\ 2K_r(\varphi, m^{-r})_{\infty,1}. \qquad \blacksquare$$

By [13, Theorem 12.4.5] we have that

$$\sigma_{m,r}(g)_\infty \le Cm^{-r}|g^{(r-1)}|_{BV}.$$

Let $s \in C^{r-2}([0,1])$ be a spline of degree $r-1$ over m pieces such that $\|g - s\|_\infty \le 2\sigma_{m,r}(g)_\infty$. Then,

$$\begin{aligned}\|\varphi - s\|_\infty &\le \|\varphi - g\|_\infty + \|g - s\|_\infty \\ &\le \|\varphi - g\|_\infty + Cm^{-r}\|g^{(r-1)}\|_{BV} \\ &\le CK_r(\varphi, m^{-r})_{\infty,1}.\end{aligned} \qquad \blacksquare$$

Corollary 5.3. *Let $b(u) := (b_1(u), b_2(u)) : [0,1] \to \mathbb{R}^2$ be a continuous planar curve. Then for each $m \ge 1$ there exists an interpolating polygon $s : [0,1] \to \mathbb{R}^2$ with m segments such that*

$$\max_{0\le u\le 1} |b(u) - s(u)| \le CK_2(b, m^{-2})_{\infty,1}. \tag{5.1}$$

Proof. By Lemma 5.2 we can find univariate polygons $\widetilde{s}_1, \widetilde{s}_2$ each with m segments, such that $\max_{0\le u\le 1} |b_i(u) - \widetilde{s}_i(u)| \le CK_2(b_i, m^{-2})_{\infty,1}$, $i = 1, 2$. It is easy to see that we can "correct" the polygons so that they interpolate the functions b_1, b_2 at the knots without hurting this estimate (by changing the constant). Merging the knot-sequences of $\widetilde{s}_1, \widetilde{s}_2$ we construct a polygon $s(u) := (s_1(u), s_2(u))$ with at most $2m$ knots for which (5.1) holds. $\blacksquare$

Proof of Theorem 1.7. Let $1 \le p < \infty$ and let $f \in L_\infty(\Omega)$. For each $m \ge 1$ there exists a partition $\Lambda \in \Lambda(m^{-1})$ such that

$$\left(\sum_{j=1}^{n_E(\Lambda)} \operatorname{len}(b_j)K_2(b_j, t_j^2)_{\infty,1} + \sum_{k=1}^{n_F(\Lambda)} K_r(f, \Omega_k)_p^p\right)^{1/p} \le 2\widetilde{K}_{2,r}(f, m^{-1})_p,$$

$0 < t_j \le 1$ and $\sum_{j=1}^{n_E(\Lambda)} t_j^{-1} \le m$. By Corollary 5.3, each boundary curve b_j of Λ can be approximated by an interpolating polygon s_j with $n_j := \lceil t_j^{-1} \rceil$ line-segments such that $|b_j(u) - s_j(u)| \le CK_2(b_j, t_j^2)_{\infty,1}$ for all $u \in [0,1]$. The total number of segments is therefore bounded by

$$\sum_{j=1}^{n_E(\Lambda)} \lceil t_j^{-1} \rceil \le Cm.$$

Observe that although the curves b_j may intersect only at their end points, the polygons s_j may self-intersect or intersect with other polygons. We now apply Theorem 3.1 that ensures the existence of polygonal piecewise disjoint regions $\{\widetilde{\Omega}_{k,l}\}$, of total complexity $\leq Cm$ such that $\widetilde{\Omega}_{k,l} \subseteq \Omega_k$, $k = 1, \dots, n_F(\Omega)$, $l = 1, \dots, n_F(k)$, and

$$|\Omega| - \sum_{k,i} |\widetilde{\Omega}_{k,i}| \leq C \sum_{j=1}^{n_E(\Lambda)} \operatorname{len}(b_j) K_2(b_j, t_j^2)_{\infty,1}.$$

We obtain that the error, in the p-norm, of replacing the domains Ω_k by the regions $\{\widetilde{\Omega}_{k,l}\}$ is bounded by

$$C\|f\|_\infty \left(\sum_{j=1}^{n_E(\Lambda)} \operatorname{len}(b_j) K_2(b_j, t_j^2)_{\infty,1} \right)^{1/p}.$$

We now triangulate the polyhedral subdomains $\{\widetilde{\Omega}_{k,l}\}$ into a total of $n_\Delta(k)$ triangles with disjoint interiors. We have that $\bigcup_{l=1}^{n_F(k)} \widetilde{\Omega}_{k,l} = \bigcup_{i=1}^{n_\Delta(k)} \Delta_{k,i}$ and $\sum_{k=1}^{n_F(\Lambda)} n_\Delta(k) \leq Cm$. On each triangle $\Delta_{k,i}$, we can find a bivariate polynomial of degree $r_2 - 1$ such that

$$\|f - P_{k,i}\|_{L_p(\Delta_{k,i})} = E_{r-1}(f, \Delta_{k,i})_p.$$

Thus, we define $\varphi \in S^r_{Cm}(\mathbb{R}^2)$ by

$$\varphi := \sum_{k=1}^{n_F(\Lambda)} \sum_{i=1}^{n_\Delta(k)} \mathbb{1}_{\Delta_{k,i}} P_{k,i}.$$

We now apply (2.1) to obtain

$$\begin{aligned}
\sigma_{Cm}(f)_p^p &\leq \|f - \varphi\|_p^p \\
&\leq C\|f\|_\infty^p \sum_{j=1}^{n_E(\Lambda)} \operatorname{len}(b_j) K_2(b_j, t_j^2)_{\infty,1} + \sum_{k=1}^{n_F(\Lambda)} \sum_{i=1}^{n_\Delta(k)} \|f - P_{k,i}\|_{L_p(\Delta_{k,i})}^p \\
&\leq C \max(\|f\|_\infty^p, 1) \left(\sum_{j=1}^{n_E(\Lambda)} \operatorname{len}(b_j) K_2(b_j, t_j^2)_{\infty,1} + \sum_{k=1}^{n_F(\Lambda)} K_r(f, \Omega_k)_p^p \right) \\
&\leq C_1^p(p, r) \max(\|f\|_\infty^p, 1) \widetilde{K}_{2,r}(f, C_2 m^{-1})_{L_{(\Omega)}}^p. \qquad \blacksquare
\end{aligned}$$

Remark. It is interesting to compare the method used in the proof of Theorem 1.7 which allocates to edges some "thickness" with the Bandlets of [26]. Also, there are variants of the Mumford–Shah functionals (see, e.g., [19]) that replace the edges' length "penalty" in (4.1) with a boundary set $B \subset \Omega$ "penalty" In this approach (4.1) takes the form

$$E(g, B) = \|f - g\|_{L_2(\Omega \setminus B)}^2 + \mu_1 \|Dg\|_{L_2(\Omega \setminus B)}^2 + \mu_2 \int_B dx.$$

Finally,

Proof of Theorem 1.9. The first statement follows from Lemma 5.1, while Theorem 1.7 yields (1.15). ■

Acknowledgments. The work by Micha Sharir has been supported by the Hermann Minkowski–MINERVA Center for Geometry at Tel Aviv University.

References

1. P. Agarwal, S. Suri (1998): *Surface approximation and geometric partitions*. SIAM J. Comput., **19**:1016–1035.
2. C. Bennett, R. Sharpley (1988): Interpolation of Operators. New York: Academic Press.
3. E. Candes, D. Donoho (1999): *Curvelets—A surprisingly effective nonadaptive representation for objects with edges*. In: Curves and Surfaces (L. L. Schumaker et al., eds.). Nashville, TN: Vanderbilt University Press.
4. E. Candès, D. Donoho (2001): *Curvelets and curvilinear integrals*. J. Approx. Theory, **113**:59–90.
5. A. Chambolle, R. DeVore, B. Lucier, N. Y. Lee (1998): *Nonlinear wavelet image processing: Variational problems, compression, and noise removal through wavelet shrinkage*. IEEE Trans. Image Proc., **7**:319–335.
6. B. Chazelle, L. J. Guibas (1986): *Fractional cascading*: II. *Applications*. Algorithmica, **1**:163–191.
7. A. Cohen, R. DeVore, P. Petrushev, H. Xu (1999): *Nonlinear approximation and the space $BV(\mathbb{R}^2)$*. Amer. J. Math., **121**:587–628.
8. A. Cohen, B. Matei (2001): *Compact representation of images by edge adapted multiscale transforms*. Preprint.
9. S. Dekel, D. Leviatan (to appear): *On the relation between piecewise polynomial and rational approximation in $L_p(\mathbb{R}^2)$*. Constr. Approx.
10. S. Dekel, D. Leviatan (2002): *The Bramble–Hilbert lemma for convex domains*. Preprint.
11. R. DeVore (1998): *Nonlinear approximation*. Acta Numerica, **7**:51–150.
12. R. DeVore, B. Jawerth, P. Popov (1992): *Compression of wavelet decompositions*. Amer. J. Math., **114**:737–785.
13. R. DeVore, G. Lorentz (1991): Constructive Approximation. New York: Springer-Verlag.
14. R. DeVore, V. Popov (1988): *Interpolation of Besov spaces*. Trans. Amer. Math. Soc., **305**:397–414.
15. M. N. Do, P. L. Dragotti, R. Shukla, M. Vetterli (2001): *On the compression of two-dimensional piecewise smooth functions*. Preprint.
16. D. L. Donoho (2000): *Beyond wavelets*: *A case study*. NSF-SIAM Conference Board in the Mathematical Sciences Lectures. Philadelphia, PA: SIAM.
17. H. Edelsbrunner, A. D. Robison, X. Shen (1990): *Covering convex sets with non-overlapping polygons*. Discrete Math., **81**:153–164.
18. J. Froment, S. Mallat (1992): *Second generation compact image coding with wavelets*. In: Wavelets: A Tutorial in Theory and Applications (C. K. Chui, ed.). New York: Academic Press.
19. G. Hewer, C. Kenney, B. Manjunath (1998): *Variational image segmentation using boundary functions*. IEEE Trans. Image Proc. **7**:1269–1282.
20. H. Johnen, K. Scherer (1976): *On the equivalence of the K-functional and the moduli of continuity and some applications*. Lecture Notes in Mathematics, Vol. 571, pp. 119–140, Berlin: Springer-Verlag.
21. B. Karaivanov, P. Petrushev (2001): *Nonlinear piecewise polynomial approximation beyond Besov spaces*. Industrial Mathematics Institute. University of South Carolina. Technical Report 01:13.
22. K. Kedem, R. Livne, J. Pach, M. Sharir (1986): *On the union of Jordan regions and collision-free translational motion amidst polygonal obstacles*. Discrete Comput. Geom., **1**:59–71.
23. G. G. Lorentz, M. Golitschek, Y. Makovoz (1996): Constructive Approximation—Advanced Problems. New York: Springer-Verlag.
24. D. Mumford, J. Shah (1985): *Boundary detection by minimizing functionals*. In: Proc. IEEE Conf. Computer Vision and Pattern Recognition, San Francisco, CA.
25. D. Mumford, J. Shah (1989): *Optimal approximations of piecewise smooth functions and associated variational problems*. Comm. Pure Appl. Math., **42**:577–685.
26. E. Pennec, S. Mallat (2001): *Image compression with geometrical wavelets*. Preprint.
27. P. Prandoni, M. Vetterli (1999): *Approximation and compression of piecewise-smooth functions*. Philos. Trans. Roy. Soc. London Ser. A. **357**:2573–2591.

28. L. RUDIN, S. OSHER, E. FATEMI (1992): *Nonlinear total variation based noise removal algorithms*. Phys. D, **60**:259–268.
29. D. SCHILLING, P. COSMAN (2001): *Feature-preserving image coding for very low bit rates*. In: Proc. IEEE DCC, Snowbird, Utah, pp. 103–112.

S. Dekel
RealTimeImage
6 Hamasger St.
Or-Yehuda 60408
Israel
shai.dekel@turboimage.com

D. Leviatan
School of Mathematical Sciences
Tel Aviv University
Tel Aviv 69978
Israel
leviatan@post.tau.ac.il

M. Sharir
School of Computer Science
Tel Aviv University
Tel Aviv 69978
Israel
michas@post.tau.ac.il