ADAPTIVE MULTIVARIATE APPROXIMATION USING BINARY SPACE PARTITIONS AND GEOMETRIC WAVELETS

S. DEKEL† AND D. LEVIATAN‡

Abstract. The Binary Space Partition (BSP) technique is a simple and efficient method to adaptively partition an initial given domain to match the geometry of a given input function. As such the BSP technique has been widely used by practitioners, but up until now no rigorous mathematical justification to it has been offered. Here we attempt to put the technique on sound mathematical foundations, and we offer an enhancement of the BSP algorithm in the spirit of what we are going to call *geometric wavelets*. This new approach to sparse geometric representation is based on recent development in the theory of multivariate nonlinear piecewise polynomial approximation. We provide numerical examples of n-term geometric wavelet approximations of known test images and compare them with dyadic wavelet approximation. We also discuss applications to image denoising and compression.

AMS subject classifications. 41A15, 41A25, 41A17, 41A63, 65T60, 68U10.

Key words. Binary Space Partitions, Geometric Wavelets, Piecewise polynomial approximation, Nonlinear approximation, Adaptive multivariate approximation.

1. Introduction. The Binary Space Partition (BSP) technique is widely used in image processing and computer graphics [15], [17], [19], and can be described as follows. Given an initial convex domain in $\mathbb{R}^d$, such as $[0,1]^d$, and a function $f \in L_p\big([0,1]^d\big)$, $0 < p < \infty$, one subdivides the initial domain into two subdomains, by intersecting it with an hyper-plane. The subdivision is performed so that a given cost function is minimized. This subdivision process then proceeds recursively on the subdomains until some exit criterion is met. To be specific, we describe the algorithm of [17], which is a BSP algorithm for the purpose of finding a compact geometric description of the target function, in this case a digital image ($d = 2$).

In [17], at each stage of the BSP process, for a given convex polytope Ω, the algorithm finds two subdomains Ω', Ω'', and two bivariate (linear) polynomials $Q_{\Omega'}, Q_{\Omega''}$, that minimize the quantity

$$\|f - Q_{\Omega'}\|^p_{L_p(\Omega')} + \|f - Q_{\Omega''}\|^p_{L_p(\Omega'')}\,,$$

over all pairs Ω', Ω'' of polyhedral domains that are the result of a binary space partition of Ω. The polynomials $Q_{\Omega'}, Q_{\Omega''}$ are found using the least-squares technique with $p = 2$. The goal in [17] is to encode a *cut* of the BSP tree, i.e., a sparse piecewise polynomial approximation of the original digital image based on a union of disjoint polytopes from the BSP tree. Also, to meet a given bit target, rate-distortion optimization strategies are used (see also [21]).

Inspired by the recent progress in multivariate piecewise polynomial approximation, made by Karaivanov, Petrushev and collaborators [13], [14], we propose a modification to the above method which can be described as a *geometric wavelets* approach. Let Ω' be a *child* of Ω in a BSP tree, i.e., $\Omega' \subset \Omega$ and Ω' has been created by a BSP partition of Ω. We use the polynomial approximations $Q_\Omega, Q_{\Omega'}$ that were found for these domains by the local optimization algorithm above and define

$$\psi_{\Omega'} := \psi_{\Omega'}(f) := \mathbf{1}_{\Omega'}\,(Q_{\Omega'} - Q_\Omega)\,, \tag{1.1}$$

†RealTimeImage, 6 Hamasger St., Or-Yehuda 60408, Israel; shai.dekel@turboimage.com

‡School of Mathematical Sciences, Sackler Faculty of Exact Sciences, Tel-Aviv University, Tel-Aviv 69978, Israel; leviatan@math.tau.ac.il

as the geometric wavelet associated with the subdomain Ω' and the function f. A reader familiar with wavelets (see, e.g., [3], [7]), will notice that $\psi_{\Omega'}$ is a 'local difference' component that belongs to the detail space between two levels in the BSP tree, a 'low resolution' level associated with Ω and a 'high resolution' level associated with Ω'. Also, these wavelets have what may be regarded as the 'zero moments' property, i.e., if f is locally a polynomial over Ω, then we get $Q_{\Omega'} = Q_\Omega = f$ and $\psi_{\Omega'} = 0$. However, the BSP method is highly nonlinear, both the partition and the geometric wavelets are so much dependent on the function f, that one cannot expect some of the familiar properties of wavelets like a two-scale relation, a partition of unity or, spanning of some a-priori given spaces.

Our modified BSP algorithm proceeds as follows. We apply the BSP algorithm and create a 'full' BSP tree $\mathcal{P}$. Obviously, in applications, the subdivision process is terminated when the leaves of the tree are subdomains of sufficiently small volume, or equivalently, in image processing, when the subdomains contain only a few pixels. We shall see that under certain mild conditions on the partition $\mathcal{P}$ and the function f we have

$$f = \sum_{\Omega \in \mathcal{P}} \psi_\Omega(f), \quad \text{a.e. in } [0,1]^d,$$

where

$$\psi_{[0,1]^d} := \psi_{[0,1]^d}(f) := \mathbf{1}_{[0,1]^d} Q_{[0,1]^d}.$$

We then compute all the geometric wavelets (1.1) and sort them according to their L_p norm, i.e.,

$$\left\|\psi_{\Omega_{k_1}}\right\|_p \geq \left\|\psi_{\Omega_{k_2}}\right\|_p \geq \left\|\psi_{\Omega_{k_3}}\right\|_p \cdots . \tag{1.2}$$

Given an integer $n \in \mathbb{N}$, we approximate f by the n-term geometric wavelet sum

$$\sum_{j=1}^{n} \psi_{\Omega_{k_j}}. \tag{1.3}$$

The sum (1.3) is, in some sense, a generalization of the classical n-term wavelet approximation (see [7] and references therein), where the wavelets are constructed over dyadic cubes.

A key observation is that the BSP algorithm described above is a *geometric greedy* algorithm. At each stage of the algorithm we try to find a locally optimal partition of a given subdomain. Indeed, the problem of finding an optimal triangulation or partition is associated with an NP-HARD problem (see the discussion in [6, Section 4] and references therein).

It is known in classical wavelet theory (see, e.g., [7]) that the energy of the wavelet basis coefficients in some l_τ-norm, $0 < \tau < p$, is a valid gauge for the 'sparseness' of the wavelet representation of the given function. We follow this idea extending it to our geometric wavelet setup. Thus we take as a reasonable *benchmark* by which to measure the efficiency of the greedy algorithm, a BSP partition that 'almost' minimizes, over all possible partitions, the sum of energies of the geometric wavelets of a given function, namely,

$$\left(\sum_{\Omega \in \mathcal{P}} \|\psi_\Omega\|_p^\tau \right)^{1/\tau}, \tag{1.4}$$

for some $0 < \tau < p$.

We note the following geometric sub-optimality of the BSP algorithm (see [12], [25] and references therein). We say that a BSP for n disjoint objects in a given convex domain is a recursive dissection of the domain into convex regions such that each object (or part of an object) is in a distinct region. Ideally, every object should be in one convex region, but sometimes it is inevitable that some of the objects are dissected. The size of the BSP is defined as the number of leaves in the resulting BSP tree.

It can be shown that for a collection of n disjoint line segments in the plane, there exists a BSP of complexity $O(n \log n)$. Recently, Tòth [24] showed a lower bound of $\Omega(n \log n / \log\log n)$, meaning that for $d = 2$, in the worst-case, the BSP algorithm might need slightly more elements to 'capture' arbitrary linear geometry. In higher dimension, the performance of the BSP, in the worst case, decreases. For example the known lower-bound for the BSP of collection of n disjoint rectangles in $\mathbb{R}^3$ is $\Omega(n^2)$.

The paper is organized as follows. In Section 2, we outline the algorithmic aspects of the geometric wavelet approach so that the reader who is less interested in the rigorous mathematics may skip Section 3 and proceed directly to Section 4. In Section 3, we review the more theoretical aspects of our approach and we provide some details on the approximation spaces that are associated with the method. It is interesting to note that, while the approximation spaces corresponding to nonlinear n-term wavelet approximation are linear Besov spaces (see [7] for details), the adaptive nature of the geometric wavelets implies that the corresponding approximation spaces are nonlinear. Nevertheless, it turns out that the problem at hand is 'tamed' enough so as to enable the application of the classical machinery of the Jackson and Bernstein. Specifically, the analysis can be carried out because we are adaptively selecting one nested fixed partition for a given function, from which we select n-term geometric wavelets for any n. (In contrast, general adaptive piecewise polynomial n-term approximation [6] allows for each n, the selection of any n pieces, with no assumptions that they are taken from a fixed partition.) We conclude the paper with Section 4, where some numerical examples of n-term geometric wavelet approximation of digital images and discussion of possible applications in image denoising and compression.

2. Adaptive BSP partitions and the geometric wavelet approximation algorithm. Let $\Pi_{r-1} := \Pi_{r-1}(\mathbb{R}^d)$ denote the multivariate polynomials of total degree $r-1$ (order r) in d variables. Given a bounded domain $\Omega \subset \mathbb{R}^d$, we denote the *degree (error) of polynomial approximation* of a function $f \in L_p(\Omega)$, $0 < p \leq \infty$,

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

Recall that the greedy BSP algorithm consists of finding, at each step, an optimal dissection of some domain Ω, and computing polynomials $Q_{\Omega'}$ and $Q_{\Omega''}$ that best approximate the target function f in the p-norm over the children $\Omega', \Omega'' \subset \Omega$. In practice, we will have a suboptimal dissection, and *near best* approximation. Thus, we are going to assume that for each $\Omega \in \mathcal{P}$, Q_Ω is a near best approximation, i.e.,

$$\|f - Q_\Omega\|_{L_p(\Omega)} \leq C E_{r-1}(f, \Omega)_p, \tag{2.1}$$

where C is independent of f and Ω, but may depend on parameters like d, r and possibly p. We shall see in Section 3 that for the purpose of analysis when $p \leq 1$, we need the stronger assumption that Q_Ω is a (possibly not unique) best approximation.

Let $\mathcal{P}$ be a partition of $[0,1]^d$, and let Ω' be a child of $\Omega \in \mathcal{P}$. For $f \in L_p\big([0,1]^d\big)$, $0 < p < \infty$, we set $\psi_{\Omega'}$ as in (1.1). As noted in the introduction, the function $\psi_{\Omega'}$ in (1.1) may be regarded as a local wavelet component of the function f that corresponds to the partition $\mathcal{P}$. For $0 < \tau \leq p$ we denote the τ-energy of the sequence of geometric wavelets by the l_τ-norm of its L_p norms,

$$\mathcal{N}_\tau\left(f,\mathcal{P}\right) := \left(\sum_{\Omega\in\mathcal{P}} \left\|\psi_\Omega\right\|_p^\tau\right)^{1/\tau}. \tag{2.2}$$

We will show that, under some mild conditions, the geometric wavelet expansion converges to the function. Namely, we introduce a weak constraint on the BSP partitions, which allows the analysis below to be carried out (see for example the proof of Theorem 3.5 below). We say that $\mathcal{P}$ is in $BSP(\rho)$, $3/4 < \rho < 1$, if for any child Ω', of Ω, we have

$$\left|\Omega'\right| \leq \rho \left|\Omega\right|, \tag{2.3}$$

where $|V|$ denotes the volume of a bounded set $V \subset \mathbb{R}^d$.

THEOREM 2.1. *Assume that $\mathcal{N}_\tau\left(f,\mathcal{P}\right) < \infty$, for some $f \in L_p\big([0,1]^d\big)$, $0 < p < \infty$, $0 < \tau < p$, and $\mathcal{P} \in \mathrm{BSP}(\rho)$. Then*

1. *$f = \sum_\Omega \psi_\Omega$, absolutely, a.e. in $[0,1]^d$,*
2. *$\left\|f\right\|_p \leq C\left(d,r,p,\tau,\rho\right)\mathcal{N}_\tau\left(f,\mathcal{P}\right)$.*

Proof. The proof is almost identical to the proof of [13, Theorem 2.17], except that here we take $\eta = p$, and we replace [13, Lemma 2.7] by Lemma 2.4 below. □

Thus, it is expedient to look for partitions (and τ) that yield finite energy, better still, that minimize the energy. Obviously, this is not always possible or too costly, and we are willing to settle for somewhat less. To this end, we define

DEFINITION 2.2. *For $f \in L_p\big([0,1]^d\big)$ and $0 < \tau < p < \infty$, we say that $\mathcal{P}_\tau(f) \in \mathrm{BSP}(\rho)$, is a* near best partition *if*

$$\mathcal{N}_\tau\left(f,\mathcal{P}_\tau(f)\right) \leq C \inf_{\mathcal{P}\in\mathrm{BSP}(\rho)} \mathcal{N}_\tau\left(f,\mathcal{P}\right). \tag{2.4}$$

Let $\mathcal{P}_D$ be the BSP partition that gives the classical subdivision of $[0,1]^d$ into dyadic cubes. This can be done for example in the case $d = 2$ by partitioning $[0,1]^2$ along the line $x_1 = 1/2$ and then partitioning the two resulting rectangles along the line $x_2 = 1/2$. We get four dyadic cubes and we proceed on each one recursively in the same manner. In Section 3 we show the following relationship between $\mathcal{N}_\tau\left(f,\mathcal{P}_\tau(f)\right)$ and the Besov seminorm of f (compare with the classical dyadic wavelet-type characterization of Besov spaces [10] and in particular the quantities $N_3\left(f\right)$ and $N_4\left(f\right)$ therein).

We will show that for $f \in L_p\big([0,1]^d\big)$, $0 < p < \infty$, $\alpha > 0$, and $1/\tau = \alpha + 1/p$, we have

$$\mathcal{N}_\tau\left(f,\mathcal{P}_\tau(f)\right) \leq C\mathcal{N}_\tau\left(f,\mathcal{P}_D\right) \approx \left|f\right|_{B_\tau^{d\alpha,r}}, \tag{2.5}$$

where $B_\tau^{\gamma,r}$, $\gamma > 0$, is the classical Besov space (see Definition 3.1 below). The proof follows from the discussion beyond (3.6), and especially from (3.16).

We note that (2.2) was already defined in [16] for the special case of partitions over dyadic boxes. Also in [16], the author gives an algorithm to find the best dyadic

box partition (see also [11]), thereby providing a complete solution to a restricted version of (2.4).

For $1 < p < \infty$, a more subtle but sharper definition of $\mathcal{P}_\tau(f)$ would be to define it as an 'almost' minimizer of the weak ℓ_τ-norm of its corresponding geometric wavelets instead of the ℓ_τ-norm (2.2). Recall that the weak ℓ_τ-norm of a sequence $\{a_k\}$ is defined by

$$\|\{a_k\}\|_{w\ell_\tau} := \inf\left\{M : \ \#\left\{k : \ |a_k| > M\varepsilon^{1/\tau}\right\} \le \varepsilon^{-1}, \ \forall \varepsilon > 0\right\},$$

and satisfies $\|\{a_k\}\|_{w\ell_\tau} \le \|\{a_k\}\|_{\ell_\tau}$. This corresponds to a well known fact that n-term wavelet approximation can be estimated using the weaker p-norm when $1 < p < \infty$ (see [13, Theorem 3.3] for details, and see [7, Theorem 7.2.5] for the case of classic dyadic wavelets).

As we shall see, $\mathcal{N}_\tau(f, \mathcal{P})$ may serve as a 'quality gauge' for partitions, when τ takes certain values strictly smaller than p. The following example demonstrates the role of τ.

EXAMPLE 2.3. *Let $\widetilde{\Omega} \subset [0,1]^d$ be a convex polytope, and denote $f(x) := \mathbf{1}_{\widetilde{\Omega}}(x)$. Assume $\mathcal{P}$ is a partition such that for each $\Omega \in \mathcal{P}$, either $\widetilde{\Omega} \subseteq \Omega$, $\Omega \subseteq \widetilde{\Omega}$ or $int(\Omega \cap \widetilde{\Omega}) = \emptyset$, where $int(E)$ denotes the interior of $E \subset \mathbb{R}^d$. Then, for $p = 2$ and $r = 1$ it is easy to see that*

$$Q_\Omega = \begin{cases} \frac{|\widetilde{\Omega}|}{|\Omega|}, & \widetilde{\Omega} \subseteq \Omega, \\ 0, & int(\Omega \cap \widetilde{\Omega}) = \emptyset. \end{cases}$$

Therefore we have $\psi_{[0,1]^d} = \left|\widetilde{\Omega}\right| \mathbf{1}_{[0,1]^d}$ and, for $\Omega' \in \mathcal{P}$, a child of Ω,

$$\|\psi_{\Omega'}\|_2^\tau = \|Q_{\Omega'} - Q_\Omega\|_{L_2(\Omega')}^\tau = \begin{cases} \left|\widetilde{\Omega}\right|^\tau \left(\frac{1}{|\Omega'|} - \frac{1}{|\Omega|}\right)^\tau |\Omega'|^{\tau/2}, & \widetilde{\Omega} \subseteq \Omega', \\ \left|\widetilde{\Omega}\right|^\tau \frac{1}{|\Omega|^\tau} |\Omega'|^{\tau/2}, & int(\widetilde{\Omega}) \subset \Omega \setminus \Omega', \\ 0, & int(\Omega \cap \widetilde{\Omega}) = \emptyset \text{ or } \Omega \subseteq \widetilde{\Omega}. \end{cases}$$

Thus, the energy of the geometric wavelets is given by the formal sum

$$\begin{aligned} \mathcal{N}_\tau^\tau(f, \mathcal{P}) &= \sum_{\Omega \in \mathcal{P}} \|\psi_\Omega\|_2^\tau \\ &= |\widetilde{\Omega}|^\tau \left(1 + \sum_{\substack{\widetilde{\Omega} \subseteq \Omega' \\ \Omega' \text{ child} \\ \text{of } \Omega}} \left(\frac{1}{|\Omega'|} - \frac{1}{|\Omega|}\right)^\tau |\Omega'|^{\tau/2} + \frac{1}{|\Omega|^\tau} \left(|\Omega| - |\Omega'|\right)^{\tau/2}\right). \quad (2.6) \end{aligned}$$

The above sum converges, for example, if $\mathcal{P}$ is in BSP(ρ), for some $\rho < 1$. In the

special case $\tau = 2$ we get

$$\begin{aligned}\mathcal{N}_2^2(f,\mathcal{P}) &= \left|\widetilde{\Omega}\right|^2\left(1+\sum_{\substack{\widetilde{\Omega}\subseteq\Omega'\\ \Omega' \text{ child of } \Omega}}\left(\frac{1}{|\Omega'|}-\frac{1}{|\Omega|}\right)^2|\Omega'|+\frac{1}{|\Omega|^2}(|\Omega|-|\Omega'|)\right)\\ &= \left|\widetilde{\Omega}\right|^2\left(1+\sum_{\substack{\widetilde{\Omega}\subseteq\Omega'\\ \Omega' \text{ child of } \Omega}}\left(\frac{1}{|\Omega'|}-\frac{1}{|\Omega|}\right)\right)\\ &= |\widetilde{\Omega}|,\end{aligned}$$

which implies that $\mathcal{N}_2(f,\mathcal{P}) = \|f\|_2$. Since this equality holds for any partition that satisfies the above conditions, it follows that $\mathcal{N}_2(f,\mathcal{P})$ is not a good sparsity gauge for adaptive partitions when $p = 2$.

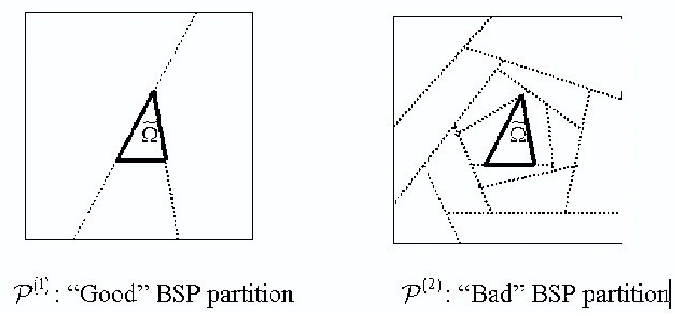

FIG. 2.1. *Two BSP partitions with* $\mathcal{N}_2\ f,\mathcal{P}^{(1)}\ = \mathcal{N}_2\ f,\mathcal{P}^{(2)}\ = \|f\|_2$

Referring to Figure 2.1, we see that the partition $\mathcal{P}^{(1)}$ is optimal since its BSP lines coincide with the hyper-planes that describe $\partial\widetilde{\Omega}$, while $\mathcal{P}^{(2)}$ contains 'unnecessary' subdomains. Nevertheless, the equality $\mathcal{N}_2\left(f,\mathcal{P}^{(1)}\right) = \mathcal{N}_2\left(f,\mathcal{P}^{(2)}\right) = \|f\|_2$ holds. However, things change dramatically when we choose a sufficiently small τ. In this case, the ℓ_τ norm serves almost as counting measure and since the sum (2.6) contains significantly less non-zero elements in the case of $\mathcal{P}^{(1)}$, we obtain that $\mathcal{N}_\tau\left(f,\mathcal{P}^{(1)}\right)$ is much smaller than $\mathcal{N}_\tau\left(f,\mathcal{P}^{(2)}\right)$. ∎

Thus, we wish to address the issue of the expected range of the parameter τ for digital images and $p = 2$. If the image contains a curve singularity that is not a straight line, then the theory of Section 3 below suggests that we should take $\tau \geq 2/5$. Since, in a way, dyadic wavelets are a special case of geometric wavelets, we can obtain an upper bound estimate on τ using the ideas of [8]. One needs to compute the discrete dyadic wavelet transform of the image and then compute the rate of convergence of the n-term wavelet approximation, by fitting the error function with the exponent $e(f,n) := C(f)\,n^{-\alpha(f)}$. Since we expect geometric wavelets to perform at least at the rate of dyadic wavelets, we should take $\tau \leq 2/(2\alpha(f)+1)$.

Going back to the greedy BSP step described in the introduction, let $(\Omega',\Omega'') \in \mathrm{BSP}(\Omega)$, and let Q_Ω ,$Q_{\Omega'}$,$Q_{\Omega''}$ be the near best polynomials approximations for their corresponding subdomains. Then, we have by (1.1)

$$\begin{aligned}&\|\psi_{\Omega'}\|_p^\tau + \|\psi_{\Omega''}\|_p^\tau\\ &\leq C\left(\|f-Q_\Omega\|_{L_p(\Omega)}^p + \|f-Q_{\Omega'}\|_{L_p(\Omega')}^p + \|f-Q_{\Omega''}\|_{L_p(\Omega'')}^p\right).\end{aligned} \tag{2.7}$$

Observing that Q_Ω has been already been determined at a previous (greedy) step, we have that the local greedy optimization step of [17], will capture the geometry in which the local geometric wavelet components of f are relatively small. If we denote the levels of a BSP partition $\mathcal{P}$ of $[0,1]^d$, by $\{\mathcal{P}_m\}_{m\in\mathbb{N}}$, and we say that $\Omega' \in \mathcal{P}_{m+1}$ is a *child* of $\Omega \in \mathcal{P}_m$ if $\Omega' \subset \Omega$. Then we note that our analysis also suggests that a significant improvement may be obtained if the local optimization step is carried out for several levels at once. Namely, given $\Omega \in \mathcal{P}_m$, to try to minimize, for some (small) $J \geq 2$,

$$\sum_{j=1}^{J} \sum_{\substack{\widetilde{\Omega}\subset\Omega \\ \widetilde{\Omega}\in\mathcal{P}_{m+j}}} \left\| f - Q_{\widetilde{\Omega}} \right\|_{L_p(\widetilde{\Omega})}^p. \tag{2.8}$$

Finally, we return to the proof of Theorem 2.1. Condition (2.3) implies that

$$(1-\rho)\,|\Omega| \leq |\Omega'| \leq \rho\,|\Omega|\,. \tag{2.9}$$

This condition for BSP partitions corresponds to the *weak locally regularity (WLR)* condition that is assumed for triangulations in [13]. Observe that a BSP-partition still allows the polytopes of the partition to be adaptive to the geometry of the function to be approximated, i.e., the polytopes may become as thin as one may wish, so long as the 'thinning' process occurs over a sequence of levels of the partition. Also, note that we have not limited the complexity of the polytopes. Indeed, polytopes at the mth level may be of complexity m.

we need the following results on norms of polynomials over convex domains.

LEMMA 2.4. *Let $P \in \Pi_{r-1}(\mathbb{R}^d)$, and let $0 < \rho < 1$ and $0 < p, q \leq \infty$.*

(a) Assume that $\Omega', \Omega \subset \mathbb{R}^d$ are bounded convex domains, such that $\Omega' \subseteq \Omega$, and $(1-\rho)|\Omega| \leq |\Omega'|$. Then

$$\|P\|_{L_p(\Omega)} \leq C(d,r,p,\rho)\|P\|_{L_p(\Omega')}.$$

(b) For any bounded convex domain $\Omega \subset \mathbb{R}^d$,

$$\|P\|_{L_q(\Omega)} \approx |\Omega|^{1/q-1/p}\|P\|_{L_p(\Omega)},$$

with constants of equivalency depending only on d, r, p, and q.
(c) If Ω' is a child of Ω in a BSP *partition $\mathcal{P} \in \mathrm{BSP}(\rho)$, then*

$$\|P\|_{L_q(\Omega)} \approx \|P\|_{L_q(\Omega')} \approx |\Omega|^{1/q-1/p}\|P\|_{L_p(\Omega')},$$

with constants of equivalency depending only on d, r, p, q, and ρ.

Proof. The proof of (a) and (b) can be found in [5, Lemma 3.1] and the first part of the proof of [5, Lemma 3.2], respectively. Assertion (c) follows from (a) and (b), since by the properties of $\mathcal{P}$, we have that all the domains concerned, are convex, and the following equivalence of volumes holds

$$(1-\rho)\,|\Omega| \leq |\Omega'| \leq (1-\rho)^{-1}\,|\Omega \setminus \Omega'|. \square$$

We conclude this section by outlining the steps of the adaptive geometric wavelet approximation algorithm:

1. Given $f \in L_p\big([0,1]^d\big)$, find a BSP partition using local steps of optimal partitions and polynomial approximations (see discussion above (2.8)).

2. For each subdomain of the partition, $\Omega \in \mathcal{P}$, compute the $p-$norm of the corresponding geometric wavelet ψ_Ω.

3. Sort the geometric wavelets according their energy as in (1.2). As in the case of classical dyadic wavelets, this step can be simplified by using thresholding (see [7, Section 7.8]).

4. For any $n \geq 1$, construct the n-term geometric wavelet sum (1.3).

3. Theoretical aspects of the geometric wavelet approach. One of the greatest challenges in approximation theory is the characterization of adaptive multivariate piecewise polynomial approximation (see the discussion in [7, Section 6.5] and [6]). Given $f \in L_p\big([0,1]^d\big)$, we wish to understand the behavior of the degree of nonlinear approximation

$$\inf_{S \in \Sigma_n^r} \|f - S\|_{L_p\left([0,1]^d\right)}, \tag{3.1}$$

where Σ_n^r is the collection $\sum_{k=1}^{n} \mathbf{1}_{\Omega_k} P_k$, and $\{\Omega_k\}$ are convex polytopes with disjoint interiors, such that $\bigcup_{k=1}^{n} \Omega_k = [0,1]^d$, and $P_k \in \Pi_{r-1}$, $1 \leq k \leq n$. Usually $\{\Omega_k\}$ are assumed to be simplices (triangles in the bivariate case), so as to keep their complexity bounded. However, when using the BSP approach, the polytopes $\{\Omega_k\}$ can be of arbitrary complexity and descendant polytopes are contained in their ancestors.

In the univariate case there is a certain equivalence between the two n-term approximation methods, wavelets and piecewise polynomials. Namely, the approximation spaces associated with the two methods are characterized by the same Besov spaces [7] . The advantage of wavelet approximation over piecewise polynomial approximation is the simplicity and efficiency with which one can implement it. When $d \geq 2$, these two methods are no longer equivalent. Wavelet approximation is still characterized by the (linear) Besov spaces, while the approximation spaces associated with piecewise polynomials are known to be nonlinear spaces [6], and their characterization remains an open problem.

While the geometric wavelet algorithm of Section 2 is highly adaptive and geometrically flexible, it is nothing but a 'tamed' version of the piecewise polynomial method (see also discussion in [13]). To explain this, for a given BSP partition $\mathcal{P}$, denote by $\Sigma_n^r(\mathcal{P})$ the collection

$$\sum_{k=1}^{n} \mathbf{1}_{\Omega_k} P_k, \quad \Omega_k \in \mathcal{P}, \quad P_k \in \Pi_{r-1}, \quad 1 \leq k \leq n. \tag{3.2}$$

Observe that the n-term geometric wavelet sum (1.3) is in $\Sigma_n^r(\mathcal{P})$, for the given partition $\mathcal{P}$. Let $\mathcal{P}_\tau(f) \in \mathrm{BSP}(\rho)$, be the near best partition of Definition 2.2 for $f \in L_p\big([0,1]^d\big)$, $0 < \tau < p$. Then, the degree of nonlinear approximation from the near best partition is given by

$$\sigma_{n,r,\tau}(f)_p := \inf_{S \in \Sigma_n^r(\mathcal{P}_\tau(f))} \|f - S\|_p. \tag{3.3}$$

We see that the main difference between (3.1) and (3.3) is that in the latter, the n-term approximations are taken from a fixed partition. This is a major advantage

as one of the main difficulties one encounters when trying to analyze the degree of approximation of n-term piecewise polynomial approximation (where the supports have disjoint interiors), is that for $S_1, S_2 \in \Sigma_n^r$ we may have, in the worst case, that $S_1 + S_2$ is of complexity $O\left(n^d\right)$, that is, supported on n^d domains with disjoint interiors. On the other hand, if we have a fixed partition $\mathcal{P}$, and two piecewise polynomials $S_1, S_2 \in \Sigma_n^r(\mathcal{P})$, then $S_1 + S_2 \in \Sigma_{2n}^r(\mathcal{P})$. Still, even for a fixed partition, it is hard to find a solution to (3.3). As we demonstrate below, a good method to compute an n-term piecewise polynomial approximation is to take the n-term geometric wavelet sum (1.3) (see the proof of Theorem 3.6).

The goal of this section is to provide some characterization of the adaptive geometric wavelet approximation, where the n-terms are taken from a near best adaptive partition $\mathcal{P}_\tau(f)$, which we consider as a benchmark to any of the greedy algorithms discussed above. To this end we denote by $A_{q,\tau}^{\gamma,r}(L_p)$, $\gamma > 0$, $0 < q \le \infty$, $0 < \tau < p$, the *approximation space* corresponding to nonlinear approximation from $\mathcal{P}_\tau(f)$. This is the collection of all functions $f \in L_p\left([0,1]^d\right)$ for which the error (3.3) roughly 'decays' at the rate $n^{-\gamma}$, i.e., $f \in L_p\left([0,1]^d\right)$ for which

$$(f)_{A_{q,\tau}^{\gamma,r}(L_p)} := \begin{cases} \left(\sum_{m=0}^{\infty} \left(2^{m\gamma}\sigma_{2^m,r,\tau}(f)_p\right)^q\right)^{1/q}, & 0 < q < \infty, \\ \sup_{m\ge 0}\left(2^{m\gamma}\sigma_{2^m,r,\tau}(f)_p\right), & q = \infty, \end{cases}$$

is finite.

Recall that for $f \in L_\tau(\Omega)$, $0 < \tau \le \infty$, $h \in \mathbb{R}^d$ and $r \in \mathrm{N}$, we denote the rth order difference operator

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

where $[x,y]$ denotes the line segment connecting the points $x, y \in \mathbb{R}^d$. The *modulus of smoothness of order* r of $f \in L_\tau(\Omega)$ (see e.g. [7], [9]), is defined by

$$\omega_r(f,t)_{L_\tau(\Omega)} := \sup_{|h|\le t} \left\|\Delta_h^r(f,\Omega,\cdot)\right\|_{L_\tau(\Omega)}, \quad t > 0,$$

where for $h \in \mathbb{R}^d$, $|h|$ denotes the length of h. We also denote

$$\omega_r(f,\Omega)_\tau := \omega_r(f, \mathrm{diam}(\Omega))_{L_\tau(\Omega)}. \tag{3.4}$$

Definition 3.1. *For $\gamma > 0$, $\tau > 0$ and $r \in N$, the* Besov space $B_\tau^{\gamma,r}$ *is the collection of functions $f \in L_\tau\left([0,1]^d\right)$ for which*

$$|f|_{B_\tau^{\gamma,r}} := \left(\sum_{m=0}^{\infty}\left(2^{\gamma m}\omega_r\left(f, 2^{-m}\right)_{L_\tau([0,1]^d)}\right)^\tau\right)^{1/\tau} < \infty.$$

Definition 3.2. *For $0 < p < \infty$, $\alpha > 0$, $\rho > 0$, and $1/\tau := \alpha + 1/p$, we define the Geometric B-space $\mathcal{GB}_\tau^{\alpha,r}$, $r \in N$, as the set of functions $f \in L_p\left([0,1]^d\right)$ for which*

$$(f)_{\mathcal{GB}_\tau^{\alpha,r}} := \left(\inf_{\mathcal{P}\in \mathrm{BSP}(\rho)} \sum_{\Omega\in\mathcal{P}} \left(|\Omega|^{-\alpha}\,\omega_r\left(f,\Omega\right)_\tau\right)^\tau\right)^{1/\tau} < \infty. \tag{3.5}$$

Note that the smoothness measure $(\,\cdot\,)_{\mathcal{GB}^{\alpha,r}_\tau}$ is not a (quasi-)seminorm since the triangle inequality, in general, is not satisfied. However, it is easy to show that for $\alpha_1 \le \alpha_2$ and $1/\tau_k = \alpha_k + 1/p$, $k = 1, 2$, we have $\mathcal{GB}^{\alpha_2,r}_{\tau_2} \subseteq \mathcal{GB}^{\alpha_1,r}_{\tau_1}$, so just as in the case of Besov spaces, a larger α implies a smaller class of functions with 'more smoothness'. Also, the smoothness measure $(\,\cdot\,)_{\mathcal{GB}^{\alpha,r}_\tau}$ of a function is bounded by the Besov (quasi-)seminorm of the function in $B^{d\alpha,r}_\tau$. Indeed, let $\mathcal{P}_D$ denote the BSP partition that gives the classical dyadic partition. If we denote the collection of dyadic cubes of side length 2^{-m} by $\mathcal{D}_m$, then

$$\begin{aligned}(f)_{\mathcal{GB}^{\alpha,r}_\tau} &\le \left(\sum_{\Omega\in\mathcal{P}_D}\left(|\Omega|^{-\alpha}\,\omega_r\left(f,\Omega\right)_\tau\right)^\tau\right)^{1/\tau}\\ &\le C\left(\sum_{m=0}^{\infty}\sum_{I\in\mathcal{D}_m}\left(2^{d\alpha m}\omega_r\left(f,I\right)_\tau\right)^\tau\right)^{1/\tau}\\ &\le C|f|_{B^{d\alpha,r}_\tau}.\end{aligned} \tag{3.6}$$

For a Geometric B-space, $\mathcal{GB}$ we introduce the (nonlinear) *K-functional* corresponding to the pair L_p and $\mathcal{GB}$

$$K(f,t) := K\left(f,t,L_p,\mathcal{GB}\right) := \inf_{g\in\mathcal{GB}}\left\{\|f-g\|_p + t\cdot(g)_{\mathcal{GB}}\right\}, \quad t>0. \tag{3.7}$$

The (nonlinear) *interpolation space* $(L_p,\mathcal{GB})_{\lambda,q}$, $\lambda>0$, $0<q\le\infty$, is defined as the set of all $f\in L_p\big([0,1]^d\big)$ such that

$$(f)_{(L_p,\mathcal{GB})_{\lambda,q}} := \begin{cases}\left(\sum\limits_{m=0}^{\infty}\left(2^{m\lambda}K\left(f,2^{-m}\right)\right)^q\right)^{1/q}, & 0<q<\infty,\\ \sup\limits_{m\ge0}2^{m\lambda}K\left(f,2^{-m}\right), & q=\infty,\end{cases}$$

is finite. Although the interpolation spaces $(L_p,\mathcal{GB})_{\lambda,q}$ are nonlinear, we can still apply the Jackson and Bernstein machinery that one usually applies in the case of linear spaces defined over fixed geometry, such as dyadic partitions [7] or fixed triangulations [13], [5]. We obtain the following characterization.

THEOREM 3.3. *Let $0<\gamma<\alpha$, $0<q\le\infty$, and $0<p<\infty$, then*

$$A^{\gamma,r}_{q,\tau}(L_p) = \left(L_p,\mathcal{GB}^{\alpha,r}_\tau\right)_{\frac{\gamma}{\alpha},q}, \tag{3.8}$$

where $1/\tau := \alpha + 1/p$.

The remainder of this section is devoted to the proof of Theorem 3.3.

In [5] we proved that for all bounded convex domains $\Omega\subset\mathbb{R}^d$ and functions $f\in L_\tau(\Omega)$, $0<\tau\le\infty$, we have the equivalence

$$E_{r-1}(f,\Omega)_\tau \approx \omega_r(f,\Omega)_\tau, \tag{3.9}$$

where the constants of equivalency depend only on d, r and τ.

To proceed with our analysis, we have to show that the polynomial approximations Q_Ω in (2.1), that are near best approximations in the p-norm are also near best approximations for some $0<\eta<p$. Indeed we show that

LEMMA 3.4. *Let $\Omega \subset \mathbb{R}^d$ be a bounded convex domain and let $f \in L_p(\Omega)$, $0 < p < \infty$. Then for any $r \in \mathrm{N}$, there exists a polynomial $Q \in \Pi_{r-1}$, such that for all $0 < \eta \le p$, if $0 < p \le 1$, and for all $1 \le \eta \le p$, if $1 < p < \infty$, we have*

$$\|f - Q\|_{L_\eta(\Omega)} \le C E_{r-1}(f, \Omega)_\eta, \tag{3.10}$$

where for $1 < p < \infty$, $C = C(r, d)$, and for $0 < p \le 1$, $C = C(r, d, \eta) \le C(r, d, \eta_0)$, $\eta_0 \le \eta \le p$.

Proof. We begin with the case $1 < p < \infty$. Given a convex domain $\Omega \subset \mathbb{R}^d$, in [4] we have constructed for any $g \in C^r(\Omega)$, a near best polynomial $\widetilde{Q} \in \Pi_{r-1}$ such that

$$\left\| g - \widetilde{Q} \right\|_{L_\eta(\Omega)} \le C(r, d) E_{r-1}(g, \Omega)_\eta, \quad 1 \le \eta < \infty. \tag{3.11}$$

Let $f \in L_p(\Omega)$, and let $\{g_n\}$ be a sequence in $C^r(\Omega)$, such that $\|f - g_n\|_p \to 0$, as $n \to \infty$. By Hölder's inequality, it follows that for all $1 \le \eta \le p$, $\|f - g_n\|_\eta \to 0$, as $n \to \infty$. Now let Q_n be the near best approximation to g_n guaranteed by (3.11). Then $\|g_n - Q_n\|_p \le C(r, d)\|g\|_p$, and since we may assume that $\|f - g_n\|_p \le \|f\|_p$, we obtain

$$\|Q_n\|_\infty \le C(r, d)|\Omega|^{-1/p}\|Q_n\|_p \le C(r, d)|\Omega|^{-1/p}\|f\|_p.$$

Hence, the set of polynomials Q_n is compact in $C(\Omega)$, and we may assume that $\{Q_n\}$ converges in the uniform norm to a polynomial Q. Now

$$\|f - Q\|_\eta \le \|f - g_n\|_\eta + \|g_n - Q_n\|_\eta + \|Q_n - Q\|_\eta, \quad 1 \le \eta \le p,$$

whence

$$\|f - Q\|_\eta \le \lim_{n \to \infty} C(r, d) E_{r-1}(g_n, \Omega)_\eta = C(r, d) E_{r-1}(f, \Omega)_\eta \quad 1 \le \eta \le p.$$

This proves (3.10) for $1 < p < \infty$.

For the case $0 < p \le 1$, we first make the following observation. Let A be a nonsingular affine mapping on $\mathbb{R}^d$, given by $A(x) := Mx + b$, where M is a nonsingular $d \times d$ matrix and let $f \in L_p(\Omega)$. Denote $\tilde{f} := f(A\cdot)$, $\tilde{Q} := Q(A\cdot)$, and $\widetilde{\Omega} := A^{-1}\Omega$. Then $\tilde{f} \in L_p(\widetilde{\Omega})$, and

$$\|f - Q\|_{L_\eta(\Omega)} = |\mathrm{det} M|^{1/\eta}\|\tilde{f} - \tilde{Q}\|_{L_\eta(\widetilde{\Omega})}, \quad 0 < \eta \le p. \tag{3.12}$$

Therefore

$$E_{r-1}(f, \Omega)_\eta = |\mathrm{det} M|^{1/\eta} E_{r-1}(\tilde{f}, \widetilde{\Omega})_\eta, \quad 0 < \eta \le p. \tag{3.13}$$

By John's Theorem (see [4], [5] and references therein), for any bounded convex domain $\Omega \subset \mathbb{R}^d$, there exists a nonsingular affine mapping A, such that

$$B(0, 1) \subseteq \widetilde{\Omega} \subseteq B(0, d), \tag{3.14}$$

where $B(x_0, R)$ denotes the ball of radius R, with center at x_0. Then we follow [1] (see also [9, Theorem 3.10.4]), and for $\tilde{f} \in L_p(\widetilde{\Omega})$ obtain $\tilde{Q} \in \Pi_{r-1}$, a so-called polynomial of best approximation in $L_1(\widetilde{\Omega})$, which satisfies

$$\|\tilde{f} - \tilde{Q}\|_{L_\eta(\widetilde{\Omega})} \le C(r, d, \eta) E_{r-1}(\tilde{f}, \widetilde{\Omega})_\eta, \quad \eta \le 1, \tag{3.15}$$

where $C(r,d,\eta) \le C(r,d,\eta_0)$, $\eta_0 < \eta \le p$. Now, (3.10) for $0 < p \le 1$, follows by virtue of (3.12) and (3.13).□

THEOREM 3.5. *For $0 < p < \infty$, $\alpha > 0$, $1/\tau = \alpha + 1/p$ and $f \in L_p\big([0,1]^d\big)$, we have the equivalence*

$$(f)_{\mathcal{GB}^{\alpha,r}_\tau} \approx \mathcal{N}_\tau\left(f, \mathcal{P}_\tau(f)\right), \tag{3.16}$$

with constants of equivalency depending only on α, d, r, p and ρ.

Proof. Let $\mathcal{P} \in \mathrm{BSP}(\rho)$ be a given partition. For $0 < \mu \le p$ and $\Omega \in \mathcal{P}$, denote by $Q_{\Omega,\mu}$ a near best polynomial approximation of $f \in L_\mu(\Omega)$. Note that with this notation, the near best polynomials used in (1.1) are $Q_\Omega = Q_{\Omega,p}$. We denote

$$\mathcal{N}_{\tau,\mu}(f,\mathcal{P}) := \left(\sum_{\Omega\in\mathcal{P}} \|\psi_{\Omega,\mu}\|_p^\tau\right)^{1/\tau},$$

where $\psi_{\Omega,\mu}$ are defined in (1.1) with the near best polynomials $Q_{\Omega,\mu}$, and

$$\widetilde{\mathcal{N}}_{\tau,\mu}(f,\mathcal{P}) := \left(\sum_{\Omega\in\mathcal{P}} \left(|\Omega|^{1/p-1/\mu}\omega_r(f,\Omega)_\mu\right)^\tau\right)^{1/\tau}.$$

By Lemma 3.4 we know that there is an $\tau < \eta < p$, such that for any $\Omega \in \mathcal{P}$, we may take $\psi_{\Omega,\eta} = \psi_{\Omega,p} = \psi_\Omega$. Therefore, in order to prove (3.16), it suffices to prove that for any $\mathcal{P} \in \mathrm{BSP}(\rho)$,

$$\mathcal{N}_{\tau,\eta}(f,\mathcal{P}) \approx \widetilde{\mathcal{N}}_{\tau,\tau}(f,\mathcal{P}), \tag{3.17}$$

holds with constants of equivalency that depend only on d, r, p, τ, η, and ρ.

To this end, take $\tau \le \mu \le \eta$, and recall that if Ω' is a child of Ω, then

$$\begin{aligned}\|\psi_{\Omega',\mu}\|_\mu &\le C\left(\|f - Q_{\Omega,\mu}\|_{L_\mu(\Omega')} + \|f - Q_{\Omega',\mu}\|_{L_\mu(\Omega')}\right)\\ &\le C\left(E_{r-1}(f,\Omega)_\mu + E_{r-1}(f,\Omega')_\mu\right),\end{aligned} \tag{3.18}$$

where $C = C(r,d,\mu)$. Hence

$$\begin{aligned}\mathcal{N}_{\tau,\mu}(f,\mathcal{P}) &= \left(\sum_{\Omega\in\mathcal{P}} \|\psi_{\Omega,\mu}\|_p^\tau\right)^{1/\tau}\\ &\le C\left(\sum_{\Omega\in\mathcal{P}} \left(|\Omega|^{1/p-1/\mu}\,\|\psi_{\Omega,\mu}\|_\mu\right)^\tau\right)^{1/\tau}\\ &\le C\left(\sum_{\Omega\in\mathcal{P}} \left(|\Omega|^{1/p-1/\mu}E_{r-1}(f,\Omega)_\mu\right)^\tau\right)^{1/\tau}\\ &\le C\left(\sum_{\Omega\in\mathcal{P}} \left(|\Omega|^{1/p-1/\mu}\omega_r(f,\Omega)_\mu\right)^\tau\right)^{1/\tau}\\ &= C\widetilde{\mathcal{N}}_{\tau,\mu}(f,\mathcal{P}),\end{aligned} \tag{3.19}$$

where for the first inequality we applied Lemma 2.4, for the second, (3.18) and (2.9), and finally for the third inequality, we applied (3.9).

Next we show that for $\tau \le \mu \le \eta$

$$\widetilde{\mathcal{N}}_{\tau,\eta}(f,\mathcal{P}) \le \mathcal{N}_{\tau,\mu}(f,\mathcal{P}). \tag{3.20}$$

We may assume that $\mathcal{N}_{\tau,\mu}(f,\mathcal{P}) < \infty$, for otherwise there is nothing to prove. Since $\mu < p$, we have that $f \in L_\mu\big([0,1]^d\big)$ and Theorem 2.1 implies

$$f = \sum_{\Omega\in\mathcal{P}} \psi_{\Omega,\mu}, \quad \text{a.e.}$$

Therefore,

$$\begin{aligned}
\omega_r(f,\Omega)_\eta^\tau &= \omega_r(f - \sum_{\widetilde{\Omega}\in\mathcal{P},\,\widetilde{\Omega}\supseteq\Omega} \psi_{\widetilde{\Omega},\mu},\Omega)_\eta^\tau \\
&\le C\Big\| \sum_{\widetilde{\Omega}\in\mathcal{P},\,\widetilde{\Omega}\subset\Omega} |\psi_{\widetilde{\Omega},\mu}|\Big\|_\eta^\tau \\
&\le C \sum_{\widetilde{\Omega}\in\mathcal{P},\,\widetilde{\Omega}\subset\Omega} \|\psi_{\widetilde{\Omega},\mu}\|_\eta^\tau \\
&\le C \sum_{\widetilde{\Omega}\in\mathcal{P},\,\widetilde{\Omega}\subset\Omega} \Big|\widetilde{\Omega}\Big|^{\tau(1/\eta-1/\tau)} \|\psi_{\widetilde{\Omega},\mu}\|_\tau^\tau,
\end{aligned}$$

where for the equality we used the fact that for $\Omega \subseteq \widetilde{\Omega}$, the geometric wavelet $\psi_{\widetilde{\Omega},\mu}$ is a polynomial of total degree $\le r-1$, for the second inequality we applied [13, Theorem 3.3], and for the third inequality we applied Lemma 2.4. Therefore,

$$\begin{aligned}
\widetilde{\mathcal{N}}_{\tau,\eta}(f,\mathcal{P})^\tau &\le C \sum_{\Omega\in\mathcal{P}} |\Omega|^{\tau(1/p-1/\eta)} \sum_{\widetilde{\Omega}\subset\Omega} \Big|\widetilde{\Omega}\Big|^{\tau(1/\eta-1/\tau)} \|\psi_{\widetilde{\Omega},\mu}\|_\tau^\tau \\
&= C \sum_{\Omega\in\mathcal{P}} \sum_{\widetilde{\Omega}\subset\Omega} \left(\frac{|\widetilde{\Omega}|}{|\Omega|}\right)^{\tau(1/\eta-1/p)} \Big(\Big|\widetilde{\Omega}\Big|^{1/p-1/\tau} \|\psi_{\widetilde{\Omega},\mu}\|_\tau\Big)^\tau \\
&= C \sum_{\widetilde{\Omega}\in\mathcal{P}} \Big(\Big|\widetilde{\Omega}\Big|^{1/p-1/\tau} \|\psi_{\widetilde{\Omega},\mu}\|_\tau\Big)^\tau \sum_{\substack{\Omega\in\mathcal{P}\\ \Omega\supset\widetilde{\Omega}}} \left(\frac{|\widetilde{\Omega}|}{|\Omega|}\right)^{\tau(1/\eta-1/p)}.
\end{aligned}$$

Now, if $\widetilde{\Omega} \in \mathcal{P}_m$ and $\Omega \in \mathcal{P}_{m-k}$, $k > 0$, is one of its ancestors, then by (2.9),

$$\Big|\widetilde{\Omega}\Big| \le |\Omega|\rho^k.$$

Hence

$$\sum_{\Omega\in\mathcal{P},\,\Omega\supset\widetilde{\Omega}} \left(\frac{|\widetilde{\Omega}|}{|\Omega|}\right)^{\tau(1/\eta-1/p)} \le C\sum_{k=1}^\infty \rho^{k\tau(1/\eta-1/p)} \le C(p,\eta,\tau,\rho).$$

We conclude that

$$\begin{aligned}
\widetilde{\mathcal{N}}_{\tau,\eta}(f,\mathcal{P})^\tau &\le C \sum_{\widetilde{\Omega}\in\mathcal{P}} \Big(\Big|\widetilde{\Omega}\Big|^{1/p-1/\tau} \|\psi_{\widetilde{\Omega},\mu}\|_\tau\Big)^\tau \\
&\le C \sum_{\widetilde{\Omega}\in\mathcal{P}} \|\psi_{\widetilde{\Omega},\mu}\|_p^\tau = C\mathcal{N}_{\tau,\mu}(f,\mathcal{P})^\tau,
\end{aligned}$$

where for the last inequality we again applied Lemma 2.4. This proves (3.20).

Now combining (3.19) with $\mu = \eta$, (3.20) with $\mu = \tau$ and then (3.19) with $\mu = \tau$, we obtain

$$\mathcal{N}_{\tau,\eta}(f,\mathcal{P}) \le C\widetilde{\mathcal{N}}_{\tau,\eta}(f,\mathcal{P}) \le C\mathcal{N}_{\tau,\tau}(f,\mathcal{P}) \le C\widetilde{\mathcal{N}}_{\tau,\tau}(f,\mathcal{P}),$$

which proves one direction in (3.17). In order to prove the opposite direction, we observe that it follows from Hölder's inequality that

$$\widetilde{\mathcal{N}}_{\tau,\tau}(f,\mathcal{P}) \le \widetilde{\mathcal{N}}_{\tau,\eta}(f,\mathcal{P}).$$

Using (3.20) with $\mu = \eta$ yields

$$\widetilde{\mathcal{N}}_{\tau,\eta}(f,\mathcal{P}) \le C\mathcal{N}_{\tau,\eta}(f,\mathcal{P}),$$

which gives

$$\widetilde{\mathcal{N}}_{\tau,\tau}(f,\mathcal{P}) \le \widetilde{\mathcal{N}}_{\tau,\eta}(f,\mathcal{P}) \le C\mathcal{N}_{\tau,\eta}(f,\mathcal{P}).$$

This completes the proof of the opposite direction in (3.17), and concludes our proof. □

In view of the above one may draw the following conclusion: There are cases of functions that are not in the Besov space of scale $d\alpha$ and therefore cannot be approximated by n-term wavelet approximation at the 'rate' $n^{-\alpha}$ (see [7]). Yet, there might exist an adaptive partition which captures the geometry (if it exists!) of the function's singularities and does lead to a finite smoothness measure (3.5) for the scale α. In fact we show that such a partition can also provide n-term geometric wavelet approximation at the "rate" $n^{-\alpha}$.

THEOREM 3.6 (Jackson estimate). *Let $0 < p < \infty$, $\alpha > 0$, and $r \in \mathrm{N}$. If $f \in \mathcal{GB}^{\alpha,r}_{\tau}$, $1/\tau = \alpha + 1/p$, then*

$$\sigma_{n,r,\tau}(f)_p \le Cn^{-\alpha}(f)_{\mathcal{GB}^{\alpha,r}_{\tau}}, \tag{3.21}$$

where $C := C(\alpha, d, r, p, \rho)$.

Proof. Given f, p and τ we select the near best adaptive partition $\mathcal{P}_\tau(f)$. Applying [13, Theorem 3.4] with the collection $\{\Phi_m\} := \{\psi_\Omega\}_{\Omega\in\mathcal{P}_\tau(f)}$, and then (3.16) we obtain

$$\begin{aligned}\sigma_{n,r,\tau}(f)_p &\le Cn^{-\alpha}\mathcal{N}_\tau\big(f,\mathcal{P}_\tau(f)\big)\\ &\le Cn^{-\alpha}(f)_{\mathcal{GB}^{\alpha,r}_{\tau}}.\end{aligned}$$

□

Let $\phi \in L_p\big([0,1]^d\big)$ and let $\mathcal{P} \in \mathrm{BSP}(\rho)$ be a *fixed* partition. Then, the smoothness of ϕ with respect to the fixed partition $\mathcal{P}$, is

$$|\phi|_{\mathcal{B}^{\alpha,r}_{\tau}(\mathcal{P})} := \left(\sum_{\Omega\in\mathcal{P}} \Big(|\Omega|^{-\alpha}\,\omega_r\left(\phi,\Omega\right)_\tau\Big)^\tau\right)^{1/\tau}.$$

For a fixed partition $\mathcal{P}$, the smoothness quantity $|\cdot|_{\mathcal{B}^{\alpha,r}_{\tau}(\mathcal{P})}$ is a quasi-seminorm. Therefore we obtain the Bernstein estimate for BSP partitions in much the same way that it was proved for triangulations in the bivariate case in [13], and in arbitrary dimension $d \ge 2$ in [5]. Namely,

THEOREM 3.7 (Bernstein estimate). *Let $\mathcal{P} \in \mathrm{BSP}(\rho)$, and let $\phi \in \Sigma_n^r(\mathcal{P})$. Then for all $0 < p < \infty$, $\alpha > 0$, and $1/\tau = \alpha + 1/p$,*

$$|\phi|_{\mathcal{B}_\tau^{\alpha,r}(\mathcal{P})} \le C n^\alpha \|\phi\|_p, \tag{3.22}$$

where $C := C(\alpha, d, r, p, \rho)$.

We are ready to prove Theorem 3.3

Proof of Theorem 3.3. The proof is similar to the proof of [9, Theorem 7.9.1]. The proof that the righthand side of (3.8) is contained in the lefthand side, readily follows by the Jackson inequality. Indeed, it is a standard technique to show that (3.21) implies that for every $f \in L_p$,

$$\sigma_{n,r,\tau}(f)_p \le CK\left(f, n^{-\alpha}, L_p, \mathcal{GB}_\tau^{\alpha,r}\right).$$

Hence by the first part of the proof of [9, Theorem 7.9.1]

$$(f)_{A_{q,\tau}^{\gamma,r}} \le C\left(\|f\|_p + (f)_{(L_p,\mathcal{GB}_\tau^{\alpha,r})_{\frac{\gamma}{\alpha},q}}\right).$$

In order to prove that the lefthand side of (3.8) is contained in the righthand side, we have to estimate the appropriate K-functional. Namely, we replace the proof of [9, Theorem 7.5.1(ii)] with the estimate

$$K\left(f, 2^{-m\alpha}, L_p, \mathcal{GB}_\tau^{\alpha,r}\right) \le C 2^{-m\alpha}\left(\sum_{j=1}^m \left(2^{j\alpha}\sigma_{2^{j-1}}(f)_p\right)^\mu + \|f\|_p^\mu\right)^{1/\mu}, \tag{3.23}$$

where $K(f, \cdot, L_p, \mathcal{GB}_\tau^{\alpha,r})$ is defined by (3.7), $\sigma_{2^j}(f)_p := \sigma_{2^j,r,\tau}(f)_p$, $m \ge 1$ and $\mu := \min(\tau, 1)$. Note that, in proving this, special attention is needed to circumvent the fact that $(\cdot)_{\mathcal{GB}_\tau^{\alpha,r}}$ is not a (quasi-)seminorm. Indeed, for each $j \ge 0$ we take a geometric wavelet sum $S_j \in \Sigma_{2^j}^r(\mathcal{P}_\tau(f))$ such that

$$\|f - S_j\|_{L_p\left([0,1]^d\right)} \le 2\sigma_{2^j}(f)_p.$$

Since $\mathcal{P}_\tau(f)$ is a fixed nested partition, we have that $\phi_j := S_j - S_{j-1} \in \Sigma_{2^{j+1}}^r(\mathcal{P}_\tau(f))$, $j \ge 1$, and

$$\|\phi_j\|_p \le \|f - S_j\|_p + \|f - S_{j-1}\|_p \le 2\sigma_{2^{j-1}}(f)_p, \quad j \ge 1.$$

We also put $\phi_0 := S_0$. Since S_0 is a single geometric wavelet component, we conclude that (3.9) implies that $\|\phi_0\|_p \le C\|f\|_p$. Now, we substitute $g := S_m = \sum_{j=0}^m \phi_j$, in (3.7)

and apply the Bernstein inequality (3.22) on the fixed partition $\mathcal{P}_\tau(f)$ to obtain

$$\begin{aligned}
K\left(f, 2^{-m\alpha}, L_p, \mathcal{GB}_\tau^{\alpha,r}\right) &\leq \|f - S_m\|_p + 2^{-m\alpha}\left(S_m\right)_{\mathcal{GB}_\tau^{\alpha,r}} \\
&\leq C\left(\sigma_{2^m}(f)_p + 2^{-m\alpha}\left|S_m\right|_{B_\tau^{\alpha,r}(\mathcal{P}_\tau(f))}\right) \\
&\leq C\left(\sigma_{2^m,r}(f)_p + 2^{-m\alpha}\left(\sum_{j=0}^{m}\left|\phi_j\right|^\mu_{B_\tau^{\alpha,r}(\mathcal{P}_\tau(f))}\right)^{1/\mu}\right) \\
&\leq C\left(\sigma_{2^m}(f)_p + 2^{-m\alpha}\left(\sum_{j=0}^{m}\left(2^{(j+1)\alpha}\left\|\phi_j\right\|_p\right)^\mu\right)^{1/\mu}\right) \\
&\leq C2^{-m\alpha}\left(\sum_{j=1}^{m}\left(2^{j\alpha}\sigma_{2^{j-1}}(f)_p\right)^\mu + \|f\|_p^\mu\right)^{1/\mu}.
\end{aligned}$$

We leave the rest of the proof to the reader. ∎

4. Simulation results and discussion. We implemented the geometric wavelet algorithm for the purpose of finding sparse representations of digital images with $r = 2$ (linear polynomials) and $p = 2$. We point out that, in our current implementation, condition (2.3) does not come into play.

To reduce the time complexity of the implementation, the images were subdivided to tiles of size 64×64 and a BSP tree was constructed over each of the tiles separately. Although JPEG-like artifacts, resulting from the tiles' boundaries, are visible in the examples below, this approach ensures that the time complexity of the algorithm is almost linear with respect to the image size. Once all the BSP trees were constructed over the 64×64 tiles, and the geometric wavelets were computed, we extracted a global n-term approximation (1.3) from the joint list of all the geometric wavelets over all the tiles. Our experiments show that in most cases increasing the tile size does not have a significant impact on the results.

To further improve the time complexity of the algorithm, we performed coarse partition searches at lower levels of the BSP tree and fine searches at the higher levels. The search for the optimal partition was done by advancing two points on a domain's boundary, computing the two subdomains created by the line that goes through these points and then, the two least-squares linear polynomials over each of these subdomains. In lower levels of the BSP tree, this march was done in larger steps and in finer levels, the step size was set to 1, the pixel resolution. In some sense, the idea of finer partitions at higher resolutions, is related to the way Curvelets [2] have 'more directions' at higher resolutions.

In Figure 3 we see an n-term geometric wavelet approximation of the known test image "peppers" of size 512×512 with 2048 elements. In Figure 4 we see an n-term dyadic wavelet approximation with twice as many elements, 4096. In all the examples below, we used a ratio of 1:2 ("peppers", "Lena"), 1:3 ("Barbara") or 1:4 ("Cameraman") between the number of geometric wavelets and dyadic wavelets, so as to make the comparison more relevant. Observe that on the more 'geometric' images, "Peppers" and "Cameraman", i.e., images that are roughly composed of smooth regions and strong distinct edges, the geometric wavelets seem to perform relatively better. For example, for the "Cameraman" image the 512-term geometric wavelet approximation gives the same PSNR as the 2048-term dyadic wavelet approximation.

For the dyadic wavelets approximation we used MatLab's wavelet toolbox, where we selected the well known biorthogonal wavelet basis $(4, 4)$ (see [3]), also known as the "nine-seven" in the engineering community. This biorthogonal wavelet has 4 zero moments, corresponding to $r = 4$. We note that we actually allowed the dyadic wavelet approximation to use even slightly more elements than claimed in the figures, so as to compensate for MatLab's handling of the image boundaries by a somewhat over-redundant wavelet decomposition. The results are summarized in the table below.

Image	**N-term dyadic**	**N-term geometric**	**Ratio**	**PSNR dyadic**	**PSNR geometric**
"Peppers"	4096	2048	2:1	29.22	31.32
"Lena"	4096	2048	2:1	30.18	31.26
"CameraMan"	2048	512	4:1	26.72	26.71
		1024			28.93
"Barbara"	12288	4096	3:1	27.54	27.10

TABLE 4.1
Comparison of n-term dyadic and geometric wavelets

In Figure 15 we see an example of image denoising using geometric wavelets. To compare with results in [22], we added Guassian white noise to the "Lena" test image with standard deviation of 20, which gives a noisy image with PSNR=22.14. Following the usual 'sparse representation' methodology [22], we applied the geometric wavelet algorithm to the noisy image from and extracted an n-term approximation (1.3), which serves as the approximation to the original image. We see that geometric features are recovered quite well in the process, in a manner which is very competitive with Curvelets. The algorithm produced a restored image with PSNR=29.76.

As with classical wavelets, the n-term strategy can be used for progressive coding and rate-distortion control where more geometric wavelets are added according to their order of appearance in (1.2). It is important to note that when trying to encode the approximation (1.3) it should be remembered that for a geometric wavelet located in a 'deep' level of the BSP tree, one needs to encode the sequence of binary partitions that created it. Thus, if the wavelet ψ_Ω is located at the mth level of the BSP partition, $O(m)$ bits are required to encode its location. Therefore, encoding geometric wavelets at higher levels is more expensive when considering bit allocation. However, this is no different from dyadic wavelet compression, where encoding the index of a dyadic wavelet located at the resolution m also requires $O(m)$ bits. Recall that at lower levels of the BSP tree we perform coarse partitions and at higher level, fine partitions. As pointed out in [17], this also improves the coding performance, since it facilitates the quantization and encoding of the partitions.

Although image coding using geometric wavelets is on-going work, we anticipate that the problem of encoding 'geometric side-information' can be solved by using zero-tree type encoding [18], [20], and rate-distortion optimization techniques [21], [23]. Furthermore, we plan to incorporate a geometric Rate-Distortion optimization technique borrowed from the wavelet coding algorithm WedgePrints [26]. Namely, at each node of the BSP tree one may allocate a flag (bit) to signal to the decoder a decision whether all further partitions of this domain are uniform (non-adaptive) or geometrically adaptive. Encoding geometric wavelets whose supports lie in a 'uniform'

ancestor domain, is similar to dyadic wavelet encoding, where only an index of the geometric wavelet in a uniform partition needs to be encoded and the support of the geometric wavelet is known from the uniform partition of the ancestor. Thus, using rate-distortion optimization techniques, one would choose at each node of the BSP whether to use an adaptive partition whose geometry needs to be encoded, or a uniform, non-adaptive partition.

REFERENCES

[1] L. Brown and B. Lucier, *Best approximations in L^1 are near best in L^p, $p < 1$*, Proc. Amer. Math. Soc. 120 (1994), 97-100.

[2] E. Candès and D. Donoho, *New Tight Frames of Curvelets and Optimal Representations of Objects with Smooth Singularities*, Technical Report, Stanford, 2002.

[3] I. Daubechies, *Ten lectures on wavelets*, CBMS-NSF Reg. Conf. Series in Applied Math., Vol. 61, SIAM, 1992.

[4] S. Dekel and D. Leviatan, *The Bramble-Hilbert lemma for convex domains*, SIAM J. Math. Anal. 35 (2004), 1203-1212.

[5] S. Dekel and D. Leviatan, *Whitney estimates for convex domains with applications to multivariate piecewise polynomial approximation*, Found. of Comp. Math. (to appear).

[6] S. Dekel, D. Leviatan and M. Sharir, *On bivariate smoothness spaces associated with nonlinear approximation*, Constr. Approx. (to appear).

[7] R. DeVore, *Nonlinear approximation*, Acta Numerica 7 (1998), 51-150.

[8] R. DeVore, B. Jawerth and B. Lucier, *Image compression through wavelet transform coding*, IEEE Trans. Inf. Theory 38 (1992), 719-746.

[9] R. DeVore and G. Lorentz, *Constructive Approximation*, Springer-Verlag, 1991.

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

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

[12] J. Hershberger and S. Suri, *Binary space partitions for 3D subdivision*, proceedings of the ACM-SIAM joint symposium on discrete algorithms, 2003, 100-108.

[13] B. Karaivanov and P. Petrushev, *Nonlinear piecewise polynomial approximation beyond Besov spaces*, Appl. Comput. Harmon. Anal., 15 (2003), No. 3, 177-223.

[14] B. Karaivanov, P. Petrushev and R. Sharpley, *Algorithms for nonlinear piecewise polynomial approximation: Theoretical aspects*, Trans., Amer. Math. Soc., 355 (2003), 2585-2631

[15] M. S. Paterson and F. F. Yao, *Efficient binary space partitions for hidden-surface removal and solid modeling*, Discrete Comput. Geom. 5 (1990), 485–503.

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

[17] H. Radha, M. Vetterli and R. Leonardi, *Image compression using binary space partitioning trees*, IEEE Trans. Image Proc. 5 (1996), 1610-1624.

[18] A. Said and W. Pearlman, *A new fast and efficient image codec based on set partitioning in hierarchical trees*, IEEE. Trans. Circuits and Systems for Video Technology 6 (1996), 243-250.

[19] P. Salembier and L. Garrido, *Binary partition tree as an efficient representation for image processing, segmentation, and information retrieval*, IEEE Trans. Image Proc. 9 (2000), 561-576.

[20] M. Shapiro, *An embedded hierarchical image coder using zerotrees of wavelet coefficients*, IEEE Trans. Signal Proc. 41 (1993), 3445-3462.

[21] R. Shukla, P. L. Dragotti, M. N. Do and M. Vetterli, *Rate-distortion optimized tree structured compression algorithms for piecewise smooth images*, IEEE Trans. Image Proc. (to appear).

[22] J. L. Starck, E. Candès and D. L. Donoho, *The Curvelet Transform for Image Denoising*, IEEE Trans. Image Proc. 11 (2000), 670-684.

[23] D. Taubman, *High performance scalable image compression with EBCOT*, IEEE Transactions on Image Processing 9 (2000), 1151-1170.

[24] C. Tòth, *A note on binary plane partitions*, in Proceedings of the 17th ACM Symposium on Computational Geometry, ACM, New York (2001), 151-156.

[25] C. Tòth, *Binary space partitions for line segments with a limited number of directions*, SIAM J. Comput. 32 (2003), 307-325.

[26] M. Wakin, J. Romberg, H. Choi, and R. Baraniuk, *Geometric methods for wavelet-based image compression*, SPIE Conference proceedings, Volume 5207, Wavelets: Applications in Signal and Image Processing X, M. Unser, A. Aldroubi, A. Laine, Editors (2003), 507-520.

FIG. 4.1. *The "peppers" image* 512×512.

FIG. 4.2. *Geometric wavelet approximation of the "peppers" image with* $n = 2048$*, PSNR=31.32.*

FIG. 4.3. *Dyadic biorthogonal wavelet approximation of the "peppers" image with* $n = 4096$, *PSNR=29.22.*

FIG. 4.4. *The "Lena" image* 512×512.

FIG. 4.5. *Geometric wavelet approximation of the "lena" image with* $n = 2048$*, PSNR=31.26.*

FIG. 4.6. *Dyadic biorthogonal wavelet approximation of the "lena" image with* $n = 4096$, *PSNR=30.18.*

FIG. 4.7. *The "cameraman" image* 256×256.

FIG. 4.8. *Geometric wavelet approximation of the "cameraman" image with* $n = 512$, *PSNR=26.71.*

FIG. 4.9. *Geometric wavelet approximation of the "cameraman" image with* $n = 1024$, *PSNR=28.93.*

FIG. 4.10. *Dyadic biorthogonal wavelet approximation of the "cameraman" image with* $n = 2048$, *PSNR=26.72.*

FIG. 4.11. *The "Barbara" image* 512×512.

FIG. 4.12. *Geometric wavelet approximation of the "Barbara" image with* $n = 4096$, *PSNR=27.10.*

FIG. 4.13. *Dyadic biorthogonal wavelet approximation of the "Barbara" image with* $n = 12288$, *PSNR=27.54.*

FIG. 4.14. *Geometric wavelet denoising. Noisy image PSNR=22.14, Restored image PSNR=29.76*