
Algorithms and Error Bounds for Multivariate Piecewise Constant Approximation

Oleg Davydov

Department of Mathematics and Statistics, University of Strathclyde,
26 Richmond Street, Glasgow G1 1XH, Scotland, UK, oleg.davydov@strath.ac.uk

Summary. We review the surprisingly rich theory of approximation of functions of many variables by piecewise constants. This covers for example the Sobolev-Poincaré inequalities, parts of the theory of nonlinear approximation, Haar wavelets and tree approximation, as well as recent results about approximation orders achievable on anisotropic partitions.

1 Introduction

Let Ω be a bounded domain in \mathbb{R}^d , $d \geq 2$. Suppose that Δ is a *partition* of Ω into a finite number of subsets $\omega \subset \Omega$ called *cells*, where the default assumptions are just these: $|\omega| := \text{meas}(\omega) > 0$ for all $\omega \in \Delta$, $|\omega \cap \omega'| = 0$ if $\omega \neq \omega'$, and $\sum_{\omega \in \Delta} |\omega| = |\Omega|$. For a finite set D we denote its cardinality by $|D|$, so that $|\Delta|$ stands for the number of cells ω in Δ . Given a function $f : \Omega \rightarrow \mathbb{R}$, we are interested in the error bounds for its approximation by piecewise constants in the space

$$S(\Delta) = \left\{ \sum_{\omega \in \Delta} c_\omega \chi_\omega : c_\omega \in \mathbb{R} \right\}, \quad \chi_\omega(x) := \begin{cases} 1, & \text{if } x \in \omega, \\ 0, & \text{otherwise.} \end{cases}$$

The best approximation error is measured in the L_p -norm $\|\cdot\|_p := \|\cdot\|_{L_p(\Omega)}$,

$$E(f, \Delta)_p := \inf_{s \in S(\Delta)} \|f - s\|_p, \quad 1 \leq p \leq \infty,$$

and various methods are known for the generation of the sequences of partitions Δ_N such that $E(f, \Delta_N)_p \rightarrow 0$ as $N \rightarrow \infty$ under certain smoothness assumptions on f , such as $f \in W_q^r(\Omega)$, where $W_q^r(\Omega)$ is the Sobolev space.

Note that the *simple functions* (measurable functions that take only finitely many values) used in the definition of Lebesgue integral are piecewise constants in the above sense. Given a function $f \in L_\infty(\Omega)$, we can generate a partition Δ_N as follows. Let $m, M \in \mathbb{R}$ be the essential infimum and essential supremum of f in Ω , respectively. Note that $\|f\|_\infty = \max\{-m, M\} \geq (M - m)/2$. Split the interval $[m, M]$ into N subintervals $I_k = [m + (k - 1)h, m + kh)$, $k = 1, \dots, N - 1$, $I_N = [m + (N - 1)h, M]$, $h = (M - m)/N$, and set

$$s_N = \sum_{k=1}^N c_k \chi_{\omega_k}, \quad \omega_k = f^{-1}(I_k), \quad c_k = m + (k - \frac{1}{2})h.$$

Then

$$\|f - s_N\|_\infty \leq \frac{M - m}{2N} \leq N^{-1} \|f\|_\infty.$$

If f is continuous on $\overline{\Omega}$ and $m = -M \neq 0$, then the above splitting of $[m, M]$ can be used to show that $E(f, \Delta)_\infty \geq N^{-1} \|f\|_\infty$ for any partition Δ with $|\Delta| \leq N$. Clearly, the above partition Δ_N is in general very complicated because the cells ω may be arbitrary measurable sets and so the above s_N cannot be stored using a finite number of real parameters.

Therefore piecewise constant approximation algorithms are practically useful only if the resulting approximation can be efficiently encoded. In the spirit of optimal recovery we will measure the complexity of an approximation algorithm by the maximum number of real parameters needed to store the piecewise constant function s it produces. If the algorithm produces an explicit partition Δ and defines s by $s = \sum_{\omega \in \Delta} c_\omega \chi_\omega$, then the constants c_ω give N such parameters, where $N = |\Delta|$. As in all ‘partition based’ algorithms discussed in this paper the partition Δ can be described using $\mathcal{O}(N)$ parameters, their overall complexity is $\mathcal{O}(N)$. The same is true for the ‘dictionary based’ algorithms such as Haar wavelet thresholding, with N being the number of basis functions that are active in an approximation.

In this paper we review a variety of algorithms for piecewise constant approximation for which the error bounds (‘Jackson estimates’) are available for functions in classical function spaces (Sobolev spaces or Besov spaces). We do not discuss ‘Bernstein estimates’ and the characterization of approximation spaces, and refer the interested reader to the original papers and the survey [14] that extensively covers this topic. However, we present a number of ‘saturation’ theorems that give a limit on the accuracy achievable by certain methods on general smooth functions. With only one exception (in the

beginning of Section 3) we do not discuss the approximation of functions of one variable, where we again refer to [14].

The paper is organized as follows. Section 2 is devoted to a simple linear approximation algorithm based on a uniform subdivision of the domain and local approximation by constants. In addition, we show that the approximation order $N^{-1/d}$ cannot be improved on isotropic partitions and give a review of the results on the approximation by constants (Sobolev-Poincaré inequalities) on general domains. Section 3 is devoted to the methods of nonlinear approximation restricted to our topic of piecewise constants. We discuss adaptive partition based methods such as Birman-Solomyak's algorithm and tree approximation, as well as dictionary based methods such as Haar wavelet thresholding and best n -term approximation. Finally, in Section 4 we present a simple algorithm with the approximation order $N^{-2/(d+1)}$ of piecewise constants on anisotropic polyhedral partitions, which cannot be further improved if the cells of a partition are required to be convex.

2 Linear approximation on isotropic partitions

Given $s = \sum_{\omega \in \Delta} c_\omega \chi_\omega$, we have

$$\|f - s\|_p = \begin{cases} \left(\sum_{\omega \in \Delta} \|f - c_\omega\|_{L_p(\omega)}^p \right)^{1/p} & \text{if } p < \infty, \\ \sup_{\omega \in \Delta} \|f - c_\omega\|_{L_\infty(\omega)} & \text{if } p = \infty. \end{cases} \quad (1)$$

Hence the best approximation on a fixed partition Δ is achieved when c_ω are the best approximating constants $c_\omega^*(f)$ such that

$$\|f - c_\omega^*(f)\|_{L_p(\omega)} = \inf_{c \in \mathbb{R}} \|f - c\|_{L_p(\omega)} =: E(f)_{L_p(\omega)}.$$

In the case $p = \infty$ obviously

$$c_\omega^*(f) = \frac{1}{2}(M_\omega f + m_\omega f), \quad E(f)_{L_\infty(\omega)} = \frac{1}{2}(M_\omega f - m_\omega f),$$

where

$$M_\omega f := \operatorname{ess\,sup}_{x \in \omega} f(x), \quad m_\omega f := \operatorname{ess\,inf}_{x \in \omega} f(x).$$

For any $1 \leq p \leq \infty$, it is easy to see that the average value of f on ω ,

$$f_\omega := |\omega|^{-1} \int_\omega f(x) dx,$$

satisfies $\|f_\omega - c\|_{L_p(\omega)} \leq \|f - c\|_{L_p(\omega)}$ for any constant c , in particular for $c = c_\omega^*$. Therefore

$$\|f - f_\omega\|_{L_p(\omega)} \leq 2E(f)_{L_p(\omega)},$$

and we conclude that the approximation

$$s_\Delta(f) := \sum_{\omega \in \Delta} f_\omega \chi_\omega \in S(\Delta) \quad (2)$$

is near best in the sense that

$$\|f - s_\Delta(f)\|_p \leq 2E(f, \Delta)_p, \quad 1 \leq p \leq \infty.$$

If $f|_\omega$ belongs to the Sobolev space $W_p^1(\omega)$, and the domain ω is sufficiently smooth then the error $\|f - f_\omega\|_{L_p(\omega)}$ may be estimated with the help of the Poincaré inequality

$$\|f - f_\omega\|_{L_p(\omega)} \leq C_\omega \text{diam}(\omega) |f|_{W_p^1(\omega)}, \quad f \in W_p^1(\omega), \quad (3)$$

where C_ω may still depend on ω in a scale-invariant way. For example, if ω is a Lipschitz domain, then C_ω can be found depending only on d, p and the Lipschitz constant of the boundary. At the end of this section we provide some more detail about the Poincaré inequality as well as the more general Sobolev-Poincaré inequalities available for various types of domains.

If the partition Δ is such that $C_\omega \leq C$, where C is independent of ω (but may depend for example on the Lipschitz constant of the boundary of Ω), then (1) and (3) imply

$$\|f - s_\Delta(f)\|_p \leq C \text{diam}(\Delta) |f|_{W_p^1(\Omega)}, \quad \text{diam}(\Delta) := \max_{\omega \in \Delta} \text{diam}(\omega).$$

This estimate suggests looking for partitions Δ that minimize $\text{diam}(\Delta)$ provided the number of cells $N = |\Delta|$ is fixed. Clearly, $\text{diam}(\Delta) \geq CN^{-1/d}$ for some constant C independent of N , and the order $N^{-1/d}$ is achieved if we for example choose a (hyper)cube Q containing Ω , split it uniformly into N^d equal subcubes Q_i , and define the cells of Δ by intersecting Ω with these subcubes, $\omega_i = \Omega \cap Q_i$, see Fig. 1. This gives a simple algorithm for piecewise constant approximation with approximation order $N^{-1/d}$ for all $f \in W_p^1(\Omega)$.

For the sake of simplicity we formulate this and all other algorithms only for the case when Ω is a cube $(0, 1)^d$.

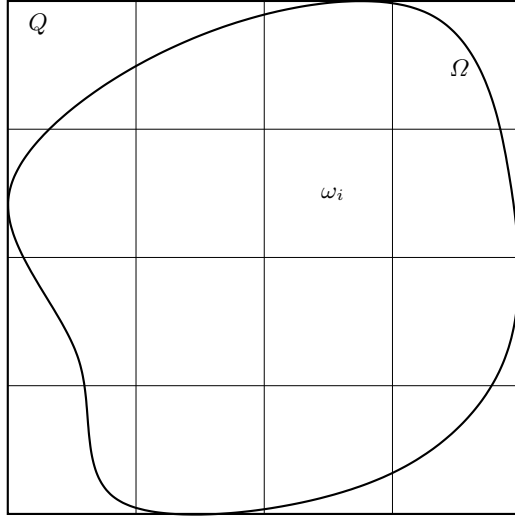


Fig. 1. Uniform partition.

Algorithm 1 Define Δ by splitting $\Omega = (0, 1)^d$ into $N = m^d$ cubes $\omega_1, \dots, \omega_N$ of edge length $h = 1/m$. Let $s_\Delta(f)$ be given by (2).

Theorem 1. The error of the piecewise constant approximation $s_\Delta(f)$ generated by Algorithm 1 satisfies

$$\|f - s_\Delta(f)\|_p \leq C(d, p)N^{-1/d} |f|_{W_p^1(\Omega)}, \quad f \in W_p^1(\Omega), \quad 1 \leq p \leq \infty. \quad (4)$$

The order $N^{-1/d}$ in (4) means that an approximation with error $\|f - s_\Delta(f)\|_p = \mathcal{O}(\varepsilon)$ is only achieved using $(\frac{1}{\varepsilon})^d$ degrees of freedom, which grows exponentially fast with the number of dimensions d . This phenomenon is often referred to as *curse of dimensionality*.

The approximation order $N^{-1/d}$ in (4) cannot be improved in general. See [14, Section 6.2] for a discussion of *saturation and inverse theorems*, where certain smoothness properties of f are deduced from appropriate assumptions about the order of its approximation by multivariate piecewise polynomials. For example, assuming that $E(f, \Delta)_\infty = o(\text{diam}(\Delta))$ as $\text{diam}(\Delta) \rightarrow 0$ for all partitions Δ , we can easily show that f is a constant function. Indeed, for any $x, y \in \Omega$ we can find a partition Δ such that x and y belong to the same cell ω , and $\text{diam}(\omega) = \text{diam}(\Delta) \leq 2\|x - y\|_2$. Then $|f(x) - f(y)| \leq |f(x) - f_\omega| + |f_\omega - f(y)| \leq 4E(f, \Delta)_\infty = o(\text{diam}(\omega))$. Hence $|f(x) - f(y)| = o(\|x - y\|_2)$ as $y \rightarrow x$, which implies that f has a zero differential at every $x \in \Omega$, that is f is a constant.

A saturation theorem in terms of the number of cells holds for any sequence of ‘isotropic’ partitions. We say that a sequence of partitions $\{\Delta_N\}$ is *isotropic* if there is a constant γ such that

$$\text{diam}(\omega) \leq \gamma\rho(\omega) \quad \text{for all } \omega \in \bigcup_N \Delta_N,$$

where $\rho(\omega)$ is the maximum diameter of d -dimensional balls contained in ω . Note that an isotropic partition may contain cells of very different sizes, see for example Fig. 2 below.

Theorem 2. *Assume that $f \in C^1(\Omega)$ and there is an isotropic sequence of partitions $\{\Delta_N\}$ with $\lim_{N \rightarrow \infty} \text{diam}(\Delta_N) = 0$ such that*

$$E(f, \Delta_N)_\infty = o(|\Delta_N|^{-1/d}), \quad N \rightarrow \infty.$$

Then f is a constant.

Proof. If f is not constant, then the gradient $\nabla f := [\partial f / \partial x_i]_{i=1}^d$ is nonzero at a point $\hat{x} \in \Omega$. Since the gradient of f is continuous, there is $\delta > 0$, a unit vector σ and a cube $Q \subset \Omega$ with edge length h containing \hat{x} such that $D_\sigma f(x) \geq \delta$ for all $x \in Q$, where $D_\sigma f = \nabla f^T \sigma$ denotes the directional derivative of f . The cube $\tilde{Q} := \{x \in Q : \text{dist}(x, \partial Q) > h/4\}$ has edge length $h/2$ and volume $(h/2)^d$. Assume that N is large enough to ensure that $\text{diam}(\Delta_N) < h/4$. Then any cell $\omega \in \Delta_N$ that has nonempty intersection with \tilde{Q} is contained in Q . If $[x_1, x_2]$ is an interval in ω parallel to σ , then $|f(x_2) - f(x_1)| \geq \delta \|x_2 - x_1\|_2$, which implies $|f(x_2) - f(x_1)| \geq \delta\rho(\omega)$ and hence

$$\varepsilon_N := E(f, \Delta_N)_\infty \geq E(f)_{L^\infty(\omega)} \geq \frac{\delta}{2}\rho(\omega) \geq \frac{\delta}{2\gamma} \text{diam}(\omega).$$

Therefore

$$\text{vol}(\omega) \leq \mu_d \left(\frac{\gamma}{\delta} \right)^d \varepsilon_N^d,$$

where μ_d denotes the volume of the d -dimensional ball of radius 1. Since \tilde{Q} is covered by such cells ω , we conclude that

$$\left(\frac{h}{2} \right)^d = \text{vol}(\tilde{Q}) \leq \sum_{\omega \cap \tilde{Q} \neq \emptyset} \text{vol}(\omega) \leq \mu_d \left(\frac{\gamma}{\delta} \right)^d \varepsilon_N^d |\Delta_N|,$$

which implies

$$E(f, \Delta_N)_\infty \geq \frac{h\delta}{2\gamma\mu_d^{1/d}} |\Delta_N|^{-1/d},$$

contrary to the assumption. \square

Sobolev-Poincaré inequalities

Sobolev-Poincaré inequalities provide bounds for the error of $f - f_\omega$. They hold on domains satisfying certain geometric conditions, for example the interior cone condition or the Lipschitz boundary condition. In some cases even a necessary and sufficient condition for ω to admit such an inequality is known. A domain $\omega \subset \mathbb{R}^d$ is called a *John domain* if there is a fixed point $x_0 \in \omega$ and a constant $c_J > 0$ such that every point $x \in \omega$ can be connected to x_0 by a curve $\gamma \subset \omega$ such that

$$\text{dist}(y, \partial\omega) \geq c_J \ell(\gamma(x, y)), \quad \text{for all } y \in \gamma,$$

where $\ell(\gamma(x, y))$ denotes the length of the segment of γ between x and y . Every domain with the interior cone condition is a John domain, but not otherwise. In particular, there are John domains with fractal boundary of Hausdorff dimension greater than $d - 1$.

The following variant of Sobolev inequality holds for all John domains $\omega \subset \mathbb{R}^d$, see [18] and references therein,

$$\|f - f_\omega\|_{L_{q^*}(\omega)} \leq C(d, q, \lambda) \|\nabla f\|_{L_q(\omega)}, \quad f \in W_q^1(\omega), \quad 1 \leq q < d, \quad (5)$$

where $q^* = dq/(d - q)$ is the Sobolev conjugate of q , and λ is the John constant of ω . Note that $\|\nabla f\|_{L_q(\omega)}$ denotes the L_q -norm of the euclidean norm of ∇f , that is $\|\nabla f\|_{L_q(\omega)}^q = \int_\omega \left(\sum_{i=1}^d |\partial f / \partial x_i|^2 \right)^{q/2} dx$, which is equivalent to the more standard seminorm of the Sobolev space $W_q^1(\omega)$ given by $|f|_{W_q^1(\omega)}^q = \sum_{i=1}^d \int_\omega |\partial f / \partial x_i|^q dx$. We prefer using $\|\nabla f\|_{L_q(\omega)}$ because of the explicit expressions for $C(d, q, \lambda)$ available in certain cases, see the end of this section.

According to [7], if the Sobolev inequality (5) holds for some $1 \leq q < d$ and certain mild separation condition (valid for example for any simply connected domain in \mathbb{R}^2) is satisfied, then ω is a John domain.

Assuming $1 \leq p < \infty$, let $\tau = \max\{\frac{d}{1+d/p}, 1\}$. Then $\tau^* \geq p$ and $1 \leq \tau < d$. If $|\omega| < \infty$, then Hölder inequality and (5) imply for any $q \geq \tau$

$$\begin{aligned}
\|f - f_\omega\|_{L_p(\omega)} &\leq |\omega|^{\frac{1}{p} - \frac{1}{\tau^*}} \|f - f_\omega\|_{L_{\tau^*}(\omega)} \\
&\leq C(d, \tau, \lambda) |\omega|^{\frac{1}{p} - \frac{1}{\tau^*}} \|\nabla f\|_{L_\tau(\omega)} \\
&\leq C(d, \tau, \lambda) |\omega|^{\frac{1}{d} + \frac{1}{p} - \frac{1}{q}} \|\nabla f\|_{L_q(\omega)},
\end{aligned}$$

and we arrive at the following Sobolev-Poincaré inequality for all p, q such that $1 \leq p < \infty$ and $\tau \leq q \leq \infty$,

$$\|f - f_\omega\|_{L_p(\omega)} \leq C(d, p, \lambda) |\omega|^{\frac{1}{d} + \frac{1}{p} - \frac{1}{q}} \|\nabla f\|_{L_q(\omega)}, \quad f \in W_q^1(\omega). \quad (6)$$

In particular, since $\tau \leq p$, we can choose $q = p$, which leads to the Poincaré inequality for bounded John domains for all $1 \leq p < \infty$ in the form

$$\|f - f_\omega\|_{L_p(\omega)} \leq C \operatorname{diam}(\omega) \|\nabla f\|_{L_p(\omega)}, \quad f \in W_p^1(\omega), \quad (7)$$

where C depends only on d, p, λ .

Poincaré inequality in the case $p = \infty$ has been considered in [25]. If $\omega \subset \mathbb{R}^d$ is a bounded path-connected domain, then

$$E(f)_{L_\infty(\omega)} \leq r(\omega) \|\nabla f\|_{L_\infty(\omega)}, \quad \text{with } r(\omega) := \inf_{x \in \omega} \sup_{y \in \omega} \rho_\omega(x, y),$$

where $\rho_\omega(x, y)$ is the geodesic distance, i.e. the infimum of the lengths of the paths in ω from x to y .

If ω is star-shaped with respect to a point, then $r(\omega) \leq \operatorname{diam}(\omega)$, and so (7) holds with $C = 2$ for all such domains if $p = \infty$. Moreover, as observed in [25], the arguments of [2, 11] can be applied to show that for any bounded star-shaped domain

$$\|f - f_\omega\|_{L_p(\omega)} \leq \frac{2}{1-d/p} \operatorname{diam}(\omega) \|\nabla f\|_{L_p(\omega)}, \quad f \in W_p^1(\omega), \quad d < p \leq \infty. \quad (8)$$

In particular, (8) applies to star-shaped domains with cusps that fail to be John domains.

If ω is a bounded convex domain in \mathbb{R}^d , then (7) holds for all $1 \leq p \leq \infty$ with a constant C depending only on d [12]. Moreover, optimal constants are known for $p = 1, 2$: $C = 1/\pi$ for $p = 2$ [22, 3] and $C = 1/2$ for $p = 1$ [1]. Since $r(\omega) = \frac{1}{2} \operatorname{diam}(\omega)$, it follows that (7) holds with $C = 1$ if $p = \infty$.

Note that similar estimates are available for the approximation by polynomials of any degree, where in the case $p = q$ the corresponding result is usually referred to as the Bramble-Hilbert lemma, see [6, Chapter 4]. Moreover, instead of Sobolev spaces, the smoothness of f can be measured in some other function spaces (e.g. Besov spaces), or with the help of a modulus of smoothness (Whitney estimates), see [14, 13].

3 Nonlinear approximation

We have seen in Section 2 that the approximation order $N^{-1/d}$ is the best achievable on isotropic partitions. Nevertheless, by using more sophisticated algorithms the estimate (4) can be improved in the sense that the norm $|f|_{W_p^1(\Omega)}$ in its right hand side is replaced by a weaker norm, for example $|f|_{W_q^1(\Omega)}$ with $q < p$. This improvement is often quite significant because the norm $|f|_{W_q^1(\Omega)}$ is finite for functions with more substantial singularities than those allowed in the space $W_p^1(\Omega)$, see [14] for a discussion.

Recall that in Algorithm 1 the partition Δ is independent of the target function f , and so $s_\Delta(f)$ depends linearly on f . A simple example of a nonlinear algorithm is given by Kahane's approximation method for continuous functions of bounded total variation on an interval, see [14, Section 3.2]. To define a partition of the interval (a, b) , the points $a = t_0 < t_1 < \dots < t_N = b$ are chosen such that $\text{var}_{(t_{i-1}, t_i)}(f) = \frac{1}{N} \text{var}_{(a,b)}(f)$, $i = 1, \dots, N-1$. By setting $\omega_i = (t_{i-1}, t_i)$, $c_i = (M_{\omega_i} f + m_{\omega_i} f)/2$, we see that the piecewise constant function $s = \sum_{i=1}^{N-1} c_i \chi_{\omega_i}$ satisfies $\|f - s\|_\infty \leq \frac{1}{2N} \text{var}_{(a,b)}(f)$. Thus, for the partition $\Delta = \{\omega_i\}_{i=1}^{N-1}$,

$$E(f, \Delta)_\infty \leq \frac{1}{2N} \text{var}_{(a,b)}(f) = \frac{1}{2N} |f|_{BV(a,b)} \leq \frac{1}{2N} |f|_{W_1^1(a,b)},$$

where the last inequality presumes that f belongs to $W_1^1(a, b)$, that is it is absolutely continuous and its derivative is absolutely integrable.

In the multivariate case the first algorithm of this type was given in [5]. It is based on *dyadic partitions* Δ of $\Omega = (0, 1)^d$ that consist of the *dyadic cubes* of the form

$$2^{-jd}(k_1, k_1 + 1) \times \dots \times (k_d, k_d + 1), \quad j = 0, 1, 2, \dots, \quad 0 \leq k_i < 2^{jd},$$

produced adaptively by successive *dyadic subdivisions* of a cube into 2^d equal subcubes with halved edge length, see Fig. 2.

The following lemma plays a crucial role in [5].

Lemma 1. *Let $\Phi(\omega)$ be a nonnegative function of sets $\omega \subset \Omega$ which is sub-additive in the sense that $\Phi(\omega') + \Phi(\omega'') \leq \Phi(\omega' \cup \omega'')$ as soon as ω', ω'' are disjoint subdomains of Ω . Given $\alpha > 0$, we set*

$$g_\alpha(\omega) := |\omega|^\alpha \Phi(\omega), \quad \omega \subset \Omega,$$

and, for any partition Δ of Ω ,

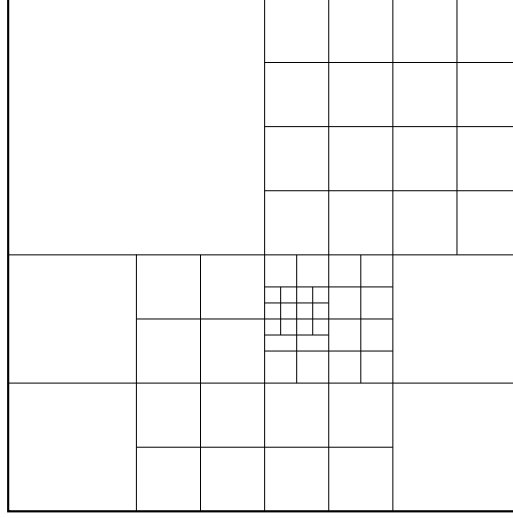


Fig. 2. Example of a diadic partition.

$$G_\alpha(\Delta) := \max_{\omega \in \Delta} g_\alpha(\omega).$$

Assume that a sequence of partitions $\{\Delta_k\}_{k=0}^\infty$ of $\Omega = (0, 1)^d$ into diadic cubes is obtained recursively as follows. Set $\Delta_0 = \{\Omega\}$. Obtain Δ_{k+1} from Δ_k by the diadic subdivision of those cubes $\omega \in \Delta_k$ for which

$$g_\alpha(\omega) \geq 2^{-d\alpha} G_\alpha(\Delta_k).$$

Then

$$G_\alpha(\Delta_k) \leq C(d, \alpha) |\Delta_k|^{-(\alpha+1)} \Phi(\Omega), \quad k = 0, 1, \dots$$

This lemma can be used with $\Phi(\omega) = |f|_{W_q^1(\omega)}^q$, $1 \leq q < \infty$, which is obviously subadditive, giving rise to the following algorithm which we only formulate for piecewise constants even though the results in [5] also apply to the higher order piecewise polynomials.

Algorithm 2 ([5]) Suppose we are interested in the approximation in L_p norm, $1 < p \leq \infty$. Choose $1 \leq q < \infty$ such that $q > \tau := \frac{d}{1+d/p}$ ($\tau = d$ if $p = \infty$), and assume that $f \in W_q^1(\Omega)$, $\Omega = (0, 1)^d$. Set $\Delta_0 = \{\Omega\}$. While $|\Delta_k| < N$, obtain Δ_{k+1} from Δ_k by the diadic subdivision of those cubes $\omega \in \Delta_k$ for which

$$g_\alpha(\omega) \geq 2^{-d\alpha} \max_{\omega \in \Delta} g_\alpha(\omega),$$

where

$$g_\alpha(\omega) := |\omega|^\alpha |f|_{W_q^1(\omega)}^q, \quad \alpha = \frac{q}{\tau} - 1.$$

Since $|\Delta_k| < |\Delta_{k+1}|$, the subdivisions terminate at some $\Delta = \Delta_m$ with $|\Delta| \geq N$ and $|\Delta| = \mathcal{O}(N)$. The resulting piecewise constant approximation $s_\Delta(f)$ of f is given by (2).

Theorem 3 ([5]). *The error of the piecewise constant approximation $s_\Delta(f)$ generated by Algorithm 2 satisfies*

$$\|f - s_\Delta(f)\|_p \leq C(d, p, q) N^{-1/d} |f|_{W_q^1(\Omega)}, \quad f \in W_q^1(\Omega). \quad (9)$$

Proof. We only consider the case $p < \infty$. For any $\omega \in \Delta$ it follows by the Sobolev-Poincaré inequality (6) for cubes that

$$\|f - f_\omega\|_{L_p(\omega)}^p \leq C_1 |\omega|^{\frac{p}{d} + 1 - \frac{p}{q}} |f|_{W_q^1(\omega)}^p = C_1 g_\alpha^{p/q}(\omega) \leq C_1 G_\alpha^{p/q}(\Delta),$$

where C_1 depends only on d, p, q . Hence

$$\|f - s_\Delta(f)\|_p^p = \sum_{\omega \in \Delta} \|f - f_\omega\|_{L_p(\omega)}^p \leq C_1 |\Delta| G_\alpha^{p/q}(\Delta).$$

Now Lemma 1 implies

$$\|f - s_\Delta(f)\|_p \leq C_2 |\Delta|^{1/p} |\Delta|^{-(\alpha+1)/q} \Phi^{1/q}(\Omega) = C_2 |\Delta|^{-1/d} |f|_{W_q^1(\Omega)}. \quad \square$$

If $q \geq p$, then the estimate (9) is also valid for the much simpler Algorithm 1. Therefore the scope of Algorithm 2 is when $f \in W_q^1(\Omega)$ for some q satisfying $\tau < q < p$ but $f \notin W_p^1(\Omega)$ or if $|f|_{W_q^1(\Omega)}$ is significantly smaller than $|f|_{W_p^1(\Omega)}$. Note that the computation of $g_\alpha(\omega)$ in Algorithm 2 requires first order partial derivatives of f . Algorithm 2 is nonlinear (in contrast to Algorithm 1) because the partition Δ depends on the target function f .

An adaptive algorithm based on the local approximation errors rather than local Sobolev norm of f was studied in [17]. We again restrict to the piecewise constant case.

Algorithm 3 *Assume $f \in L_p(\Omega)$, $\Omega = (0, 1)^d$, for some $0 < p \leq \infty$ and choose $\varepsilon > 0$. Set $\Delta_0 = \{\Omega\}$. For $k = 0, 1, \dots$, obtain Δ_{k+1} from Δ_k by the dyadic subdivision of those cubes $\omega \in \Delta_k$ for which*

$$\|f - f_\omega\|_{L_p(\omega)} > \varepsilon.$$

Since $\|f - f_\omega\|_{L_p(\omega)} \rightarrow 0$ as $|\omega| \rightarrow 0$, the subdivisions terminate at some $\Delta = \Delta_m$. The resulting piecewise constant approximation $s_\Delta(f)$ of f is given by (2).

Now in contrast to [5], $0 < p < 1$ is also allowed. The error bounds are obtained for functions in Besov spaces rather than Sobolev spaces. Recall that f belongs to the Besov space $B_{q,\sigma}^\alpha(\Omega)$, $\alpha > 0$, $0 < q, \sigma \leq \infty$, if

$$|f|_{B_{q,\sigma}^\alpha(\Omega)} = \begin{cases} \left(\int_0^\infty (t^{-\alpha} \omega_r(f, t)_q)^\sigma \frac{dt}{t} \right)^{1/\sigma} & \text{if } 0 < \sigma < \infty, \\ \sup_{t>0} t^{-\alpha} \omega_r(f, t)_q & \text{if } \sigma = \infty. \end{cases}$$

is finite, where $r = [\alpha] + 1$ is the smallest integer greater than α , and $\omega_r(f, t)_q$ denotes the r -th modulus of smoothness of f in L_q . In particular, $B_{q,\infty}^\alpha(\Omega) = \text{Lip}(\alpha, L_q(\Omega))$ for $0 < \alpha < 1$.

Theorem 4 ([17]). *Let $0 < \alpha < 1$, $q > \frac{d}{\alpha+d/p}$ and $0 < \sigma \leq \infty$. If $f \in B_{q,\sigma}^\alpha(\Omega)$, then for any N there is an $\varepsilon > 0$ such that the partition Δ produced by Algorithm 3 satisfies $|\Delta| \leq N$ and*

$$\|f - s_\Delta(f)\|_p \leq C(d, p, q) N^{-\alpha/d} |f|_{B_{q,\sigma}^\alpha(\Omega)}.$$

The set of all dyadic cubes is a tree \mathcal{T}^{dc} , where the children of a cube ω are the cubes $\omega_1, \dots, \omega_{2^d}$ obtained by its dyadic subdivision. The only root of \mathcal{T}^{dc} is $\Omega = (0,1)^d$. Clearly, Algorithms 2 and 3 produce a *complete subtree* \mathcal{T} of \mathcal{T}^{dc} in the sense that for any node in \mathcal{T} its parent and all siblings are also in \mathcal{T} . The corresponding partition Δ consists of all leaves of \mathcal{T} . If we set $e(\omega) = \|f - f_\omega\|_{L_p(\omega)}^p$, then $E(\mathcal{T}) := \sum_{\omega \in \Delta} e(\omega) = \|f - s_\Delta(f)\|_p^p$. It is easy to see that $|\Delta| = 1 + (2^d - 1)n(\mathcal{T})$, where $n(\mathcal{T})$ denotes the number of subdivisions used to create \mathcal{T} . The quantity $n(\mathcal{T})$ measures the complexity of a tree, and $E_n := \inf_{n(\mathcal{T}) \leq n} E(\mathcal{T})$ gives the optimal error achievable by a tree of a given complexity. It is natural to look for optimal or near optimal trees. The concept of *tree approximation* was introduced in [8] in the context of n -term wavelet approximation. General results applicable in particular to the piecewise constant approximations on dyadic partitions are given in [4]. The idea is that replacing $\|f - f_\omega\|_{L_p(\omega)} > \varepsilon$ in Algorithm 3 by a more sophisticated refinement criterion leads to an algorithm that produces a near optimal tree.

Algorithm 4 ([4]) *Assume $f \in L_p(\Omega)$, $\Omega = (0,1)^d$, for some $1 \leq p < \infty$ and choose $\varepsilon > 0$. Generate a sequence of complete subtrees \mathcal{T}_k of \mathcal{T}^{dc} as follows. Set $\mathcal{T}_0 = \{\Omega\}$ and $\alpha(\Omega) = 0$. For $k = 0, 1, \dots$, obtain \mathcal{T}_{k+1} from \mathcal{T}_k by the dyadic subdivision of those leaves ω of \mathcal{T}_k for which*

$$e(\omega) := \|f - f_\omega\|_{L_p(\omega)}^p > \varepsilon + \alpha(\omega),$$

and define $\alpha(\omega_i)$ for all children $\omega_1, \dots, \omega_{2^d}$ of ω by

$$\alpha(\omega_i) = \frac{e(\omega_i)}{\sigma(\omega)} \left[\alpha(\omega) + (\varepsilon - e(\omega) - \sigma(\omega))_+ \right], \quad \sigma(\omega) := \sum_j e(\omega_j),$$

assuming that $\sigma(\omega) \neq 0$. The algorithm terminates at some tree $\mathcal{T} = \mathcal{T}_m$ since $\|f - f_\omega\|_{L_p(\omega)} \rightarrow 0$ as $|\omega| \rightarrow 0$. The resulting piecewise constant approximation $s_\Delta(f)$ of f is given by (2), where Δ is the diadic partition defined by the leaves of \mathcal{T} .

Theorem 5 ([4]). *The tree \mathcal{T} produced by Algorithm 4 is near optimal as it satisfies*

$$E(\mathcal{T}) \leq 2(2^d + 1)E_{[n/2]}, \quad n = n(\mathcal{T}).$$

Further results on tree approximation are reviewed in [15].

The above algorithms generate piecewise constant approximations by constructing an appropriate partition of Ω . A different approach is to look for an approximation as linear combination of a fixed set of piecewise constant ‘basis functions’, for example characteristic functions of certain subsets of Ω .

More general, let $\mathcal{D} \subset L_p(\Omega)$ be a set of functions, called *dictionary*, such that the finite linear combinations of the elements in \mathcal{D} are dense in $L_p(\Omega)$. Note that the set \mathcal{D} does not have to be linearly independent. Given $f \in L_p(\Omega)$, the error of the *best n -term approximation* is defined by

$$\sigma_n(f, \mathcal{D})_p = \inf_{s \in \Sigma_n} \|f - s\|_p,$$

where $\Sigma_n = \Sigma_n(\mathcal{D})$ is the set of all linear combinations of at most n elements of \mathcal{D} ,

$$\Sigma_n(\mathcal{D}) := \left\{ \sum_{g \in D} c_g g : D \subset \mathcal{D}, |D| \leq n, c_g \in \mathbb{R} \right\}.$$

If the functions in \mathcal{D} are piecewise constants, then the approximants in $\Sigma_n(\mathcal{D})$ are piecewise constants as well. If each element $g \in \mathcal{D}$ can be described using a bounded number of parameters, then $s = \sum_{g \in D} c_g g \in \Sigma_n(\mathcal{D})$ requires $\mathcal{O}(n)$ parameters even though the number of cells in the partition Δ such that $s \in S(\Delta)$ may in general grow superlinearly (even exponentially) with n .

Piecewise constant approximation $s_\Delta(f)$ produced by Algorithm 2 or 3 belongs to $\Sigma_n(\mathcal{D}^c)$, with $n = |\Delta|$, where the dictionary \mathcal{D}^c consists of the characteristic functions χ_ω of all dyadic cubes $\omega \subseteq (0, 1)^d$. Therefore Theorems 3 and 4 imply that

$$\sigma_n(f, \mathcal{D}^c)_p \leq \begin{cases} C(d, p, q)n^{-\alpha/d}|f|_{W_q^\alpha(\Omega)} & \text{if } f \in W_q^\alpha(\Omega), \alpha = 1, \\ C(d, p, q)n^{-\alpha/d}|f|_{B_{q,\sigma}^\alpha(\Omega)} & \text{if } f \in B_{q,\sigma}^\alpha(\Omega), 0 < \alpha < 1. \end{cases}$$

as soon as $q > \frac{d}{\alpha+d/p}$.

Clearly, $\Sigma_n(\mathcal{D}^c)$ includes many piecewise constants with more than n dyadic cells, for example, $\chi_{(0,1)^2} - \chi_{(0,1/2^m)^2} \in \Sigma_2(\mathcal{D}^c)$ is piecewise constant with respect to a partition of $(0, 1)^2$ into $3m+1$ dyadic squares. A larger dictionary \mathcal{D}^r consisting of the characteristic functions of all ‘dyadic rings’ (differences between pairs of embedded dyadic cubes) has been considered in [9, 20]. An adaptive algorithm proposed in [9] produces a piecewise constant approximation $s_\Delta(f)$ of any function f of bounded variation, where Δ is a partition of $(0, 1)^2$ into N dyadic rings, such that

$$\|f - s_\Delta(f)\|_2 \leq 18\sqrt{3}N^{-1/2}|f|_{BV((0,1)^2)},$$

where $|f|_{BV(\Omega)}$ is the variation of f over Ω . Recall that the space $BV(\Omega)$ coincides with $\text{Lip}(1, L_1(\Omega))$ and contains the Besov space $B_{1,1}^1(\Omega)$. In [20] this result is generalized to certain spaces of bounded variation with respect to dyadic rings, and to the Besov spaces, leading in particular to the estimate

$$\sigma_n(f, \mathcal{D}^r)_p \leq C(d, p)n^{-\alpha/d}|f|_{B_{\tau,\tau}^\alpha(\Omega)}, \quad f \in B_{\tau,\tau}^\alpha(\Omega), \quad 0 < \alpha < 1,$$

where $\tau = \frac{d}{\alpha+d/p}$. Recall for comparison that $q > \tau$ in Theorems 3 and 4.

An important ‘piecewise constant’ dictionary \mathcal{D}^h is given by the multivariate Haar wavelets. Let $\psi^0 = \chi_{(0,1)}$ and $\psi^1 = \chi_{(0,1/2)} - \chi_{(1/2,1)}$. For any $e = (e_1, \dots, e_d) \in V^d$, where V^d is the set consisting of the nonzero vertices of the cube $(0, 1)^d$, let

$$\psi^e(x_1, \dots, x_d) := \psi^{e_1}(x_1) \cdots \psi^{e_d}(x_d).$$

Then the set

$$\Psi^h = \{\psi_{j,k}^e := 2^{jd/2}\psi^e(2^j(\cdot - k)) : j \in \mathbb{Z}, k \in \mathbb{Z}^d, e \in V^d\}$$

of *Haar wavelets* $\psi_{j,k}^e$ is an orthonormal basis for $L_2(\mathbb{R}^d)$. Therefore every $f \in L_2(\mathbb{R}^d)$ has an L_2 -convergent Haar wavelet expansion

$$f = \sum_{j,k,e} \langle f, \psi_{j,k}^e \rangle \psi_{j,k}^e, \quad \langle f, \psi_{j,k}^e \rangle := \int_{\mathbb{R}^d} f(x) \psi_{j,k}^e(x) dx.$$

If $f \in L_2(\Omega)$, $\Omega = (0, 1)^d$, then $f - f_\Omega$ has zero mean on $(0, 1)^d$ and hence by extending it to \mathbb{R}^d by zero and taking a Haar wavelet decomposition, we obtain an L_2 -convergent series

$$f = f_\Omega + \sum_{(j,k,e) \in \Lambda_\Omega} f_{j,k,e} \psi_{j,k}^e, \quad (10)$$

where Λ_Ω denotes the set of indices (j, k, e) such that $\text{supp } \psi_{j,k}^e \subseteq \Omega$, and $f_{j,k,e}$ are the *Haar wavelet coefficients* of f ,

$$f_{j,k,e} = \int_{(0,1)^d} f(x) \psi_{j,k}^e(x) dx.$$

Clearly, the Haar wavelet coefficients are well defined for any function $f \in L_1(\Omega)$. The series (10) converges unconditionally in L_p -norm if $f \in L_p(\Omega)$, $1 < p < \infty$. This implies in particular that every subset of $\{\|f_{j,k,e} \psi_{j,k}^e\|_p : (j, k, e) \in \Lambda_\Omega\}$ has a largest element. The dictionary of Haar wavelets on $\Omega = (0, 1)^d$ is given by

$$\mathcal{D}^h = \{\psi_{j,k}^e : (j, k, e) \in \Lambda_\Omega\}.$$

A standard approximation method for this dictionary is *thresholding*, also called *greedy approximation*.

Algorithm 5 (Haar wavelet thresholding) *Assume $f \in L_p(\Omega)$, $\Omega = (0, 1)^d$, for some $1 < p < \infty$. Let $\int_\Omega f(x) dx = 0$. (Otherwise, replace f by $f - f_\Omega$.) Given $n \in \mathbb{N}$, choose n largest elements in the sequence $\{\|f_{j,k,e} \psi_{j,k}^e\|_p : (j, k, e) \in \Lambda_\Omega\}$ and denote the set of their indices by Λ_Ω^n . The resulting approximation of f is given by*

$$G_n(f) = \sum_{(j,k,e) \in \Lambda_\Omega^n} f_{j,k,e} \psi_{j,k}^e.$$

If $p = 2$, then clearly $G_n(f)$ is the best n -term approximation of f with respect to the dictionary \mathcal{D}^h . The following theorem gives an error bound in this case.

Theorem 6 ([9]). *Let $f \in BV(\Omega)$, $\Omega = (0, 1)^2$, and $\int_{\Omega} f(x) dx = 0$. Then the approximation $G_n(f)$ produced by Algorithm 5 satisfies*

$$\|f - G_n(f)\|_2 \leq C n^{-1/2} |f|_{BV(\Omega)},$$

where $C = 36(480\sqrt{5} + 168\sqrt{3})$.

It turns out that $G_n(f)$ is also near best for any $1 < p < \infty$.

Theorem 7 ([24]). *Let $f \in L_p(\Omega)$, $\Omega = (0, 1)^d$, for some $1 < p < \infty$, and $\int_{\Omega} f(x) dx = 0$. The approximation $G_n(f)$ produced by Algorithm 5 satisfies*

$$\|f - G_n(f)\|_p \leq C(d, p) \sigma_n(f, \mathcal{D}^h)_p.$$

An estimate for $\sigma_n(f, \mathcal{D}^h)_p$ follows from the results of [16] by using the extension theorems for functions in Besov spaces, see [14, Section 7.7].

Theorem 8 ([16]). *Let $1 < p < \infty$, $0 < \alpha < 1/p$, $\tau = \frac{d}{\alpha+d/p}$. If $f \in B_{\tau, \tau}^{\alpha}(\Omega)$, $\Omega = (0, 1)^d$, and $\int_{\Omega} f(x) dx = 0$, then*

$$\sigma_n(f, \mathcal{D}^h)_p \leq C(d, p) n^{-\alpha/d} |f|_{B_{\tau, \tau}^{\alpha}(\Omega)}.$$

The best n -term approximation by piecewise constants (and by piecewise polynomials of any degree) on hierarchical partitions of \mathbb{R}^d or $(0, 1)^d$ into anisotropic diadic boxes of the form

$$\left(\frac{k_1}{2^{j_1 d}}, \frac{k_1 + 1}{2^{j_1 d}} \right) \times \cdots \times \left(\frac{k_d}{2^{j_d d}}, \frac{k_d + 1}{2^{j_d d}} \right), \quad j_s, k_s \in \mathbb{Z},$$

has been studied in [23]. Here, the smoothness of the target function is expressed in terms of certain Besov-type spaces defined with respect to a given hierarchical partition. In [21], results of the same type are obtained for even more flexible anisotropic hierarchical triangulations. Let $\mathcal{T} = \cup_{m \in \mathbb{Z}} \Delta_m$, where each Δ_m is a locally finite triangulation of \mathbb{R}^2 such that Δ_{m+1} is obtained from Δ_m by splitting each triangle $\omega \in \Delta_m$ into at least two and at most M subtriangles (children). The hierarchical triangulation \mathcal{T} is called *weak locally regular* if there are constants $0 < r < \rho < 1$ ($r \leq \frac{1}{4}$), such that for any $\omega \in \mathcal{T}$ it holds $r|\omega'| \leq |\omega| \leq \rho|\omega'|$, where $\omega' \in \mathcal{T}$ is the parent triangle of ω . Clearly, the triangles in \mathcal{T} may have arbitrarily small angles. The *skinny B-space* $B_q^{\alpha, k}(\mathcal{T})$, $0 < q < \infty$, $\alpha > 0$, $k \in \mathbb{N}$, is the set of all $f \in L_q^{\text{loc}}(\mathbb{R}^2)$ such that

$$|f|_{B_q^{\alpha,k}(\mathcal{T})} := \left(\sum_{\omega \in \mathcal{T}} |\omega|^{-\alpha q} w_k(f, \omega)_q^q \right)^{1/q},$$

where

$$w_k(f, \omega)_q := \sup_{h \in \mathbb{R}^2} \|\delta_h^k(f)\|_{L_q(\omega)},$$

$\delta_h^k(f)$ being the k -th finite difference of f , in particular

$$\delta_h^1(f, x) := \begin{cases} f(x+h) - f(x), & \text{if } [x, x+h] \subset \omega, \\ 0, & \text{otherwise.} \end{cases}$$

It is shown in [21] that if \mathcal{T} is *regular*, i.e. there is a positive lower bound for the minimum angles of all triangles in \mathcal{T} , then $B_q^{\alpha,k}(\mathcal{T}) = B_{q,q}^{2\alpha}(\mathbb{R}^2)$ with equivalent norms whenever $0 < 2\alpha < \min\{1/q, k\}$.

Consider the dictionary $\mathcal{D}^{\mathcal{T}} = \{\chi_\omega : \omega \in \mathcal{T}\}$.

Theorem 9 ([21]). *Let $0 < p < \infty$, $\alpha > 0$, $\tau = \frac{2}{\alpha+2/p}$. If $f \in B_{\tau}^{\frac{\alpha}{2},1}(\mathcal{T}) \cap L_p(\mathbb{R}^2)$, then*

$$\sigma_n(f, \mathcal{D}^{\mathcal{T}})_p \leq C(p, \alpha, \rho, r) n^{-\alpha/2} |f|_{B_{\tau}^{\frac{\alpha}{2},1}(\mathcal{T})}.$$

Note that certain Haar type bases can be introduced on the anisotropic dyadic partitions and on hierarchical triangulations obtained by a special refinement rule, see [23, 21] for their definition and approximation properties. An extension of Theorem 9 to \mathbb{R}^d with $d > 2$ is given in [13].

4 Anisotropic partitions

We have seen in Theorem 2 that piecewise constants on isotropic partitions cannot approximate nontrivial smooth functions with order better than $N^{-1/d}$. We now turn to the question what approximation order can be achieved on anisotropic partitions. An argument similar to that in the proof of Theorem 2 shows that it is not better than $N^{-2/(d+1)}$ if we assume that the partition is convex, i.e. all its cells are convex sets.

Theorem 10 ([10]). *Assume that $f \in C^2(\Omega)$ and the Hessian of f is positive definite at a point $\hat{x} \in \Omega$. Then there is a constant C depending only on f and d such that for any convex partition Δ of Ω ,*

$$E(f, \Delta)_\infty \geq C |\Delta|^{-2/(d+1)}.$$

The order of piecewise constant approximation on anisotropic partitions in two dimensions has been investigated in [19]. It is shown that for any $f \in C^2([0, 1]^2)$ there is a sequence of partitions Δ_N of $(0, 1)^2$ into polygons with the cell boundaries consisting of totally $\mathcal{O}(N)$ straight line segments, such that $\|f - s_{\Delta_N}(f)\|_\infty = \mathcal{O}(N^{-2/3})$. Moreover, the approximation order $N^{-2/3}$ cannot be improved on such partitions. Note that by triangulating each polygonal cell of Δ_N one obtains a convex partition with $\mathcal{O}(N)$ triangular cells, so that Theorem 10 also applies, giving the same saturation order $N^{-2/3}$. Another result of [19] is that for any $f \in C^3([0, 1]^2)$ there is a sequence of partitions Δ_N of $(0, 1)^2$ into cells with piecewise parabolic boundaries defined by a total of $\mathcal{O}(N)$ parabolic segments (pieces of graphs of univariate quadratic polynomials) such that $\|f - s_{\Delta_N}(f)\|_\infty = \mathcal{O}(N^{-3/4})$.

The following algorithm achieves the approximation order $N^{-2/(d+1)}$ on convex polyhedral partitions with totally $\mathcal{O}(N)$ facets.

Algorithm 6 ([10]) *Assume $f \in W_1^1(\Omega)$, $\Omega = (0, 1)^d$. Split Ω into $N_1 = m^d$ cubes $\omega_1, \dots, \omega_{N_1}$ of edge length $h = 1/m$. Then split each ω_i into N_2 slices ω_{ij} , $j = 1, \dots, N_2$, by equidistant hyperplanes orthogonal to the average gradient $g_i := |\omega_i|^{-1} \int_{\omega_i} \nabla f$ on ω_i . Set $\Delta = \{\omega_{ij} : i = 1, \dots, N_1, j = 1, \dots, N_2\}$, and define the piecewise constant approximation $s_\Delta(f)$ by (2). Clearly, $|\Delta| = N_1 N_2$ and each ω_{ij} is a convex polyhedron with at most $2(d+1)$ facets.*

This algorithm is illustrated in Fig. 3.

Theorem 11 ([10]). *Assume that $f \in W_p^2(\Omega)$, $\Omega = (0, 1)^d$, for some $1 \leq p \leq \infty$. For any $m = 1, 2, \dots$, generate the partition Δ_m by using Algorithm 6 with $N_1 = m^d$ and $N_2 = m$. Then*

$$\|f - s_{\Delta_m}(f)\|_p \leq C(d, p) N^{-2/(d+1)} (|f|_{W_p^1(\Omega)} + |f|_{W_p^2(\Omega)}), \quad (11)$$

where $N = |\Delta_m| = m^{d+1}$.

Proof. For simplicity we assume $d = 2$ and $p = \infty$. (The general case is treated in [10].) Let us estimate the error of the best approximation of f by constants on ω_{ij} ,

$$E(f)_{L_\infty(\omega_{ij})} = \frac{1}{2} \left(\max_{x \in \omega_{ij}} f(x) - \min_{x \in \omega_{ij}} f(x) \right).$$

Let σ_i be a unit vector orthogonal to g_i . Since ∇f is continuous, there is $\tilde{x} \in \omega_i$ such that $g_i = \nabla f(\tilde{x})$. Then $D_{\sigma_i} f(\tilde{x}) = 0$ and hence $\|D_{\sigma_i} f\|_{L_\infty(\omega_i)} \leq$

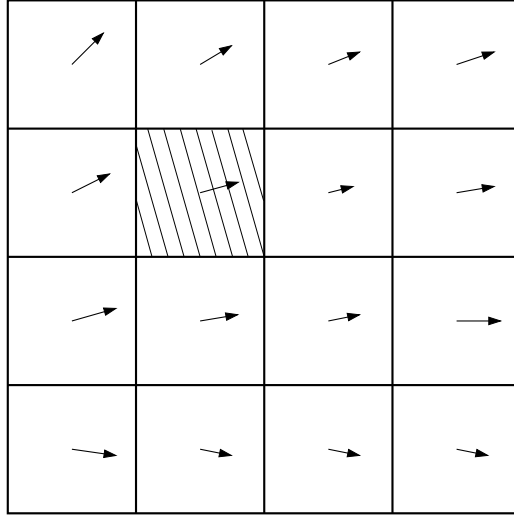


Fig. 3. Algorithm 6 ($d = 2, m = 4$). The arrows stand for the average gradients g_i on the cubes ω_i . The cells ω_{ij} are shown only for one cube.

$c_1 h |f|_{W_\infty^2(\omega_i)}$. Given $x, y \in \omega_{ij}$, choose a point $x' \in \omega_{ij}$ such that $y - x'$ and $x' - x$ are collinear with g_i and σ_i , respectively. (This is always possible if we swap the roles of x and y when necessary, see Fig. 4.)

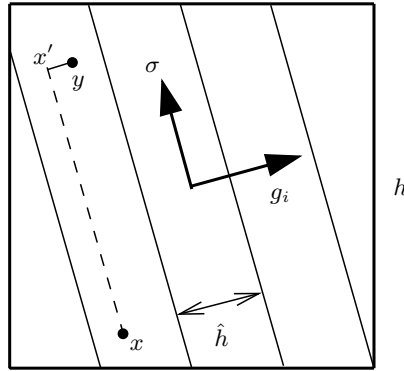


Fig. 4. Illustration of the proof of Theorem 11, showing a single ω_i .

Hence, denoting by \hat{h} the distance between the hyperplanes that split ω_i , we obtain

$$\begin{aligned}
|f(y) - f(x)| &\leq |f(y) - f(x')| + |f(x') - f(x)| \\
&\leq \hat{h} \|\nabla f\|_{L_\infty(\omega_{ij})} + c_2 h \|D_\sigma f\|_{L_\infty(\omega_{ij})} \\
&\leq c_3 (\hat{h} |f|_{W_\infty^1(\omega_{ij})} + h^2 |f|_{W_\infty^2(\omega_i)}) \\
&\leq c_4 m^{-2} (|f|_{W_\infty^1(\omega_{ij})} + |f|_{W_\infty^2(\omega_i)}).
\end{aligned}$$

Thus,

$$\|f - f_{\omega_{ij}}\|_{L_\infty(\omega_{ij})} \leq 2E(f)_{L_\infty(\omega_{ij})} \leq c_4 m^{-2} (|f|_{W_\infty^1(\Omega)} + |f|_{W_\infty^2(\Omega)}),$$

and (11) follows. \square

The improvement of the approximation order by piecewise constants from $N^{-1/d}$ on isotropic partitions to $N^{-2/(d+1)}$ on convex partitions does not extend to higher degree piecewise polynomials. Given a partition Δ , let $E_1(f, \Delta)_p$ denote the best error of (discontinuous) piecewise linear approximation in L_p -norm. Then the approximation order on isotropic partitions is $N^{-2/d}$ for sufficiently smooth functions, and it cannot be improved in general on any convex partitions.

Theorem 12 ([10]). *Assume that $f \in C^2(\Omega)$ and the Hessian of f is positive definite at a point $\hat{x} \in \Omega$. Then there is a constant C depending only on f and d such that for any convex partition Δ of Ω ,*

$$E_1(f, \Delta)_\infty \geq C |\Delta|^{-2/d}.$$

References

1. G. Acosta and R. G. Durán: An optimal Poincaré inequality in L^1 for convex domains, Proc. Amer. Math. Soc. **132**, 2004, 195–202.
2. R. Arcangeli and J. L. Gout: Sur l'évaluation de l'erreur d'interpolation de Lagrange dans un ouvert de \mathbb{R}^n , R.A.I.R.O. Analyse Numerique **10**, 1976, 5–27.
3. M. Bebendorf: A note on the Poincaré inequality for convex domains, J. Anal. Appl. **22**, 2003, 751–756.
4. P. Binev and R. DeVore: Fast computation in adaptive tree approximation, Numer. Math. **97**, 2004, 193–217.
5. M. S. Birman and M. Z. Solomyak: Piecewise polynomial approximation of functions of the classes W_p^α , Mat. Sb. **73 (115)**, no. 3, 1967, 331–355 (in Russian). English translation in Math. USSR-Sb. **2**, no. 3, 1967, 295–317.
6. S. Brenner and L. R. Scott: *The Mathematical Theory of Finite Element Methods*. Springer-Verlag, Berlin, 1994.
7. S. Buckley and P. Koskela: Sobolev-Poincaré implies John, Math. Res. Lett. **2**, 1995, 577–594.

8. A. Cohen, W. Dahmen, I. Daubechies, and R. DeVore: Tree approximation and optimal encoding, *Appl. Comp. Harm. Anal.* **11**, 2001, 192–226.
9. A. Cohen, R. DeVore, P. Petrushev, and H. Xu: Nonlinear approximation and the space $BV(\mathbb{R}^2)$, *Amer. J. Math.* **121**, 1999, 587–628.
10. O. Davydov, Approximation by piecewise constants on convex partitions, in preparation.
11. L.T. Dechevski and E. Quak: On the Bramble-Hilbert lemma, *Numer. Funct. Anal. Optim.* **11**, 1990, 485–495.
12. S. Dekel and D. Leviatan: The Bramble-Hilbert lemma for convex domains, *SIAM J. Math. Anal.* **35**, 2004, 1203–1212.
13. S. Dekel and D. Leviatan: Whitney estimates for convex domains with applications to multivariate piecewise polynomial approximation, *Found. Comput. Math.* **4**, 2004, 345–368.
14. R. A. DeVore: Nonlinear approximation, *Acta Numerica* **7**, 1998, 51–150.
15. R. A. DeVore: Nonlinear approximation and its applications, in *R. DeVore and A. Kunoth (eds.): Multiscale, Nonlinear and Adaptive Approximation - Dedicated to Wolfgang Dahmen on the Occasion of his 60th Birthday*. Springer-Verlag, Berlin, 2009, 169–201.
16. R. DeVore, B. Jawerth and V. Popov: Compression of wavelet decompositions, *Amer. J. Math.* **114**, 1992, 737–785.
17. R. A. DeVore and X. M. Yu: Degree of adaptive approximation, *Math. Comp.* **55**, 1990, 625–635.
18. P. Hajlasz: Sobolev inequalities, truncation method, and John domains, in *Papers on Analysis: A volume dedicated to Olli Martio on the occasion of his 60th birthday*. Report. Univ. Jyväskylä **83**, 2001, 109–126.
19. A. S. Kochurov: Approximation by piecewise constant functions on the square, *East J. Approx.* **1**, 1995, 463–478.
20. Y. Hu, K. Kopotun and X. Yu: On multivariate adaptive approximation, *Constr. Approx.* **16**, 2000, 449–474.
21. B. Karaivanov, P. Petrushev, Nonlinear piecewise polynomial approximation beyond Besov spaces, *Appl. Comput. Harmon. Anal.* **15** (2003), No. 3, 177–22
22. L. E. Payne and H. F. Weinberger: An optimal Poincaré inequality for convex domains, *Arch. Rational Mech. Anal.* **5**, 1960, 286–292.
23. P. Petrushev: Multivariate n -term rational and piecewise polynomial approximation, *J. Approx. Theory* **121**, 2003, 158–197.
24. V. Temlyakov: The best m -term approximation and greedy algorithms, *Adv. Comput. Math.* **8**, 1998, 249–265.
25. S. Waldron: Minimally supported error representations and approximation by the constants, *Numer. Math.* **85**, 2000, 469–484.