JOURNAL OF APPROXIMATION THEORY 25, 337–359 (1979)

An Adaptive Algorithm for Multivariate Approximation Giving Optimal Convergence Rates*

CARL DE BOOR

Mathematics Research Center, University of Wisconsin, Madison, Wisconsin 53706

AND

JOHN R. RICE

Division of Mathematical Sciences, Purdue University, W. Lafayette, Indiana 47907

Communicated by O. Shisha

Received October 10, 1977

1. INTRODUCTION

We consider approximation from the class

$$\mathbb{P}_{n,K}$$

of functions on (some domain in) $\mathbb{R}^N$ which consist of no more than K polynomial pieces each of order n, i.e., of total degree less than n. Approximation from $\mathbb{P}_{n,K}$ for $N > 1$ was studied as early as 1967 by Birman and Solomiak [2]. Improvements of their results were obtained by Brudnyi [3], and much of the work is presented in his survey article [4]. Birman and Solomiak make use of a kind of adaptive partition algorithm involving m-cubes in some of their work, but their results do not contain ours nor ours theirs. Their results go beyond the statement that $\operatorname{dist}(f, \mathbb{P}_{n,K}) = O(K^{-n})$ for smooth functions. But, their description of certain function classes for which such an optimal rate of approximation is achievable is in terms of certain moduli of smoothness. This description is difficult to apply to a specific function not in $C^{(n)}$. By contrast, our analysis includes explicitly functions of the form

$$f(x) = g(x) \operatorname{dist}(x, S)^{\alpha},$$

* Sponsored by the United States Army under Contract No. DAAG29-75-C-0024.

0021-9045/79/040337–23$02.00/0

where $g \in C^{(n)}$ and S is a smooth manifold. We show that the optimal convergence rate on $D \subseteq \mathbb{R}^N$ can be achieved if $\alpha > mn/N - (N - m)/p$, where $m := \dim(S)$ and $\mathbb{L}_p$-approximation is used. We also show by an example that this restriction on α is in general necessary for optimal convergence rates. These examples help to establish the boundary of the class of functions which saturate piecewise polynomial approximation. Our analysis is not restricted to piecewise polynomial approximation and includes, for example, approximation by blending function methods.

We now describe the subject matter of our paper in some detail.

We are interested in gauging the efficiency of an adaptive algorithm for the approximation of functions, or of functionals on functions, on some domain D in $\mathbb{R}^N$. The algorithm produces a subdivision of D into K nonoverlapping cells $C_1, \ldots, C_K$ and, on each such cell C_i, an appropriate approximation.

The ingredients for the algorithm are:

a1. A collection $\mathbb{C}$ of *allowable* cells.

a2. A (nonnegative) function $E: \mathbb{C} \to \mathbb{R}$, with $E(C)$ giving the *error* (bound) for the approximation on the cell C.

a3. An initial subdivision of the domain D into allowable cells.

a4. A *division algorithm* for subdividing any allowable cell C into two or more allowable cells. C is called the *parent* for these latter cells.

The adaptive algorithm consists in producing, for each current subdivision, a new subdivision by dividing some cell in the current subdivision à la a4, until $E(C) \leqslant \epsilon$ for all cells in the current subdivision, with ϵ some prescribed positive number.

Existing adaptive algorithms for quadrature or for piecewise polynomial approximation (in one variable) are considerably more sophisticated than this simple algorithm. Yet our simple algorithm allows us to analyze quite satisfactorily the *efficiency* of the approximations produced by these more complex algorithms, i.e., the relationship between the prescribed tolerance ϵ and the number $K = K(\epsilon)$ of cells in the final subdivision.

Here, we visualize the work of constructing the appropriate approximation on an allowable cell to be the same for all cells so that the work for constructing the final approximation is proportional to the number K of cells in the final subdivision. This is still true even if we count all the intermediate approximations constructed as well, since the total number of cells considered cannot be bigger than $2K$. On the other hand, ϵ, or at times $K\epsilon$, or some other function of ϵ, measures the accuracy achieved by the final approximation.

For the analysis of the relationship between K and ϵ, we make the following assumptions regarding $\mathbb{C}$ and E.

c1. $\mathbb{C}$ consists of bounded, closed, convex sets.

c2. Cells are not too different from balls: Associated with each cell $\mathbb{C} \in C$ are two closed balls, b_C and B_C, for which $b_C \subseteq C \subseteq B_C$, and

$$\eta := \inf_{C \in \mathbb{C}} \operatorname{diam}(b_C)/\operatorname{diam}(B_C) > 0.$$

c3. Invariance under scaling: If $C \in \mathbb{C}$ and c is the center of its inscribed ball b_C, then, for all positive ρ,

$$C_\rho := c + \rho(C - c) \in \mathbb{C}.$$

d1. Parent and children have comparable size: For some positive β and all $C \in \mathbb{C}$, each child C' of C produced by the division algorithm a4 satisfies

$$|C'|/|C| \geqslant \beta.$$

Here, and below, $|B|$ denotes the N-dimensional volume of B.

e1. Monotonicity: $C \subseteq C'$ implies $E(C) \leqslant E(C')$.

The basic tool in our analysis is the number

$$|E|_\epsilon := \int_D dx/\theta(x, \epsilon) \tag{5}$$

with

$$\theta(x, \epsilon) := \inf\{|C| : x \in C \in \mathbb{C}, E(C) > \epsilon\}.$$

We will show that

$$K(\epsilon) \leqslant |E|_\epsilon/\beta \tag{6}$$

and, in this way, obtain quite explicit bounds on K for specific choices of E. The following lemma gives a hint as to why (6) might hold.

LEMMA 1. *If $(C_i)_1^K$ is a subdivision for D with $E(C_i) > \epsilon$ for all i, then $K \leqslant |E|_\epsilon$.*

"*Proof*". We have

$$K = \sum_{i=1}^{K} |C_i|/|C_i| = \sum_{i=1}^{K} \int_{C_i} dx/|C_i| \leqslant \sum_{i=1}^{K} \int_{C_i} dx/\theta(x, \epsilon) = |E|_\epsilon$$

since $\theta(x, \epsilon) \leqslant |C_i|$ for all $x \in C_i$. ∎

Of course, this argument fails to establish that $1/\theta(\cdot, \epsilon)$ is even integrable. But, as we said, it gives a hint as to why (6) might be true.

In the next section, we derive various basic properties of the function θ and the number $|E|_\epsilon$ and prove (6).

2. The Function θ

We begin our discussion of the function θ by establishing some properties of the allowable cells.

Lemma 2. *For all $C \in \mathbb{C}$ and all y, the allowable cell $C_\rho := c + \rho(C - c)$ contains y for all $\rho \geqslant 1 + \mathrm{dist}(y, C)/\mathrm{radius}(b_C)$. Here, c is the center of the associated inscribed ball b_C.*

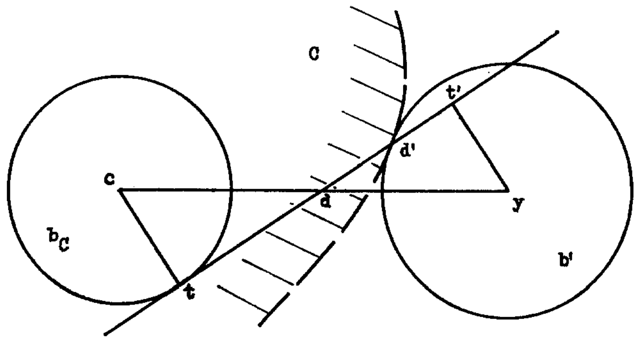

Fig. 1. Determination of ρ for which $y \in c + \rho(C - c)$.

Proof. Any such cell is allowable by c3, so that we only have to prove that $y \in C_\rho$ for the specified values of ρ. This is obvious for $\mathrm{dist}(y, C) = 0$. Let $\mathrm{dist}(y, C) > 0$ and let b' be the ball around y of radius $\mathrm{dist}(y, C)$, and let d' be a point common to C and b'. In a plane containing d', y and c, let l be the straight line through d' which intersects the segment $[c, y]$, at the point d, say, and is tangent to b_C, at the point t, say. Then $d \in C$ by convexity, hence y is contained in the cell

$$c + \frac{|y - c|}{|d - c|}(C - c).$$

But then, $r := \mathrm{dist}(y, l) \leqslant \mathrm{dist}(y, C)$, and, with t' the point of l closest to y, the two triangles (c, d, t) and (y, d, t') are similar. Therefore

$$\frac{|y - d|}{|d - c|} = \frac{|y - t'|}{|c - t|} = \frac{r}{\mathrm{radius}(b_C)} \leqslant \frac{\mathrm{dist}(y, C)}{\mathrm{radius}(b_C)}$$

and the lemma now follows since $|y - c|/|d - c| = 1 + |y - d|/|d - c|$ (see Fig. 1). ∎

Corollary. *If $\theta(x, \epsilon) < \infty$ for some x, then θ is bounded on bounded sets.*

Proof. By assumption, there exists $C \in \mathbb{C}$ with $x \in C$ and $E(C) > \epsilon$. If now $y \in C$, then $\theta(y, \epsilon) \leqslant |C|$, while, if $y \notin C$, then, by the lemma, $y \in C_\rho = c + \rho(C - c)$ for $\rho := 1 + \text{dist}(y, C)/\text{radius}(b_C)$. But then, $C_\rho \supseteq C$ (by convexity of C), hence, by $e1$, $E(C_\rho) \geqslant E(C) > \epsilon$. This shows that, for every y,

$$\theta(y, \epsilon) \leqslant |C_\rho| = \rho^N |C|. \quad \blacksquare \tag{7}$$

A little more work proves that θ is locally Lipschitz continuous.

THEOREM 2.1. *If $\theta(\cdot, \epsilon)$ is finite at some point, then*

$$|\theta(x, \epsilon) - \theta(y, \epsilon)| \leqslant L(x, y) |x - y|, \qquad \text{all } x, y, \tag{8}$$

with L bounded on bounded sets.

Proof. Let $x \in C \in \mathbb{C}$ with $E(C) > \epsilon$ and let $r := \text{radius}(b_C)$. By (7),

$$\theta(y, \epsilon) - |C| \leqslant (\rho^N - 1) |C|$$

with

$$\rho = 1 + \text{dist}(y, C)/r \leqslant 1 + |y - x|/r,$$

hence

$$(\rho^N - 1) |C| \leqslant (\rho - 1) |C| \sum_{n=1}^{N-1} \rho^n$$

$$\leqslant |y - x| (|C|/r^N) \sum_{n=1}^{N-1} (r + |y - x|)^n r^{N-1-n}.$$

By c2, $|b_C| \leqslant |C| \leqslant |B_C| \leqslant |b_C|/\beta^N$ while $|b_C| = r^N \text{const}_N$. Therefore $|C|/r^N \leqslant \text{const}_N/\beta^N$ and $r \leqslant (|C|/\text{const}_N)^{1/N}$, showing that

$$\theta(y, \epsilon) - |C| \leqslant (\rho^N - 1) |C| \leqslant F_N(|C|, |y - x|) |y - x|$$

for some function F_N which depends only on N and is monotone increasing in its two arguments. By taking the infimum over all such C, we get

$$\theta(y, \epsilon) - \theta(x, \epsilon) \leqslant F_N(\theta(x, \epsilon), |y - x|) |y - x|,$$

which proves (8) with $L(x, y) := F_N(\max\{\theta(x, \epsilon), \theta(y, \epsilon)\}, |y - x|)$. But such L is bounded on bounded sets by the corollary to Lemma 2 and the monotonicity of F_N. $\blacksquare$

COROLLARY. *If $G \subseteq \mathbb{R}^N$ is bounded and $m_G := \inf_{x \in G} \theta(x, \epsilon) > 0$, then $1/\theta(\cdot, \epsilon)$ is Lipschitz continuous on $\bar{G}$.*

It is obvious that $\theta(x, \cdot)$ is monotone (even if E were not). Further, one would expect $\lim_{\epsilon \to 0} \theta(x, \epsilon) = 0$ for each fixed x, but this need not happen. Consider, e.g., the case when $N = 1$, $\mathbb{C}$ consists of all closed intervals, and

$E(C) = \text{dist}_{\infty,C}(f, \mathbb{P}_1)$ with f the step function having just two jumps, both of positive size 2, at -1 and 1, say. Then

$$E(C) = \text{number of jumps of } f \text{ contained in } C.$$

Therefore, $\theta(x, \epsilon)$ is infinite for $\epsilon \geqslant 2$. Further, for $1 \leqslant \epsilon < 2$,

$$\theta(x, \epsilon) = \begin{cases} |[x, 1]| = |x| + 1 & \text{if} \quad x \leqslant -1, \\ |[-1, 1]| = 2 & \text{if} \quad -1 \leqslant x \leqslant 1, \\ |[-1, x]| = x + 1 & \text{if} \quad 1 \leqslant x. \end{cases}$$

Finally, for $0 < \epsilon < 1$,

$$\theta(x, \epsilon) = \text{dist}(x, \{-1, 1\}).$$

Note that the failure of $\theta(x, \epsilon)$ to go to zero with ϵ implies that, because of (6), unusually good approximation rates are possible. This will be taken up again in Section 5.

The example also illustrates the possibility that, even for positive ϵ, $\theta(\cdot, \epsilon)$ may vanish at some points. In such a case, though, our algorithm will not terminate since then, by the monotonicity of the error E, $E(C) \geqslant \epsilon$ for all C containing a point x with $\theta(x, \epsilon) = 0$. Since $\theta(y, \epsilon)$ cannot grow faster than const $|y - x|$ near such a point, by Theorem 2.1, it follows that $\int_G dy/\theta(y, \epsilon)$ is infinite for any G containing x. Thus, (6) *holds trivially in case* $\theta(x, \epsilon) = 0$ *for some* $x \in D$, both sides then being infinite for all small ϵ.

We are now ready to prove (6).

THEOREM 2.2. *Let* $(C_i)_1^K$ *be a subdivision of* D *produced by the adaptive algorithm from an initial subdivision* (C_i^0) *with* $E(C_i^0) > \epsilon$, *all* i. *Then, with* β *the constant in assumption* d1,

$$K \leqslant |E|_\epsilon/\beta. \tag{6}$$

Proof. By assumption, $\theta(\cdot, \epsilon)$ is finite everywhere and, since the algorithm stopped, $1/\theta(\cdot, \epsilon)$ must be finite everywhere in D. Hence, by Theorem 2.1, $1/\theta(\cdot, \epsilon)$ is continuous on D (and positive), thus integrable, and

$$\text{for all } C \in \mathbb{C}, \qquad \int_C dx/\theta(x, \epsilon) = |C|/\theta(x_C, \epsilon) \qquad \text{for some } x_C \in C. \tag{9}$$

Now, for each i, let J_i be the parent of C_i (in the sense of a4). Then $E(J_i) > \epsilon$, therefore $\theta(y, \epsilon) \leqslant |J_i| \leqslant |C_i|/\beta$, using d1, for all $y \in J_i$. Hence

$$\begin{aligned} K = \sum_{i=1}^K |C_i|/|C_i| &\leqslant \beta^{-1} \sum_{i=1}^K |C_i|/\theta(x_{C_i}, \epsilon) \\ &= \beta^{-1} \sum_{i=1}^K \int_{C_i} dx/\theta(x, \epsilon) = |E|_\epsilon/\beta, \end{aligned}$$

which finishes the proof. ∎

We will show in many specific circumstances that, as $\epsilon \to 0$, both $K = K(\epsilon)$ and $|E|_\epsilon$ go to infinity at the same rate, so there is then no doubt as to the sharpness of (6). Still, it is nice to know in general whether the two quantities are comparable. For this reason, we now prove a converse inequality.

THEOREM 2.3. *Suppose $1/\theta(\cdot, \epsilon)$ is Riemann integrable uniformly in ϵ, i.e., there exists ω with $\omega(0^+) = 0$ so that for any subdivision (C_i) of the bounded domain D and any choice of points $x_i \in C_i$,*

$$\left| \int_D dx/\theta(x, \epsilon) - \sum_i |C_i|/\theta(x_i, \epsilon) \right| \leqslant \omega(\max_i |C_i|)\, |E|_\epsilon, \quad \textit{all } \epsilon.$$

Then, for any subdivision $(C_i)_1^K$ of D with $E(C_i) \leqslant \epsilon$, all i,

$$(1 - \omega(|D|/K))\, |E|_\epsilon \leqslant \text{const}_{N,\beta,\eta}\, K. \tag{10}$$

Hence, if $\lim_{\epsilon\to 0} K(\epsilon) = \infty$, then $|E|_\epsilon$ and $K(\epsilon)$ approach infinity at the same rate.

For the proof, we need the following lemma.

LEMMA 3. *If $C \in \mathbb{C}$ and $E(C) \leqslant \epsilon$, then, at the center x_C of the inscribed ball b_C for C,*

$$\theta(x_C, \epsilon) \geqslant (\eta^2/2)^N\, |C|. \tag{11}$$

Proof. If $x_C \in C' \in \mathbb{C}$ with circumscribed ball $B_{C'}$ and $|B_{C'}| \leqslant |b_C|/2^N$, then $C' \subseteq B_{C'} \subseteq b_C$, and so $E(C') \leqslant E(b_C) \leqslant E(C) \leqslant \epsilon$, by e1. Hence, $x_C \in C' \in \mathbb{C}$ and $E(C') > \epsilon$ implies that $|b_C|/2^N < |B_{C'}|$, while, by c2,

$$|\hat{C}| \leqslant |B_{\hat{C}}| \leqslant |b_{\hat{C}}|/\eta^N \leqslant |\hat{C}|/\eta^N \qquad \text{for all} \quad \hat{C} \in \mathbb{C}.$$

Therefore, $|C|\,(\eta/2)^N \leqslant |b_C|/2^N < |B_{C'}| \leqslant |C'|/\eta^N$, and (11) now follows since C' was arbitrary. ∎

Proof of Theorem 2.3. Let $(C_i)_1^K$ be any subdivision for D with $E(C_i) \leqslant \epsilon$, all i, and let x_i be the center of the ball b_{C_i} inscribed in the cell C_i, all i. Then, by Lemma 3 and the assumption on uniform Riemann integrability, we have

$$\begin{aligned} K = \sum_{i=1}^K |C_i|/|C_i| &\geqslant (2/\eta^2)^N \sum_{i=1}^K |C_i|/\theta(x_i, \epsilon) \\ &= (2/\eta^2)^N \left[|E|_\epsilon + \left(\sum_{i=1}^K |C_i|/\theta(x_i, \epsilon) - |E|_\epsilon \right) \right] \\ &\geqslant (2/\eta^2)^N [1 - \omega(\max_i C_i)]\, |E|_\epsilon . \end{aligned}$$

There is, of course, no guarantee that $\max_i |C_i| \to 0$ as $\epsilon \to 0$. Still, we may refine our subdivision sufficiently while keeping the number of its cells within $O(K)$ as follows. Starting with the subdivision $(C_i)_1^K$ under discussion, we carry on with the adaptive algorithm until a new (finer) subdivision $(C_i)_1^{K'}$ is reached with $\max_i |C_i'| \leqslant |D|/K$. Then, the parent of each new cell must have had volume greater than $|D|/K$, hence there must have been less than K such parents, and each such parent could not have had more than $1/\beta$ children, by d1. Consequently, $K' \leqslant (1+\beta^{-1})K$ (in particular, the algorithm must have stopped), and from our earlier argument, now applied to the refined partition (C_i'),

$$(1+\beta^{-1})K \geqslant K' \geqslant (2/\beta^2)^N[1-\omega(|D|/K)]\,|E|_\epsilon . \quad \blacksquare$$

We conclude this section with a short discussion of our error measure

$$|E|_\epsilon = \int_D dx/\theta(x,\epsilon).$$

$|E|_\epsilon$ increases with ϵ^{-1} and, usually, $\lim_{\epsilon\to 0}|E|_\epsilon = \infty$. In particular instances, we are able to state quite precisely how $|E|_\epsilon$ goes to infinity with ϵ^{-1}.

In general, $|\cdot|_\epsilon$ *is monotone*, i.e.,

$$E \leqslant F \quad \text{implies} \quad |E|_\epsilon \leqslant |F|_\epsilon . \tag{12}$$

Also, $|E|_\epsilon = |\alpha E|_{\alpha\epsilon}$ for $\alpha > 0$. Finally,

$$|E+F|_{2\epsilon} \leqslant |E|_\epsilon + |F|_\epsilon . \tag{13}$$

For the proof of (13), note that for any $C \in \mathbb{C}$ with $(E+F)(C) > 2\epsilon$, we must have $\max\{E(C), F(C)\} > \epsilon$, hence $\theta_{E+F}(x, 2\epsilon) \geqslant \min\{\theta_E(x,\epsilon), \theta_F(x,\epsilon)\}$, and so

$$|E+F|_{2\epsilon} \leqslant \int_D \max\left\{\frac{1}{\theta_E(x,\epsilon)}, \frac{1}{\theta_F(x,\epsilon)}\right\} dx \leqslant |E|_\epsilon + |F|_\epsilon .$$

3. An Example: Best Approximation in $\mathbb{L}_p[a, b]$ from $\mathbb{P}_{n,K}$

In this section, we bound $|E|_\epsilon$ for a specific choice of E in order to illustrate the use of Theorems 2.2 and 2.3.

We are given a function f on some interval $[a, b]$, and f is in $C^{(n)}$ on $[a,b]\backslash\{s\}$. But we know that

$$|f^{(n)}(x)| \leqslant \text{const}_f\, |x-s|^{\alpha-n} \tag{15}$$

for some α with $\alpha p > -1$. We intend to approximate f from $\mathbb{P}_{n,K}$, i.e., by piecewise polynomial functions consisting of at most K polynomial pieces, each piece of order n, i.e., of degree less than n. We take the $\mathbb{L}_p$-norm

$$\| g \|_p := \left(\int_a^b | g(x)|^p \, dx \right)^{1/p}$$

as our measure of function size. Our assumptions on f then imply that $f \in \mathbb{L}_p[a, b]$.

Rice [8] has shown some time ago that, for such an f,

$$\operatorname{dist}_p(f, \mathbb{P}_{n,K}) = O(K^{-n}).$$

We are about to reprove this result. In fact, we will prove that the approximation to such a function f constructed by the adaptive algorithm approaches f at the rate $O(K^{-n})$.

As collection $\mathbb{C}$ of allowable cells we choose all finite closed intervals on $\mathbb{R}$. Thus, c1, c2, c3 are satisfied, and $\eta = 1$ in c2. For the division algorithm a4, we take interval halving, so d1 is satisfied with $\beta = 1/2$.

Ideally, we would take for the error measure on the interval C the distance of $f|_C$ from $\mathbb{P}_n$,

$$E_f(C) := \operatorname{dist}_{p,C}(f, \mathbb{P}_n).$$

But it is simpler, and corresponds better to actual practice, to work with some bound $E(C)$ for $E_f(C)$. Such a bound we now derive.

If $f \in C^{(n)}(C)$ for some interval $C \subseteq [a, b]$, then $\operatorname{dist}_{p,C}(f, \mathbb{P}_n) \leqslant \operatorname{const}_n \| f^{(n)} \|_{p,C} \, | C |^{n+1/p}$. Thus, with our assumption (15), we have $E_f(C) \leqslant \operatorname{const}_{f,n} F(C)$, with

$$F(C) := \operatorname{dist}(s, C)^{\alpha-n} \, | C |^{n+1/p}, \qquad \text{all } C \subseteq [a, b] \tag{16}$$

for such C. Note that $F(C) = \infty$ in case $s \in C$ if, as we assume,

$$\alpha < n.$$

Hence, for such C, and for C "near" s, we need an alternative bound. If s is not in the interior of C, e.g., $C = [u, v]$ with $v \leqslant s$, then we have

$$f(x) = \sum_{j<n} f^{(j)}(u)(x - u)^j/j! + \int_u^x (x - t)^{n-1} f^{(n)}(t) \, dt/(n - 1)!$$

and so

$$\operatorname{dist}_{p,C}(f, \mathbb{P}_n) \leqslant \operatorname{const}_n \left(\int_u^v \left| \int_u^x | x - t |^{n-1} f^{(n)}(t) \, dt \right|^p dx \right)^{1/p}.$$

But, by (15),

$$\left| \int_u^x (x - t)^{n-1} f^{(n)}(t)\, dt \right| \leqslant \text{const}_{n,f} \int_u^x | x - t |^{n-1} | t - s |^{\alpha - n}\, dt$$
$$\leqslant \text{const}_{n,f} \int_u^x | s - t |^{\alpha - 1}\, dt$$
$$\leqslant \text{const}_{n,f} | s - x |^{\alpha}.$$

Therefore,

$$\text{dist}_{p,C}(f, \mathbb{P}_n) \leqslant \text{const} \left(\int_C | s - x |^{\alpha p}\, dx \right)^{1/p} \leqslant \text{const} \left| (s - x)^{\alpha + 1/p} \Big|_u^v \right|.$$

This shows that, for such C, $\text{dist}_{p,C}(f, \mathbb{P}_n) \leqslant \text{const}_{f,n}\, G(C)$, with

$$G(C) := (\text{dist}(s, C) + | C |)^{\alpha + 1/p}, \qquad \text{all } C \subseteq [a, b]. \tag{17}$$

Finally, if the singularity s lies in the interior of C, then the error might be as bad as $\text{const}_f | C |^{1/p}$ without additional hypotheses on f. Hence, if $\alpha > 0$, then we *assume* that $\text{dist}_{p,C}(f, \mathbb{P}_n) \leqslant \text{const}_{n,f}\, G(C)$ also for C with $s \in \text{int}(C)$.

To summarize, we have

$$E_f \leqslant \text{const}_{n,f}\, E := \text{const}_{n,f} \min\{F, G\}, \tag{18}$$

with F and G given by (16), (17). Both F and G are monotone, and continuous where they are finite, and $F(C) = \infty$ implies $G(C) < \infty$. Hence E is monotone and continuous. Extending E to all of $\mathbb{C}$ by

$$E(C) := E(C \cap [a, b])$$

clearly changes nothing in this.

Proposition 3.1. *For the function E defined in* (18),

$$| E |_\epsilon \leqslant \text{const}\, \epsilon^{-1/(n+1/p)}.$$

Proof. For each $x \in [a, b]$, let C_x be an interval with $x \in C_x$ and $| C_x | = \theta(x, \epsilon)$, hence $E(C_x) = \epsilon$. Such surely exists for all sufficiently small ϵ by the continuity of E. Then $C_x \subseteq [a, b]$.

If now $\text{dist}(s, C_x) \leqslant | C_x |$, then

$$F(C_x) = \text{dist}(s, C_x)^{\alpha - n} | C_x |^{n+1/p} \geqslant | C_x |^{\alpha + 1/p}$$

while, for any C,

$$G(C) \geqslant | C |^{\alpha + 1/p}.$$

Therefore, $\text{dist}(s, C_x) \leqslant |C_x|$ implies that $\epsilon \geqslant |C_x|^{\alpha+1/p}$, hence $|s - x| \leqslant \text{dist}(s, C_x) + |C_x| \leqslant 2|C_x| \leqslant 2\epsilon^{1/(\alpha+1/p)}$. This shows that

$$A := \{x \in [a, b]: \text{dist}(s, C_x) \leqslant |C_x|\}$$

has $|A| \leqslant 4\epsilon^{1/(\alpha+1/p)}$. Since

$$\epsilon \leqslant G(C_x) \leqslant (2|C_x|)^{\alpha+1/p} \qquad \text{for} \quad x \in A,$$

it follows that

$$\int_A dx/\theta(x, \epsilon) \leqslant |A| \, 2\epsilon^{-1/(\alpha+1/p)} \leqslant 8.$$

On the other hand, since $\epsilon \leqslant F(C_x) = \text{dist}(s, C_x)^{\alpha-n} |C_x|^{n+1/p}$, we have

$$1/\theta(x, \epsilon) = 1/|C_x| \leqslant (\text{dist}(s, C_x)^{\alpha-n}/\epsilon)^{1/(n+1/p)}$$

and so, as $|s - x| \leqslant \text{dist}(s, C_x) + |C_x| < 2\,\text{dist}(s, C_x)$ for $x \notin A$,

$$\int_{\setminus A} dx/\theta(x, \epsilon) \leqslant \epsilon^{-1/(n+1/p)} \int_a^b (|s - x|/2)^{(\alpha-n)/(n+1/p)} \, dx$$
$$\leqslant \epsilon^{-1/(n+1/p)} \text{const}_{a,b,\alpha,p}\,.$$

Thus

$$|E|_\epsilon = \int_a^b dx/\theta(x, \epsilon) = \left(\int_A + \int_{\setminus A}\right) dx/\theta(x, \epsilon) \leqslant 8 + \text{const}\ \epsilon^{-1/(n+1/p)}$$

which finishes the proof. ∎

It follows with Theorem 2.2 that, for such a function f, the adaptive algorithm working either with E_f or with the bound $\text{const}_{n,f}\, E$ for it, produces an approximation $g \in \mathbb{P}_{n,K}$ to f for which

$$\|f - g\|_p^p \leqslant K\epsilon^p \leqslant \text{const}\ \epsilon^{p-1/(n+1/p)} = \text{const}\ \epsilon^{pn/(n+1/p)},$$

while, again by the proposition, $\epsilon^{1/(n+1/p)} \leqslant \text{const}/\int (dx/\theta(x, \epsilon)) \leqslant \text{const}/K$. This shows that then

$$\|f - g\|_p \leqslant \text{const}\ \epsilon^{n/(n+1/p)} \leqslant \text{const}\ K^{-n},$$

the promised bound.

Note that the argument also covers functions having finitely many singularities of algebraic type no worse than α. Precisely, if $f = \sum_1^r f_j$, with $f_j \in C^{(n)}([a, b]\setminus\{s_j\})$ and $|f_j^{(n)}(x)| \leqslant a_j |x - s_j|^{\alpha-n}$, all j, then, from the argument for (13),

$$|E_f|_{r\epsilon} \leqslant \sum_{j=1}^r |a_j E_j|_\epsilon$$

with $E_j := \min\{F_j, G_j\}$ and F_j and G_j defined by (16) and (17) with s replaced by s_j. Thus

$$| E_f |_\epsilon \leqslant \sum_{j=1}^{r} | E_j |_{\epsilon/(ra_j)} \leqslant \text{const } \epsilon^{-1/(n+1/p)} \sum_{j=1}^{r} (ra_j)^{1/(n+1/p)}.$$

Therefore, the adaptive algorithm produces approximations of optimal order for such functions, too.

We have carried out this last argument in such detail in order to show that it will not, by itself, support the analysis of an f with infinitely many singularities. For this, a more refined version of (13) would be needed. Also, our argument comes close to, but does not recapture, the result by Burchard [5–6] and others that

$$\text{dist}_p(f, \mathbb{P}_{n,K}) \leqslant \text{const } K^{-n} \| f^{(n)} \|_{1/(n+1/p)} \,. \tag{18a}$$

4. The Adaptive Approximation of a Function on $\mathbb{R}^N$ with Singularities on a Smooth Manifold

In this section, we investigate the approximation of a function f on some bounded domain D in $\mathbb{R}^N$ when f is in $C^{(n)}(D\backslash S)$ for some smooth manifold S of dimension m. We do not specify the collection $\mathbb{C}$ of allowable cells beyond the requirements made in Section 1. Then we have

$$\text{dist}_{p,C}(f, \mathbb{P}_n) \leqslant \text{const}_n f^{(n)}(C) \mid C \mid^{1/p} (\text{diam } C)^n \qquad \text{for} \quad C \cap S = \varnothing \tag{19}$$

with

$$f^{(n)}(C) := \sup_{x \in C} \max_{|\gamma|=n} |(D^\gamma f)(x)|,$$

as is well known (see, e.g., Morrey [7; p. 85]). Here, $\mathbb{P}_n$ stands for the collection of polynomials on $\mathbb{R}^N$ of total degree less than n. If now f were smooth enough, i.e., if $f^{(n)}(D) < \infty$, then, for any partition $(C_i)_1^K$ of D, we would get an approximation g to f with

$$\begin{aligned} \| f - g \|_p^p &= \sum_i \text{dist}_{p,C}(f, \mathbb{P}_n)^p \\ &\leqslant (\text{const}_n f^{(n)}(D))^p \sum_i \mid C_i \mid (\text{diam } C_i)^{np} \\ &\leqslant (\text{const}_n f^{(n)}(D))^p (\max_i (\text{diam } C_i)^n)^p \mid D \mid \end{aligned}$$

hence

$$\| f - g \|_p \leqslant \text{const}_{n,f,D} \max_i (\text{diam } C_i)^n.$$

This expression is $O(K^{-n/N})$ if the C_i are chosen to be more or less uniform. We intend to show that this same *optimal* order of approximation can be achieved even for a function with certain singularities when the approximation is constructed by our adaptive algorithm.

We now specify the singular behavior of f. We assume that

$$f^{(n)}(C) \leqslant \text{const}_f \operatorname{dist}(S, C)^{\alpha-n}. \tag{20}$$

Here and below, we take the distance between two sets in $\mathbb{R}^N$ to be the shortest distance between them, i.e.,

$$\operatorname{dist}(S, C) := \inf_{\substack{s\in S\\ c\in C}} | s - c |,$$

with $| \cdot |$ denoting Euclidean distance. Our assumption (20) does not imply much about $\operatorname{dist}_{p,C}(f, \mathbb{P}_n)$ in case $S \cap C \neq \varnothing$. We make the assumption that

$$\operatorname{dist}_{p,C}(f, \mathbb{P}_n) \leqslant \text{const}_f | C |^{1/p}(\operatorname{diam} C)^\alpha \qquad \text{for} \quad C \cap S \neq \varnothing. \tag{21}$$

Consequently,

$$E_f(C) := \operatorname{dist}_{p,C}(f, \mathbb{P}_n) \leqslant \text{const}_f\, E(C)$$

with

$$\begin{aligned} E(C) &:= \min\{F(C), G(C)\},\\ F(C) &:= \operatorname{dist}(S, C)^{\alpha-n}(\operatorname{diam} C)^n \, | C |^{1/p},\\ G(C) &:= (\operatorname{dist}(S, C) + \operatorname{diam} C)^\alpha \, | C |^{1/p}. \end{aligned}$$

Finally, we assume that the m-dimensional manifold S of singularities of f is *smooth* in the following sense:

s1. It is possible to subdivide D into finitely many (nonoverlapping) pieces $D_1, ..., D_r$ so that, for each i, either $\operatorname{dist}(S, D_i) > 0$ or else there exists a continuously differentiable map φ_i which maps the cylinder

$$Z_m := \left\{x \in \mathbb{R}^N : 0 \leqslant x_i \leqslant 1 \text{ for } i = 1, ..., m;\ \sum_{j>m} x_j^2 \leqslant 1\right\}$$

one-one onto some neighborhood V_i of D_i so that

$$\operatorname{dist}(\varphi_i(x), S) = \left(\sum_{j>m} x_j^2\right)^{1/2} \qquad \text{for all} \quad \varphi_i(x) \in D_i\,.$$

THEOREM 4.1. *If $f \in C^{(n)}(D\backslash S)$ and (20), (21) both hold for some α with*

$$\alpha > mn/N - (N - m)/p, \tag{22}$$

and the manifold S of singularities of f is smooth in the sense that it satisfies s1 *above; then the adaptive algorithm, starting from a subdivision* (C_i') *for D with* $C_i' \in \mathbb{C}$ *and* $E(C_i') > \epsilon$, *all i, produces a subdivision* $(C_i)_1^K$ *of D for which*

$$K \leqslant \text{const}_f\, \epsilon^{-1/(n/N+1/p)}.$$

Proof. The proof parallels the one for Proposition 3.1. We assume without loss of generality that $\alpha \leqslant n$, since (20), (21) hold for $\alpha = n$ if they hold for some $\alpha > n$. By Theorem 2.2, $K \leqslant \beta^{-1} \int_D dx/\theta(x, \epsilon)$. To estimate $\theta(x, \epsilon)$, let $x \in C_x \in \mathbb{C}$ with $\theta(x, \epsilon) = |C_x|$ and $E(C_x) = \epsilon$. By c2, there exists a positive const $= \text{const}_{N,n}$ so that

$$\text{const diam } C \geqslant |C|^{1/N} \geqslant \text{diam } C/\text{const}, \qquad \text{for all } C \in \mathbb{C}.$$

Set

$$A := \{x \in D: \text{dist}(S, C_x) \leqslant \text{diam } C_x\}.$$

Then, for $x \in A$,

$$F(C_x) \geqslant |\text{diam } C_x|^\alpha\, |C_x|^{1/p} \geqslant \text{const}\, |C_x|^{\alpha/N+1/p}$$

while, for any x,

$$G(C_x) \geqslant |\text{diam } C_x|^\alpha\, |C_x|^{1/p} \geqslant \text{const}\, |C_x|^{\alpha/N+1/p}.$$

Hence,

$$\epsilon = \min\{F(C_x), G(C_x)\} \geqslant \text{const}\, |C_x|^{\alpha/N+1/p} \qquad \text{for} \quad x \in A.$$

This implies that $\epsilon^{N/(\alpha+N/p)} \geqslant \text{const}\, |C_x|$ and so

$$\begin{aligned} \text{dist}(x, S) &\leqslant \text{dist}(C_x, S) + \text{diam } C_x \\ &\leqslant 2 \text{ diam } C_x \leqslant \text{const}\, \epsilon^{1/(\alpha+N/p)} \qquad \text{for} \quad x \in A. \end{aligned} \tag{23}$$

Further, for $x \in A$,

$$\epsilon \leqslant G(C_x) \leqslant (2 \text{ diam } C_x)^\alpha\, |C_x|^{1/p} \leqslant \text{const}\, |C_x|^{\alpha/N+1/p}$$

which proves that

$$1/\theta(x, \epsilon) = 1/|C_x| \leqslant \text{const}\, \epsilon^{-N/(\alpha+N/p)} \qquad \text{for} \quad x \in A. \tag{24}$$

By assumption, s1 holds, hence S has finite m-dimensional volume. Therefore, (23) and (24) combine to give

$$\begin{aligned} \int_A dx/\theta(x, \epsilon) &\leqslant \text{const}\, \epsilon^{-N/(\alpha+N/p)}\, |A| \\ &\leqslant \text{const}\, \epsilon^{-N/(\alpha+N/p)}\, \text{const}_S\, \epsilon^{(N-m)/(\alpha+N/p)} \\ &= \text{const}\, \epsilon^{-m/(\alpha+N/p)} \\ &\leqslant \text{const}\, \epsilon^{-N/(\alpha+N/p)}. \end{aligned}$$

Next, since

$$\epsilon \leqslant F(C_x) \leqslant \text{dist}(S, C_x)^{\alpha-n}(\text{diam } C_x)^n \mid C_x \mid^{1/p}$$
$$\leqslant \text{const dist}(S, C_x)^{\alpha-n} \mid C_x \mid^{n/N+1/p}$$

we have

$$1/\theta(x, \epsilon) = 1/\mid C_x \mid \leqslant \text{const}(\text{dist}(S, C_x)^{\alpha-n}/\epsilon)^{N/(n+N/p)}.$$

Now, for $x \notin A$, $\text{dist}(x, S) \leqslant \text{dist}(S, C_x) + \text{diam } C_x < 2 \text{ dist}(S, C_x)$. Thus, with $\alpha \leqslant n$,

$$1/\theta(x, \epsilon) \leqslant \epsilon^{-N/(n+1/p)} \text{const}(\text{dist}(x, S)/2)^{(\alpha-n)/(n/N+1/p)}, \qquad x \notin A. \quad (25)$$

It follows that

$$\int_{D\setminus A} dx/\theta(x, \epsilon) \leqslant \epsilon^{-N/(n+N/p)} \text{ const} \int_D \text{dist}(x, S)^{(\alpha-n)/(n/N+1/p)}\, dx.$$

Finally, we show that the last integral is finite under our assumptions. For this, we make use of the smoothness assumption s1 on S. For each D_i in the postulated subdivision of D, we have

$$\int_{D_i} \text{dist}(x, S)^\gamma\, dx \leqslant \mid D_i \mid \text{dist}(D_i, S)^\gamma$$

with

$$\gamma := (\alpha - n)/(n/N + 1/p)$$

since γ is negative by (22). Hence, $\int_{D_i} \text{dist}(x, S)^\gamma\, dx$ is finite in case $\text{dist}(D_i, S) > 0$. If, on the other hand, $\text{dist}(D_i, S) = 0$, then, by s1,

$$\int_{D_i} \text{dist}(x, S)^\gamma\, dx = \int_{\varphi_i^{-1}(D_i)} \left(\sum_{j>m} x_j^2\right)^{\gamma/2} \det \varphi_i'(x)\, dx$$
$$\leqslant \text{const} \int_{Z_m} \left(\sum_{j>m} x_j^2\right)^{\gamma/2} dx$$
$$\leqslant \text{const} \int_{S_{N-m}} \mid x \mid^\gamma dx,$$

with S_k the unit sphere in $\mathbb{R}^k$. Since

$$\int_{S_k} \mid x \mid^\gamma dx = \text{const}_k \int_{\partial S_k} \int_0^1 r^{\gamma+k-1}\, dr\, ds,$$

we have $\int_{D_i} \operatorname{dist}(x, S)^\gamma \, dx < \infty$ provided $\gamma > -(N - m)$, i.e., provided

$$(\alpha - n)/(n/N + 1/p) > m - N. \tag{26}$$

But this is exactly the second inequality in (22). ▮

We know ([2; Lemma 3.2]) that, for $f \in W^a_{q,D\setminus S}$, (19) may be replaced by

$$\operatorname{dist}_{p,C}(f, \mathbb{P}_n) \leqslant \operatorname{const}_n \mid C \mid^{a/N+1/p-1/q} \| f \|_{L^a_{q,C}} \tag{19a}$$

where $\| \cdot \|_{L^a_{q,C}}$ is the L_q-norm on the cell C of the a-th derivative (possibly fractional). This bound reduces to (19) when $q = \infty$ and $a = n$. We replace assumption (20) by

$$\| f \|_{L^a_{q,C}} \leqslant \operatorname{const}_f \cdot \operatorname{dist}(S, C)^{\alpha - a} \mid C \mid^{1/q} \tag{20a}$$

and clearly the only case of interest is $\alpha < a$. Assumption (21) is unchanged as are the functions E, F, and G. Then we have

COROLLARY 1. *Assume $f \in W^a_{q,D\setminus S}$ and that both* (20a) *and* (21) *hold for some α with*

$$\alpha > ma/N - (N - m)/p.$$

Then, with the remaining hypotheses of Theorem 4.1, *we have*

$$K \leqslant \operatorname{const}_f \cdot \epsilon^{-1/(a/N+1/p)}.$$

Denote by

$$\mathbb{P}_{n,K}$$

the collection of piecewise polynomial functions of order n consisting of no more than K pieces, with the corresponding subdivision $(C_i)_1^r$ of D with $r \leqslant K$ taken from $\mathbb{C}$.

COROLLARY 2. *Under the assumptions of the theorem,*

$$\operatorname{dist}_{p,D}(f, \mathbb{P}_{n,K}) = O(K^{-n/N}).$$

Proof. For each small enough ϵ, we can find a subdivision $(C_i)_1^K$ for D so that $E_f(C_i) \leqslant E(C_i) \leqslant \epsilon$, while $K \leqslant \operatorname{const} \epsilon^{-N/(n+N/p)}$ for some const $=$ $\operatorname{const}_{f,N,m,n,\alpha,D,S}$ but independent of ϵ. This shows that

$$\epsilon^{N/(n+N/p)} \leqslant \operatorname{const} K^{-1} \tag{27}$$

and implies the existence of an approximation $g \in \mathbb{P}_{n,K}$ for f for which

$$\|f - g\|_p^p \leqslant K\epsilon^p \leqslant \text{const } \epsilon^{p-N/(n+N/p)}$$
$$= \text{const } \epsilon^{pn/(n+N/p)} \leqslant \text{const } K^{-pn/N},$$

the last inequality by (27). ▌

Theorem 4.1 and its Corollary only treat functions with a smooth manifold of singularities in the sense of s1. Thus, a function like $f(x) := x_1^\alpha + x_2^\beta$ on $\mathbb{R}^2$ is not covered. But, by (13) and its obvious generalization, we clearly get $\text{dist}_{p,C}(f, \mathbb{P}_{n,K}) = O(K^{-n/N})$ for any f which can be written as a finite sum of functions satisfying the assumptions of Theorem 4.1. Thus, the function $f(x) = x_1^\alpha + x_2^\beta$ is covered. Still, Theorem 4.1 does not apply to functions like

$$f(x) := (x_1x_2)^\alpha \qquad \text{for} \quad x \in \mathbb{R}^2.$$

Finally, Theorem 4.1 has the assumption that the domain D be representable as the finite union of nonoverlapping allowable cells. This, is offhand, a severe restriction. E.g., we require all allowable cells to be convex, and, typically, $\mathbb{C}$ consists of just hyperrectangles. But, if D is not so representable, then it is sufficient to start off with some domain $\hat{D}$ containing D which is the union of nonoverlapping allowable cells (C_i') *provided* we can extend f suitably to this larger domain $\hat{D}$. The possibility of such an extension is already implicit in the discussion of the error function E at the beginning of this section. Our estimate (19) for $\text{dist}_{p,C}(f, \mathbb{P}_n)$ makes sense only if $C \subseteq \text{dom} f$. Without trying to squeeze the most general statement out of our arguments, we can say that Theorem 4.1 applies to the approximation of any function f on D which can be suitably extended to some bounded convex domain $\hat{D}$ containing D. The definition of E, offhand defined only for $C \subseteq \hat{D}$, is then extended to all $C \in \mathbb{C}$ by $E(C) := E(C \cap \hat{D})$, and the condition of Theorem 4.1 is relaxed to require initially only a finite covering (C_i') for D of allowable cells with $E(C_i) > \epsilon$, all i.

Note that, for $m \neq 0$, (26) is stronger than the inequality

$$\alpha > (m - N)/p$$

needed to conclude that $f \in \mathbb{L}_p(D)$. One might, for this and other reasons, raise the question of whether (26) is necessary. We now show that (26) is necessary in general to achieve the optimal approximation rate $O(K^{-n/N})$.

THEOREM 4.2. *If $m > 0$, then there exist $\mathbb{C}$, D, S and f satisfying all assumptions of Theorem* 4.1 *except that α is not an integer and satisfies*

$$-(N - m)/p < \alpha < mn/N - (N - m)/p, \tag{28}$$

and, for this f,

$$\operatorname{dist}_p(f, \mathbb{P}_{n,K}) \neq O(K^{-n/N}).$$

Proof. We choose

$$S := \{x \in \mathbb{R}^N : x_i = 0 \text{ for } i > m\}$$

and take

$$f(x) := \operatorname{dist}(x, S)^\alpha.$$

We choose D to be the unit cube $\{x \in \mathbb{R}^N : 0 \leqslant x_i \leqslant 1, \text{all } i\}$ and take $\mathbb{C}$ to be the collection of all scaled translates of D.

We claim that

$$\operatorname{dist}_{p,C}(f, \mathbb{P}_n) \geqslant \text{const} \mid C \mid^{\alpha/N+1/p}, \qquad \text{for all } C \in \mathbb{C}, C \cap S \neq \varnothing, \quad (29)$$

for some positive const independent of the particular C. For the proof, let $\varphi : \mathbb{R}^N \to \mathbb{R}^N : x \mapsto \rho x + s$ for some positive scalar ρ and some $s \in S$. Then $f\varphi = \rho^\alpha f$ while, for any $g \in \mathbb{P}_n$, $g\varphi \in \mathbb{P}_n$. Therefore, if $g \in \mathbb{P}_n$ is a best $\mathbb{L}_p$-approximation to f on $C \in \mathbb{C}$, then

$$\begin{aligned}\operatorname{dist}_{p,C}(f, \mathbb{P}_n)^p &= \int_C |f(x) - g(x)|^p \, dx \\ &= \int_{\varphi^{-1}(C)} |\rho^\alpha f(y) - g(\rho y + s)|^p \rho^N \, dy \\ &\leqslant \rho^{\alpha p+N} \operatorname{dist}_{p,\varphi^{-1}(C)}(f, \mathbb{P}_n).\end{aligned}$$

But since $\varphi^{-1}(x) = \rho^{-1}x + s'$ with $s' = -s/\rho \in S$, this implies that

$$\operatorname{dist}_{p,C}(f, \mathbb{P}_n) = \rho^{\alpha+N/p} \operatorname{dist}_{p,\varphi^{-1}(C)}(f, \mathbb{P}_n).$$

Associate now with each $C \in \mathbb{C}$ a specific map $\varphi : x \mapsto \rho x + s$ for which $s \in S$, $\varphi^{-1}(C) \cap S^\perp \neq \varnothing$ and $|\varphi^{-1}(C)| = 1$. Then $\rho = C^{1/N}$ and $\varphi^{-1}(C) \cap S \neq \varnothing$. Hence, for all $C \in \mathbb{C}$ with $C \cap S \neq \varnothing$,

$$\begin{aligned}\operatorname{dist}_{p,C}(f, \mathbb{P}_n) &= | C |^{\alpha/N+1/p} \operatorname{dist}_{p,\varphi^{-1}(C)}(f, \mathbb{P}_n) \\ &\geqslant \text{const} \mid C \mid^{\alpha/N+1/p}\end{aligned}$$

with

$$\text{const} := \inf\{\operatorname{dist}_{p,C}(f, \mathbb{P}_n) : C \in \mathbb{C}, C \cap S^\perp \neq \varnothing \neq C \cap S, | C | = 1\}.$$

If now const $= 0$, then, since $\operatorname{dist}_{p,C}(f, \mathbb{P}_n)$ is a continuous function of C on $\mathbb{C}$ and the infimum is taken over a compact subset of $\mathbb{C}$, we would obtain a $C \in \mathbb{C}$ with $|C| = 1$ for which $f|_C \in \mathbb{P}_n$, which is absurd since α is not an integer.

With (29) thus established, let $(C_i)_1^K$ be any collection of nonoverlapping cubes which cover $S \cap D$. Then, by (29),

$$\text{error}^p := \sum_i \operatorname{dist}_{p,C_i}(f, \mathbb{P}_n)^p \geqslant \text{const}^p \sum_i |C_i|^{(p\alpha/N)+1}$$

while

$$\sum_i |C_i|^{m/N} \geqslant 1$$

since they cover $S \cap D$. But this implies that

$$\begin{aligned} \text{error} &\geqslant \text{const} \inf \left\{ \sum_{i=1}^K |C_i|^{(p\alpha/N)+1} : \sum_{i=1}^K |C_i|^{m/N} \geqslant 1 \right\}^{1/p} \\ &\geqslant \text{const}\ \delta \end{aligned}$$

with

$$\delta^p := \inf \left\{ \sum_{i=1}^K |c_i|^\gamma : \sum_{i=1}^K |c_i| = 1 \right\}$$

and

$$\gamma := (p\alpha + N)/m.$$

Since $\gamma > 1$ by the first inequality in (28), the last infimum is taken on when $|c_i| = 1/K$, all i. Thus

$$\delta^p = \sum_{i=1}^K K^{-\gamma} = K^{1-\gamma}.$$

This proves that, for some positive const,

$$\begin{aligned} \operatorname{dist}_{p,D}(f, \mathbb{P}_n) &\geqslant \text{const}\ K^{(1-\gamma)/p} \\ &= \text{const}\ K^{-\alpha/m+(m-N)/(pm)} \\ &\neq O(K^{-n/N}) \end{aligned}$$

since, by assumption (28), $-\alpha/m + (m - N)/(pm) > -n/N$. ∎

5. SUPERCONVERGENCE

In this section, we give one more application of Theorem 2.2, this time to illustrate how it deals with superconvergence.

First, consider the step function f on $\mathbb{R}$ with just one jump, of size 1, say, at some point $x_0 \in (-1, 1)$. The function is to be approximated from $\mathbb{P}_{1,K}$ in the $\mathbb{L}_1$-norm. Clearly, the placement of just one breakpoint, at x_0, would provide exact approximation. But we are dealing with an adaptive algorithm which only knows (a bound for) the function E_f, and does not know the point x_0. We want to show that, even without the exact knowledge of the jump point x_0, our adaptive algorithm performs in this case much better than the "optimal" rate $O(K^{-1})$ would indicate.

It is easy to see that

$$E_f(C) := \text{dist}_{1,C}(f, \mathbb{P}_1) = \text{dist}(x_0, \mathbb{R}\backslash C)$$

for any particular interval C. Thus

$$\theta(x, \epsilon) = \text{dist}(x, x_0) + \epsilon$$

and so

$$|E_f|_\epsilon = \int_{-1}^{1} dx/(|x - x_0| + \epsilon) = 2\ln(1/\epsilon) + \ln[(1 + x_0 + \epsilon)(1 - x_0 + \epsilon)].$$

Consequently, the adaptive algorithm produces a subdivision with

$$K \leqslant (2\ln(1/\epsilon) + \text{const})/\beta$$

intervals for a total error of no more than

$$K\epsilon \leqslant K\epsilon^{-(\beta K - \text{const})/2} = O(e^{-\beta K/2}). \tag{30}$$

In fact, the total error is $\leqslant\epsilon$ since the approximation fails to be perfect only in the one interval which contains x_0 in its interior. Further, if interval halving is used, i.e., $\beta = 1/2$, then, at each stage, only the interval containing x_0 is subdivided. Assuming x_0 to be in general position, i.e., $x_0 \notin \{r2^s: r, s \in \mathbb{Z}\}$, the error with K intervals behaves therefore like $2^{-K} = e^{-K\ln 2}$. Thus, the error goes to zero even faster than our estimate $O(e^{-K/4})$ in (30) would indicate.

As a second example, we consider $\mathbb{L}_1$-approximation from $\mathbb{P}_{1,K}$ to the function

$$f(x_1, x_2) := (x_1 - x_2)_+^0 .$$

We take D to be the unit square and take for $\mathbb{C}$ all scaled translates of D. It is now not possible to have f approximated exactly; still, it can be approximated better than the "optimal" order $O(K^{-1/2})$ possible for general smooth functions.

Let $C \in \mathbb{C}$. If the line $S := \{x \in \mathbb{R}^2 : x_1 = x_2\}$ intersects C at all, it cuts it into a triangle T and another piece, and then $E_f(C) = |T|$. Otherwise, $E_f(C) = 0$. Thus, if x is the vertex of C farthest from S and h is its side, then

$$E_f(C) = \begin{cases} 2(h - |x_1 - x_2|/2)_+^2 & \text{for } h \leqslant |x_1 - x_2| \\ |x_1 - x_2|^2/2 & \text{for } h > |x_1 - x_2|. \end{cases}$$

This shows that, for $|x_1 - x_2|^2/2 \leqslant \epsilon$, $\theta(x, \epsilon) = h^2$ with h such that $(h - |x_1 - x_2|/2)^2 = (\epsilon/2)^{1/2}$. Thus,

$$\theta(x, \epsilon) = \begin{cases} ((\epsilon/2)^{1/2} + |x_1 - x_2|/2)^2 & \text{if } |x_1 - x_2| \geqslant (2\epsilon)^{1/2} \\ 2\epsilon & \text{otherwise.} \end{cases}$$

Consequently, $\theta(x, \epsilon) \geqslant 2\epsilon$ for all x, and so $|E_f|_\epsilon \leqslant |D|/(2\epsilon)$. But the resulting estimate $K\epsilon \leqslant (\text{const}/(2\epsilon))\epsilon = \text{const}$ for the total error is not too encouraging.

We get a sharper bound as follows. Set

$$A := \{x \in D : |x_1 - x_2| \leqslant (2\epsilon)^{1/2}\}.$$

Then $|A| \leqslant \sqrt{2}\,(2\epsilon)^{1/2}$, hence

$$\int_A dx/\theta(x, \epsilon) \leqslant |A|/(2\epsilon) = 1/\epsilon^{1/2}.$$

Also,

$$\int_{D \setminus A} dx/\theta(x, \epsilon) = \int ((\epsilon/2)^{1/2} + |x_1 - x_2|/2)^{-2}\, dx$$

$$\leqslant 2 \int_0^1 \int_{\epsilon/2}^{1/2} ((\epsilon/2)^{1/2} + s)^{-2}\, 8\, ds\, dt < 16/(2\epsilon)^{1/2}.$$

Thus $|E_f|_\epsilon \leqslant \text{const}/\epsilon^{1/2}$, hence $\epsilon \leqslant \text{const}/K^2$ for the number K of squares in the partition constructed by the adaptive algorithm. The error achieved is therefore no bigger than $K\epsilon \leqslant \text{const}/K$, or, $O(K^{-1})$ as compared to the "optimal" rate $O(K^{-1/2})$.

6. Algorithm Realizations for Smooth Approximation

The obvious concrete realizations of the adaptive algorithm are for piecewise polynomial approximation such as analyzed in Sections 3–5. Most of these realizations would produce discontinuous approximations. This is perfectly acceptable for applications such as quadrature, i.e., $\mathbb{L}_1$-approximation, or in situations where only the accuracy of the approximation matters. Other applications require smooth approximations and, in one variable, this may be achieved by either using a local approximation scheme that preserves smoothness (see Rice [9] for two such methods) or else by "smoothing" the original discontinuous approximation by "pulling apart the knots." In principle, one can also "smooth" a multivariate piecewise polynomial approximation, but it is not clear that one can do it in practice. The simple mechanism of "pulling apart knots" is not available, and the problem of carrying out some reasonable local "smoothing" on a piecewise polynomial function on a nonuniform subdivision seems insurmountable.

The difficulty of preserving smoothness with piecewise polynomial approximations is illustrated in Fig. 2. Near the point A, the polynomial piece $q = q(x, y)$ for $x, y \leqslant 1/2$ may remain fixed while squares above A are continually refined. Unless f is exactly equal to q near A, the enforcement of continuity with q at A limits the accuracy of the approximations obtained above A.

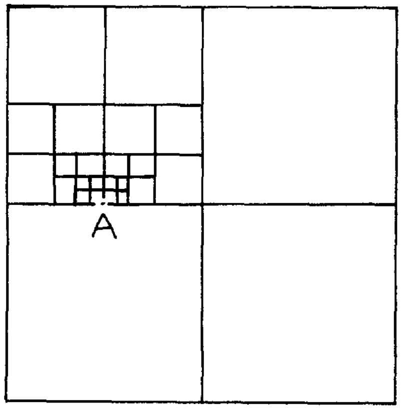

FIG. 2. A subdivision of the unit square by quadrisection.

There are four independent properties of an algorithm realization: localness, accuracy, smoothness and shape preservation (of the cells) which we desire. The only schemes we know which have all these properties are *blending function schemes* such as Coon's patches (see Barnhill [1] for a survey of such schemes in $\mathbb{R}^2$). One may interpret "blending function" to mean interpolation to the interior of a cell of data from *all* of the cell's boundary, i.e., the data functionals have values in $\mathbb{R}^{N-1}$. Thus, only in $\mathbb{R}^1$ does one obtain ordinary piecewise polynomial approximation, using local Hermite interpolation.

The analysis of Sections 1 and 2 applies directly to adaptive blending function approximations and we conjecture that these are the only realizations of our adaptive algorithm that produce smooth approximations in $\mathbb{R}^N$ for $N > 1$.

Our algorithm can be modified to include a constraint on the "generation gap" between neighboring cells (recall the terminology of "parent" of a cell introduced in a4). We say that a subdivision is *r-graded* if the difference in generations between neighboring cells is at most r. A 0-graded subdivision is uniform. One can easily construct situations where this constraint makes the subdivision of one cell the cause for subdivision of almost all the remaining cells. In general, a graded algorithm will produce a larger K than our algorithm does. We conjecture, however, that this constraint does not destroy the optimal rate of convergence obtained in Section 4. In fact, it seems plausible that (for almost all f) there is an r depending on f, D, E and the local approximation scheme, but independent of ϵ, so that all subdivisions produced by the algorithm are r-graded.

We close by noting that adaptive tensor product algorithms can be devised which preserve smoothness for piecewise polynomials, but they cannot achieve the optimal convergence rate.

Note added in proof. Ron DeVore has pointed out to us that adaptive algorithms of the kind we are considering here cannot be used to prove results like those at the end of Section 3, as the following simple example shows. Consider the hat function

$$f(x) := \begin{cases} 1 - |x|/\alpha, & |x| \leqslant \alpha \\ 0 & \text{, otherwise} \end{cases}$$

and take $p = \infty$, $n = 1$. Then the right side of (18a) becomes const/K regardless of α. Yet, to obtain an approximation from $\mathbb{P}_{1,K}$ to this f to within $<1/2$ requires cells of size α and, because of assumption $d1$, the algorithm reaches cells of such size only after $\ln\alpha/\ln\beta$ subdivisions, a number which goes to infinity as $\alpha \to 0$.

References

1. R. E. Barnhill, Representation and approximation of surfaces, *in* "Mathematical Software III" (J. R. Rice, Ed.), pp. 69–120, Academic Press, New York, 1977.
2. M. S. Birman and M. Z. Solomiak, Piecewise-polynomial approximations of functions of the classes W_p^α, *Mat. Sb.* **73** (1967), 295–317; transl. as *Math. USSR-Sb.* **2** (1967), 295–317.
3. Ju. A. Brudnyi, On a property of functions from the space H_p^λ, *Dokl. Akad. Nauk SSSR* **215** (1974); transl. as *Soviet Math. Dokl.* **15** (1974), 624–627.
4. Ju. A. Brudnyi, Piecewise polynomial approximation, embedding theorem and rational approximation, *in* "Approximation Theory, Bonn 1976" (R. Schaback and K. Scherer, Eds.), pp. 73–98, Lecture Notes in Mathematics No. 556, Springer–Verlag, Heidelberg, 1976.
5. H. G. Burchard, Splines (with optimal knots) are better, *J. Appl. Anal.* **3** (1973/1974), 309–319.
6. H. G. Burchard, On the degree of convergence of piecewise polynomial approximation on optimal meshes II, *Trans. Amer. Math. Soc.* **234** (1977), 531–559.
7. C. B. Morrey, Jr., "Multiple Integrals in the Calculus of Variations," Springer–Verlag, New York/Berlin, 1966.
8. J. R. Rice, On the degree of convergence of nonlinear spline approximation, *in* "Approximations with Special Emphasis on Spline Functions" (I. J. Schoenberg, Ed.), pp. 349–365, Academic Press, New York, 1969.
9. J. R. Rice, Adaptive approximation, *J. Approximation Theory* **16** (1976), 329–337.