DEGREE OF ADAPTIVE APPROXIMATION

RONALD A. DeVORE AND XIANG MING YU

ABSTRACT. We obtain various estimates for the error in adaptive approximation and also establish a relationship between adaptive approximation and free-knot spline approximation.

1. INTRODUCTION

Nonlinear methods of approximation are important because they enable us to approximate functions with singularities much better than do their linear counterparts. This speaks for the great interest in rational and free-knot spline approximation. On the other hand, adaptive approximation, which is a variant of free-knot spline approximation, is much more useful in applications of approximation to numerical computation. However, there are only a few error estimates for this type of approximation. The earliest results were obtained by Birman and Solomjak [3] who gave an estimate for the degree of adaptive approximation for functions in Sobolev spaces. Estimates of a different flavor were given by Rice [11] and in the multivariate case by de Boor and Rice [2]. More recently, with a simple application of the Hardy-Littlewood maximal function, DeVore [5] proved a limiting version of the Birman-Solomjak theorem.

In this paper, we obtain a variety of estimates for the degree of adaptive approximation in L_q, $0 < q \le \infty$. Our first results are for functions in the smoothness spaces C_p^α $(\alpha > 0,\ 0 < p \le \infty)$ which measure smoothness of order α in L_p through certain maximal functions. Since, in the case that α is an integer and $1 < p \le \infty$, they are closely related to the Sobolev spaces, our results are a generalization of the Birman-Solomjak theorem and DeVore's theorem. The advantage here, however, is that the smoothness of functions can be measured for all $\alpha > 0$ and in L_p spaces with $p < 1$. It is well understood that the description of approximation classes by nonlinear methods ultimately involves smoothness spaces in L_p, $p < 1$.

We shall also establish a relationship between adaptive approximation and free-knot spline approximation, which allows the estimation of the degree of

adaptive approximation by means of known results for the degree of free-knot spline approximation. Moreover, we prove for adaptive approximation an analogue of Burchard's theorem [4] for free-knot spline approximation.

2. Adaptive approximation

We begin with a brief description of what we mean by adaptive approximation. Other variants are possible and would produce similar results. Let $\Omega = [0, 1]^d$ be the unit cube in $\mathbb{R}^d$. If $f \in L_q(\Omega)$, $0 < q \leq \infty$, we denote by

$$E(Q) := E_r(f, Q)_q := \inf_{P \in \mathscr{P}_r} \|f - P\|_{L_q(Q)}$$

the local error in approximating f on the cube $Q \subseteq \Omega$ by elements of the space $\mathscr{P}_r$, the space of polynomials of degree $< r$. The following adaptive procedure constructs a piecewise polynomial which approximates f to a prescribed tolerance $\varepsilon > 0$. We say that a cube Q is *good* if $E(Q) \leq \varepsilon$; otherwise Q is called *bad*. We want to generate a partition $\mathscr{G}_\varepsilon$ of Ω into *good* cubes. If Ω is *good* then $\mathscr{G}_0 := \{\Omega\}$ is the desired partition. In this case, we define $\mathscr{B}_0 := \varnothing$. On the other hand, if Ω is *bad*, we let $\mathscr{G}_0 := \varnothing$ and $\mathscr{B}_0 := \{\Omega\}$. In this latter case, we proceed further and divide Ω into 2^d cubes of equal size and test these whether they are *good*, in which case they are put into the set $\mathscr{G}_1$, or *bad*, in which case they are put into the set $\mathscr{B}_1$. We continue in this way. At each step k we have a set of *good* cubes $\mathscr{G}_k$ and a set of *bad* cubes $\mathscr{B}_k$. We divide the cubes in $\mathscr{B}_k$ into their children and test whether these children Q are *good* or *bad* to obtain the sets $\mathscr{G}_{k+1}$ and $\mathscr{B}_{k+1}$. This process will terminate because $E(Q) \to 0$ as $|Q| \to 0$. We let $\mathscr{G}_\varepsilon := \bigcup_k \mathscr{G}_k$. Then $\mathscr{G}_\varepsilon$ forms a partition of Ω. For each $Q \in \mathscr{G}_\varepsilon$ we have a polynomial P_Q which is a best $L_q(Q)$ approximation to f from $\mathscr{P}_r$. The piecewise polynomial

$$S(x) := P_Q(x), \qquad x \in \text{int}(Q), \ Q \in \mathscr{G}_\varepsilon,$$

is the adaptive approximation to f, and the error of approximation of S to f is

$$\|f - S\|_{L_q(\Omega)} = \left(\sum_{Q \in \mathscr{G}_\varepsilon} \|f - P_Q\|_{L_q(Q)}^q \right)^{1/q} \leq |\mathscr{G}_\varepsilon|^{1/q} \varepsilon,$$

where $|\mathscr{G}_\varepsilon|$ is the number of cubes in $\mathscr{G}_\varepsilon$.

We would like to compare adaptive methods of approximation with other methods of approximation which depend on $O(n)$ parameters. Accordingly, we change our viewpoint slightly and ask, given n, how good of an error of approximation can we achieve by the adaptive method with the stipulation that the number of resulting intervals be $\leq n$. This leads us to define

(2.1) $$a_n(f)_q := \inf_{|\mathscr{G}_\varepsilon| \leq n} |\mathscr{G}_\varepsilon|^{1/q} \varepsilon.$$

It follows that for each n, the adaptive procedure generates a piecewise polynomial with n pieces which satisfies

(2.2) $$\|f - S\|_q \leq a_n(f)_q.$$

The results that follow will give estimates of $a_n(f)_q$ in terms of the smoothness of f. In contrast to linear approximation, the smoothness of f will be measured in spaces L_p with $p < 1$. It will be useful for the reader to keep in mind the Sobolev embedding theorem in order to understand our results. This theorem says that spaces defined by smoothness of order α in L_p embed into L_q provided $\alpha > (d/p - d/q)$.

3. ERROR ESTIMATES FOR FUNCTIONS IN C_p^α

We first recall the definition of the spaces C_p^α. The reader is referred to the monograph [9] of DeVore and Sharpley for a detailed study of these spaces. Let $f \in L_p(\Omega)$ $(0 < p \le \infty)$. If $\alpha > 0$, we let $r := [\alpha] + 1$. For a cube $Q \subset \Omega$, we denote by $P_Q f$ a best polynomial approximation to f of degree $< r$ in $L_p(Q)$. For $0 < p < \infty$, we define

$$f_{\alpha,p}^{\#}(x) := \sup_{Q \ni x} \frac{1}{|Q|^{\alpha/d}} \left(\frac{1}{|Q|} \int_Q |f - P_Q f|^p \right)^{1/p},$$

where the sup is taken over all cubes Q which contain x and are contained in Ω. For $0 < p < \infty$ we say that $f \in C_p^\alpha(\Omega)$ if $f_{\alpha,p}^{\#} \in L_p(\Omega)$, and define

$$|f|_{C_p^\alpha(\Omega)} := \|f_{\alpha,p}^{\#}\|_{L_p(\Omega)}.$$

The definition in [9] for these spaces uses $f_{\alpha,1}^{\#}$ for the definition in the case $p \ge 1$ but as was shown in [9] the definition we have used gives the same space with an equivalent norm. When $p = \infty$, the same definition with $f_{\alpha,q}^{\#}$ for any $q < \infty$ gives the space C_∞^α and its norm. These spaces are related to Sobolev spaces and Lipschitz spaces (see [9, §6]). For example, when $p = \infty$, $C_\infty^\alpha = \mathrm{Lip}\,\alpha$, $\alpha \ne$ integer.

For $g \in L_p(\Omega)$, we shall also use the Hardy-Littlewood maximal function:

$$(3.1) \qquad M_p(g)(x) := \sup_{Q \ni x} \left(\frac{1}{|Q|} \int_Q |g|^p \right)^{1/p},$$

where the sup is taken over all cubes Q which contain x and are contained in Ω.

Theorem 3.1. *Let* $0 < q < \infty$ *and* $\alpha > 0$, $p := (\alpha/d + 1/q)^{-1}$. *If* $f \in C_p^\alpha(\Omega)$ *and* $M_p(f_{\alpha,p}^{\#}) \in L_p(\Omega)$, *then*

$$(3.2) \qquad a_n(f)_q \le C n^{-\alpha/d} \|M_p(f_{\alpha,p}^{\#})\|_{L_p(\Omega)},$$

where C *is a constant independent of* f *and* n.

Proof. By the embedding theorems [9, Corollary 9.5], we know $f \in L_q(\Omega)$. From the definition (2.1) of $a_n(f)_q$, there is an $\varepsilon > 0$ such that $|\mathscr{G}_\varepsilon| \le n$ and

$$(3.3) \qquad a_n(f)_q \le 2|\mathscr{G}_\varepsilon|^{1/q}\varepsilon.$$

Suppose that $Q \in \mathscr{G}_\varepsilon$ is a *good* cube; then the parent $\widetilde{Q}$ of Q is a *bad* cube. Therefore, the error of best L_q approximation $E_r(f, \widetilde{Q})_q$ to f on $\widetilde{Q}$ by polynomials of degree $< r$ satisfies

$$E_r(f, \widetilde{Q})_q > \varepsilon.$$

On the other hand, from the definition of $f_{\alpha,q}^{\#}$, for any $x \in \widetilde{Q}$ we have

$$f_{\alpha,q}^{\#}(x) \geq \frac{1}{|\widetilde{Q}|^{\alpha/d}} \left(\frac{1}{|\widetilde{Q}|} \int_{\widetilde{Q}} |f - P_{\widetilde{Q}} f|^q \right)^{1/q} = |\widetilde{Q}|^{-(\alpha/d + 1/q)} E_r(f, \widetilde{Q})_q.$$

Since $|\widetilde{Q}| = 2^d |Q|$ and $Q \subset \widetilde{Q}$, we have

$$(3.4) \qquad \varepsilon \leq |\widetilde{Q}|^{(\alpha/d + 1/q)} f_{\alpha,q}^{\#}(x) \leq C |Q|^{(\alpha/d + 1/q)} f_{\alpha,q}^{\#}(x), \qquad x \in Q.$$

From the properties of functions $f_{\alpha,q}^{\#}$, we know (see Theorem 4.3 in [9]) that

$$f_{\alpha,q}^{\#}(x) \leq C M_p(f_{\alpha,p}^{\#})(x),$$

where $p = (\alpha/d + 1/q)^{-1}$. Hence, by (3.4), for any $x \in Q$ we have

$$\varepsilon \leq C |Q|^{1/p} M_p(f_{\alpha,p}^{\#})(x),$$

which gives

$$(3.5) \qquad \varepsilon \leq C \left(\int_Q |M_p(f_{\alpha,p}^{\#})(x)|^p \, dx \right)^{1/p}.$$

Because (3.5) is true for any $Q \in \mathscr{G}_\varepsilon$, we have

$$(3.6) \quad \begin{aligned} |\mathscr{G}_\varepsilon| \varepsilon^p &\leq C \sum_{Q \in \mathscr{G}_\varepsilon} \int_Q |M_p(f_{\alpha,p}^{\#})(x)|^p \, dx \\ &= C \int_\Omega |M_p(f_{\alpha,p}^{\#})(x)|^p \, dx = C \|M_p(f_{\alpha,p}^{\#})\|_{L_p(\Omega)}^p. \end{aligned}$$

Now, we take $\varepsilon := C^{1/p} n^{-1/p} \|M_p(f_{\alpha,p}^{\#})\|_{L_p(\Omega)}$ with C the constant of (3.6). Then (3.6) shows that $|\mathscr{G}_\varepsilon| \leq n$. We therefore obtain

$$\begin{aligned} a_n(f)_q &\leq C |\mathscr{G}_\varepsilon|^{1/q} \varepsilon \leq C n^{1/q} n^{-1/p} \|M_p(f_{\alpha,p}^{\#})\|_{L_p(\Omega)} \\ &= C n^{-\alpha/d} \|M_p(f_{\alpha,p}^{\#})\|_{L_p(\Omega)}. \quad \square \end{aligned}$$

We mention next some simple consequences of Theorem 3.1. Let $p' > p$. From the definition of $f_{\alpha,p}^{\#}(x)$ and Hölder's inequality we know

$$f_{\alpha,p}^{\#}(x) \leq f_{\alpha,p'}^{\#}(x).$$

Hence, $f \in C_{p'}^{\alpha}$ implies $f_{\alpha,p}^{\#} \in L_{p'}$, which gives $M_p(f_{\alpha,p}^{\#}) \in L_p$. Therefore, we have

Corollary 3.2. *Let* $0 < q < \infty$, $\alpha > 0$, *and* $p' > (\alpha/d + 1/q)^{-1}$. *Then for* $f \in C_{p'}^{\alpha}(\Omega)$ *we have*

$$(3.7) \qquad\qquad a_n(f)_q \leq Cn^{-\alpha/d}|f|_{C_{p'}^{\alpha}(\Omega)}.$$

For $0 < p, \sigma \leq \infty$ and $0 < \alpha < r$, the Besov space $B_{\sigma}^{\alpha}(L_p(\Omega))$ consists of all functions f in $L_p(\Omega)$ such that

$$(3.8) \qquad |f|_{B_{\sigma}^{\alpha}(L_p(\Omega))} := \begin{cases} (\int_0^{\infty}[t^{-\alpha}\omega_r(f,t)_p]^{\sigma}\frac{dt}{t})^{1/\sigma}, & 0 < \sigma < \infty, \\ \sup_{t \geq 0} t^{-\alpha}\omega_r(f,t)_p, & \sigma = \infty, \end{cases}$$

is finite. Here, $\omega_r(f,\cdot)_p$ is the modulus of smoothness of f as measured in $L_p(\Omega)$. By the embedding theorems for the C_p^{α} spaces and Besov spaces [9], we have

Corollary 3.3. *Let* $0 < q < \infty$, $\alpha > 0$, *and* $p' > (\alpha/d + 1/q)^{-1}$. *Then for* $f \in B_{\sigma}^{\alpha}(L_{p'}(\Omega))$, $0 < \sigma \leq \infty$, *we have*

$$(3.9) \qquad\qquad a_n(f)_q \leq Cn^{-\alpha/d}|f|_{B_{\sigma}^{\alpha}(L_{p'}(\Omega))}.$$

In fact, for $p' > (\alpha/d + 1/q)^{-1}$, we can choose p'' such that $p' > p'' > (\alpha/d + 1/q)^{-1}$. For $0 < \sigma \leq \infty$, $f \in B_{\sigma}^{\alpha}(L_{p'}(\Omega))$ implies $f \in B_{p''}^{\alpha}(L_{p''}(\Omega))$ and then by the embedding [9, Theorem 7.1], we have $f \in C_{p''}^{\alpha}$. Then, by Corollary 3.2, we obtain the conclusion.

4. Adaptive approximation and free-knot spline approximation

For further results, we restrict our attention to the univariate case. We shall give a comparison between adaptive and free-knot spline approximation. Let $\Sigma_n := \Sigma_{n,r}$ be the set of all piecewise polynomial functions S of order $\leq r$ with at most n pieces on I. If $f \in L_q(I)$, $I = [0,1]$, we let

$$(4.1) \qquad\qquad \sigma_n(f)_q := \inf_{S \in \Sigma_n} \|f - S\|_q$$

be the error of approximation of f by the elements of Σ_n. Throughout this section, all norms are on $\Omega = [0,1]$ unless otherwise noted. Obviously, we have $\sigma_n(f)_q \leq a_n(f)_q$. We next obtain an inverse inequality between adaptive approximation and free-knot spline approximation.

For this we shall use the Besov spaces

$$B^{\alpha} := B_{\lambda}^{\alpha}(L_{\lambda}),$$

where $\lambda := \lambda(\alpha, p) := (\alpha + 1/p)^{-1}$. The spaces B^{α} are embedded into C_{λ}^{α} and then into L_p provided $0 < p < \infty$ (see [9, §7]). Moreover (see [7]), for any

interval $J \subseteq I$ we have

$$(4.2) \qquad\qquad E_r(f, J)_p \leq C|f|_{B^\alpha(J)}, \qquad 0 < \alpha < r,$$

where the constant C depends only on r and p.

The importance of the spaces B^α in nonlinear approximation is that sometimes we can precisely characterize the approximation order of the elements of B^α. For example, we have the following theorem of DeVore and Popov [8] (using heavily the results of Petrushev [10]):

Theorem 4.1. *A function f is in B^α, $\alpha < r$, if and only if*

$$(4.3) \qquad\qquad \left(\sum_{k=1}^{\infty} k^{\alpha\lambda-1} \sigma_k(f)_p^\lambda\right)^{1/\lambda}$$

is finite, and (4.3) *is an equivalent semi-(quasi)norm for B^α.*

A main step in the proof of this theorem is the inverse inequality of Petrushev [10], which says that for each $S_n \in \Sigma_n$:

$$(4.4) \qquad\qquad |S_n|_{B^\alpha} \leq C n^\alpha \|S_n\|_p.$$

As a special case of Corollary 3.3 and (4.3), we obtain

$$(4.5) \qquad a_n(f)_q \leq C n^{-\alpha} \left(\sum_{k=1}^{\infty} k^{\alpha\lambda-1} \sigma_k(f)_p^\lambda\right)^{1/\lambda}, \qquad p > q.$$

The results that follow can be considered an improvement of this inequality.

We begin with the following simple lemma.

Lemma 4.2. *Let $0 < p < \infty$, $\alpha > 0$, and $\lambda := \lambda(\alpha, p) := (\alpha + 1/p)^{-1}$. Then for any collection $\{I_j\}_{j=1}^m$ of pairwise disjoint intervals we have*

$$\left(\sum_{j=1}^{m} E_r(f, I_j)_p^\lambda\right)^{1/\lambda} \leq C \left(\sum_{k=1}^{m} k^{\alpha\lambda-1} \sigma_k(f)_p^\lambda\right)^{1/\lambda}.$$

Proof. Let S_m be a best approximation to f from Σ_m, $\|f - S_m\|_p = \sigma_m(f)_p$. From (4.2), we have

$$(4.6) \qquad \sum_{j=1}^{m} E_r(S_m, I_j)_p^\lambda \leq \sum_{j=1}^{m} |S_m|_{B^\alpha(I_j)}^\lambda \leq C^\lambda |S_m|_{B^\alpha}^\lambda.$$

Here we have used the fact that as a set function (on intervals), B^α has an equivalent seminorm $|\cdot|$ for which $|\cdot|^\lambda$ is subadditive. Now, $\sigma_k(S_m)_p = 0$, $k \geq m$, and $\sigma_k(S_m)_p \leq \sigma_k(f)_p + \|f - S_m\|_p \leq 2\sigma_k(f)_p$, $k \leq m$. Hence, from Theorem 4.1 we obtain

$$(4.7) \qquad\qquad |S_m|_{B^\alpha}^\lambda \leq C^\lambda \sum_{k=1}^{m} k^{\alpha\lambda-1} \sigma_k(f)_p^\lambda.$$

Therefore,

$$(4.8) \qquad \sum_{j=1}^{m} E_r(f, I_j)_p^{\lambda} \leq C^{\lambda} \left\{ \sum_{j=1}^{m} E_r(S_m, I_j)_p^{\lambda} + \sum_{j=1}^{m} \|f - S_m\|_p^{\lambda}(I_j) \right\}.$$

The first sum can be estimated by (4.7), and by Hölder's inequality the second sum on the right does not exceed $[m^{1/\lambda - 1/p} \sigma_m(f)_p]^{\lambda} = m^{\alpha\lambda} \sigma_m(f)_p^{\lambda}$. $\square$

For our next theorem, we need to make some further simple observations about the adaptive procedure. We let $N(\varepsilon) := N(\varepsilon, f) := |\mathscr{G}_{\varepsilon}(f)|$. Then $N(\varepsilon)$ is a nonincreasing, integer-valued function which is continuous from the right. Indeed, if $\varepsilon > 0$ and $I \in \mathscr{G}_{\varepsilon}$, then the parent $\tilde{I}$ of I is a bad interval and therefore $E(\tilde{I}) := E_r(f, \tilde{I})_q > \varepsilon$. We choose $\eta > \varepsilon$ so small that $E(\tilde{I}) > \eta$ for all $I \in \mathscr{G}_{\varepsilon}$. Then, $\mathscr{G}_{\eta} = \mathscr{G}_{\varepsilon}$ and therefore $N(\eta) = N(\varepsilon)$. Hence, N is constant on $[\varepsilon, \eta]$, which shows the right continuity of N.

Now suppose that we are given n; then the above remarks show that there is an ε such that $a_n(f)_q = N(\varepsilon)^{1/q} \varepsilon$. It may happen that $N(\varepsilon) < n/2$. In this case, we shall construct a second partition $\widetilde{\mathscr{G}}_{\varepsilon}$ which is obtained from $\mathscr{G}_{\varepsilon}$ by further subdivision and satisfies $n/2 \leq |\widetilde{\mathscr{G}}_{\varepsilon}| \leq n$. If $|\mathscr{G}_{\varepsilon}| \geq n/2$, then we simply take $\widetilde{\mathscr{G}}_{\varepsilon} = \mathscr{G}_{\varepsilon}$. Otherwise, we first observe that there are always intervals $I \in \mathscr{G}_{\varepsilon}$ for which $E(I) = \varepsilon$, since otherwise we could decrease ε slightly without changing $N(\varepsilon)$ and thereby contradict the definition of $a_n(f)_q$. Now, we subdivide all intervals $I \in \mathscr{G}_{\varepsilon}$ for which $E(I) = \varepsilon$. We continue in this way by further dividing any resulting cubes which satisfy our criterion $E(I) = \varepsilon$. We stop this process at the first stage k (i.e., when we are processing cubes of sidelength 2^{-k}), where the total number of intervals, i.e., those in $\mathscr{G}_{\varepsilon}$ and those just constructed, number $n/2$ or more. It may happen that there are no intervals left with $E(I) = \varepsilon$ and still the total number of cubes is not $\geq n/2$. In this case, if ε' is the largest value of $E(I)$ for our current set of cubes, then the current set of cubes is $\mathscr{G}_{\varepsilon'}$. We can therefore repeat our procedure for ε'. If we continue in this way, we shall eventually arrive at a partition, which we denote by $\widetilde{\mathscr{G}}_{\varepsilon}$, which satisfies $n/2 \leq |\widetilde{\mathscr{G}}_{\varepsilon}| \leq n$, and for some $\varepsilon_0 \leq \varepsilon$ all the intervals $I \in \widetilde{\mathscr{G}}_{\varepsilon}$ satisfy $E(I) \leq \varepsilon_0$ and all their parents satisfy $E(I) \geq \varepsilon_0$. As usual, we call the cubes in $\widetilde{\mathscr{G}}_{\varepsilon}$ good cubes and denote by $\widetilde{\mathscr{G}}_k$ the good cubes of sidelength 2^{-k}. The parents of these good cubes are bad cubes. We denote by $\widetilde{\mathscr{B}}_k$ the bad cubes of sidelength 2^{-k}. We also note that

$$(4.9) \qquad a_n(f)_q = \varepsilon |\mathscr{G}_{\varepsilon}|^{1/q} \leq \varepsilon_0 |\mathscr{G}_{\varepsilon_0}|^{1/q} \leq \varepsilon_0 |\widetilde{\mathscr{G}}_{\varepsilon}|^{1/q}$$

because $|\mathscr{G}_{\varepsilon_0}| \leq n$ and $|\mathscr{G}_{\varepsilon_0}| \leq |\widetilde{\mathscr{G}}_{\varepsilon}|$.

We can now prove the following.

Theorem 4.3. *Let* $\alpha > 0$, $0 < q < p < \infty$, *and* $\lambda = (\alpha + 1/p)^{-1}$. *Then for* $f \in L_p$ *we have*

$$a_n(f)_q \leq C n^{-\alpha} \left(\sum_{j=1}^{n} j^{\alpha\lambda - 1} \sigma_j(f)_p^{\lambda} \right)^{1/\lambda}, \qquad n = 1, 2, \dots,$$

where C *is independent of* f *and* n.

Proof. Let n be given and let $\varepsilon > 0$ be chosen so that $|\mathscr{G}_\varepsilon|^{1/q} \varepsilon = \alpha_n(f)_q$, and let $\widetilde{\mathscr{G}_\varepsilon}$ be the associated set of intervals described above and let $\varepsilon_0 \leq \varepsilon$ be as above. Then for any interval $I \in \widetilde{\mathscr{B}_k}$ we have

$$E_r(f, I)_q \geq \varepsilon_0$$

and $|I| = 2^{-k}$. The Hölder inequality gives

$$E_r(f, I)_q \leq E_r(f, I)_p |I|^{1/q - 1/p}.$$

Therefore, from the above information and Lemma 4.2 we have

$$|\widetilde{\mathscr{B}_k}| \varepsilon_0^\lambda \leq \sum_{I \in \widetilde{\mathscr{B}_k}} E_r(f, I)_q^\lambda \leq 2^{-k(1/q - 1/p)\lambda} \sum_{I \in \widetilde{\mathscr{B}_k}} E_r(f, I)_p^\lambda$$

(4.10)

$$\leq C 2^{-k(1/q - 1/p)\lambda} \sum_{j=1}^{|\widetilde{\mathscr{B}_k}|} j^{\alpha\lambda - 1} \sigma_j(f)_p^\lambda \leq C 2^{-k(1/q - 1/p)\lambda} A,$$

where $A := \sum_{j=1}^{n} j^{\alpha\lambda - 1} \sigma_j(f)_p^\lambda$. Here, for the last inequality we have used the fact that $|\widetilde{\mathscr{B}_k}| \leq |\widetilde{\mathscr{G}_\varepsilon}| \leq n$.

We also have $|\widetilde{\mathscr{G}_k}| \leq 2^k$ and $|\widetilde{\mathscr{G}_k}| \leq 2|\widetilde{\mathscr{B}_{k-1}}|$. Hence, from these two inequalities and (4.10), we obtain for any positive integer m,

$$|\widetilde{\mathscr{G}_\varepsilon}| \leq \sum_{k=0}^{m-1} 2^k + C \varepsilon_0^{-\lambda} A \sum_{k=m}^{\infty} 2^{-k(1/q - 1/p)\lambda}$$

(4.11)

$$\leq C \{ 2^m + \varepsilon_0^{-\lambda} A 2^{-m(1/q - 1/p)\lambda} \}.$$

We want to choose m so that the two terms on the right side of (4.11) are about equal; this can be accomplished with m the smallest integer such that

$$2^m \geq \varepsilon_0^{-\lambda} A 2^{-m(1/q - 1/p)\lambda}.$$

Then (4.11) gives

(4.12) $$|\widetilde{\mathscr{G}_\varepsilon}|^{1 + (1/q - 1/p)\lambda} \leq C 2^{m(1 + (1/q - 1/p)\lambda)} \leq C \varepsilon_0^{-\lambda} A$$

and

(4.13) $$\varepsilon_0 \leq C |\widetilde{\mathscr{G}_\varepsilon}|^{-1/\lambda + 1/p - 1/q} A^{1/\lambda} \leq C n^{-\alpha - 1/q} \left(\sum_{j=1}^{n} j^{\alpha\lambda - 1} \sigma_j(f)_p^\lambda \right)^{1/\lambda},$$

where we used the fact that $n/2 \leq |\widetilde{\mathscr{G}}_\varepsilon| \leq n$. From this last inequality and (4.9) we obtain our conclusion:

$$a_n(f)_q \leq |\widetilde{\mathscr{G}}_\varepsilon|^{1/q} \varepsilon_0 \leq C n^{-\alpha} \left(\sum_{j=1}^n j^{\alpha\lambda-1} \sigma_j(f)_p^\lambda \right)^{1/\lambda}. \quad \square$$

To apply Theorem 4.3, it is enough to use discrete Hardy inequalities. For example, it follows from this theorem that if $\sigma_n(f)_p = O(n^{-\beta})$ for some $\beta > 0$, then for all $q < p$, $a_n(f)_q = O(n^{-\beta})$. Indeed, it is enough to take an $\alpha > \beta$ and apply Theorem 4.3. More generally, if ψ is a nondecreasing continuous function on $[0, 1]$ with $\psi(0) = 0$ and

$$t^\alpha \left(\int_t^1 \frac{\psi(u)^\lambda}{u^{\alpha\lambda+1}} \, du \right)^{1/\lambda} = O(\psi(t)), \qquad t > 0,$$

then $\sigma_n(f)_p = O(\psi(n^{-1}))$ implies $a_n(f)_q = O(\psi(n^{-1}))$ provided $q < p$.

5. Adaptive approximation in L_∞

We consider the case $q = \infty$. For free-knot spline approximation, Burchard [4] proved

Theorem A. *Suppose that $f^{(r-1)} \in C(0, 1)$ and $|f^{(r)}(x)| \leq \phi(x)$, where $\phi \in L_{1/r}$ is a monotone function. Then we have*

$$(5.1) \qquad \sigma_n(f)_\infty \leq C n^{-r} \|\phi\|_{1/r}.$$

The proof of this theorem (see de Boor [1]) depends on the estimate

$$(5.2) \qquad E_r(f, I)_\infty \leq C_0 \left(\int_I \phi(x)^{1/r} \, dx \right)^r$$

for any $I \subset (0, 1)$. We can prove an analogous estimate to (5.1) for adaptive approximation, but we shall need a further restriction on ϕ. We assume that

$$(5.3) \qquad \int_I \phi(x)^{1/r} \, dx \leq C_1/n \quad \text{if } |I| \leq 2^{-n},$$

where C_1 is an absolute constant.

Theorem 5.1. *Suppose that $f^{(r-1)} \in C(0, 1)$ and $|f^{(r)}(x)| \leq \phi(x)$, where $\phi \in L_{1/r}$ is a monotone function satisfying condition (5.3). Then we have*

$$(5.4) \qquad a_n(f)_\infty \leq C n^{-r} \|\phi\|_{1/r}.$$

Proof. We may assume that ϕ is increasing and $\|\phi\|_{1/r} = 1$. We take $\varepsilon := C_0 C_1^r n^{-r}$ with C_0 the constant of (5.2) and C_1 the constant of (5.3). We apply the adaptive process using the right side of (5.2) in place of $E_r(f, I)_\infty$ and obtain an adaptive partition. We enumerate the intervals $I_1, I_2, \ldots, I_N$, $N := |\mathscr{G}_\varepsilon|$, of $\mathscr{G}_\varepsilon$ in order from left to right. Since ϕ is increasing, we have

$$(5.5) \qquad |I_j| \geq |I_{j+1}|, \qquad j = 1, \ldots, N-1.$$

In order to prove (5.4), we need only show that $N \le Cn$. We write $\mathscr{G}_\varepsilon$ as a union of subsets S_i, $i = 1, \ldots, m$, where S_i is a collection of all the intervals of $\mathscr{G}_\varepsilon$ with a given common length. By (5.5) we have

$$S_1 = \{I_j\}_{j=1}^{i_1}, \qquad S_2 = \{I_j\}_{j=i_1+1}^{i_2}, \ldots, S_m = \{I_j\}_{j=i_{m-1}+1}^{N}.$$

From (5.2) and (5.3), there are no intervals of length $< 2^{-n}$ and therefore $m \le n$. It is therefore enough to count the intervals in the S_i which contain three or more intervals.

Denote by $|S_i|$ the number of intervals in S_i. Since all the intervals in S_i have the same length, they are generated at the same step k of the adaptive process. Hence, except for one or two of these intervals, we can separate them into pairs whose parent is an interval $\tilde{I} \in \mathscr{B}_{k-1}$. Then, since $E_r(f, \tilde{I})_\infty > \varepsilon$, (5.2) gives $\int_{\tilde{I}} |\phi(x)|^{1/r} dx \ge C_1/n$. Since the number of these pairs is $\ge (|S_i| - 2)/2$, for $|S_i| \ge 3$ we have

$$\sum_{I \in S_i} \int_I |\phi(x)|^{1/r} dx \ge \frac{|S_i| - 2}{2} \cdot \frac{C_1}{n} \ge \frac{|S_i|}{6n} \cdot C_1.$$

From this inequality we obtain

$$\sum_{|S_i| \ge 3} |S_i| \le \frac{6n}{C_1} \sum_i \sum_{I \in S_i} \int_I |\phi(x)|^{1/r} dx \le \frac{6n}{C_1},$$

which completes our proof. □

For example, the functions $f(x) = x^\alpha$ ($\alpha > 0$) and $f(x) = [\log x/2]^{-2r+1}$ can be approximated with order $O(n^{-r})$ in the uniform norm by the adaptive process.

Bibliography

1. C. de Boor, *Good approximation by splines with variable knots*, Spline Functions and Approximation Theory (A. Meir and A. Sharma, eds.), Birkhäuser, 1973, pp. 57–72.

2. C. de Boor and J. R. Rice, *An adaptive algorithm for multivariate approximation giving optimal convergence rates*, J. Approx. Theory **25** (1979), 337–359.

3. M. Š. Birman and M. Z. Solomjak, *Piecewise polynomial approximation of functions of the class W_p^α*, Math. USSR-Sb. **2** (1967), 295–317.

4. H. G. Burchard and D. F. Hale, *Piecewise polynomial approximation on optimal meshes*, J. Approx. Theory **14** (1975), 128–147.

5. R. A. DeVore, *A note on adaptive approximation*, Approx. Theory Appl. **3** (1987), 74–78.

6. ____, *Degree of nonlinear approximation*, in Approximation Theory 4 (C. K. Chui, L. L. Schumaker, and J. D. Ward, eds.), Academic Press, 1990, pp. 175–201.

7. R. A. DeVore and V. Popov, *Interpolation of Besov spaces*, Trans. Amer. Math. Soc. **305** (1988), 397–414.

8. ____, *Interpolation spaces and non-linear approximation*, Function Spaces and Applications (M. Cwikel, J. Peetre, Y. Sagher, and H. Wallin, eds.), Lecture Notes in Math., vol. 1302, Springer, 1988, pp. 191–205.

9. R. A. DeVore and R. C. Sharpley, *Maximal functions measuring smoothness*, Mem. Amer. Math. Soc. No. 293 (1984).

10. P. P. Petrushev, *Direct and converse theorems for spline and rational approximation and Besov spaces*, Function Spaces and Applications (M. Cwikel, J. Peetre, Y. Sagher, and H. Wallin, eds.), Lecture Notes in Math., vol. 1302, Springer, 1988, pp. 363–377.

11. J. R. Rice, *Adaptive approximation*, J. Approx. Theory **16** (1976), 329–337.

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF SOUTH CAROLINA, COLUMBIA, SOUTH CAROLINA 29208

DEPARTMENT OF MATHEMATICS, NANJING NORMAL UNIVERSITY, PEOPLE'S REPUBLIC OF CHINA