JOURNAL OF APPROXIMATION THEORY 42, 173–180 (1984)

Best Approximation in L^∞ via Iterative Hilbert Space Procedures

MOHAMED S. BANI AND BRUCE L. CHALMERS

Department of Mathematics, University of California, Riverside, California 92521, U.S.A.

Communicated by Oved Shisha

Received October 5, 1981, revised January 25, 1984

In this note it is indicated that the problem of best approximation with respect to the supremum (L^∞) norm may be solvable by iterative Hilbert space techniques. The validity of this approach for L^p, p (even) $< \infty$, has been established by L. A. Karlovitz (*J. Approx. Theory* **3** (1970), 123–127). In the Haar case (L^∞ norm) this approach yields a convergent algorithm which is a slight modification of the Remez algorithm. © 1984 Academic Press, Inc.

1. INTRODUCTION AND PRELIMINARIES

Consider the problem of approximating an element x in a Banach space X, with norm $\| \ \|$, by an element in a finite dimensional subspace V. In fact consider $X \subset L^p(Q)$, $1 \leqslant p \leqslant \infty$, where $L^\infty = C$ (the continuous functions) and Q is compact. Here $\|x\| = \|x\|_{L^p} = (\int_Q |x|^p \, dQ)^{1/p}$, where dQ denotes integration with respect to a nonnegative measure on Q which is compatible with the L^∞ norm, i.e., $\|x\|_{L^p} \to \|x\|_{L^\infty}$, $\forall x \in C(Q)$. (An arbitrary X can be identified with a subspace of $C(Q)$, where Q is the set of extreme points of the ball of the dual.)

We first make some well-known observations about the best approximation operator B_p in this setting. For all $1 \leqslant p \leqslant \infty$, insight into $B_p: X \to V$ is gained by considering B_p as a "perturbation" of B_2, the completely understood Hilbert space setting. In fact, for $1 < p < \infty$, $x - B_p x$ is completely characterized by being *p-orthogonal* to V, i.e.,

$$\int_Q |x - B_p x|^{p-1} \operatorname{sgn}(x - B_p x) \, v \, dQ = 0, \qquad \forall v \in V. \tag{1}$$

Note that (1) can be rewritten by use of the duality operator J_p ($J_p x = |x|^{p-1} \operatorname{sgn} x$ is the (unnormalized) extremal function in L^q for $x \in L^p$), where $\langle x, y \rangle = \int_Q xy \, dQ$:

$$\langle J_p(x - B_p x), v \rangle = 0, \qquad \forall v \in V. \tag{2}$$

0021-9045/84 $3.00

A complete characterization for $B_\infty x$ (in general, set-valued) is obtained from (1) by taking a limit as $p \to \infty$ (see Theorem 1 of Sect. 2) to obtain the well known "0 in the convex hull"-criterion for best approximation:

$$\sum_{t_\alpha \in \mathscr{C}} \lambda_\alpha \operatorname{sgn}[(x - B_\infty x)(t_\alpha)]\, v(t_\alpha) = 0, \qquad \forall v \in V, \tag{3}$$

where $\lambda_\alpha \geqslant 0$, $\sum \lambda_\alpha = 1$, and $\mathscr{C} = \{t \in Q: |(x - B_\infty x)(t)| = \|x - B_\infty x\|_{L^\infty}\}$, the so-called critical set of the error.

Note 1. By the Carathéodory theorem (see [4, p. 17]), for $\mathscr{C}$ as in (3) we can take a minimal $\mathscr{C}_* = \mathscr{C} \cap \{t_\alpha; \lambda_\alpha \neq 0\}$ to have no more than $n+1$ points.

Note 2. Given any set $\mathscr{C}$, the possible dependencies

$$\sum_{t_\alpha \in \mathscr{C}} \gamma_\alpha e_{t_\alpha} = 0, \qquad \sum |\gamma_\alpha| = 1, \tag{3$'$}$$

of point evaluations e_{t_α} ($e_t(x) = x(t)$) which vanish on V are determined linear-algebraically. Thus given a set $\mathscr{C}_*$ (see Note 1), the γ_α are unique (up to a sign (σ)): If also $\sum_{t_\alpha \in \mathscr{C}_*} \gamma'_\alpha e_{t_\alpha} = 0$, then $\sum [\lambda_\alpha \operatorname{sgn}[(x - B_\infty x)(t_\alpha)] - \lambda \gamma'_\alpha]\, e_{t_\alpha} = 0$, where $|\lambda|^{-1} = \max |\gamma'_\alpha|/\lambda_\alpha$, and this contradicts the minimality of $\mathscr{C}_*$. Therefore $\lambda_\alpha = |\gamma_\alpha|$ and $\sigma \operatorname{sgn} \gamma_\alpha = \operatorname{sgn}[(x - B_\infty x)(t_\alpha)]$ are algebraic consequences and hence independent of x and $B_\infty x$.

Furthermore we can let $p \to 1$ in (1) to get criteria for $B_1 x$ (provided $x - B_1 x$ does not vanish on more than a set of measure 0). In particular if $[x, V]$ is an $(n+1)$-dimensional Haar space (see Sect. 2 for definition) on $Q = [a, b]$, (1) yields immediately the existence of the Hobby–Rice points [5] $\{t_i\}_{i=1}^n$ (independent of x) such that ($t_0 = a$, $t_{n+1} = b$):

$$\sum_{i=1}^{n+1} (-1)^i \int_{t_{i-1}}^{t_i} v = 0, \qquad \forall v \in V, \tag{4}$$

and $B_1 x$ is the unique element of V interpolating x at $\{t_i\}_{i=1}^n$.

For $1 < p < \infty$ an iterative algorithm for finding $B_p x$ arises from considering (1) at each stage of the algorithm as a statement about weighted Hilbert space. That is, note that (1) can be written $\int_Q (x - B_p x)\, vw\, dQ = 0$, $\forall v \in V$, where $w = |x - B_p x|^{p-2}$. Then, given an estimate v_ν for $B_p x$, let $w = |x - v_\nu|^{p-2}$ and generate $v_{\nu+1}$ by solving $\int_Q (x - v_{\nu+1})\, vw\, dQ = 0$, $\forall v \in V$, by the usual Gram–Schmidt process for finding best (weighted) Hilbert space approximations. Karlovitz [6] formulated and investigated this algorithm for $\infty > p \geqslant 2$ (actually only p even) and showed that it converges provided $v_{\nu+1}$ is mollified appropriately, i.e., at each stage $v_{\nu+1}$ is replaced by $\lambda_\nu v_{\nu+1} + (1 - \lambda_\nu) v_\nu$ for some (easily determined) λ_ν. In fact λ_ν yields the infimum of the one parameter convex functional $d(\lambda) = \|x - [\lambda v_{\nu+1} + (1-\lambda) v_\nu]\|$.

In [1] Bani extends Karlovitz' result to all real $p \geqslant 2$ and, if $[x, V]$ is an extended Chebyshev system (see, e.g., [1] for definition), to all $1 < p < 2$. Moreover, in the case $[x, V]$ is extended Chebyshev on $[a, b]$, by letting $p \to 1$ an algorithm is established for finding the Hobby–Rice points.

In the present note the authors obtain the "0 in the convex hull"-criterion (3) for best L^{∞} approximation as a weak* limit as $p \to \infty$ of criterion (1) for best L^p-approximation. This suggests that the Karlovitz algorithm itself has a limiting interpretation for the case $p = \infty$. Indeed for $p = \infty$ and V Haar, (3) gives rise to the well-known Remez algorithm which we will show can be viewed as a limit of the Karlovitz algorithm. In fact since the Karlovitz algorithm is an iterative (weighted) Hilbert space algorithm, we will demonstrate that the interpolation performed to find the best approximation on $n+1$ points at each stage of the Remez (single or multiple) exchange algorithm can be replaced by the standard Hilbert space technique for finding a best (weighted) L^2-approximation on $n+1$ points. Finally, because of the aforementioned results and since for $2 \leqslant p < \infty$ no Haar assumption is necessary for the convergence of the Karlovitz algorithm, it is expected that also for $p = \infty$ with no Haar assumption, a Remez exchange-type algorithm can be developed where the interpolation step is replaced by a (weighted) Hilbert space procedure. Analogously it is also expected that this (weighted) Hilbert space approach can be applied to constrained approximation in L^{∞} in the absence of Haar conditions just as it has been shown to apply in L^p, $p < \infty$ [3]. The application of this technique in the absence of Haar conditions in L^{∞} will be the subject of future investigations.

2. Main Results

Let B denote the general best approximation operator. Duality theory provides the existence of an important (Hahn–Banach) *separating functional* x^* in the dual X^* of X such that

(i) $x^*(v) = 0$, $\forall v \in V$,

(ii) $\|x^*\|_{X^*} = 1$,

(iii) $x^*(x) = \|x - Bx\|$.

LEMMA 1. *If* $X = L^p(Q)$, $1 < p < \infty$, *then* $x^* = x_p^* =$

$$x_p^*(\cdot) = \int_Q \frac{|x - B_p x|^{p-1} \operatorname{sgn}(x - B_p x)}{\|x - B_p x\|_{L^p}^{p-1}} (\cdot)\, dQ. \tag{5}$$

Proof. This follows immediately via an easy calculation from (1) and the fact that $(L^p)^* = L^q$, where $1/q + 1/p = 1$. ∎

Note 3. In formula (5) and in (6) below we will identify x_p^* with its Riesz representation in L^q and C^*, respectively.

Knowing just the general form of x^* is often sufficient to motivate the formulation of and to guide the proof of a convergent algorithm providing Bx.

LEMMA 2. *If $X = C(Q)$, then $x^* = x_\infty^* =$*

$$x_\infty^*(\cdot) = \sum_{t_\alpha \in \mathscr{C}} \lambda_\alpha \operatorname{sgn}[(x - B_\infty x)(t_\alpha)](\cdot)(t_\alpha), \tag{6}$$

where $\lambda_\alpha \geqslant 0$, $\sum \lambda_\alpha = 1$, and $\mathscr{C}$ is as in (3).

Proof. This follows immediately from (3). ∎

Observe that formally (6) is a limit of (5) as $p \to \infty$. That is, suppose $B_p x \to B_\infty x$ (certainly this holds for some subsequence and some one of $B_\infty x$ (recall B_∞ is in general set-valued)). Then $(|x - B_p x|/\|x - B_p x\|_{L^p})^{p-1}$ tends to 0 off $\mathscr{C}$ and our heuristic conclusion holds. We make this statement rigorous in Theorem 1 below. First recall

LEMMA 3 (Alaoglu's theorem). *A bounded sphere S^* of X^* is compact in the weak* topology.*

THEOREM 1. *Every subsequence of x_p^* has in turn a subsequence converging weak* to an x_∞^* in $C(Q)^*$.*

Proof. From Lemma 1 and Note 3, for $1 < p < \infty$, $x_p^* = |x - B_p x|^{p-1} \operatorname{sgn}(x - B_p x)/\|x - B_p x\|_{L^p}^{p-1}$ is unique since $B_p x$ is unique and by the well-known Polyá result (see [4, p. 42; and 7, p. 8]) every subsequence of $B_p x$ has a subsequence $B_{p_k} x$ converging uniformly to some $B_\infty x$. Now assume without loss that $\int_Q dQ = 1$; then $\|x_p^*\|_{C^*} = \|x_p^*\|_{L^1} \leqslant \|x_p^*\|_{L^q} = 1$ and thus by Lemma 3, $x_{p_k}^*$ has a subsequence which we will again refer to by $x_{p_k}^*$ which converges weak* to some $z^* \in C(Q)^*$. We must check that z^* satisfies properties (i–iii) for a separating functional for x. First $x_p^*(v) = 0$, $\forall v \in V$, shows that $z^*(v) = 0$, $\forall v \in V$. Next we show that $\|z^*\|_{C^*} = 1$. To see this, note that $y_p = x - B_p x/\|x - B_p x\|_{L^p}$ is the extremal for x_p^* (i.e., $\|y_p\|_{L^p} = 1$ and $x_p^*(y_p) = 1 = \|x_p^*\|_{L^q}$). Further $y_{p_k} \to y_\infty = x - B_\infty x/\|x - B_\infty x\|_{L^\infty}$ uniformly in L^∞. We claim that y_∞ is an extremal for z^*. Note that $1 = \|y\|_{L^\infty} \geqslant \|y\|_{L^p}$ implies that $x_p^*(y) \leqslant 1$. Hence $x_{p_k}^*(y) \to z^*(y) \leqslant 1$ and so $\|z^*\|_{C^*} \leqslant 1$. On the other hand given $\varepsilon > 0$, there exists k_0 such that $k > k_0$ implies $z^*(y_\infty) \geqslant x_{p_k}^*(y_\infty) - \varepsilon$; but further there is an l_0 such that $l > l_0$ implies $\|y_{p_l} - y_\infty\|_{L^\infty} < \varepsilon$ and thus $|x_{p_k}^*(y_\infty) - x_{p_k}^*(y_{p_l})| \leqslant$

$\|x_{p_k}^*\|_{L^{q_k}} \|y_\infty - y_{p_l}\|_{L^{p_k}} \leqslant \|y_\infty - y_{p_l}\|_{L^\infty} < \varepsilon$. We have therefore that $z^*(y_\infty) \geqslant x_{p_k}^*(y_{p_l}) - 2\varepsilon$ for all $k > k_0$ and $l > l_0$; in particular, $z^*(y_\infty) \geqslant x_{p_k}^*(y_{p_k}) - 2\varepsilon$ for $k > \max(k_0, l_0)$. That is, $z^*(y_\infty) \geqslant 1 - 2\varepsilon$ and we conclude that $\|z^*\|_{C^*} = 1$. Finally, $z^*(x) = \lim_{k\to\infty} x_{p_k}^*(x) = \lim_{k\to\infty} \|x - B_{p_k}x\|_{L^{p_k}} = \|x - B_\infty x\|_{L^\infty}$. Hence we can write $z^* = x_\infty^*$. ∎

COROLLARY 1. *If x_∞^* is unique then x_p^* converges weak* to x_∞^*.*

EXAMPLE. If V is Haar (see below for definition) and $x - B_\infty x$ has exactly $n+1$ critical points (which always occurs if, e.g., $[x, V]$ is Haar), then x_∞^* is unique.

DEFINITION. V is said to be *Haar* on Q iff no nonzero member of v of V has more than $n-1$ zeros. Equivalently, V is Haar iff any n point evaluation functionals $\{e_{t_i}\}_{i=1}^n$ in $C(Q)^*$ are independent over V.

Note 4. It is well known that if V is Haar then $B_\infty x$ is unique.

Motivation. Theorem 1, Corollary 1, and Lemmas 1 and 2 assert that (at least some subsequence of) $|x - B_p x|^{p-1} \operatorname{sgn}(x - B_p x)/\|x - B_p x\|_{L^p}^{p-1}$ converges weak* in $C(Q)^*$ to the distribution function $\sum_{t_\alpha \in \mathscr{C}} \lambda_\alpha \operatorname{sgn}[(x - B_\infty x)(t_\alpha)]\, \delta_{t_\alpha}$, where as usual $\delta_{t_\alpha}(t) = 1$ if $t = t_\alpha$ and $=0$, otherwise. Now we observe that $(x - B_\infty x)(t_\alpha) = |(x - B_\infty x)(t_\alpha)| \operatorname{sgn}[(x - B_\infty x)(t_\alpha)] = \|x - B_\infty x\|_\infty \operatorname{sgn}[(x - B_\infty x)(t_\alpha)]$. We can therefore write that $(x - B_p x)\, w_p$ converges weak* in $C(Q)^*$ to $(x - B_\infty x)\, w'_\infty$, where

$$w_p = \frac{|x - B_p x|^{p-2}}{\|x - B_p x\|_{L^p}^{p-1}} \tag{7}$$

and

$$w'_\infty = \sum_{t_\alpha \in \mathscr{C}} \lambda'_\alpha \delta_{t_\alpha}, \tag{8}$$

where $\lambda'_\alpha = \lambda_\alpha / \|x - B_\infty x\|_\infty$. We are therefore motivated to make the following important reformulation of the "0 in the convex hull"-criterion for best uniform approximation.

THEOREM 2. *The best L^∞-approximation $B_\infty x$ is also a best weighted L^2-approximation with weight $w_\infty = \sum_{t_\alpha \in \mathscr{C}} \lambda_\alpha \delta_{t_\alpha}$, where λ_α and $\mathscr{C}$ are as in* (3).

Conversely, given any set $\mathscr{C} = \{t_\alpha\}$ such that the dependence (3′) *is unique, then a best w_∞-weighted L^2-approximation, where $w_\infty = \sum_{t_\alpha \in \mathscr{C}} \lambda_\alpha \delta_{t_\alpha}$ ($\lambda_\alpha = |\gamma_\alpha|$ in* (3′)), *is also a best L^∞-approximation on $\mathscr{C}$.*

Proof. By the "0 in the convex hull"-criterion (3), $v_0 = B_\infty x$ being a best L^∞-approximation to x implies that v_0 is a best L^∞-approximation on $\mathscr{C}$ which is equivalent to

$$\sum_{t_\alpha \in \mathscr{C}} \lambda_\alpha \operatorname{sgn}[(x - v_0)(t_\alpha)]\, v(t_\alpha) = 0, \qquad \forall v \in V \tag{9}$$

$$\Rightarrow \sum_{t_\alpha \in \mathscr{C}} \lambda_\alpha |x - v_0|(t_\alpha) \operatorname{sgn}[(x - v_0)(t_\alpha)]\, v(t_\alpha) = 0 \qquad \forall v \in V \tag{10}$$

$$(\text{since } |x - v_0|(t_\alpha) = \|x - v_0\|_{L^\infty(\mathscr{C})}\ \forall t_\alpha \in \mathscr{C}),$$

$$\Leftrightarrow \sum_{t_\alpha \in \mathscr{C}} \lambda_\alpha (x - v_0)(t_\alpha)\, v(t_\alpha) = 0 \qquad \forall v \in V,$$

$$\Leftrightarrow \int (x - v_0)\, v w_\infty = 0 \qquad \forall v \in V,$$

$$\Leftrightarrow v_0 \text{ is a best } w_\infty\text{-weighted } L^2\text{-approximation to } x,$$

according to (1) applied to the support of w_∞.

For the converse, reverse the logic above by observing that in (10) all $|x - v_0|(t_\alpha)$ must be constant ($\forall \alpha$) since the dependence $\sum_{t_\alpha \in \mathscr{C}} \gamma_\alpha e_{t_\alpha} = 0$ on V is unique with $|\gamma_\alpha| = \lambda_\alpha$, $\sum \lambda_\alpha = 1$. Thus divide (10) by the constant $|x - v_0|(t_\alpha) = \|x - v_0\|_{L^\infty(\mathscr{C})}$ to obtain (9). ∎

COROLLARY 2. *If V is Haar and the critical set $\mathscr{C}$ is given, then $B_\infty x$ can be determined by the usual Hilbert space Gram–Schmidt process.*

Proof. Since V is Haar, by (3) and Note 1 we may take $\mathscr{C}_* = \mathscr{C} \cap \{t_\alpha ; \lambda_\alpha \neq 0 \text{ in } (3)\}$ to have exactly $n+1$ points and (see Note 2) the λ_α are determined linear-algebraically (in fact, by inverting an $(n+1) \times (n+1)$-matrix). Hence w_∞ is a positive weight on $\mathscr{C}$ which distinguishes among independent members of V and the Gram–Schmidt process produces an orthonormal basis $v_1, \ldots, v_n$ for V (i.e., $\langle v_i, v_j \rangle_{w_\infty} = \int v_i v_j w_\infty = \delta_{ij}$, $1 \leqslant i, j \leqslant n$). Then $B_\infty x$ is given uniquely by $B_\infty x = \sum_{i=1}^n \langle x, v_i \rangle_{w_\infty} v_i$. ∎

Notation. As in the above proof let $\langle x, y \rangle_w = \int xyw$ and let $\|x\|_{L^2_w} = \langle x, x \rangle_w$.

Motivated by the preceding, we now state an iterative Hilbert space algorithm for finding a best L^∞-approximation and show that it works (in fact omitting (the "mollifier") step (b)) in the case V is Haar.

ALGORITHM (*)—Weighted L^2-algorithm for best L^∞-approximation.

(a) Given v_ν and w_ν, determine $v_{\nu+1}$ so that

$$\|x - v_{\nu+1}\|_{L^2, w_\nu} \leqslant \|x - v\|_{L^2, w_\nu} \qquad \forall v \in V, \text{ where } w_\nu = \sum \lambda_\alpha^\nu \delta_{t_\alpha^\nu}$$

and the λ_α^v are determined algebraically via Note 2. $\mathscr{C}_{v+1} = \{t_\alpha^{v+1}\}$ is a set obtained from $\mathscr{C}_v$ with no more than $n+1$ points containing a global maximum and as many other local maxima of $|x - v_v|$ as possible so that the "sign pattern" in dependence (3) is preserved.

(b) Replace v_{v+1} by $\lambda_v v_{v+1} + (1-\lambda_v) v_v$, where λ_v yields the infimum of the convex functional $d(\lambda) = \|x - [\lambda v_{v+1} + (1-\lambda) v_v\|$.

(c) Iterate.

ALGORITHM (**). Same as Algorithm (*) except that step (a) is modified so that $\mathscr{C}_{v+1}$ is obtained from $\mathscr{C}_v$ by replacing one point of $\mathscr{C}_v$ with a global maximum point of $|x - v_v|$ so that the "sign pattern" in dependence (3) is preserved.

THEOREM 3. *If V is Haar then Algorithms* (*) *and* (**) *converge to the best approximation* (*without step* (b)), *provided* $\mathscr{C}_1 = \{t_i^1\}_{i=0}^n$ *is any set where* v_1 *does not interpolate* x.

Proof. By use of Theorem 2 we recognize that Algorithms (*) and (**) are the familiar Remez multiple point exchange algorithm (see [7, p. 176]) and single point exchange algorithm (see [2, p. 111; 4, p. 96; 7, p. 173]), respectively, where the (de la Vallée Poisson) interpolation on $n+1$ points at each stage has been replaced by the standard Hilbert space technique for the best weighted L^2-approximation on the $n+1$ points outlined in the proof of Corollary 2. ∎

Note 5. The linear-algebraic procedure for determining which point in $\mathscr{C}_v$ is to be replaced in step (a) of Algorithm (**) in the case V is Haar is called the *single point exchange procedure* and is described in [2, p. 109]. The corresponding procedure for Algorithm (*) can be viewed as successive single point exchanges.

Note 6. The fact that step (b) is not needed in Theorem 3 indicates that the analogous step may be superfluous in the Karlovitz algorithm for $p < \infty$.

REFERENCES

1. M. S. BANI, Iterative algorithms for best or near best approximations, Ph. D. Thesis.
2. B. L. CHALMERS, The Remez exchange algorithm for approximation with linear restrictions, *Trans. Amer. Math. Soc.* **223** (1976), 103–131.
3. B. L. CHALMERS, A. EGGER, AND G. D. TAYLOR, Convex L^p approximation, *J. Approx. Theory* **37** (1983), 326–334.

4. E. W. CHENEY, "Introduction to Approximation Theory," McGraw–Hill, New York, 1966.
5. C. R. HOBBY AND J. R. RICE, A moment problem in L_1 approximation, *Proc. Amer. Math. Soc.* **16** (1965), 665–670.
6. L. A. KARLOVITZ, Construction of nearest points in the L^p, p even, and L^∞ norms, *J. Approx. Theory* **3** (1970), 123–127.
7. J. R. RICE, "The Approximation of Functions," Addison–Wesley, Reading, Mass., 1964.