JOURNAL OF APPROXIMATION THEORY 55, 1–11 (1988)

Convex L^p Approximation

G. A. WATSON

Department of Mathematics and Computer Science, University of Dundee, Dundee DD1 4HN, Scotland, U.K.

Communicated by E. W. Cheney

Received December 31, 1985; revised August 10, 1986

1. INTRODUCTION

The problem of approximating 0 by elements of a compact nonempty convex subset K which is contained in a finite-dimensional subspace and which does not contain 0 was recently considered by Chalmers, Egger, and Taylor [4]. The study of this problem (which clearly subsumes a more general approximation problem) was motivated by earlier work of Karlovitz [9], who developed an iterative method based on the solution of a sequence of weighted L^2 problems for finding best L^p approximations from certain finite dimensional subspaces for p an even integer. It was shown in [4] that for the problem considered there, Karlovitz' algorithm was convergent for $2 \leqslant p < \infty$; it was also remarked there that convergence occurred for $1 < p < 2$ under an additional assumption.

An essential ingredient of the algorithm studied in both papers is a line search procedure. In an analogous algorithm for the L^∞ problem, Bani and Chalmers [2] showed that with an additional Haar condition assumption, convergence is possible without this subproblem. It is the purpose of this paper to point out that for the cases $1 < p < 2$, under conditions which permit convergence of the Karlovitz algorithm, the line search is also unnecessary. Under similar conditions, it is also shown that the simpler algorithm is locally convergent for the special case when K is defined by a finite number of linear constraints and $2 < p < 3$. Closely related results are available for analogous methods applied to (mainly finite) unconstrained L^p approximation problems (see, for example, [3, 5, 7, 8, 10, 11]).

In so far as it is appropriate, the notation of this paper follows that of [4]. Let (T, Σ, μ) be a finite positive measure space, and define $L^p \equiv L^p(T, \Sigma, \mu)$ to be the Banach space of all μ-equivalence classes of p-summable real-valued functions defined on T. Let $L^\infty \equiv L^\infty(T, \Sigma, \mu)$ be

0021-9045/88 $3.00

the Banach space of real-valued measurable essentially bounded functions defined on T. Then for $f \in L^p$, the norm $\|f\|_p$ may be defined as usual by

$$\|f\|_p^p = \int_T |f|^p \, d\mu, \qquad 1 \leqslant p < \infty$$

$$\|f\|_\infty = \inf_{S \in \Sigma, \mu(S) = 0} \sup_{x \in T \setminus S} |f(x)|.$$

Let K be a compact convex subset of L^p satisfying

$$0 \notin K,$$

$$\dim(\mathrm{span}(K)) < \infty,$$

$\mu(\mathrm{supp}(h_1) \cap \mathrm{supp}(h_2)) \neq 0$ for each pair of nonzero elements

$$h_1 \in K,$$

$$h_2 \in \mathrm{span}(K),$$

and

each $h \in K$ is also in L^∞.

Assume now that $1 < p < \infty$. Then because the L^p norm is strictly convex on any convex subset of L^p, the best L^p approximation from K to 0, say g^*, is unique. The algorithm for finding g^* which is analysed here consists of the following iteration:

given $g_n \in K$ define $g_{n+1} \in K$ as the solution to the problem

$$\min_{g \in K} W(g_n, g) \tag{1.1}$$

where

$$W(g_n, g) = \int_T |g_n|^{p-2} g^2 \, d\mu.$$

Because

$$W(g_n, g_{n+1}) \leqslant \|g_n\|_p^p, \tag{1.2}$$

the problem (1.1) is well-defined for $1 < p < \infty$. When $p < 2$, $W(g_n, g)$ may not be finite. However, the set

$$X = \{g \in L^p;\ W(g_n, g) < \infty\}$$

is a linear subspace of L^p and on that subspace $W(g_n, g)^{1/2}$ is a weighted least squares norm. Then the ball

$$\{g \in L^p\colon W(g_n, g)^{1/2} \leqslant \min_{g \in K} W(g_n, g)^{1/2}\}$$

is strictly convex and meets K at a unique point. In addition it follows from the definition that $W(g_n, g)$ is Gateaux differentiable on X.

The following characterization of g^* is well known.

THEOREM 1. *$g^* \in K$ is the best L^p approximation to 0 if and only if*

$$\int_T |g^*|^{p-1} \operatorname{sign}(g^*)(g^* - g)\, d\mu \leqslant 0, \qquad \text{for all} \quad g \in K. \tag{1.3}$$

2. Convergence Properties

It is first shown that if $1 < p < 2$, then the algorithm (1.1) is a descent process from *any* initial approximation $g_0 \in K$.

THEOREM 2. *Let $1 < p < 2$ and let $\{g_n\}$ be defined by (1.1) with $g_0 \in K$ arbitrary. Then*

$$\|g_{n+1}\|_p \leqslant \|g_n\|_p$$

with equality only if $g_{n+1} = g_n$.

Proof. For any real a, b with $b \neq 0$, it is straightforward to show that if $1 < p \leqslant 2$,

$$|a|^p \leqslant |b|^p + \tfrac{1}{2} p\, |b|^{p-2} (a^2 - b^2).$$

For any n, let

$$Z_n = \{x \in T, g_n(x) = 0\}.$$

Then for any $x \in T \backslash Z_n$

$$|g_{n+1}(x)|^p \leqslant |g_n(x)|^p + \tfrac{1}{2} p\, |g_n(x)|^{p-2} (g_{n+1}(x)^2 - g_n(x)^2).$$

It follows that

$$\begin{aligned} \int_{T/Z_n} |g_{n+1}(x)|^p\, d\mu &\leqslant \|g_n\|_p^p + \tfrac{1}{2} p \int_{T/Z_n} |g_n|^{p-2} g_{n+1}^2\, d\mu - \tfrac{1}{2} p\, \|g_n\|_p^p \\ &\leqslant \|g_n\|_p^p \text{ using (1.2).} \end{aligned} \tag{2.1}$$

Because

$$\int_T |g_n|^{p-2} g_{n+1}^2 \, d\mu < \infty$$

we must have

$$\mu(x \in Z_n, \, g_{n+1}(x) \neq 0) = 0$$

so (2.1) gives

$$\| g_{n+1} \|_p^p \leqslant \| g_n \|_p^p$$

with equality only if $g_n = g_{n+1}$, by uniqueness. ∎

The question remains as to whether points of termination of the algorithm, or any of the limit points of the (bounded and finite dimensional) sequence $\{g_n\}$, solve (1.1). That the answer to this is generally in the negative is shown by the following example.

EXAMPLE. Let $(T, \Sigma, \mu) = [0, 1]$ with Lebesgue measure, let $p = \frac{3}{2}$, and let

$$K = \{c - x^2, \, c \in [0, 1]\}.$$

If the algorithm (1.1) is used with $g_0 = -x^2$, then

$$W(g_0, g) = \int_0^1 x^{-1}(c - x^2)^2 \, dx,$$

which is minimized by $g = g_0$ (the only element of K which makes $W(g_0, g) < \infty$). However g_0 does not satisfy (1.3).

It follows that to establish a useful convergence result it is necessary to impose conditions on the sequence $\{g_n\}$, and a crucial requirement is that $W(g_n, g)$ be Gateaux differentiable with continuous derivative at g_{n+1}, when g_{n+1} is characterized by

$$\int_T |g_n|^{p-2} g_{n+1}(g_{n+1} - g) \, d\mu \leqslant 0, \qquad \text{for all} \quad g \in K \tag{2.2}$$

(for example, Ekeland and Temam [6] p. 37). A sufficient condition for this is that

$$\int_T |g_n|^{\alpha - 1} \, d\mu < \infty, \qquad 0 < \alpha < 1, \tag{2.3}$$

which is just the condition for convergence of the original Karlovitz algorithm in this case. (For example, in the context of $C[0, 1]$, (2.3) is satisfied if each g_n has a finite number of simple zeros in $[0, 1]$ (see Chalmers, Egger, and Taylor [4], Bani [1]).)

THEOREM 3. *Let* $1 < p < 2$, *let* $\{g_n\}$ *be defined by* (1.1), *and let* (2.3) *be satisfied for all n. Then either the algorithm terminates at* g^* *or*

$$g_n \to g^* \text{ as } n \to \infty.$$

Proof. If $g_n = g_{n+1}$, it follows from (2.2) that g_n satisfies (1.3), so that $g_n = g^*$.

Let $F(g) = \|g\|_p$. Then $\{F(g_n)\}$ is a decreasing sequence, bounded below, and so convergent to F^*, say. Further $\{g_n\}$ is bounded and finite dimensional and so has limit points. Let

$$\|g_{i_j} - v\|_\infty \to 0 \qquad \text{as} \quad j \to \infty$$

and (going to a subsequence if necessary, which is not renamed)

$$\|g_{i_j+1} - w\|_\infty \to 0 \qquad \text{as} \quad j \to \infty.$$

By continuity of F,

$$F(v) = F(w) = F^*.$$

Now by definition of g_{i_j+1}, (2.2) gives

$$\int_T |g_{i_j}|^{p-2} g_{i_j+1}(g_{i_j+1} - g)\, d\mu \leqslant 0,$$

$$\text{for all } g \in K,\ j = 1, 2, ...$$

Let $j \to \infty$. Then by continuity of the Gateaux derivative

$$\int_T |v|^{p-2} w(w-g)\, d\mu \leqslant 0, \qquad \text{for all } g \in K,$$

and so $v = w$, otherwise $F(w) < F(v)$, a contradiction. It follows from Theorem 1 that $v = g^*$, and since this is true for all limit points, the theorem is proved. ∎

The above results hinge on the inequality (2.1), and if $p > 2$, this is reversed. In order to say something about the algorithm when $p > 2$, it seems necessary to be more specific about the set K, and in what follows it will be assumed that K is defined by a finite number of linear constraints. Let V be an r-dimensional subspace of $L^p(T, \Sigma, \mu)$ and let $\{v_1, v_2, ..., v_r\}$ be a basis for V. Then the problem to be considered may be written as

$$\text{find } c \in R^r \text{ to minimize } \int_T \left| \sum_{i=1}^{r} c_i v_i \right|^p d\mu \tag{2.4}$$

$$\begin{aligned} \text{subject to } c^T a_i &= b_i, \qquad i = 1, 2, ..., s \\ c^T a_i &\geqslant b_i, \qquad i = s+1, ..., t, \end{aligned} \tag{2.5}$$

where $b_i \in R$, $a_i \in R^r$, $i = 1, 2, ..., t$ are given, and $s \leqslant r$.

Assume first that $s = t$ so that only equality constraints are present and let these be written

$$A^T c = b, \tag{2.6}$$

where A is an $r \times s$ matrix. If A has rank s, then without loss of generality it may be expressed as

$$A^T = [B^T \vdots C^T],$$

where B is an $s \times s$ nonsingular matrix. It follows that the first s components of c may be eliminated using (2.6) and the minimization problem reduced to an equivalent unconstrained problem in R^m, where $m = r - s$. This problem may be written

$$\text{find } y \in R^m \text{ to minimize } \left\| \sum_{i=1}^{m} y_i w_i - f \right\|_p^p, \tag{2.7}$$

where

$$w_i = v_{i+s} - \sum_{j=1}^{s} M_{ij} v_j, \qquad i = 1, 2, ..., m,$$

$$f = \sum_{j=1}^{s} d_j v_j,$$

with M_{ij} the (i, j) component of $M = CB^{-1}$ and d_j the jth component of $d = -B^{-T} b$. Similarly the problem (1.1) solved by g_{n+1} is equivalent to

$$\text{find } z \in R^m \text{ to minimize } \int_T |g_n|^{p-2} \left(\sum_{i=1}^{m} z_i w_i - f \right)^2 d\mu. \tag{2.8}$$

The solution of (2.8) satisfies a nonsingular system of linear equations, say

$$F(y, z) = 0, \tag{2.9}$$

where $g_n = \sum_{i=1}^m y_i w_i - f$. This may be regarded as a simple iteration function, and provided that F is a continuously differentiable function of y in the neighbourhood of a fixed point, a standard local convergence analysis may be performed. There is no difficulty when $p > 3$; however, when $2 < p \leqslant 3$ it is necessary to impose some conditions and the following lemma is required. Let $g = g(y) = \sum_{i=1}^m y_i w_i - f$, $g^* = g(y^*)$.

LEMMA 1. *Let $0 < \alpha \leqslant 1$, and for all y in an open neighbourhood $N(y^*)$ of y^*, let*

$$\mu(x : g = 0) = 0$$

$$\int_T |g|^{\alpha - 1}\, d\mu < \infty, \qquad \text{if} \quad \alpha < 1.$$

Then $\int_T |g|^\alpha\, d\mu$ is a differentiable function of y for all $y \in N(y^*)$.

Proof. Let $y \in N(y^*)$ and define

$$p(y, T) = \left(\int_T |g|^\alpha\, d\mu \right)^{1/\alpha},$$

$$Z = Z(\varepsilon) = \{x \in T \colon |g| \leqslant \varepsilon\}.$$

Let d, $\|d\| = 1$, be arbitrary and let

$$M = \max_{x \in T} \left| \sum_{i=1}^m d_i w_i \right|.$$

Then for $\varepsilon > 0$, $0 < \gamma \leqslant \varepsilon / M$,

$$P(y + \gamma d, T - Z) = P(y, T - Z) + \gamma \sum_{i=1}^m d_i G_i(y, T - Z) + O(\gamma^2),$$

where

$$G_i(y, T - Z) = \left(\int_{T-Z} |g|^\alpha \right)^{1/\alpha - 1} \int_{T-Z} |g|^{\alpha - 1} \operatorname{sign}(g) v_i\, d\mu, \qquad i = 1, 2, \ldots, m. \tag{2.10}$$

Thus

$$P(y + \gamma d, T) - P(y, T) = \gamma \sum_{i=1}^m d_i G_i(y, T - Z) + P(y + \gamma d, Z) - P(y, Z) + O(\gamma^2). \tag{2.11}$$

Now

$$\int_Z |g(y+\gamma\, d)|^\alpha\, d\mu \leqslant (\varepsilon+\gamma M)^\alpha\, \mu(Z)$$

so that

$$P(y+\gamma\, d, Z) \leqslant (\varepsilon+\gamma M)\, \mu(Z)^{1/\alpha}.$$

It follows that

$$\left|\frac{P(y+\gamma\, d, T)-P(y, T)}{\gamma} - \sum_{i=1}^{m} d_i G_i(y, T-Z)\right| \leqslant \left(\frac{2\varepsilon}{\gamma}+M\right) \mu(Z)^{1/\alpha} + O(\gamma). \tag{2.12}$$

Now choose $\varepsilon = \gamma M$ and let $\varepsilon \to 0$ in (2.12). By continuity $\mu(Z) \to 0$. Therefore $P(y, T)$ is differentiable at y and the result follows. ∎

The rather stronger result of continuous differentiability in $N(y^*)$ is in fact necessary and it is not clear that this holds without additional assumptions. However, in the context of $C[0, 1]$, if g^* has a finite number of simple zeros in $[0, 1]$, then the conditions of the lemma are satisfied and also $\int_0^1 |g|^{\alpha-1}$ sign $(g)\, dx$ is continuous at g^* for $0<\alpha<1$.

THEOREM 4. *Let $p>2$ and let the algorithm* (1.1) *be applied to* (2.4) *subject to* (2.6) *with A an $s \times r$ matrix with full rank s. Then if the conditions of Lemma* 1 *are satisfied, and in addition $\int_T |g|^{p-3}\, \mathrm{sign}(g)\, d\mu$ is continuous at g^*, the algorithm is locally convergent to g^* if $p<3$.*

Proof. Since the problem is equivalent to (2.7) it is only necessary to establish the result for the sequence generated by (2.8). Let y be the current approximation to the solution of (2.7) with $g = \sum_{i=1}^m y_i w_i - f$. Define $h = \sum_{i=1}^m z_i w_i - f$ where z solves (2.8) with $g_n = g$, so that

$$\int_T |g|^{p-2}\, h w_i\, d\mu = 0, \qquad i = 1, 2, ..., m, \tag{2.13}$$

or

$$F(y, z) = 0, \text{ say.} \tag{2.14}$$

Now for any i, j, $1 \leqslant i, j \leqslant m$,

$$\frac{\partial F_i}{\partial z_j} = \int_T |g|^{p-2}\, w_i w_j\, d\mu,$$

$$\frac{\partial F_i}{\partial y_j} = \int_T (p-2)\, |g|^{p-3}\, \mathrm{sign}(g)\, h w_i w_j\, d\mu,$$

using Lemma 1. If $g = h(y = z)$ then

$$\frac{\partial F_i}{\partial y_j} = (p-2)\frac{\partial F_i}{\partial z_j}, \qquad i, j = 1, 2, ..., m.$$

Thus at a fixed point of the iteration function, the Jacobian matrix of z regarded as a function of y is given by

$$J = (2-p)I$$

and the result of the theorem follows. ∎

Now return to the more general problem (2.4), (2.5) and let c^* be the solution with $g^* = \sum_{i=1}^{r} c_i^* v_i$. Let $\mathscr{A}^*$ denote the active set of indices such that

$$c^{*T} a_i = b_i, \qquad i \in \mathscr{A}^*.$$

Then standard Kuhn–Tucker theory gives the existence of Lagrange multipliers λ_i^*, $i \in \mathscr{A}^*$, such that

$$\phi(g^*, g^*) - \sum_{i \in \mathscr{A}^*} \lambda_i^* a_i = 0, \tag{2.15}$$

where $\phi(g, h) \in R^r$ has the ith component $\int_T |g|^{p-2} h v_i \, d\mu$, $i = 1, 2, ..., r$, and $\lambda_i^* \geqslant 0$ if $i \geqslant s+1$.

THEOREM 5. *Let $p > 2$ and let the algorithm* (1.1) *be applied to the problem* (2.4), (2.5), *whose solution is characterized by* (2.15). *Let the conditions of Lemma* 1 *be satisfied and let $\int_T |g|^{p-3} \operatorname{sign}(g) \, d\mu$ be continuous at g^*. Then if*

(i) $\{a_i, i \in \mathscr{A}^*\}$ *is a linearly independent set,*

(ii) $\lambda_i^* \neq 0$, $i \in \mathscr{A}^*$,

the algorithm is locally convergent to g^* if $p < 3$.

Proof. Write (2.15) as

$$\phi^* - A\lambda^* = 0,$$

where A is assumed to be an $r \times k$ matrix. Then if (i) holds, λ^* is uniquely defined by the expression

$$\lambda^* = A^+ \phi^*, \tag{2.16}$$

where the superscript + denotes the usual generalized inverse. If (ii) holds, a small perturbation of ϕ^* will not zero any component λ_i^*, $i \in \mathscr{A}^*$.

Now consider the problem

$$\text{find } d \in R^r \text{ to minimize } \int_T |g|^{p-2} \left(\sum_{i=1}^r d_i v_i\right)^2 d\mu$$
$$\text{subject to } A^T d = b, \tag{2.17}$$

where $g = \sum_{i=1}^r c_i v_i$. Then a feasible $d \in R^r$ solves the problem if and only if there exists $\lambda \in R^k$ such that

$$\phi(g, h) - A\lambda = 0,$$

where $h = \sum_{i=1}^r d_i v_i$, and from (2.16)

$$\lambda - \lambda^* = A^+(\phi(g, h) - \phi^*). \tag{2.18}$$

Without loss of generality, it may be assumed that the first k columns of A form a nonsingular matrix, and so the problem (2.17) can be replaced by an unconstrained (weighted least squares) problem in R^{r-k} just as before, for which an explicit solution may be obtained. It follows that d solving (2.17) may be written explicitly as a function of c and a continuous dependence may be established. Thus if $\|g - g^*\|$ is small enough, $\|h - g^*\|$ will also be small enough so that d solving (2.17) will also solve the problem with constraint set given by (2.5), for (2.18) shows that the active set must be given by $\mathscr{A}^*$. Therefore, locally the situation reduces to that considered in Theorem 3 and the result follows. ∎

ACKNOWLEDGMENTS

I am grateful to A. L. Brown for helpful comments in connection with this work. I also thank the referee for the example and for other important remarks.

REFERENCES

1. M. S. BANI, "Iterative Algorithms for Best or Near Best Approximations," Ph. D. thesis, University of California, 1984.
2. M. S. BANI AND B. L. CHALMERS, Best approximation in L^∞ via iterative Hilbert space procedures, *J. Approx. Theory* **42** (1984), 173–180.
3. R. H. BYRD AND D. A. PYNE, On the convergence of iteratively reweighted least squares algorithms, *SIAM J. Sci. Statist. Comput.*, in press.
4. B. L. CHALMERS, A. G. EGGER, AND G. D. TAYLOR, Convex L^p approximation, *J. Approx. Theory* **37** (1983), 326–334.
5. H. EKBLOM, Calculation of linear best L_p approximations, *BIT* **13** (1973), 292–300.
6. I. EKELAND AND R. TEMAM, "Convex Analysis and Variational Problems," North-Holland, Amsterdam, 1976.
7. R. FLETCHER, J. A. GRANT, AND M. D. HEBDEN, The calculation of linear best L_p approximations, *Comput. J.* **14** (1971), 276–279.

8. S. W. KAHNG, Best L_p approximation, *Math. Comp.* **26** (1972), 505–508.
9. L. A. KARLOVITZ, Construction of nearest points in the L^p, p even, and L^∞ norms, *J. Approx. Theory* **3** (1970), 123–127.
10. G. MERLE AND H. SPÄTH, Computational experience with discrete L_p approximation, *Computing* **12** (1974), 315–321.
11. J. M. WOLFE, On the convergence of an algorithm for discrete L_p approximation, *Numer. Math.* **32** (1979), 439–459.