MATHEMATICS OF COMPUTATION, VOLUME 24, NUMBER 111, JULY, 1970

Computation of Best One-Sided L_1 Approximation*

By James T. Lewis

Abstract. A computational procedure based on linear programming is presented for finding the best one-sided L_1 approximation to a given function. A theorem which ensures that the computational procedure yields approximations which converge to the best approximation is proved. Some numerical examples are discussed.

1. Introduction. Recently there has been an interest in the problem of finding a best one-sided approximation to a given function; that is, an approximation which is everywhere below (or everywhere above) the function. In [6], [7], [8] the measure of approximation is the uniform norm; in [1], [3], [8] the integral L_1 norm. A computational procedure for the latter problem is presented and analyzed in this paper.

Let $f(x)$ be a real-valued continuous function on $[a, b]$, $\alpha(x)$ a positive continuous weight function on $[a, b]$. Assume the set $\{\phi_1, \cdots, \phi_n\}$ of continuous functions on $[a, b]$ is a Chebyshev system; that is, any nontrivial linear combination has at most $n - 1$ zeros in $[a, b]$. Of course the most commonly used Chebyshev system is the set of powers $\{1, x, \cdots, x^{n-1}\}$. Let $\mathcal{L}(f)$ denote the set of approximations to f from below; i.e., the set of all real linear combinations $\sum_{i=1}^n a_i\phi_i(x)$ such that $\sum_{i=1}^n a_i\phi_i(x) \leqq f(x)$ for all $x \in [a, b]$. Then $p_* \in \mathcal{L}(f)$ is a *best weighted L_1 approximation to f from below on $[a, b]$* if

$$\int_a^b \alpha(x)(f(x) - p_*(x))\,dx = \inf\left\{\int_a^b \alpha(x)(f(x) - p(x))\,dx : p \in \mathcal{L}(f)\right\}.$$

Best approximation from above is defined analogously.

Existence of a best approximation is straightforward; the uniqueness question is more involved. It has been shown [3] that if f is differentiable on $[a, b]$ and $\{\phi_1, \cdots, \phi_n\}$ is a differentiable Chebyshev system (each ϕ_i is differentiable and any nontrivial linear combination of the derivatives has at most $n - 2$ zeros in $[a, b]$), then the best approximation from below is unique. The following elegant theorem from [1] shows that in certain circumstances the best approximation from below by ordinary algebraic polynomials may be found by interpolation of f and f' at suitable points.

THEOREM 1. *Let $n = 2k$, f continuous on $[a, b]$, $f^{(n)}(x) \geqq 0$ for all x in (a, b). Then the algebraic polynomial p_* of best weighted L_1 approximation of f from below on $[a, b]$ of degree less than or equal to $n - 1$ is the unique polynomial defined by the equations:*

$$(1.1)\qquad p_*(y_i) = f(y_i), \quad p_*'(y_i) = f'(y_i), \quad i = 1, \cdots, k,$$

Received August 14, 1969, revised January 22, 1970.

AMS Subject Classifications. Primary 4140, 6520.

Key Words and Phrases. One-sided approximation, L_1 approximation, convex constraints, computation of best approximation, linear programming.

* This work was supported in part by the Office of Naval Research under contract NONR 562(36).

where the y_i are the zeros of the polynomial of degree k orthogonal to $1, x, \cdots, x^{k-1}$ *with inner product*

$$(g, h) = \int_a^b \alpha(x)g(x)h(x)\,dx.$$

Proof. [1, p. 152].

Analogous theorems hold for n odd, for approximation from above, and also if $f^{(n)}(x) \leqq 0$ for all x in (a, b). The polynomial satisfying (1.1) is given explicitly by (cf. [2, p. 37])

$$p_*(x) = \sum_{i=1}^{k} f(y_i)\left(1 - \frac{\pi''(y_i)}{\pi'(y_i)}(x - y_i)\right)l_i^2(x) + \sum_{i=1}^{k} f'(y_i)(x - y_i)l_i^2(x) \tag{1.2}$$

where $\pi(x) = (x - y_1) \cdots (x - y_k)$, $l_i(x) = \pi(x)/((x - y_i)\pi'(y_i))$.

If $f^{(n)}(x)$ is not one-signed on (a, b), the problem of computing the best one-sided approximation remains. In Section 2 a computational procedure based on linear programming is presented and analyzed. This method is analogous to one way of computing the best unconstrained L_1 approximation. In Section 3 numerical experience with the method is discussed.

2. The Computational Procedure. We again consider the problem of approximation from below from a Chebyshev system $\{\phi_1, \cdots, \phi_n\}$. The basic idea of the computational procedure is to obtain the best weighted L_1 approximation to f from below on $[a, b]$ as a limit of best weighted (with a different weight function) L_1 approximations to f on finite point subsets of $[a, b]$. Let $X_m = \{x_j; j = 1, \cdots, m\}$ be a subset of $[a, b]$ with $a \equiv x_0 < x_1 < \cdots < x_m = b$ and mesh size $\mu_m = \max_{1 \leq j \leq m} |x_j - x_{j-1}|$. The appropriate weighted problem on X_m to consider is

$$\min_p \sum_{j=1}^{m} \alpha(x_j)(f(x_j) - p(x_j))(x_j - x_{j-1}) \tag{2.1}$$

$$\text{subject to} \quad p(x_j) \leqq f(x_j), \qquad j = 1, \cdots, m. \tag{2.2}$$

The existence of a solution to this problem can be easily established; however, in general the solution will not be unique, as in the following example.

Example (nonuniqueness). Let $X_m = \{-1, 0, 1\}$, $x_0 = -2$; $f(-1) = 1$, $f(0) = 0$, $f(1) = 1$; $\alpha(-1) = \alpha(0) = \alpha(1) = 1$. It can be verified that for any $\beta \in [-1, 1]$ the polynomial βx is a best weighted L_1 approximation to f from below on X_m of the form $a_1 + a_2x$.

For a given set X_m a solution of the problem (2.1), (2.2) can be obtained by linear programming techniques; this will be discussed in more detail in Section 3. Now let X_m, $m = 1, 2, \cdots$, be a sequence of subsets of $[a, b]$ such that $\mu_m \to 0$ as $m \to \infty$ and let p_m be a solution of the problem (2.1), (2.2) on X_m. We wish to analyze the behavior of p_m as $m \to \infty$. The following lemma dealing with the approximation of an integral by a Riemann sum will be useful. Let

$$\omega(h; \mu) = \sup\{|h(x) - h(y)|: x, y \in [a, b], |x - y| < \mu\}$$

be the modulus of continuity of the function h on $[a, b]$.

LEMMA 1. *Let* $\alpha(x)$, $g(x)$ *be continuous on* $[a, b]$ *and* X_m *as described above. Then*

$$\left|\int_a^b \alpha(x)g(x)\,dx - \sum_{j=1}^m \alpha(x_j)g(x_j)(x_j - x_{j-1})\right| \leqq (b-a)\omega(\alpha g; \mu_m).$$

Proof. Cf. [11, p. 79].

Notice that since $\omega(\alpha g; \mu_m) \to 0$ as $m \to \infty$ the difference between the integral and the sum tends to zero as $m \to \infty$. The main result of this paper, which is the analogue of a theorem due to Motzkin and Walsh [9, p. 394] on unconstrained approximation, is now stated and proved. The proof is patterned after that of [11, p. 80].

THEOREM 2. *Let f be a continuous function on $[a, b]$; $X_m = \{x_j: j = 1, \cdots, m\}, m = 1, 2, \cdots$, a sequence of discrete subsets of $[a, b]$ with $a \equiv x_0 < x_1 < \cdots < x_m = b$ for each m. Suppose $\mu_m = \max_{1 \leq j \leq m} |x_j - x_{j-1}| \to 0$ as $m \to \infty$. Let p_m be a best L_1 approximation from below to f on X_m with weights $\alpha(x_j)(x_j - x_{j-1})$. Then there exists a subsequence of $\{p_m\}$ which converges uniformly to a best weighted L_1 approximation to f from below on $[a, b]$. If the latter best approximation is unique, then p_m converges to this best approximation as $m \to \infty$.*

Proof. First it will be shown that $\{p_m\}$ is uniformly bounded on $[a, b]$. Let p_* be a best weighted L_1 approximation from below to f on $[a, b]$. Set

$$\sigma_* = \int_a^b \alpha(x)(f(x) - p_*(x))\,dx,$$

$$\sigma_m = \sum_{j=1}^m \alpha(x_j)(f(x_j) - p_*(x_j))(x_j - x_{j-1}).$$

By the definition of p_m and the fact that $p_*(x_j) \leqq f(x_j)$ for all $x_j \in X_m$ we have

$$\begin{aligned}\sigma_m &= \sum_{j=1}^m \alpha(x_j)(f(x_j) - p_m(x_j))(x_j - x_{j-1}) \\ &\leqq \sum_{j=1}^m \alpha(x_j)(f(x_j) - p_*(x_j))(x_j - x_{j-1}) \\ &\leqq \sigma_* + 1 \quad \text{for all } m \geqq \text{some } M_1 \text{ by Lemma 1.}\end{aligned}$$

Hence

$$\sum_{j=1}^m \alpha(x_j)\,|p_m(x_j)|\,(x_j - x_{j-1}) \leqq \sigma_* + 1 + (b-a) \max_{x \in [a,b]} |\alpha(x)f(x)| \equiv C$$

some constant.

Now let $I_1, \cdots, I_n$ be n closed disjoint subintervals of $[a, b]$, each of length $(b-a)/2n$ say. Choose M_2 so that $\mu_m \leqq (b-a)/6n$ for all $m \geqq M_2$. Set $\mathcal{J}_i = \{j: [x_{j-1}, x_j] \subset I_i\}$, $i = 1, \cdots, n$. Then

$$\begin{aligned}\sum_{j \in \mathcal{J}_i} \alpha(x_j)\,|p_m(x_j)|\,(x_j - x_{j-1}) &\geqq \left(\min_{j \in [a,b]} \alpha(x)\right)\left(\min_{j \in \mathcal{J}_i} |p_m(x_j)|\right) \sum_{j \in \mathcal{J}_i} (x_j - x_{j-1}) \\ &\geqq A\left(\min_{j \in \mathcal{J}_i} |p_m(x_j)|\right)\left(\frac{b-a}{6n}\right),\end{aligned}$$

where

$$A \equiv \min_{x \in [a,b]} \alpha(x) > 0 \quad \text{since } \alpha(x) > 0 \text{ for all } x \in [a, b].$$

Hence there exists some constant, say E, such that $\min_{j\in\mathcal{I}_i} |p_m(x_j)| \leqq E$ for $i=1,\cdots,n$ and all m. Now let $t_i \in X_m \cap I_i$ be a point such that $\min_{j\in\mathcal{I}_i} |p_m(x_j)| = |p_m(t_i)|$, $i = 1, \cdots, n$. Set

$$D(z_1, \cdots, z_n) = \begin{vmatrix} \phi_1(z_1) & \cdots & \phi_n(z_1) \\ \vdots & & \vdots \\ \phi_1(z_n) & \cdots & \phi_n(z_n) \end{vmatrix}.$$

Then the generalized Lagrange interpolation polynomial which assumes the value y_i at z_i $(i = 1, \cdots, n)$ is

$$p(x) = \sum_{i=1}^{n} y_i \frac{D(z_1, \cdots, z_{i-1}, x, z_{i+1}, \cdots, z_n)}{D(z_1, \cdots, z_n)}.$$

Hence

$$p_m(x) = \sum_{i=1}^{n} p_m(t_i) \frac{D(t_1, \cdots, t_{i-1}, x, t_{i+1}, \cdots, t_n)}{D(t_1, \cdots, t_n)}.$$

Since $t_i \in I_i$ $(i = 1, \cdots, n)$ and the I_i are closed and disjoint, $\{\phi_1, \cdots, \phi_n\}$ being a Chebyshev system implies there exists D such that $|D(t_1, \cdots, t_n)| \geqq D > 0$ independent of m. Hence

$$\max_{x\in[a,b]} |p_m(x)| \leqq \frac{E}{D} \sum_{i=1}^{n} \max_{x\in[a,b]} |D(t_1, \cdots, t_{i-1}, x, t_{i+1}, \cdots, t_n)|$$

$$\leqq F \quad \text{some constant, independent of } m.$$

Thus $\{p_m\}$ is uniformly bounded.

Now let $\{p_{m_k}\}$ be a uniformly convergent subsequence with limit, say, p_0. We wish to show p_0 is an approximation from below to f on $[a, b]$. Suppose not, that is, assume there exists $x^* \in [a, b]$ with $f(x^*) - p_0(x^*) = -\gamma < 0$. By the continuity of f at x^* we can find M_3 so that $|x^* - y| < \mu_{M_3}$ implies $|f(x^*) - f(y)| < \gamma/4$. By the uniform convergence of p_{m_k} to p_0 we can find M_4 so that $m_k \geqq M_4$ implies $\max_{x\in[a,b]}|p_{m_k}(x) - p_0(x)| < \gamma/4$. By the continuity of p_0 at x^* we can find M_5 so that $|y - x^*| < \mu_{M_5}$ implies $|p_0(y) - p_0(x^*)| < \gamma/4$. Now let M be an index from the subsequence $\{m_k\}$ such that $M \geqq M_4$, $\mu_M \leqq \mu_{M_3}$, $\mu_M \leqq \mu_{M_5}$. Let $y \in X_M$ with $|x^* - y| < \mu_M$. Then

$$\begin{aligned} -\gamma &= f(x^*) - p_0(x^*) \\ &= (f(x^*) - f(y)) + (f(y) - p_M(y)) + (p_M(y) - p_0(y)) + (p_0(y) - p_0(x^*)) \\ &\geqq -|f(x^*) - f(y)| + 0 - |p_M(y) - p_0(y)| - |p_0(y) - p_0(x^*)| \\ &\geqq -3\gamma/4 \quad \text{a contradiction, since } \gamma > 0. \end{aligned}$$

Hence $f(x) - p_0(x) \geqq 0$ for all $x \in [a, b]$.

Next set $\sigma_0 = \int_a^b \alpha(x)(f(x) - p_0(x))dx$. To show $\sigma_0 = \sigma_*$. Let $\epsilon > 0$ be given. Since p_{m_k} converges to p_0 we can find M_6 such that $m_k \geqq M_6$ implies

$$\left|\sigma_0 - \int_a^b \alpha(x)\, |f(x) - p_{m_k}(x)|\, dx\right| < \epsilon/3. \tag{2.3}$$

By Lemma 1 there exists M_7 such that $m_k \geqq M_7$ implies

$$\left| \sigma_* - \sum_{j=1}^{m_k} \alpha(x_j)(f(x_j) - p_*(x_j))(x_j - x_{j-1}) \right| < \epsilon/3. \tag{2.4}$$

Again by Lemma 1 we can find N with $N \geqq M_6$, $N \geqq M_7$ such that

$$\left| \int_a^b \alpha(x) \, |f(x) - p_N(x)| \, dx - \sigma_N \right| < \epsilon/3. \tag{2.5}$$

Hence $|\sigma_0 - \sigma_N| < 2\epsilon/3$ by (2.3), (2.5). Now

$$\sigma_N \leqq \sum_{j=1}^{N} \alpha(x_j)(f(x_j) - p_*(x_j))(x_j - x_{j-1}) \leqq \sigma_* + \epsilon/3$$

from (2.4). Thus $\sigma_* \leqq \sigma_0 \leqq \sigma_N + 2\epsilon/3 \leqq \sigma_* + \epsilon$. Since ϵ was arbitrary we have $\sigma_0 = \sigma_*$ and p_0 is a best approximation to f from below on $[a, b]$. If p_* is unique then $p_0 = p_*$. Since any convergent subsequence of the uniformly bounded sequence p_m converges to p_*, the sequence p_m converges to p_*.

3. Numerical Examples. In this section numerical experience with the computational procedure is discussed. We consider the special case of approximation by ordinary polynomials ($\phi_i(x) = x^{i-1}$) with weight function $\alpha(x) \equiv 1$. We will take sets X_m of the form

$$X_m = \{x_j : j = 1, \cdots, m\} = \{a + (j/m)(b - a) : j = 1, \cdots, m\},$$

i.e. evenly spaced partitions of $[a, b]$. The problem (2.1), (2.2) is then the linear programming problem:

$$\min_{a_1, \cdots, a_n} \frac{b-a}{m} \sum_{j=1}^{m} \left(f(x_j) - \sum_{i=1}^{n} a_i (x_j)^{i-1} \right)$$

subject to

$$\sum_{i=1}^{n} a_i (x_j)^{i-1} \leqq f(x_j), \qquad j = 1, \cdots, m.$$

This is equivalent to the problem:

$$\max_{a_1, \cdots, a_n} \sum_{i=1}^{n} a_i \left(\sum_{j=1}^{m} (x_j)^{i-1} \right) \tag{3.1}$$

$$\text{subject to} \quad \sum_{i=1}^{n} a_i (x_j)^{i-1} \leqq f(x_j), \qquad j = 1, \cdots, m. \tag{3.2}$$

The techniques for solving linear programming problems are well developed, cf. [4] or [5]. In the problem (3.1), (3.2) n is the number of variables and m the number of constraints. Since m will be much larger than n and since the amount of computation depends primarily on the number of constraints, it is better to solve the dual problem of (3.1), (3.2), cf. [5, Chapter 8]. This is the problem:

$$\min_{u_1, \cdots, u_m} \sum_{j=1}^{m} u_j f(x_j) \tag{3.3}$$

(3.4) $$\text{subject to} \quad \sum_{j=1}^{m} u_j (x_j)^{i-1} = \sum_{j=1}^{m} (x_j)^{i-1}, \qquad i = 1, \cdots, n$$

and $u_j \geqq 0, j = 1, \cdots, m$.

This problem is in standard form for the application of the simplex method. For various reasons, including the amount of storage space required, it is advantageous to use the revised simplex method on a computer. The revised simplex method was programmed as a subroutine and used to solve the above linear programming problem (3.3), (3.4). Although degeneracy was present, that is the objective function did not change during some iterations, no cycling was encountered in the examples run. All computations were in double precision on the IBM System 360/50. A tolerance of 10^{-10} was used in deciding whether a basic feasible solution had been reached and in deciding if the optimal solution had been obtained. In the following examples the problem is to find the best L_1 approximation from below on $[-1, 1]$ to the given function by a cubic polynomial $a_1 + a_2 x + a_3 x^2 + a_4 x^3$.

Example 1. $f(x) = \exp(x)$.

Since $f^{(4)}(x) \geqq 0$ on $[-1, 1]$ an explicit solution is available using Theorem 1. This solution is obtained by interpolating f and f' at the zeros $\pm\sqrt{3}/3$ of the quadratic polynomial (here the Legendre polynomial) orthogonal to $1, x$ on $[-1, 1]$. After some algebraic manipulation we obtain from (1.2) (setting $R = \sqrt{3}/3$):

$$p_*(x) = \frac{(2 - R)}{4} e^R + \frac{(2 + R)}{4} e^{-R} + x\left(\frac{(-1 + 9R)}{4} e^R - \frac{(1 + 9R)}{4} e^{-R}\right)$$

$$+ x^2\left(\frac{3R}{4} e^R - \frac{3R}{4} e^{-R}\right) + x^3\left(\frac{(3 - 9R)}{4} e^R + \frac{(3 + 9R)}{4} e^{-R}\right).$$

Table I gives the results (rounded to five decimal places) of solving the problem on the finite point subset X_m of $[-1, 1]$ by the revised simplex method. According to the theory in Section 2 the solution on X_m tends to the solution on $[-1, 1]$ as $m \to \infty$. Observe that with a mesh size of .002 ($m = 1000$) each parameter of the solution on X_m agrees with the corresponding parameter of the solution on $[-1, 1]$ to four decimal places.

TABLE I

m	20	80	200	1000	*Exact Solution*
$m/b - a$	.1	.025	.01	.002	
a_1	.99244	.99493	.99501	.99527	.99527
a_2	.99850	.99900	.99901	.99906	.99906
a_3	.53619	.52926	.59901	.52821	.52824
a_4	.17386	.17247	.17244	.17228	.17229
# *iterations*	6	10	13	19	

Example 2. $f(x) = x^4$.

Since $f^{(4)}(x) \geqq 0$ on $[-1, 1]$ the exact solution can be obtained as in Example 1. It is $p_*(x) = -1/9 + 2x^2/3$. Convergence is not as rapid as in Example 1.

TABLE II

m	20	80	200	1000	*Exact Solution*
$m/b-a$	.1	.025	.01	.002	
a_1	−.17634	−.11902	−.11710	−.11084	−.11111
a_2	0	0	0	0	0
a_3	.85000	.69062	.68450	.66586	.66667
a_4	0	0	0	0	0
# *iterations*	6	10	13	18	

Example 3.

$$f(x) = \exp(x), \qquad -1 \leqq x < 0,$$
$$= 1 + x, \qquad 0 \leqq x \leqq 1.$$

f is differentiable so the best approximation from below is unique. However, f'' does not exist at zero; hence no explicit solution is available.

TABLE III

m	20	80	200	1000
$m/b-a$	.1	.025	.01	.002
a_1	.99569	.99588	.99584	.99586
a_2	.92841	.93212	.93188	.93207
a_3	.18705	.18588	.18598	.18590
a_4	−.11355	−.11824	−.11794	−.11818
# *iterations*	4	7	8	13

Example 4.

$$f(x) = \exp(-x^2), \qquad -1 \leqq x < 0,$$
$$= 1 + x^2, \qquad 0 \leqq x < 1.$$

Again, f' exists but f'' does not.

TABLE IV

m	20	80	200	1000
$m/b-a$	.1	.025	.01	.002
a_1	.94370	.95465	.95464	.95463
a_2	.38473	.35291	.35298	.35299
a_3	.20238	.17109	.17098	.17100
a_4	.46919	.52135	.52140	.52139
# *iterations*	6	9	8	14

This procedure demonstrates the utility of linear programming in solving problems in approximation theory, see also [10]. Since the constraint matrix is "thin", one dimension being the number of parameters in the approximating function, storage problems were not encountered. Computing time for $m = 1000$ was about 30–40 seconds.

Acknowledgement. This research forms part of the author's doctoral thesis at Brown University. The author would like to thank Professor Philip J. Davis for his helpful advice and guidance.

Department of Mathematics
University of Rhode Island
Kingston, Rhode Island 02881

1. R. Bojanic & R. DeVore, "On polynomials of best one sided approximation," *Enseignement Math.* (2), v. 12, 1966, pp. 139–164. MR **35** #4647.
2. P. J. Davis, *Interpolation and Approximation*, Blaisdell, Waltham, Mass., 1963. MR **28** #393.
3. R. DeVore, "One-sided approximation of functions," *J. Approximation Theory*, v. 1, 1969, pp. 11–25.
4. S. Gass, *Linear Programming: Methods and Applications*, McGraw-Hill, New York, 1958. MR **20** #3037.
5. G. Hadley, *Linear Programming*, Addison-Wesley Series in Industrial Management, Addison-Wesley, Reading, Mass., 1962. MR **24** #B1669.
6. W. Kammerer, *Optimal Approximations of Functions: One-Sided Approximation and Extrema Preserving Approximations*, Doctoral Thesis, Univ. of Wisconsin, Madison, Wisconsin, 1959.
7. P. J. Laurent, "Approximation uniforme des fonctions continues sur un compact avec constraintes de type inégalité," *Rev. Française Informat. Recherche Opérationnelle*, v. 1, 1967, no. 5, pp. 81–95. MR **37** #1855.
8. J. T. Lewis, *Approximation with Convex Constraints*, Doctoral thesis, Brown University, Providence, R. I., 1969.
9. T. S. Motzkin & J. L. Walsh, "The least *p*th power polynomials on a finite point set," *Trans. Amer. Math. Soc.*, v. 83, 1956, pp. 371–396. MR **18**, 479.
10. P. Rabinowitz, "Applications of linear programming to numerical analysis," *SIAM Rev.*, v. 10, 1968, pp. 121–159. MR **37** #2397.
11. T. J. Rivlin, *An Introduction to the Approximation of Functions*, Blaisdell, Waltham, Mass., 1969.