The Lawson Algorithm and Extensions

By John R. Rice* and Karl H. Usow†

1. Introduction and Background. The primary objective of this paper is to present the Lawson algorithm for computing best Tchebycheff (L_∞) approximations in the general mathematical literature. Other objectives are to present some extensions of the algorithm, to discuss some possible modifications of it and to report on some computational experience.

This interesting algorithm has been proposed on heuristic grounds by several individuals, but the only thorough analysis of it is contained in Lawson's thesis [1]. A simplified version of that analysis is to appear in [3]. Thus we do not present any of Lawson's proofs here, but only state some of his results. Lawson's original algorithm computes best Tchebycheff approximations as the limit of a special sequence of best weighted L_p approximations *with p fixed*. The interesting case is for $p = 2$. We extend this algorithm to compute L_p approximations for $2 < p < \infty$ as the limit of best weighted L_2 approximations. This extension is defined and convergence established in the next section. The final two sections discuss some modifications of this algorithm and report on some computational experience with both the original and extended version. In particular, a useful convergence acceleration scheme is presented for the original algorithm.

The possibility that such algorithms might exist follows from the work of Motzkin and Walsh [2]. Their work in this area is presented in detail in [3]. The following theorems summarize some results pertinent to this paper.

THEOREM 1 (MOTZKIN AND WALSH). *Let $\{\phi_i(x)\}$ be a Tchebycheff set** and define*

$$L(A, x) = \sum_{i=1}^{n} a_i\phi_i(x) ,$$

where A denotes the parameter vector $(a_1, a_2, \cdots, a_n)$.

Then, given $f(x)$ continuous on [0, 1] *and $1 < q < p \leqq \infty$, we have three pairs of identical sets:*

1. $\{A \mid L(A, x)$ *is a best weighted L_p approximation to $f(x)$ on* $[0, 1]\}$
   $\{A \mid L(A, x)$ *strongly interpolates $f(x)$ on* $[0, 1]\}$
2. $\{A \mid L(A,x)$ *is a best weighted L_1 approximation to $f(x)$ on* $[0, 1]\}$
   $\{A \mid L(A, x)$ *weakly interpolates $f(x)$ on* $[0, 1]\}$
3. $\{A \mid L(A, x)$ *is a best weighted L_p approximation to $f(x)$ on* $[0, 1]\}$
   $\{A \mid L(A, x)$ *is a best weighted L_q approximation to $f(x)$ on* $[0, 1]\}$.

THEOREM 2. *The conclusions of Theorem* 1 *are valid if the interval* [0, 1] *is replaced by a finite point set $X \subset [0, 1]$ upon which the approximation is made.*

Received April 17, 1967.

* Supported in part by the Air Force Office of Scientific Research under AF49(639)-1616.

† Supported in part by the National Science Foundation under Grant GP-4052.

** The set $\{\phi_i(x)\}$, $i = 1, 2, \cdots, n$, is a Tchebycheff set if the matrix $(\phi_i(x))$, $i, j = 1, 2, \cdots, n$ is nonsingular for arbitrary distinct $x_j \in [0, 1]$.

Actually only the third result is pertinent here, but the first two results are presented as they are not widely known among numerical analysts. $L(A, x)$ is said to strongly (weakly) interpolate $f(x)$ n times if

$$(-1)^i[L(A, x_i) - f(x_i)] > 0 \qquad (\text{or } (-1)^i[L(A, x_i) - f(x_i)] \geqq 0)$$

for some $n + 1$ points x_i in the interval [0, 1].

From the third conclusion of these theorems, we see that we can compute best Tchebycheff approximations by computing a certain weighted least-squares approximation. This is inviting, as the second computation is substantially simpler than the first. Furthermore, there are several areas (vector-valued functions and functions of a complex variable) where there are no known algorithms for L_∞ approximation, but where least squares can be used. Lawson's algorithm consists, then, of generating the required weight function. We *only* consider approximation on *a finite point set* X.

We wish to approximate the values $f(x_i) = f_i$, $i = 1, 2, \cdots, m$, on the set $X = \{x_i | i = 1, 2, \cdots, m\}$ by

$$L(A, x) = \sum_{j=1}^{n} a_j \phi_j(x)$$

where $\{\phi_j(x)\}$ is a Tchebycheff set.

Lawson's Algorithm for L_∞ Approximation. We define a sequence of weight functions $w^k(x_i) = w_i^k$ with $\sum_X w_i^k = 1$ and corresponding approximations $L(A_k, x)$ as follows. Choose $w_i^1 > 0$ arbitrary.

a. *$L(A_k, x)$ is the best L_2 approximation to $f(x)$ on X with the weights w_i^k.*

b. $w_i^{k+1} = w_i^k |f(x_i) - L(A_k, x_i)| / \sum_X w_i^k |f(x_i) - L(A_k, x_i)|$.

THEOREM 3 (LAWSON). *The sequence $L(A_k, x)$ converges to $L(A_0, x)$, which is the best L_∞ approximation to $f(x)$ on a set $X_2 \subset X$. The sequence $\{\sigma^k\}$*

$$\sigma^k = \left[\sum_X w_i^k [f(x) - L(A_k, x)]^2\right]^{1/2}$$

is monotonically increasing (strictly so unless convergence takes place in a finite number of steps), and

$$\lim_{k\to\infty} \sigma^k = \max_{x \in X_2} |f(x) - L(A_0, x)| = \sigma^* .$$

THEOREM 4 (LAWSON). *If X_2 is a proper subset of X, then the algorithm may be restarted with*

$$\overline{w}_i^1 = (1 - \lambda) \lim_{k\to\infty} w_i^k + \lambda u(x) , \qquad 0 \leqq \lambda < 1 ,$$

where $u(x) = 0$ for $x \neq z$ and $u(z) = 1$, where $z \in X - X_2$ and $|f(z) - L(A_0, z)| > \sigma^*$. For λ sufficiently small $\bar{\sigma}^1 > \sigma^*$ and after a finite number of restarts, we obtain the best L_∞ approximation to $f(x)$ on X.

In practice we use the last weight function actually calculated rather than $\lim_{k\to\infty} w_i^k$. The fact that the algorithm must be restarted sometimes is not as serious as it first appears, as it is very rare that this occurs. The proofs of these theorems are not easy, and it is an open question whether they remain true if the interval [0, 1] replaces the finite set X.

2. The Lawson Algorithm for L_p Approximation. We define a sequence of weight functions w_i^k with $\sum_X w_i^k = 1$ and corresponding approximations as follows. Choose $w_i^1 > 0$ arbitrary.

a. *$L(A_k, x)$ is the best L_2 approximation to $f(x)$ on X with the weight w_i^k.*

b. $w_i^{k+1} = (w_i^k|e_i^k|)^{(p-2)/(p-1)} / \sum_X (w_i^k|e_i^k|)^{(p-2)/(p-1)}$ *where* $e_i^k = e^k(x_i) = f(x_i) - L(A_k, x_i)$.

Note that the formula in b restricts us to $p > 2$ and, that as p tends to infinity, we obtain Lawson's original algorithm.

We now establish five lemmas in preparation for the proof of the convergence theorem. We introduce the notation

$$\sigma^k = \left[\sum w_i^k|e_i^k|^2\right]^{1/2} / \left[\sum (w_i^k)^{p/(p-2)}\right]^{(p-2)/2p} .$$

All summations are over the set X unless otherwise indicated. Define $W_k = \{x_i | w_i^k > 0\}$.

LEMMA 1. *If $\sigma^1 > 0$, then $\sigma^k > 0$ for all k.*

Proof. The proof is by induction; i.e., assume that $\sigma^k > 0$. This implies that W_k is not empty. If $W_{k+1} = W_k$, then no function $L(A, x)$ agrees with $f(x)$ on $W_{k+1} = W_k$, and hence $\sigma^{k+1} > 0$. If $W_k - W_{k+1}$ is not empty, then it is seen that $L(A_k, x)$ is a best L_2 approximation to $f(x)$ on W_{k+1} as well as W_k with the weight $w^k(x)$. Again, $\sigma^k > 0$ implies that $f(x)$ is not of the form $L(A, x)$ on W_{k+1} and $\sigma^{k+1} > 0$. This concludes the proof.

This lemma implies, by the Tchebycheff set assumption, that $\bigcap W_k$ contains at least $n + 1$ points. For the remainder of this discussion we assume that $\sigma^1 > 0$.

LEMMA 2. *If $w_i^{k+1} = w_i^k$ for all i, then $\sigma^{k+1} = \sigma^k$, otherwise $\sigma^{k+1} > \sigma^k$.*

Proof. We introduce the inner product notation

$$(f, g)_w = \sum w(x_i) f(x_i) g(x_i) .$$

The first assertion is clear, therefore we assume $w^{k+1}(x) \neq w^k(x)$. Since

$$\sum w_i^{k+1} e_i^{k+1} \phi_j(x_i) = 0 \quad \text{for } j = 1, 2, \cdots, n ,$$

we have

$$\sigma^{k+1} = \frac{\sum f_i e_i^{k+1} w_i^{k+1}}{\left[\sum |e_i^{k+1}|^2 w_i^{k+1}\right]^{1/2} \left[\sum (w_i^{k+1})^{p/(p-2)}\right]^{(p-2)/2p}} . \tag{1}$$

Consider $g_i = e_i^{k+1} / [\sum |e_i^{k+1}|^2 w_i^{k+1}]^{1/2}$ and recall that it is a property of least-squares approximation that

1. $(g, g)_{w^{k+1}} = 1$.

2. $g \perp [\phi_1, \cdots, \phi_n]$ in the L_2 norm with weight w^{k+1}. Here $[\phi_1, \cdots, \phi_n]$ denotes the linear subspace spanned by the $\{\phi_j\}$.

3. g maximizes $\sum f_i g_i w_i^{k+1}$ over all g satisfying 1 and 2. Since $\sum \phi_j(x_i) e_i^k w_i^k = 0$ for $j = 1, 2, \cdots, n$, we have

$$\sum \phi_j(x_i) (e_i^k w_i^k / w_i^{k+1}) w_i^{k+1} = 0 , \quad \text{for } w_i^{k+1} > 0 .$$

Note that with $\phi_j = \phi_j(x_i)$,

$$0 = \sum \phi_j e_i^k w_i^k = \sum_W \phi_j e_i^k w_i^k = \sum_{W_{k+1}} \phi_j e_i^k w_i^k .$$

Let

$$\hat{g}_i = (e_i^k w_i^k / w_i^{k+1}) / \sum (e_i^k w_i^k)^2 / w_i^{k+1} \quad \text{for } w_i^{k+1} > 0 ,$$
$$= 0 \quad \text{otherwise} .$$

Then $\hat{g}_i$ satisfies 1 and 2 above. Hence, replacing g by $\hat{g}$ in (1) does not increase the left-hand side. Thus we have

$$\sigma^{k+1} \geqq \frac{\sum f_i e_i^k w_i^k}{[\sum (e_i^k w_i^k)^2 / w_i^{k+1}]^{1/2} [\sum (w_i^{k+1})^{p/(p-2)}]^{(p-2)/2p}} . \tag{2}$$

Now substitute for w_i^{k+1} in the denominator terms of (2) as given by the recurrence relation to obtain

$$[\sum (w_i^{k+1})^{p/(p-2)}]^{(p-2)/2p} = \frac{[\sum (w_i^k |e_i^k|)^{p/(p-1)}]^{(p-2)/2p}}{[\sum (w_i^k |e_i^k|)^{(p-2)/(p-1)}]^{1/2}}$$

and

$$\left[\sum \frac{(e_i^k w_i^k)^2}{w_i^{k+1}}\right]^{1/2} = [\sum (|e_i^k| w_i^k)^{p/(p-1)}]^{1/2} [\sum (|e_i^k| w_i^k)^{(p-2)/(p-1)}]^{1/2} .$$

Combining these factors, we see the denominator of (2) is

$$[\sum (w_i^k |e_i^k|)^{p/(p-1)}]^{(p-1)/p} . \tag{3}$$

We apply Hölder's inequality

$$\sum a_i b_i \leqq [\sum a_i^r]^{1/r} [\sum b_i^s]^{1/s}$$

with

$$a_i = |e_i^k|^{p/(p-1)} (w_i^k)^{p/2(p-1)} , \qquad r = 2(p-1)/p ,$$
$$b_i = (w_i^k)^{p/2(p-1)} , \qquad s = 2(p-1)/(p-2) ,$$

and obtain

$$[\sum (w_i^k |e_i^k|)^{p/(p-1)}]^{(p-1)/p} \leqq [\sum |e_i^k|^2 w_i^k]^{1/2} [\sum (w_i^k)^{p/(p-2)}]^{(p-2)/2p}$$

with equality if and only if there is an $\alpha > 0$ such that

$$|e_i^k|^2 w_i^k = \alpha (w_i^k)^{p/(p-2)} \quad \text{for all } i .$$

That is to say, if and only if

$$|e_i^k| = \sqrt{\alpha} (w_i^k)^{1/(p-2)} \quad \text{for all } i .$$

But if this were the case, we would have

$$w_i^{k+1} = \frac{(w_i^k |e_i^k|)^{(p-2)/(p-1)}}{\sum (w_i^k |e_i^k|)^{(p-2)/(p-1)}} = \frac{w_i^k}{\sum w_i^k} = w_i^k ,$$

which contradicts the assumption that $w^{k+1}(x) \neq w^k(x)$. Hence we have strict inequality and

$$\sigma^{k+1} > \frac{\sum f_i e_i^k w_i^k}{[\sum |e_i^k|^2 w_i^k]^{1/2} [\sum (w_i^k)^{p/(p-2)}]^{(p-2)/2p}} = \sigma^k .$$

LEMMA 3. *Let $L(A^*, x)$ be the best L_p approximation to $f(x)$ on X. Then*

$$\sigma^k \leqq \zeta^* = \left[\sum |f_i - L(A^*, x_i)|^p\right]^{1/p}.$$

Proof. We have

$$(\sigma^k)^2 = \sum w_i^k |e_i^k|^2 / \left[\sum (w_i^k)^{p/(p-2)}\right]^{(p-2)/p}$$
$$\leqq \sum w_i^k |f_i - L(A^*, x_i)|^2 / \left[\sum (w_i^k)^{p/(p-2)}\right]^{(p-2)/p}.$$

Again we apply Hölder's inequality with

$$a_i = w_i^k, \qquad r = p/(p-2),$$
$$b_i = |f_i - L(A^*, x_i)|^2, \qquad s = p/2,$$

and obtain

$$\sigma^k \leqq \left[\sum |f_i - L(A^*, x_i)|^p\right]^{1/p} \left[\sum (w_i^k)^{p/(p-2)}\right]^{(p-2)/2p} / \left[\sum (w_i^k)^{p/(p-2)}\right]^{(p-2)/2p} = \zeta^*.$$

For the next lemma we set

$$\sigma^* = \lim_{k\to\infty} \sigma^k$$

and define the set W_0 as follows: All of the sequences w_i^k lie in a bounded region of E_m. Therefore, $\{w_i^k\}$ has one or more convergent subsequences as k tends to ∞. Pick one such subsequence (also denoted by $\{w_i^k\}$) and set

$$w_i^0 = w^0(x_i) = \lim_{k\to\infty} w_i^k, \qquad W_0 = \{x | w^0(x_i) > 0\}.$$

LEMMA 4. *Let $L(A_0, x)$ be the best weighted L_2 approximation to $f(x)$ on W_0 (and hence on X). Then $\sigma^0 > 0$ and*

$$\lim_{k\to\infty} L(A_k, x) = L(A_0, x).$$

Proof. It is known [1] that the error of the best L_2 approximation is a continuous function of the weights and hence so is σ^k. Thus

$$\sigma^0 = \sigma^* = \lim_{k\to\infty} \sigma^k.$$

We have $\sigma^1 > 0$ by assumption, and it follows from Lemma 1 that $\sigma^0 > 0$. Since $\{\phi_i(x)\}$ is a Tchebycheff set, the set W_0 must contain at least $n + 1$ points. This implies that the best weighted L_2 approximation to $f(x)$ on W_0 is also a continuous function of the weights, and the lemma is established.

LEMMA 5. *$L(A_0, x)$ is also the best L_p approximation to $f(x)$ on W_0 and*

$$\sigma^* = \left[\sum_{W_0} [f(x_i) - L(A_0, x_i)]^p\right]^{1/p}.$$

Proof. We start the algorithm with $w_i^0 = w^0(x_i)$. Then by Lemma 2 either $w^1(x) = w^0(x)$ or $\sigma^1 > \sigma^*$. We have $\lim_{k\to\infty} w^k(x) = w^0(x)$, $\lim_{k\to\infty} \sigma^k = \sigma^*$, and $\sigma^{k+1}(w_i^k)$ is a continuous function of w_i^k. Hence $\sigma^1(w^0) = \sigma^*$, otherwise σ^k does not converge to σ^*. Therefore $w^1(x) = w^0(x)$. Thus by the recurrence relation

$$w_j^1 = w_j^0 = \frac{(w_j^0 |e_j^0|)^{(p-2)/(p-1)}}{\sum (w_i^0 |e_i^0|)^{(p-2)/(p-1)}}. \tag{4}$$

We solve (4) for the w_j^0 to obtain for $x_j \in W_0$

(5) $$w_j^0 = |e_j^0|^{p-2} \Big/ \left[\sum_{W_0} (w_i^0 |e_i^0|)^{(p-2)/(p-1)} \right]^{p-1} .$$

Now, since $L(A_0, x)$ is the best L_2 approximation with weight $w^0(x)$ (on both X and W_0), we have

$$\sum_{W_0} w_i^0 e_i^0 L(A, x_i) = 0 \qquad \text{for all } A .$$

We substitute for w_i^0 as given in (5) and multiply out the denominator to obtain

$$\sum_{W_0} |e_i^0|^{p-2} e_i^0 L(A, x_i) = 0 .$$

Since

$$|e_i^0|^{p-2} e_i^0 = |e_i^0|^{p-1} \operatorname{sgn} [e_i^0] ,$$

we have

$$\sum_{W_0} |e_i^0|^{p-1} \operatorname{sgn} [e_i^0] L(A, x) = 0 \qquad \text{for all } A ,$$

which is precisely the condition for $L(A_0, x)$ to be the best L_p approximation to $f(x)$ on W_0. This establishes the first part of the lemma.

We have

$$\sigma^* = \operatorname*{Lim}_{k\to\infty} [\sum |e_i^k|^2 w_i^k]^{1/2} / [\sum (w_i^k)^{p/p-2}]^{(p-2)/2p} .$$

We know that σ^k depends continuously on the weights $w^k(x)$ and hence

$$\sigma^* = [\sum |e_i^0|^2 w_i^0]^{1/2} / [\sum (w_i^0)^{p/(p-2)}]^{(p-2)/2p} .$$

We substitute (5) into the numerator and denominator of this equation to obtain, respectively

$$\left[\sum_{W_0} |e_i^0|^p \right]^{1/2} \Big/ \left[\sum_{W_0} (w_i^0 |e_i^0|)^{(p-2)/(p-1)} \right]^{(p-1)/2} ,$$

$$\left[\sum_{W_0} |e_i^0|^p \right]^{(p-2)/2p} \Big/ \left[\sum_{W_0} (w_i^0 |e_i^0|)^{(p-2)/(p-1)} \right]^{(p-1)/2} .$$

Thus we have

$$\sigma^* = \left[\sum_{W_0} |e_i^0|^p \right]^{1/2} \Big/ \left[\sum_{W_0} |e_i^0|^p \right]^{(p-2)/2p} = \left[\sum_{W_0} |e_i^0|^p \right]^{1/p}$$
$$= \left[\sum_{W_0} [f(x_i) - L(A_0, x_i)] \right]^{1/p} .$$

This concludes the proof.

In the last two lemmas we only considered a particular subsequence of $\{L(A_k, x)\}$ and its corresponding limit. We now establish the major convergence theorem related to the entire sequence $\{L(A_k, x)\}$.

THEOREM 5. *The sequence* $\{L(A_k, x)\}$ *converges to* $L(A_0, x)$, *which is the best* L_p *approximation to* $f(x)$ *on* W_0.

Proof. We first establish that

(6) $$\operatorname*{Lim}_{k\to\infty} w_i^{k+1} - w_i^k = 0 .$$

Assume the contrary. Then there is a subsequence, denoted by $\{\bar{w}_i^{k+1} - \bar{w}_i^k\}$, which converges to a nonzero limit. Let $\{w_i^l\}$ be a subsequence of $\{\bar{w}_i^k\}$ which converges to $w^0(x)$ as in Lemma 5. We know that if the algorithm is started with $w_i^1 = w_i^0$, then we have $\sigma^2 = \sigma^0$ and $w_i^2 = w_i^0$. Therefore

$$\lim_{l\to\infty} w_i^{l+1} = \lim_{l\to\infty} (w_i^l|e_i^l|)^{(p-2)/(p-1)}/\sum (w_i^l|e_i|^l)^{(p-2)/(p-1)}$$
$$= (w_i^0|e_i^0|)^{(p-2)/(p-1)}/\sum (w_i^0|e_i^0|)^{(p-2)/(p-1)} = w_i^0 .$$

Thus for any convergent subsequence of $\bar{w}_i^k$, we have that $\bar{w}_i^{k+1} - \bar{w}_i^k$ converges to zero, which then must be true for the entire sequence.

Denote by $\mathcal{W}$ the limit points of $\{w^k(x)\}$ in E_m. It is clear that $\mathcal{W}$ is not empty, closed and bounded, i.e., $\mathcal{W}$ is compact. Furthermore (6) shows that $\mathcal{W}$ is connected. We now assert that every $w(x) \in \mathcal{W}$ gives the same best L_p approximation to $f(x)$.

The set $\mathcal{W}$ may be decomposed into equivalence classes by defining two weight functions to be equivalent if they lead to the same approximation. If $L(A, x)$ is a best L_2 approximation to $f(x)$ with weight $w(x)$, then it is the unique best L_p approximation to $f(x)$ on the set W_0 where $w(x) > 0$. This follows from Lemma 5. Since X is finite, there are at most a finite number of equivalence classes, each of which is compact and distinct. Since $\mathcal{W}$ is connected, there is at most one such equivalence class, and every $w(x) \in \mathcal{W}$ gives the same best L_p approximation to $f(x)$.

To complete the proof we note that $\{L(A_k, x)\}$ is bounded and hence contains convergent subsequences. If there are two such subsequences with different limits, then consider the corresponding sequences of weight functions. These sequences have convergent subsequences, which, as just established, lead to the same weighted L_2 approximation. This is impossible and shows that $\{L(A_k, x)\}$ converges, say to $L(A_0, x)$. It follows from Lemma 5 that $L(A_0, x)$ is the best L_p approximation to $f(x)$ on W_0. This concludes the proof.

There is a distinct difference between the L_∞ and L_p Lawson algorithms as follows. The L_∞ algorithm tends to drive the weight function $w^k(x)$ to zero everywhere but at the critical points of the error curve of the best L_∞ approximation. This implies that the analogous set W_0 in Theorem 5 does not contain many points. This is not the case for the L_p algorithm, and normally the set W_0 is all of X. However, it is possible that the error curve "accidentally" becomes zero at a point x_0 of X in the early stages of the algorithm. This means that this x_0 does not belong to W_0, and hence $L(A_0, x)$ might not be a best L_p approximation to $f(x)$ on X.

If this occurs, then the Lawson algorithm can be restarted with a specific choice for $w^1(x)$ which ensures that larger values for σ are obtained. Since X is finite, one must obtain $L(A^*, x)$ after a finite number of restarts. This is established in

THEOREM 6. *If W_0 is a proper subset of X, then the algorithm may be restarted with*

$$w_i^\lambda = (1 - \lambda)w^0(x) + \lambda u(x) , \qquad 0 \leqq \lambda < 1 ,$$

where $u(x) = 0$ for $x \neq z$ and $u(z) = 1$, where $z \in X - W_0$ and $L(A_0, z) - f(z) \neq 0$. For λ sufficiently small, we have

$$\sigma^1 > \sigma^*$$

and after a finite number of restarts we obtain the best L_p approximation $L(A^*, x)$ to $f(x)$ on X.

Proof. Let us denote by $L(A_\lambda, x)$ the best L_2 approximation to $f(x)$ on X (also on $W_0 \cup \{z\}$) with the weight function w_i^λ. Set

$$e_i^\lambda = (f(x_i) - L(A_\lambda, x_i))$$

and denote the corresponding σ value by

$$\sigma(\lambda)^2 = \sum w_i^\lambda |e_i^\lambda|^2 / [\sum (w_i^\lambda)^{p/(p-2)}]^{(p-2)/p} .$$

Now

$$\sigma^2(\lambda) = \frac{\lambda |e_i^\lambda(z)|^2 + (1-\lambda) \sum_{W_0} w_i^0 |e_i^\lambda|^2}{[\lambda^{p/(p-2)} + (1-\lambda)^{p/(p-2)} \sum_{W_0} (w_i^0)^{p/(p-2)}]^{(p-2)/p}} .$$

For λ sufficiently small, say $0 < \lambda \leqq \lambda_0$, we have that $L(A_\lambda, x)$ and $L(A_0, x)$ are arbitrarily close, and hence $|e^\lambda(z)| > 0$. Furthermore, we have

$$\sum_{W_0} w_i^0 |e^\lambda|^2 \geqq \sum_{W_0} w_i^0 |e_i^0|^2$$

and hence, after manipulation,

$$\begin{aligned} \sigma^2(\lambda) \geqq & \frac{\sum_{W_0} w_i^0 |e_i^0|^2}{[\sum_{W_0} (w_i^0)^{p/(p-2)}]^{(p-2)/p} (1 + (\lambda/(1-\lambda))^{p/(p-2)})^{(p-2)/p}} \\ & + \frac{\lambda |e^\lambda(z)|^2}{[\sum_{W_0} (w_i^0)^{p/(p-2)}]^{(p-2)/p} (1 + (\lambda/(1-\lambda))^{p/(p-2)})^{(p-2)/p}} \\ \geqq & \frac{(\sigma^*)^2 + \lambda v^2(\lambda)}{[1 + (\lambda/(1-\lambda))^{p/(p-2)}]^{(p-2)/p}} \\ \geqq & \frac{(\sigma^*)^2 + \lambda v^2(\lambda)}{1 + (p-2)(\lambda)/p^{p/(p-2)} + O(\lambda^{2p/(p-2)})} \end{aligned}$$

where

$$v(\lambda) = \frac{|e^\lambda(z)|}{[\sum_{W_0} (w_i^0)^{p/(p-2)}]^{(p-2)/p}} .$$

Since $v(\lambda)$ is not zero for $0 < \lambda \leqq \lambda_0$ and $p/(p-2) > 1$, we have, for λ sufficiently small,

$$\sigma^2(\lambda) > (\sigma^*)^2 .$$

For any specific choice of λ in this range we have that $\sigma^1 = \sigma(\lambda)$, and hence the first relation is established.

Thus the second start of Lawson's algorithm generates another approximation, say $L(A_{01}, x)$, a corresponding σ_1^* and W_1 where $\sigma_1^* > \sigma^*$. Since X is finite, there are only a finite number of possibilities for the set W_j, $j = 0, 1, \cdots$. One of these corresponds to the best L_p approximation $L(A^*, x)$. Since the σ_j^* values obtained are strictly increasing, the sets W_j are distinct and the last statement of the theorem follows.

3. Modifications and Acceleration of the Lawson Algorithm. One can "accidentally" set $w^k(x_i) = 0$ in both the L_∞ and L_p versions of the Lawson algorithm.

This might prevent one from obtaining a best approximation, and hence one can consider modifying the algorithm (particularly in the early stages) so as to avoid this. Two possible modifications are

$$(7)\qquad \begin{aligned} w^{k+1}(x) &= w^k(x)(|e^k(x)|)^{(p-2)/(p-1)}, & e^k(x) &\neq 0, \\ &= w^k(x), & e^k(x) &= 0, \end{aligned}$$

or

$$(8)\qquad \begin{aligned} w^{k+1}(x) &= w^k(x)(|e^k(x)|)^{(p-2)/(p-1)}, & e^k(x) &\neq 0, \\ &= a(k), & e^k(x) &= 0. \end{aligned}$$

(The normalizing factors are omitted from (7) and (8) for simplicity.) In (8) one might consider for $a(k)$ functions like $1/k$, $1/k^4$, 2^{-k}, etc. The convergence proofs break down (in Lemma 2) for both of these modifications. In view of the rarity of these accidents observed so far, it is probably more efficient to use the restarting procedure rather than make such a modification.

While one wants to prevent setting $w^k(x) = 0$ by "accident," one is interested in the L_∞ algorithm with making $w^k(x)$ tend to zero as rapidly as possible except at the extremal points of the error curve of the best L_∞ approximation. At those points where $|e^*(x)|$ is nearly maximum, the corresponding weights do not tend to zero very rapidly. Indeed, set

$$\rho^* = \max\,[\rho = |e^*(x)|/\max_{x\in X}|e^*(x)|\,,\rho < 1]$$

then Lawson reports that the algorithm converges linearly with ratio ρ^*. We also have observed this and ρ^* is usually rather close to 1. This is slow convergence and leads one to look for convergence acceleration schemes. Modifications which might make $w^k(x)$ tend to zero faster for the L_∞ case are

$$(9)\qquad w^{k+1}(x) = w^k(x)|e^k(x)|^2\,,$$

$$(10)\qquad w^{k+1}(x) = (w^k(x))^2|e^k(x)|\,.$$

These modifications make $w^k(x)$ tend to zero like $(\rho^*)^{2k}$ and $(\rho^*)^{2^k}$, respectively (if the algorithm converges). It has been observed by Lawson and us that (9) sometimes leads to divergence. However, when it does converge, we observe that it does accelerate the convergence.

We have found the following acceleration scheme effective:

1. Do l steps of the Lawson algorithm.
2. Set $w_i^k = 0$ if

$$|f_i - L(A_k, x_i)| \leqq \lambda_k\sigma^k, \quad \text{where} \quad \lambda_k = \sigma^k/\max_x |f(x) - L(A_k, x)|\,.$$

3. Go back to step 1.

One may verify that the convergence proofs of Section 2 are valid if the L_p algorithm is defined by

$$(11)\qquad w_i^{k+1} = (w_i^k)^\alpha|e_i^k|^\beta/\sum (w_i^k)^\alpha|e_i^k|^\beta\,,$$

where α and β are positive and satisfy

$$\alpha(p-2) + \beta = p - 2\,.$$

The choice presented in Section 2 corresponds to

$$\alpha = \beta = (p - 2)/(p - 1).$$

4. Computational Remarks. We first discuss the L_p algorithm.

A. The method presented in Section 2 was compared to the special case of $\alpha = \frac{1}{2}\beta$ in (11), i.e.,

$$w_i^{k+1} = (w_i^k e_i^2)^{(p-2)/p} / \sum (w_i^k e_i^2)^{(p-2)/p}.$$

Our experience indicates that this case converges somewhat slower than the $\alpha = \beta$ case.

B. For typical functions and ranges of $p \leqq 20$ the algorithm converged so that

$$\left[\sum |f(x) - L(A_k, x)|^p\right]^{1/p}$$

agreed to 5 or 6 digits of the best value within 15 iterations or less. For larger values of p, e.g., $p = 100$, $p = 1000$ the convergence is slower.

C. A useful convergence criterion is

$$|(\sigma^k - [\sum |e_i^k|^p]^{1/p})/\sigma^k| \leqq \epsilon.$$

In general we observed that

$$\left[\sum |e^k|^p\right]^{1/p} - \left[\sum |e^*|^p\right]^{1/p} \ll \left[\sum |e^*|^p\right]^{1/p} - \sigma^k.$$

For the L_∞ case we observed the following.

A. Without acceleration the convergence is slow, as indicated by Lawson.

B. For a typical problem involving $n = 4$ parameters and $m = 50$ points, the acceleration scheme reduced the number of iterations from over 250 to less than 15 using values of l with $1 \leqq l \leqq 4$. This is for convergence to 7 significant digits.

C. An increase in the number n of parameters or number m of points increases the number of iterations required. Typically, $n = 10$ and $m = 100$ required about 40 iterations for 7 significant digits.

If one has a reliable least-squares approximation program, then one can write and debug a program for either one of these algorithms rather quickly (in a few days).

Division of Mathematical Sciences
Purdue University
Lafayette, Indiana

Department of Mathematics
The University of Colorado
Boulder, Colorado

1. C. L. Lawson, *Contributions to the Theory of Linear Least Maximum Approximations*, Ph.D. Thesis, UCLA, 1961.

2. T. S. Motzkin & J. L. Walsh, "Polynomials of best approximation on an interval," *Proc. Nat. Acad. Sci. U.S.A.*, v. 45, 1959, pp. 1523–1528.

3. T. S. Motzkin & J. L. Walsh, "Polynomials of best approximation on a real finite point set I," *Trans. Amer. Math. Soc.*, v. 91, 1959, pp. 231–245. MR 21 #7388.

4. J. R. Rice, *The Approximation of Functions*, Vol. II, Addison-Wesley, Reading, Mass., 1968. (To appear.)