JOURNAL OF APPROXIMATION THEORY 6, 231–241 (1972)

A Descent Algorithm for Linear Continuous Chebyshev Approximation

P. D. SCOTT

Department of Electrical Engineering, State University of New York at Buffalo, Buffalo, New York 14214

AND

J. S. THORP

Department of Electrical Engineering, Cornell University, Ithaca, NY 14850

Communicated by John R. Rice

DEDICATED TO PROFESSOR J. L. WALSH ON THE OCCASION OF HIS 75TH BIRTHDAY

This paper presents a direct, parameter-space descent algorithm for the linear continuous Chebyshev approximation problem. After suitable definition and characterization of edges and vertices, the search proceeds on a vertex-to-vertex basis. The advantage of the procedure is its generality, since the approximating set need not be a Chebyshev set, and a somewhat quicker time-to-convergence, at least on the examples attempted, than comparison algorithms. For approximation with non-Chebyshev sets the algorithm is defined up to a stop rule.

1. INTRODUCTION

The descent method of analysis has proved useful for a wide range of minimization-type problems. Suppose a functional d is defined over a set of functions characterized by a parameter n-vector $\mathbf{A}$. In the direct product space $\mathbf{A}xd$ a mapping M is defined which takes the space into itself in the following way: M: $(\mathbf{A}_i, d_i) \to (\mathbf{A}_{i+1}, d_{i+1})$ such that $d_{i+1} \leqslant d_i$, with equality if and only if $d_i \leqslant d_j$ for all $(\mathbf{A}_j, d_j) \in \mathbf{A}xd$. If an initial state is chosen, repeated application of M may be viewed as descent along the surface of the functional until the minimum is located.

The usefulness of this approach is largely determined by the structure of d in its parameters $\mathbf{A}$. Efficient mappings have been developed for the two special cases of strict convexity and polytope structure. In the first case gradients and Hessians may be calculated, and in the latter case a vertex-to-vertex descent used. The concept of descent along edges of polytopes, introduced by Zukovskii in 1951, has been exploited successfully in the Chebyshev solution of finite linear inconsistent systems [1].

The linear continuous Chebyshev approximation problem possesses neither of these special structures. While the functional is convex in the approximating coefficients **A**, it is characterized by the existence of edges (intersections of smooth regions of the functional) which are themselves curved hypersurfaces. We will present a descent algorithm which adapts the vertex-to-vertex and steepest descent philosophies to this setting.

2. Problem Statement and Notation

Let $f(t)$, $\phi_1(t),..., \phi_n(t)$ be elements of $C[0, 1]$ (the space of realvalued continuous functions defined on the closed interval [0, 1]) and let the space be normed by

$$\| g \| \Delta \max_{t \in [0,1]} | g |.$$

Define the subspace $K \subset C[0, 1]$ which is spanned by linear sums of the basis set $\{\phi_i(t)\}$. From the subspace select the point of minimum distance (norm) from the point $f \in C[0, 1]$. The approximating function is written as

$$L(\mathbf{A}, t) = \sum_{i=1}^{n} a_i \phi_i(t)$$
$$= \mathbf{A}^T \mathbf{\Phi}(t),$$

where $\mathbf{A}^T$ and $\mathbf{\Phi}(t)^T$ are row vectors, $\mathbf{A}^T = (a_1 ,..., a_n)$, $\mathbf{\Phi}^T(t) = (\phi_1(t),..., \phi_n(t))$. The error function is then the difference

$$e(\mathbf{A}, t) = f(t) - L(\mathbf{A}, t)$$

and the solution is a point $\mathbf{A}^*$ which satisfies

$$\| e(\mathbf{A}^*, t) \| \leqslant e(\mathbf{A}, t) \| \qquad \text{for all } \mathbf{A}.$$

(All boldface symbols indicate n-vectors. Special notational conventions, such as that for the directional derivative, will be defined as they are needed.)

3. General Approach

The procedure to be followed is best introduced through an example. Suppose it is desired to find the best Chebyshev approximation (in the interval $0 \leqslant t \leqslant 1$) to the parabola $f(t) = t^2$ by the straight line

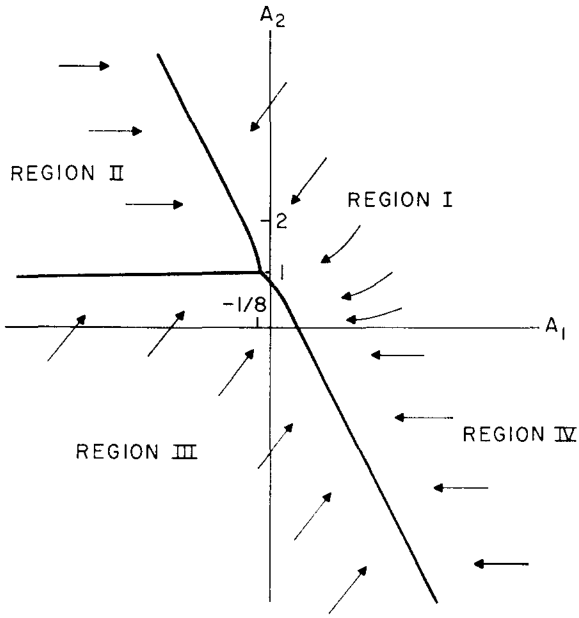

FIG. 1. Dependence of the norm on its parameters an example.

$L(A, t) = a_1 + a_2 t$. Figure 1 indicates the dependence of the maximum value of the error

$$d(\mathbf{A}) \Delta \max_{t \in [0,1]} | f(t) - L(\mathbf{A}, t) |,$$

on the approximating coefficients a_1 and a_2. The d-axis extends out of the page and arrows indicate the direction of the negative gradient. The minimum value of d occurs at $a_1 = -1/8$, $a_2 = 1$. In Region III, $d = | 1 - a_1 - a_2 |$ while in Region IV $d = | a_1 |$. In the fourth quadrant these surfaces intersect along the straight line $a_2 + 2a_1 - 1 = 0$, and this locus of the points of intersection defines an edge.

Since edges occupy lower-dimensional subspaces of $\mathbf{A} \times d$ an arbitrary initial state is likely to fall in one of the smooth regions of the functional. The initial step in the search is evaluation of the negative gradient at the initial point $\mathbf{A}_0$ and location of the nearest edge lying in that direction. The algorithm then follows the edge in a "downhill" direction (decreasing d) until the nearest vertex (intersection of edges) is located. Thereafter the search proceeds from vertex to vertex, in a downhill direction, until the minimum is attained. For the example illustrated in Fig. 1, the search should require at most two iterations, independent of initial state.

It should be clear that a direct gradient search would prove inefficient.

Barring a fortuitous choice of initial state, a negative gradient search applied to Fig. 1 would take many more than two iterations, even with an optimum choice of step size. Clearly the general convex functional is not well structured for a gradient algorithm.

4. Calculation of the Gradient

As illustrated in Fig. 1, there is subspace of $\mathbf{A} \times d$ on which the gradient is undefined. As this subspace of edges and vertices is crucial to the search, it is necessary not only to calculate the gradient where it exists but to develop a simple criterion for identifying any point $\mathbf{A}_i$ along the search as an element of the subspace. Since edges and vertices are simply collections of points where at least one of the partial derivatives, and thus the gradient, fails to exist, both needs are met by the following theorem:

Theorem 1. *Let $f(t)$, $\{\phi_i(t)\}$, $i = 1,..., n$ be continuous on* [0, 1] *and let $e(\mathbf{A}_0, t)$ attain its maximum absolute value on the set $T \in [0, 1]$. If the function $-\operatorname{sgn}[e(\mathbf{A}_0, t)]\, \phi_k(t)$ is constant on T, then*

$$\partial/\partial a_k \| e \| |_{\mathbf{A}_0} = -\operatorname{sgn}[e(\mathbf{A}_0, t)]\, \phi_k(t), \qquad t \in T.$$

Otherwise the kth *partial derivative is undefined.*

Proof. The proof will only be outlined here; for a more complete development, see [2].

For a single-term approximation $e(a_0, t) = f(t) - a_0\phi(t)$ a change Δa in the parameter yields a new error function

$$e(a_0 + \Delta a, t) = e(a_0, t) - \Delta a\, \phi(t). \tag{1}$$

Defining $g(a) \,\Delta\, \| e(a, t) \|$, bounding (1) at the critical points of both error curves yields the inequality

$$\left| \frac{g(a_0 + \Delta a) - g(a_0)}{\Delta a} \right| \leqslant \| \phi \|. \tag{2}$$

For fixed a_0 this is a bounded function of the single variable Δa and must have at least one limit point as $\Delta a \to 0$. The lim sup and lim inf of this expression are the upper and lower derivatives of $g(a)$ at a_0. If they are equal the derivative is then the common value; if not, the partial is undefined.

Define two disjoint exhaustive subsets of [0, 1]:

$$F(\Delta a) = \{t: | e(a_0, t)| \geqslant g(a_0) - 2 \cdot | \Delta a | \cdot \| \phi \| \},$$
$$G(\Delta a) = \{t: | e(a_0, t)| < g(a_0) - 2 \cdot | \Delta a | \cdot \| \phi \| \}.$$

$F(\Delta a)$ is a set of neighborhoods around T, the critical point set of the error curve $e(a_0, t)$. Note that $T \subseteq F(\Delta a)$ for all Δa, and $F(0) = T$. Absolute bounds derived from Eq. (1) and evaluated first on $G(\Delta a)$ and then on T demonstrate that the critical point set T' of $e(a_0 + \Delta a, t)$ is also in $F(\Delta a)$ for all Δa.

For Δa sufficiently small,

$$| e(a_0 + \Delta a, t) | = | e(a_0, t) | - \Delta a \operatorname{sgn}[e(a_0, t)] \phi(t), \qquad t \in F(\Delta a) \tag{3}$$

since $F(\Delta a)$ is not local to any zero-crossings. Then

$$g(a_0 + \Delta a) = \max_{t \in F(\Delta a)} \{| e(a_0, t) | + \Delta a \cdot s(t)\} \tag{4}$$

with $s(t) \,\Delta\, - \operatorname{sgn}[e(a_0, t)] \phi(t)$, and

$$\frac{g(a_0 + \Delta a) - g(a_0)}{\Delta a} = \max_{t \in F(\Delta a)} \left\{\frac{| e(a_0, t) | - g(a_0)}{\Delta a} + s(t)\right\}, \qquad \Delta a > 0 \tag{5}$$

$$= \min_{t \in F(\Delta a)} \left\{\frac{| e(a_0, t) | - g(a_0)}{\Delta a} + s(t)\right\}, \qquad \Delta a < 0 \tag{6}$$

Then

$$\frac{g(a_0 + \Delta a) - g(a_0)}{\Delta a} \leqslant \max_{t \in F(\Delta a)} \{s(t)\}, \qquad \Delta a > 0 \tag{7}$$

$$\geqslant \min_{t \in F(\Delta a)} \{s(t)\}, \qquad \Delta a < 0 \tag{8}$$

and since $s(t)$ is continuous on F, passing to limits and bounding the upper and lower derivatives,

$$\min_{t \in T} s(t) \leqslant \underline{D}g(a) |_{a_0} \leqslant \bar{D}g(a) |_{a_0} \leqslant \max_{t \in T} s(t), \tag{9}$$

whence the sufficiency of Theorem 1.

To show necessity, restrict (5) and (6) to T, a subset of $F(\Delta a)$. Then the derivative from the left is at most the minimum and the derivative from the right at least the maximum of $s(t)$.

A direct corollary of Theorem 1 will be useful:

COROLLARY. *With $e(\mathbf{A}_0, t)$ as in Theorem 1, suppose $T = t_0$, a single point. Then all partials of the Chebyshev norm with respect to its parameters* $\mathbf{A}$ *exist, and*

$$\partial / \partial a_k \| e \| |_{\mathbf{A}_0} = -\operatorname{sgn}[e(\mathbf{A}_0, t_0)] \phi_k(t_0). \tag{10}$$

5. Edge Directions

Once an edge has been attained, an edge direction must be calculated and the edge followed "downhill" to the nearest vertex. The edge directions are directions in which the directional derivatives exist. For instance, the edge formed by the intersection of the two planes in three-space is a straight line, and at any point on the line there is an unique direction on which the directional derivative exists. This is the direction in which any step will remain on the edge.

Theorem 2 introduces conditions for the existence of the directional derivative in an arbitrary direction $\Delta \mathbf{A}$. First, however, a lemma is necessary.

Lemma. *Let $\{f_j(x)\}$ be in $C_1[0, 1]$, $j = 1,..., n$, Define $F(x) \Delta \max_j \{f_j(x)\}$. Suppose at some point x_0, $f_i(x_0) = f_j(x_0)$ for all $i, j \leqslant n$. Then a necessary and sufficient condition for the existence of $d/dx\, F(x)|_{x_0}$ is*

$$d/dx\, f_i(x)|_{x_0} = d/dx\, f_j(x)|_{x_0} \tag{11}$$

for all $i, j \leqslant n$, and if this is the case, the common value (11) *is equal to $d/dx\, F(x)|_{x_0}$.*

Proof. Suppose all $d/dx\, f_i(x)$ x_0 are not equal. Then there is a function with largest derivative at x_0,

$$d/dx\, f_1(x)|_{x_0} \geqslant d/dx\, f_i(x)|_{x_0}, \qquad i \leqslant n, \tag{12}$$

and a function with smallest derivative

$$d/dx\, f_s(x)|_{x_0} \leqslant d/dx\, f_i(x)|_{x_0}, \qquad i \leqslant n \tag{13}$$

such that the derivative (12) is strictly larger than (13). Since all functions are continuous and all are equal at x_0 there is some incremental region to the right of x_0 where $F(x) = f_1(x)$, $x \in x_0 +$, and some region to the left where $F(x) = f_s(x)$, $x \in x_0 -$. Thus from the left,

$$\lim_{\delta \to 0-} \frac{1}{\delta} \{F(x_0 + \delta) - F(x_0)\} = d/dx\, f_s(x)\, |_{x_0}, \tag{14}$$

and from the right,

$$\lim_{\delta \to 0+} \frac{1}{\delta} \{F(x_0 + \delta) - F(x_0)\} = d/dx\, f_1(x)\, |_{x_0}. \tag{15}$$

These are left- and right-hand derivatives of $F(x)$ at x_0. But $d/dx\, f_1(x)|_{x_0} \neq d/dx\, f_s(x)|_{x_0}$, and the derivative of $F(x)$ at x_0 is undefined.

Assume next that all $d/dx\, f_j(x)|_{x_0}$ are equal. Then $d/dx\, f_1(x)|_{x_0} = d/dx\, f_s(x)|_{x_0}$ and

$$d/dx\, F(x)|_{x_0} = d/dx\, f_j(x)|_{x_0}\,. \tag{16}$$

The corollary to Theorem 1 demonstrates existence of all partial derivatives of the Chebyshev norm if there is but one critical point. This result and the above lemma will be used to develop criteria for the existence of a directional derivative where at least one of the partials, and thus the gradient, is not defined.

THEOREM 2. *Let $f(t)$, $\{\phi_i(t)\}$ be continuous on $\{0, 1\}$ and let $e(\mathbf{A}, t) = f(t) - \sum_{i=1}^{n} a_i \phi_i(t)$ attain its maximum absolute value on the set $T = \{t_i\}$ of n isolated points in $\{0, 1\}$. If the function $-\operatorname{sgn}\{e(\mathbf{A}, t)\}\, \mathbf{\Delta A}^T\, \mathbf{\Phi}(\mathbf{t})$ is constant on T, then the directional derivative of the Chebyshev norm in the direction $\mathbf{\Delta A}$, to be designated $D(\mathbf{A}, \mathbf{\Delta A})$, exists and is equal to*

$$D(\mathbf{A}, \mathbf{\Delta A}) = -\operatorname{sgn} e(\mathbf{A}, t)\, \mathbf{\Delta A}^T\, \mathbf{\Phi}(\mathbf{t}), \qquad t \in T. \tag{17}$$

Otherwise $D(\mathbf{A}, \mathbf{\Delta A})$ is undefined.

Proof. Divide [0, 1] into N disjoint exhaustive subintervals T^j each containing one and only one critical point $t_j \in T^j$. The Chebyshev norms of the error functions $e(\mathbf{A}, t)^j$ defined over each subinterval are then all equal to the norm $\bar{e}$ over the whole interval [0, 1]. For any other value of the parameter vector, say $\mathbf{B}$, the norm over the whole interval will be the maximum norm over the subintervals T^j

$$\| e(\mathbf{B}, t)\| = \max_j \| e(\mathbf{B}, t)^j \|. \tag{18}$$

The directional derivative may be written

$$D(\mathbf{A}, \mathbf{\Delta A})^j = \lim_{\delta \to 0+} \frac{1}{\delta} \{\max_j \| e(\mathbf{A} + \delta \cdot \mathbf{\Delta A}, t)^j \| - \bar{e}\}. \tag{19}$$

Since there is only one critical point within each subinterval, by the previous corollary $\partial/\partial a_i \| e(\mathbf{A}, t)^j \|$ exists for all i and j, and all directional derivatives of these functions exist. Using Theorem 1 to calculate the partials,

$$\begin{aligned} D(\mathbf{A}, \mathbf{\Delta A})^j &= -\sum_{i=1}^{n} \Delta a_i \operatorname{sgn}\{e(\mathbf{A}, t_j)\}\, \phi_i(t_j) \\ &= -\operatorname{sgn}\{e(\mathbf{A}, t_j)\}\, \mathbf{\Delta A}^T\, \mathbf{\Phi}(\mathbf{t}_j), \end{aligned} \tag{20}$$

where $\mathbf{\Phi}(t)^T = (\phi_1(t),\ldots, \phi_n(t))$. By the lemma, the directional derivative, which is the derivative of the maximum over j of the norms, exists if and only if

$$D(\mathbf{A}, \Delta\mathbf{A})^i = D(\mathbf{A}, \Delta\mathbf{A})^j, \qquad \text{all } i, \quad j \leqslant N. \tag{21}$$

This result may be given a simple interpretation. The problem of finding a directional derivative is really one of determining the coefficients of the time function $\Delta\mathbf{A}^T \mathbf{\Phi}(t)$ such that this function interpolates to ± 1 (with the proper sign) on the critical point set T. With $T = t_0$ a single point, this is generally possible in any direction $\Delta\mathbf{A}$. As the number of critical points increases, there are less and less independent directions $\Delta\mathbf{A}$ in which this interpolation is possible. When no such direction can be found, the minimum has been obtained.

6. The Step Size: Addition of Critical Points

Suppose a step of size δ is taken in the direction $\Delta\mathbf{A}$. At any point t where the error function does not change sign, the change in the absolute value of the error function is given by

$$\begin{aligned} &\operatorname{sgn}[e(\mathbf{A}, t)] \cdot \{e(\mathbf{A} + \delta \cdot \Delta\mathbf{A}, t) - e(\mathbf{A}, t)\} \\ &\quad = \delta \cdot \{-\operatorname{sgn}[e(\mathbf{A}, t)]\} \, \Delta\mathbf{A}^T \mathbf{\Phi}(t). \end{aligned} \tag{22}$$

Then the rate of change of the error function per unit step in the direction $\Delta\mathbf{A}$ is

$$r(t) = -\operatorname{sgn}[e(\mathbf{A}, t)] \, \Delta\mathbf{A}^T \mathbf{\Phi}(t). \tag{23}$$

By comparison with Theorem 2, $\Delta\mathbf{A}$ is a downhill edge direction if the rate of change on all critical points is negative and equal.

Though the absolute value at the critical points of $e(\mathbf{A}, t)$ is decreasing uniformly as the step size increases, a point will be reached when the increasing absolute value at some other point will just equal the decreasing value on the old critical point set. Further excursion in the direction $\Delta\mathbf{A}$ will then increase d. A new step direction must be calculated based on the new critical point set. This intersection of edges is a vertex of the problem.

Ideally then, at each step the norm is reduced and a new critical point added. When no further critical points may be added, the minimum has been attained.

Digital computation of the error function makes it extremely unlikely that an error function with more than one point of absolute maximum will even be identified. For this reason, any point may be classed as a critical point if it is a local absolute maximum within some ϵ of the largest local absolute maximum.

Similarly, an approximation to the optimal step size may be calculated directly from the rate of change function r. Let $\Theta = \{\theta_i\}$ be the set of local absolute maxima of $e(\mathbf{A}, t)$ which are not within ϵ of the largest local absolute maxima and have been excluded from the critical point set $T = \{t_i\}$. We assume an edge direction of improvement $\Delta\mathbf{A}$ has been chosen, and the rate of change of the error function $r(t_i) = \bar{r} < 0$ is negative and constant on T. On Θ, however, the error is likely to be increasing. Then suppose $r(\theta_k) > 0$. After a step of size δ_k the decreasing error on T will be just equal to the increasing error at θ_k (to first-order variations), where

$$\| e(\mathbf{A}, t)\| - \delta_k \cdot | \bar{r} | = | e(\mathbf{A}, \theta_k)| + \delta_k \cdot | r(\theta_k)|, \tag{24}$$

or

$$\delta_k = \frac{\| e(\mathbf{A}, t)\| - | e(\mathbf{A}, \theta_k)|}{| \bar{r} | + | r(\theta_k)|}.$$

A step of size δ_k is certainly to be preferred to any larger step, since for $\delta > \delta_k$ the norm of the error is manifestly increasing. If we label the subset of Θ upon which $r(t_i) > 0$, $t_i \in \Theta$, as I, then the step

$$\delta* = \min_{i \in I} \{\delta_i\} \tag{25}$$

has the effect (again, to first variations) of equating the error which is decreasing uniformly over T with the most quickly increasing error in Θ and a single new critical point is added to T.

If ϵ is made fairly large, the resultant step will uniformly suppress the value of $e(\mathbf{A}, t)$ not only at its critical points but at other local absolute maxima. This seems well suited to the first stages of the search, while in the latter stage ϵ may be decreased as the accuracy is refined and the step size is reduced.

7. Characterization of the Near Region of a Solution Point

A stop rule for the search is necessary. If $\{\phi_i(t)\}$ forms a Chebyshev set [every set of n vectors $\mathbf{\Phi}(t_i)$ $i = 1,..., n$ $t_i \neq t_j$ is independent] use may be made of the following theorem (3) due to de la Vallée Poussin, here repeated without proof:

THEOREM. *Suppose there are n ordered points $t_1 < t_2 < \cdots < t_n$ such that $e(\mathbf{A}, t_i) = -e(\mathbf{A}, t_{i+1})$ $i = 1,..., n - 1$. Then the best Chebyshev error d^* satisfies*

$$| e(\mathbf{A}, t_i) | \leqslant d^* \leqslant \| e(\mathbf{A}, t) \|. \tag{26}$$

As the error function approaches n alternations these bounds draw closer, and when the difference is small the search may be terminated.

This rule is inapplicable in the general case, since it is based on an alteration property peculiar to approximation with Chebyshev sets. Suppose $\{\phi_i(t)\}$ is not a Chebyshev set, and for some $\mathbf{A}$ the error function $e(\mathbf{A}, t)$ attains its maximum value on the point set $T = \{t_1, ..., t_m\}$. The rate function $r(t)$ measures the rate of change in the absolute value of $e(\mathbf{A}, t)$ at each point $t \in [0, 1]$ due to a step in the direction $\Delta\mathbf{A}$. Then an incremental step in the direction $\Delta\mathbf{A}$ will result in a decreased norm if and only if $r(t_i) < 0$, $i = 1,..., m$. The value $\mathbf{A}$ is a solution point if and only if, for all $\Delta\mathbf{A} \in R^n$, $r(t_i) \geqslant 0$ for some $i \leqslant m$.

For some problems this test for a solution point is itself a useful stop rule. For instance, the single term approximation $L(a, t) = at$ to the constant $f(t) = 1$ on $[0, 1]$ has a single critical point at the origin for $0 < a < 2$. Using Eq. (23), $r(0) = 0$ for all such values of a. The computation would stop when the region $0 < a < 2$ has been reached.

In problems with non-Chebyshev sets when the solution region is of lower dimension than the parameter space the numerical search will never attain it. In this case the near region of the minimum must be identified. Approximation of $f(t) = 0$ by $L(a, t) = at$ in $[0, 1]$ results in a rate function of $+1$ for a negative, -1 for a positive and zero for $a = 0$. Since the point $a = 0$ is not likely to be attained using an iterative numerical procedure a more general stop rule would be useful. Certainly an initial approximation $L(a_0, t)$ can be improved by using a descent routine, and design considerations such as reduction of the norm below a maximum limit used to terminate calculation.

8. Examples Using Chebyshev Sets

A descent algorithm was programmed for use on Cornell University's CDC 1604 digital computer, as were two comparison algorithms, those due to Stiefel and Remes (second algorithm) [3]. The results are noted in Table 1. While none of the three algorithms is clearly most efficient, the descent search has certain advantages. First it is more general, since both comparison algorithms are based upon approximation with Chebyshev sets. The descent approach has no such restriction in the general case and has been defined up to a stop rule. Secondly it has the advantage of unifying L_p approximation theory to include its limiting case. While it has long been known that a descent approach is very useful for $1 < p < \infty$, the cases $p = 1$ and $p = \infty$ have been handled with separate theories. Thorp and Lewine [4] have

TABLE I
A Comparison of Three Algorithms

	Final value of parameters			Best norm	Number of iterations	Avg. Time per iter. (sec)	Total time (sec)
	a_1	a_2	a_3				
Problem 1							
Remes	2.841	−0.1192	−2.620	0.1016	9	3.75	24
Stiefel	2.839	−0.1192	−2.621	0.1016	11	2.00	22
Descent	2.840	−0.1192	−2.620	0.1016	16	1.30	21
Problem 2							
Remes	2.840	−0.1192	−2.620	0.1016	6	2.67	16
Stiefel	2.840	−0.1193	−2.620	0.1016	8	2.00	16
Descent	2.840	−0.1192	−2.620	0.1016	10	1.60	16
Problem 3							
Remes	0.2573	−2.922	3.520	0.2573	5	3.00	15
Stiefel	0.2573	−2.922	3.520	0.2573	7	2.30	16
Descent	0.2573	−2.922	3.520	0.2573	9	1.55	14
Problem 4							
Remes	7.603	−22.23	15.08	0.4595	11	3.75	41
Stiefel	7.604	−22.22	15.08	0.4595	11	3.37	37
Descent	7.604	−22.22	15.08	0.4595	19	1.80	34

Problem 1: $f(t) = 1 - e^{-10t}$; $\{\Phi(t)\} = e^{-t}, e^{-2t}, e^{-3t}$; $\mathbf{A}_0 = \mathbf{0}$
Problem 2: $f(t) = 1 - e^{-10t}$; $\{\Phi(t)\} = e^{-t}, e^{-2t}, e^{-3t}$; $\mathbf{A}_0 = (2.193, 0.5050, -2.452)$
Problem 3: $f(t) = t \cos 2\pi t$; $\{\Phi(t)\} = 1, t, t^2$; $\mathbf{A}_0 = \mathbf{0}$
Problem 4: $f(t) = t \cos 2\pi t$; $\{\Phi(t)\} = e^{-t}, e^{-2t}, e^{-3t}$; $\mathbf{A}_0 = \mathbf{0}$

successfully developed the L_1 approximation problem by means of a second-variational descent approach, and this investigation has shown the feasibility of a descent approach for L_∞.

References

1. E. W. Cheney, "Introduction to Approximation Theory," pp. 51–54, McGraw-Hill, New York, 1966.
2. P. D. Scott, "A Descent Approach to the Chebyshev Approximation Problem," pp. 25–31, Master's dissertation, Cornell University, Ithaca, NY, 1967.
3. J. R. Rice, "The Approximation of Functions," Addison–Wesley, Reading, MA, 1964.
4. J. S. Thorp and R. N. Lewine, "Exponential Approximation with Piecewise Linear Error Criteria," *J. Franklin Inst.* **286** (1968), 308–320.