An algorithm for minimax approximation in the nonlinear case

By M. R. Osborne and G. A. Watson*

An algorithm for nonlinear minimax approximation is described, and shown to be convergent under conditions which are often assumed in practice. The algorithm is illustrated by the calculation of several approximations to the solution of the Blasius equation.

(First received November 1967, and in revised form June 1968)

1. Introduction

The classical problems

(i) find numbers a_j to minimise the maximum value of

$$\left|f_i - \sum_{j=1}^{p} A_{ij}a_j\right|, \; i = 1, 2, \ldots, n$$

where $n > p$ (the discrete T-problem), and

(ii) find numbers a_j to minimise the maximum value of

$$\left|f(x) - \sum_{j=1}^{p} a_j\phi_j(x)\right|, \; a \leqslant x \leqslant b$$

where $f(x)$ and $\phi_j(x)$, $j = 1, 2, \ldots, p$, are continuous in $a \leqslant x \leqslant b$ (the continuous T-problem)

are now well understood. In particular the equivalence of (i) with a linear programming problem permits its solution under very general conditions (see, for example, Kelley (1959), Stiefel (1960), Rice (1964), Osborne and Watson (1967, 1968)).

In this paper, we consider the solution of the corresponding nonlinear minimax approximation problems, by solving a sequence of linear discrete T-problems. Certain properties of the solution of the linear problem are required, in particular, properties relevant to its solution as a linear programming problem, and these we now summarise.

If we write

$$r = f - Aa, \tag{1.1}$$

where r is the residual vector, then the discrete T-problem (i) is to find a vector a such that $\max|r_i|$, $i = 1, 2, \ldots, n$, is a minimum. We assume that A has rank p.

Any $(p + 1)$ equations of (1.1) are said to form a *reference*, and we can write a particular reference as

$$r^\sigma = f^\sigma - A^\sigma a \tag{1.2}$$

If A^σ has rank p, there exists a unique vector (to within a scaling factor) such that

$$\lambda^{\sigma T}A^\sigma = 0. \tag{1.3}$$

This vector is the λ-vector for the reference. If

$$\lambda_i^\sigma r_i^\sigma \geqslant 0 \qquad i = 1, 2, \ldots, p + 1,$$

then a is called a *reference vector*. The vector which solves the discrete T-problem is called the *levelled reference vector*. The solution to the dual linear programming formulation of (1.1) by the simplex method is characterised by the property that the residual of maximum modulus occurs in the $(p + 1)$ equations of the optimal reference (Osborne and Watson, 1968). This reference can be written

$$h^\sigma\theta = f^\sigma - A^\sigma a, \tag{1.4}$$

where h^σ is the maximum residual, called the *levelled reference deviation*, and $\theta_i = \pm 1$, $i = 1, 2, \ldots, p + 1$.

It is a property of the simplex method that A^σ has rank p, and this is possible because of the assumption on the rank of A.

Using equations (1.3) and (1.4), we see that the levelled reference deviation is given by

$$h = \frac{\lambda^{\sigma T} f^\sigma}{\sum_{i=1}^{p+1} |\lambda_i{}^\sigma|}. \tag{1.5}$$

The nonlinear approximation problems which correspond to the linear problems (i) and (ii) are:

(iii) find numbers a_j to minimise the maximum value of

$$|f_i - F_i(a_1, a_2, \ldots, a_p)| \quad i = 1, 2, \ldots, n,$$

and

(iv) find numbers a_j to minimise the maximum value of

$$|f(x) - F(x, a_1, a_2, \ldots, a_p)| \text{ in } a \leqslant x \leqslant b.$$

We will assume the existence of at least one bounded minimum for each problem, and that F is continuous as a function of x. If problem (iv) is considered on a set of points x_i, $i = 1, 2, \ldots, n$, instead of on the interval $a \leqslant x \leqslant b$, then it reduces to a problem of type (iii) with the identification $F_i(a) = F(x_i, a)$. Here we consider only problems of type (iii) and we make three further assumptions. These are:

A1. $F_i(a+\delta a)=F_i(a)+\nabla F_i\delta a+0(||\delta a||^2)$, $i=1, 2, \ldots, n$

where

∇F_i is the row vector with components $\dfrac{\partial F_i}{\partial a_j}$ $j=1,2,\ldots,p$.

This is a smoothness assumption on the F_i which permits at least a local linearisation of the nonlinear problem.

A2. The rank of the matrix M, $M = \nabla F$, is p.

(∇F is the matrix with rows ∇F_i $i = 1, 2, \ldots, n$.)

This corresponds to the assumption, made for problem (i), that the matrix A has rank p.

* *Computer Centre, The Australian National University, Canberra, A.C.T.*

A3. The system of equations $f_i - F_i(a) = 0, i = 1, 2, \ldots, n$ is inconsistent, i.e. no exact solution exists. (This assumption is also convenient in the linear case: see Osborne and Watson (1967).)

In Section 2, we describe an algorithm for solving the nonlinear approximation problem (iii) by an iterative technique. We show that the sequence of maximum residuals is convergent, and that the successive approximations to the solution vector a^* also converge under conditions which are often assumed in practice.

2. The nonlinear problem

The nonlinear problem (iii) can be formulated in a manner analogous to the linear programming formulation of problem (i). The solution is obtained by minimising h subject to the constraints

$$|f_i - F_i(a)| \leqslant h, \quad i = 1, 2, \ldots, n. \tag{2.1}$$

This problem is solved iteratively, as follows:

(1) Calculate δa^j to minimise h^j subject to the constraints

$$|f_i - F_i(a^j) - \nabla F_i(a^j)\delta a^j| \leqslant h^j, i = 1, 2, \ldots, n. \tag{2.2}$$

(This is a discrete T-problem, and can be solved by linear programming, because of assumption A2 of Section 1.) Denote the minimum value of h^j by $\hat{h}^j$.

(2) Calculate γ^j to minimise the maximum value of

$$|f_i - F_i(a^j + \gamma^j \delta a^j)|, \quad i = 1, 2, \ldots, n.$$

Let the minimum value be $\bar{h}^{j+1}$.

(3) Set $\quad a^{j+1} = a^j + \gamma^j \delta a^j. \quad$ (2.3)

Lemma 2.1 $\quad \hat{h}^j \leqslant \bar{h}^j$.

Proof We assume for simplicity that the equations determining δa^j have been ordered so that the first $(p + 1)$ make up the optimal reference.

Then, by equation (1.5),

$$\hat{h}^j = \sum_{i=1}^{p+1} \lambda_i^j (f_i - F_i(a^j)) / \sum_{i=1}^{p+1} |\lambda_i^j| \leqslant \bar{h}^j. \tag{2.4}$$

Remark If $\bar{h}^j > 0$ then equality can hold only if, for each equation in the optimal reference for which $\lambda_i^j \neq 0$,

(i) $|f_i - F_i(a^j)| = \bar{h}^j$,

(ii) $\text{sign}(\lambda_i^j) = \text{sign}(f_i - F_i(a^j))$.

(Note that $\bar{h}^j > 0$ is a consequence of assumption A3.)

Definition A unit vector y is *downhill* at the point a if

$$\max_i |f_i - F_i(a)| > \max_i |f_i - F_i(a + \gamma y)|,$$

where $\gamma > 0$ is sufficiently small.

Lemma 2.2 If $||\nabla F_i(a^j)|| > 0$ for all i in the reference, then there is a downhill direction at the point a^j if and only if $\hat{h}^j < \bar{h}^j$.

Proof Let $\hat{h}^j < \bar{h}^j$.

Then the vector δa^j is downhill by assumption A1 of the previous section. This proves sufficiency.

Now let $\hat{h}^j = \bar{h}^j$, and assume that there exists a downhill direction y. Then since equality holds in the $(p + 1)$ equations of the reference defining $\hat{h}^j$, y must satisfy

$$\nabla F_i(a^j) y = \text{sign}(f_i - F_i(a^j))\xi_i, \; i = 1, 2, \ldots, p + 1$$

for some numbers $\xi_i > 0$.

If each equation is multiplied by the corresponding λ_i, summation gives

$$0 = \sum_{i=1}^{p+1} \lambda_i \,\text{sign}(f_i - F_i(a^j))\xi_i = \sum_{i=1}^{p+1} |\lambda_i| \xi_i,$$

where the second remark following Lemma 2.1 has been used.

This is a contradiction and so no downhill direction exists. This proves necessity.

Corollary 1 $\quad$ If $\hat{h}^j < \bar{h}^j$, then $\bar{h}^{j+1} < \bar{h}^j$.

Corollary 2 Since the sequence $\bar{h}^j$ is monotonically decreasing and bounded below by zero, it is convergent.

Remark The condition $||\nabla F_i(a^j)|| > 0$ is necessary, for consider the example †

$$f_1 - F_1(a) \equiv 1 - a_1^2 - a_2^2,$$
$$f_2 - F_2(a) \equiv \tfrac{1}{2} - a_1,$$
$$f_3 - F_3(a) \equiv \tfrac{1}{2} - a_2,$$

at the point $a^j = (0, 0)$. Then $\hat{h}^j = \bar{h}^j$, and conditions A1 and A2 hold, but every direction is downhill.

Definition If every $p \times p$ submatrix of M, $M = \nabla F$, is non-singular, then we say that the matrix M satisfies the Haar condition. When required, it will be assumed to hold uniformly in the region of interest. It is clear that in this case $||\nabla F_i|| > 0, i = 1, 2, \ldots, n$.

Remark If the Haar condition holds, then *all* the components of the λ-vector are different from zero.

Lemma 2.3 A sufficient condition for the minimum to be isolated is that the matrix M satisfies the Haar condition.

Proof Suppose we have a solution a^* with

$$||f - F(a^*)|| = h^*$$

(All norms are assumed to be maximum norms.)

Then if the Haar condition on M is satisfied, there exists a reference such that

$$f_i - F_i(a^*) = \theta_i h^*, \quad i = 1, 2, \ldots, p + 1,$$

where $\theta_i = \pm 1$, where the second remark following Lemma 2.1 has been used.

Let $\psi(a^*) = - \nabla F^\sigma(a^*) t$

where the suffix σ means that we consider only the rows of M which form the reference, and we assume t to be any vector such that $||t|| = 1$.

Then $\lambda^{\sigma T}\psi(a^*) = 0$, where λ^σ is the λ-vector for the reference.

Since the Haar condition is satisfied there exists at least two of the ψ_i which are non-zero. We infer that at least one of the $\theta_i\psi_i$ is positive, and consequently the expression $\max(\theta_i\psi_i)$ is a positive function of t. Since it is also a continuous function and the domain of t is compact,

$$\delta = \min_{||t||=1}, \max_i (\theta_i \psi_i) > 0.$$

† We are indebted to the referee for this example and for other helpful suggestions which have greatly improved the paper.

We have

$$||f - F(a^* + \rho t)|| \geq \max_i \{\theta_i(f_i - F_i(a^* + \rho t))\}$$
$$= \max_i \{\theta_i(f_i - F_i(a^*)) + \theta_i(F_i(a^*) - F_i(a^* + \rho t))\}$$
$$= ||f - F(a^*)|| + \max_i \{- \theta_i \rho \nabla F_i(a^*)t - \rho^2 K_i\},$$

by assumption A1 where K_i are appropriately chosen constants.

Let $K^* = \max_i |K_i|$ for $||a - a^*|| < R$.

Then $||f - F(a^* + \rho t)|| \geq h^* + \rho\delta - \rho^2 K^*$.

Choosing $0 < \rho < \delta/K^*$, we have

$$||f - F(a^* + \rho t)|| > h^*. \tag{2.5}$$

This completes the proof, as t is an arbitrary vector of unit norm.

Remark Cheney (1966, p. 81) has given an example which shows that the inequality (2.5) cannot be proved without assuming that the Haar condition holds.

Lemma 2.4 Let the equations (2.2) be ordered so that the first $(p + 1)$ form the optimal reference defining $\hat{h}^j$, and let λ be the λ-vector for the reference scaled so that

$$\sum_{i=1}^{p+1} |\lambda_i| = 1.$$

Also let K be a positive constant. Then

(i) $||\delta a^j|| \leq 2K\bar{h}^j$

(ii) If $|\lambda_i| > 0$ and $\lambda_i(f_i - F_i(a^j)) \geq 0$, $i = 1, 2, \ldots, p+1$ and $m = \min_i |\lambda_i|$, then

$$||\delta a^j|| \leq \frac{K}{m}(\bar{h}^j - \hat{h}^j).$$

Proof

(i) We have
$\nabla F_i(a^j).\delta a^j = f_i - F_i(a^j) - \hat{h}^j\theta_i$,
where $\theta_i = \pm 1$, $i = 1, 2, \ldots, p + 1$, and
$\theta_i = \operatorname{sgn}(\lambda_i)$ if $|\lambda_i| > 0$. Further, we can arrange that the matrix formed by the first p rows ∇F_i is nonsingular, and so

$$||\delta a^j|| \leq K \max_{1 \leq i \leq p} |f_i - F_i(a^j) - \hat{h}^j\theta_i| \leq 2K\bar{h}^j.$$

(ii) If $|\lambda_i| > 0$ and $\lambda_i(f_i - F_i(a^j)) \geq 0$, $i = 1, 2, \ldots, p + 1$,

then $\hat{h}^j = \sum_{i=1}^{p+1} |\lambda_i||f_i - F_i(a^j)|$,

whence $\hat{h}^j \leq m \min_i |f_i - F_i(a^j)| + (1 - m)\bar{h}^j$,

so that $\bar{h}^j - \hat{h}^j \geq m(\bar{h}^j - \min_i |f_i - F_i(a^j)|)$.

This gives

$$||\delta a^j|| \leq K|\bar{h}^j - \min_i |f_i - F_i(a^j)||$$
$$\leq \frac{K}{m}(\bar{h}^j - \hat{h}^j).$$

Lemma 2.5 Let $q_i(\gamma) = |f_i - F_i(a^j)| - \gamma(|f_i - F_i(a^j)| - \hat{h}^j)$.

Then

$$|f_i - F_i(a^j + \gamma\delta a^j)| \leq q_i(\gamma) + W||\delta a^j||^2\gamma^2, 0 \leq \gamma \leq 1,$$

where $W > 0$ is a constant independent of i.

Proof By assumption A1 of Section 1, we have

$$f_i - F_i(a^j + \gamma\delta a^j) = f_i - F_i(a^j) - \gamma\nabla F_i(a^j)\delta a^j + W_i(\gamma)||\delta a^j||^2\gamma^2,$$

where $W_i(\gamma)$ is bounded in a finite region R.

Let $g_i(\gamma) = f_i - F_i(a^j) - \gamma\nabla F_i(a^j)\delta a^j$,

and let W be chosen so that

$$|W_i(\gamma)| \leq W, \quad 0 \leq \gamma \leq 1.$$

Then

$$|f_i - F_i(a^j + \gamma\delta a^j)| \leq |g_i(\gamma)| + W||\delta a^j||^2\gamma^2,$$

and it only remains to show that $|g_i(\gamma)| \leq q_i(\gamma)$ in $0 \leq \gamma \leq 1$.

Now each equation of the set (2.2) can be written as an equality in the form, either

(i) $f_i - F_i(a^j) - \nabla F_i(a^j)\delta a^j = \hat{h}^j - \phi_i$,

or

(ii) $f_i - F_i(a^j) - \nabla F_i(a^j)\delta a^j = -\hat{h}^j + \phi_i$,

where $0 \leq \phi_i \leq \hat{h}^j$.

We will only consider equations of type (i), as type (ii) can be treated in a similar manner. Then we have

$$g_i(\gamma) = f_i - F_i(a^j) - \gamma(f_i - F_i(a^j) - \hat{h}^j + \phi_i),$$

and we distinguish two possibilities:

(1) $f_i - F_i(a^j) \geq 0$.
In this case
$$g_i(\gamma) = (1 - \gamma)(f_i - F_i(a^j)) + \gamma(\hat{h}^j - \phi_i) \geq 0 \text{ in } 0 \leq \gamma \leq 1.$$
Also $q_i(\gamma) - g_i(\gamma) = \gamma\phi_i \geq 0$ in $0 \leq \gamma$, and so $|g_i(\gamma)| \leq q_i(\gamma)$, $0 \leq \gamma \leq 1$.

(2) $f_i - F_i(a^j) < 0$.
Here, $g_i(\gamma) = 0$ at $\gamma = \gamma^* < 1$, and we have
$$q_i(\gamma) - g_i(\gamma) = 2(1 - \gamma)|f_i - F_i(a^j)| + \gamma\phi_i \geq 0 \text{ in } 0 \leq \gamma \leq 1,$$
and
$$q_i(\gamma) + g_i(\gamma) = \gamma(2\hat{h}^j - \phi_i) \geq 0 \text{ in } 0 \leq \gamma \leq 1.$$
Thus again
$$|g_i(\gamma)| \leq q_i(\gamma), 0 \leq \gamma \leq 1.$$

This completes the proof of the lemma.

Theorem 2.1
Assume that the iteration is confined to a bounded region R. Then

$$|\bar{h}^j - \hat{h}^j| \to 0 \text{ as } j \to \infty.$$

Proof Let $q_i(\gamma)$ and W be defined as in Lemma 2.5, and let $Q_i(\gamma) = q_i(\gamma) + \gamma^2 W||\delta a^j||^2$,

so that $Q_i(\gamma) \geq 0$ in $0 \leq \gamma \leq 1$.

Then $Q_i(\gamma)$ satisfies

(i) $Q_i(0) = |f_i - F_i(a^j)|$,

(ii) $\max_i Q_i(\gamma) \leqslant \max_i Q_i(0)$, for sufficiently small $\gamma > 0$,

(iii) $Q_i(\gamma) \geqslant |f_i - F_i(a^j + \gamma\delta a^j)|$, $0 \leqslant \gamma \leqslant 1$,

where Lemma 2.5 has been used in (iii).

Let $Q = \min_{0 \leqslant \gamma \leqslant 1} \max_i Q_i(\gamma)$.

Then

$$h^j \geqslant Q \geqslant \min_{0\leqslant\gamma\leqslant 1} \max_i |f_i - F_i(a^j + \gamma\delta a^j)|$$
$$\geqslant h^{j+1}.$$

Thus Q is an upper bound for h^{j+1}. Now the curves $Q_i(\gamma)$, $i = 1, 2, \ldots, n$, all intersect at the point $\gamma = 1$, or are coincident. Also, for at least one $i = s$,

$$|f_s - F_s(a^j)| = h^j.$$

Thus Q is determined by either

(*a*) $Q = Q_s(\gamma)$, $\dfrac{dQ_s(\gamma)}{d\gamma} = 0$.

This gives $\quad \gamma = \dfrac{h^j - \hat{h}^j}{2W||\delta a^j||^2}$,

or

(*b*) $Q = Q_s(1)$,

whichever gives the smaller value of γ (see **Fig. 1**).

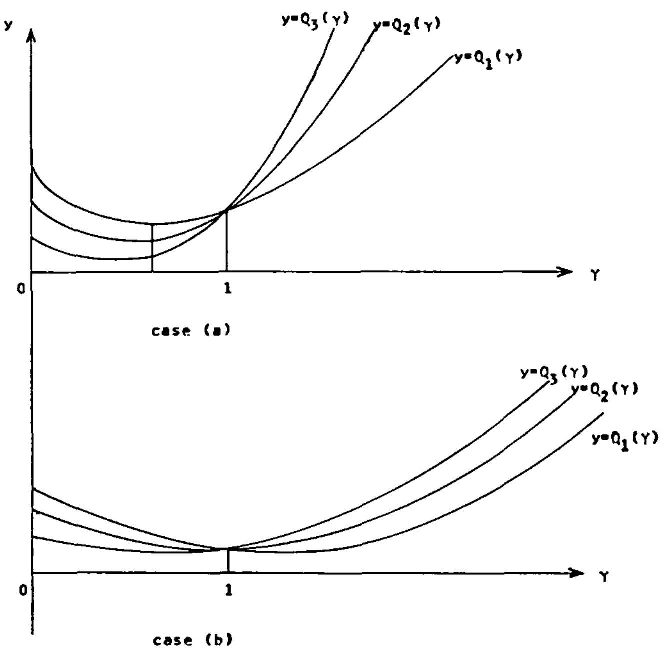

Fig. 1

If $\quad \dfrac{h^j - \hat{h}^j}{2W||\delta a^j||^2} < 1$, then

$$Q = h^j - \frac{(h^j - \hat{h}^j)^2}{4W||\delta a^j||^2}$$
$$\leqslant h^j - \frac{(h^j - \hat{h}^j)^2}{16K^2W(h^0)^2} \text{ by Lemma 2.4,}$$

where $\quad h^0 > h^j, j > 0$, and so

$$(h^j - \hat{h}^j)^2 \leqslant 16K^2W(h^0)^2(h^j - h^{j+1}).$$

Otherwise

$$\frac{h^j - \hat{h}^j}{2W||\delta a^j||^2} \geqslant 1, \text{ giving}$$

$$W||\delta a^j||^2 \leqslant \frac{h^j - \hat{h}^j}{2}, \text{ and } \gamma = 1, \text{ so that}$$

$$Q = \hat{h}^j + W||\delta a^j||^2 \leqslant \hat{h}^j + \frac{h^j - \hat{h}^j}{2},$$

and so

$$h^j - \hat{h}^j \leqslant 2(h^j - h^{j+1}).$$

Thus, in either case, $|h^j - \hat{h}^j| \to 0$ as $j \to \infty$, by Corollary 2 to Lemma 2.2.

Corollary 1 At a limit point of the iteration, $h^j = \hat{h}^j$.

Corollary 2 If the conditions of Lemma 2.2 are fulfilled, then a limit point of the sequence a^j is a stationary point of

$$\max_i |f_i - F_i(a)|.$$

Theorem 2.2 If the Haar condition is satisfied on the matrix M in the region R, then the sequence a^j converges as $j \to \infty$.

Proof If λ is the λ-vector for the optimal reference defining $\hat{h}^j$, then all elements λ_i are nonzero. In this case, convergence of the $|h^j - \hat{h}^j|$ to zero implies that for j sufficiently large

$$\lambda_i(f_i - F_i(a^j)) > 0, \quad i = 1, 2, \ldots, p + 1,$$

where assumption A3 of Section 1 has been used.

Thus, from Lemma 2.4 we have

$$||\delta a^j|| \leqslant \frac{K}{m}(h^j - \hat{h}^j).$$

Further, in case (*a*) of Theorem 2.1,

$$Q = h^j - \frac{(h^j - \hat{h}^j)^2}{4W||\delta a^j||^2} \leqslant h^j - \frac{m^2}{4WK^2}$$

and this implies

$$h^j - h^{j+1} \geqslant \frac{m^2}{4WK^2},$$

which contradicts the convergence of the sequence h^j.

Thus, for sufficiently large j, we must have $Q = Q_s(1)$ in Theorem 2.1, and this gives

$$||\delta a^j|| \leqslant \frac{2K}{m}(h^j - h^{j+1}).$$

The convergence of the sequence a^j is thus a consequence of the convergence of the h^j in this case. Note that the Haar condition ensures that m is bounded away from zero in R.

3. Application of the algorithm

As an example of the application of the algorithm described in the previous section, we consider the solution to the differential equation

$$y^{(3)} + yy^{(2)} = 0 \tag{3.1}$$

subject to the boundary conditions

$$y(0) = y^{(1)}(0) = 0, \; y^{(1)}(x) \to k \text{ as } x \to \infty, \tag{3.2}$$

where k is a constant.

This is the original equation of Blasius, and arises in the consideration of the flow of a fluid which streams past a plate placed edgeways in it (see, e.g., Davis, 1960).

We assume a trial solution containing free parameters and minimise the maximum residual on a discrete set of points. The form of solution is particularly important (polynomials, for example, give large residuals), and we use a form suggested by Mason (1965), where y is approximated by

$$\phi = A + Bx + \psi(x),\ A, B \text{ constants.} \tag{3.3}$$

The boundary conditions are best dealt with by assuming that they are satisfied by ϕ, and so we have

$$A + \psi(0) = 0,$$
$$B + \psi^{(1)}(0) = 0,$$
$$B + \psi^{(1)}(x) \to k \text{ as } x \to \infty.$$

The last condition suggests a $\psi(x)$ of the form

$$\psi(x) = C/(P(x))^r,$$

where C and r are constants and $P(x)$ is a polynomial.

Table 1

Successive iterates, $r = 4$

$\hat{h}^j$	γ^j	$\bar{h}^{j+1}$
0·000 445	0·008	0·420
0·017 738	0·1	0·379
0·058 304	0·4	0·286
0·013 659	0·7	0·156
0·005 463	1	0·018 536
0·003 450	1	0·003 975
0·001 343	0·9	0·001 574
0·000 366	1	0·000 503
0·000 019	0·9	0·000 038
0·000 004 2	1	0·000 004 4
0·000 004 2	1	0·000 004 2

Table 2

Successive iterates, $r = 8$

$\hat{h}^j$	γ^j	$\bar{h}^{j+1}$
0·000 341	0·07	1·227
0·003 053	0·7	0·409
0·027 909	1·1	0·129
0·034 045	1	0·036
0·008 308	0·9	0·012 375
0·001 111	1	0·002 056
0·000 252	1	0·000 275
0·000 009 57	1	0·000 011
0·000 000 89	1	0·000 000 91
0·000 000 85	1	0·000 000 85

Table 3

Maximum residuals for various values of r

r	MAXIMUM RESIDUAL
1	0·012 6
2	0·000 118 1
3	0·000 017 4
4	0·000 004 2
5	0·000 001 7
6	0·000 001 06
7	0·000 000 901
8	0·000 000 850
9	0·000 000 834
10	0·000 000 828
11	0·000 000 826

Table 5

Comparison of coefficients, $r = 4$

	O. & W.	MASON
a_1	0·205 459 87	0·290 567 60
a_2	0·057 292 53	0·114 587 94
a_3	0·015 490 48	0·043 815 29
a_4	0·003 430 83	0·013 736 44
a_5	0·000 695 72	0·003 869 42
a_6	0·000 089 01	0·000 889 98
a_7	0·000 041 65	0·000 204 75
a_8	−0·000 012 17	0·000 042 57
a_9	0·000 006 06	0·000 011 98
a_{10}	−0·000 001 06	0·000 002 55
a_{11}	0·000 000 12	0·000 000 63

Table 4

Computed solution values

r \ x	0·75	1·5	2·25	3	3·75	4·25	5
1	0·130 910 7	0·512 448 4	1·093 292	1·788 004	2·524 424	3·021 671	3·769 888
3	0·131 641 4	0·515 030 5	1·098 369	1·795 565	2·534 712	3·033 496	3·783 228
4	0·131 641 6	0·515 031 2	1·098 370	1·795 567	2·534 716	3·033 499	3·783 232
8	0·131 641 7	0·515 031 5	1·098 371	1·795 568	2·534 717	3·033 501	3·783 234
11	0·131 641 7	0·515 031 5	1·098 371	1·795 568	2·534 717	3·033 501	3·783 234

In fact, the conditions (3.2) are satisfied if we set

$$\psi(x) = \frac{k}{ra_1(1 + a_1x + a_2x^2 + \ldots)^r}, A = -\frac{k}{ra_1}, \quad B = k.$$

That is,

$$\phi = -\frac{k}{ra_1} + kx + \frac{k}{ra_1(1 + a_1x + a_2x^2 + \ldots)^r}. \quad (3.4)$$

Solutions were obtained, for $k = 1$, on 31 equispaced points in the range $0 \leqslant x \leqslant 5$. $P(x)$ was taken to have degree 11 (i.e. there were 11 unknowns), and a range of integral values of r from 1 to 11 was considered.

In **Table 1** we tabulate successive values of $\hat{h}$, $\bar{h}$ and γ for the solution with $r = 4$. It was found to be sufficient to evaluate γ to one decimal place, except where this gave the value zero, when a more precise value was computed.

Incidence of a small γ occurs here (though certainly not always) when the starting point is far from the solution. Approaching the solution, we see that γ tends to 1, and takes this value when convergence is finally obtained. This feature is exhibited again in **Table 2** for the case $r = 8$.

It is interesting to note that different values of r give a large range of maximum residual. This is shown in **Table 3.** Computed solutions for some of those values of r, over a range of values of x, are given in **Table 4.** The solutions for $r = 11$ are correct to about one figure in the last decimal place.

Finally, in **Table 5**, we list the coefficients obtained in the case $r = 4$ along with those obtained by Mason for the same value of r, by collocation methods.

4. Conclusion

The algorithm described is applicable to a large range of nonlinear approximation problems. In particular, problems with linear constraints are easily handled. Nonlinearly constrained problems are best tackled by adjusting the form of solution to satisfy these constraints, but failing this, it may be possible to make progress by linearising the outstanding constraints.

Convergence is obtained provided the Haar condition is satisfied on the matrix M. However, it may happen that the values of γ^j become very small, and in this case steps towards the solution are small, and progress can become intolerably slow. Consequently, the initial approximation is often critical, expecially when a large number of unknowns is involved. Despite these drawbacks, the method has been used in a variety of nonlinear approximation problems with considerable success.

References

CHENEY, E. W. (1966). *Introduction to Approximation Theory*, McGraw-Hill.

DAVIS, H. T. (1960). *Introduction to Nonlinear Differential and Integral Equations*, U.S. Atomic Energy Commission.

KELLEY, JAMES E. (Jnr.) (1959). An application of Linear Programming to Curve Fitting, *J. Soc. Indust. Appl. Maths.*, Vol. 6, pp. 15–22.

MASON, J. (1965). *Some New Approximations for the Solution of Differential Equations*, D. Phil. Thesis, Oxford.

OSBORNE, M. R., and WATSON, G. A. (1967). On the best linear Chebyshev approximation, *Computer Journal*, Vol. 10, pp. 172–177.

OSBORNE, M. R., and WATSON, G. A. (1968). A note on singular minimax approximation problems (to appear in *J. Math. Anal. and Applications*).

RICE, J. R. (1964). *The Approximation of Functions*, Vol. 1. Addison-Wesley Publishing Co.

STIEFEL, E. (1960). Note on Jordan Elimination, Linear Programming, and Tschebyscheff Approximation, *Numerische Mathematik*, Vol. 2, pp. 1–17.