A Newton-method for nonlinear Chebyshev approximation

R. Hettich

Department of Applied Mathematics
Twente University of Technology
P.O. Box 217, Enschede, The Netherlands

1. Introduction. In this paper the following approximation problem is considered:

Let $B \subset R^m$ be a compact set, $P \subseteq R^n$ an open set, $f(x)$ and $a(p,x)$ twice continuously differentiable, real-valued functions defined on B and $P \times B$ resp. Then a $p_o \in P$ is to be found, such that

$$\| f - a(p_o,.)\| \leq \| f - a(p,.)\|, \quad p \in U_o \cap P, \tag{1.1}$$

with U_o a neighbourhood of p_o. $\|.\|$ denotes the maximumnorm: $\|g\| = \max_{x \in B} |g(x)|$, g continuous on B. A p_o satisfying (1.1) is called a locally best approximation.

Note that best approximations are defined with respect to the parameter set P and not the function set $\{a(p,.) \mid p \in P\}$. From a practical point of view, this seems more appropriate to us, since otherwise global information about the function set is required, e.g. about points $p^1 \neq p^2$ with $a(p^1,.) = a(p^2,.)$. In practice, apart from special cases, to obtain that sort of information will be very difficult or even impossible.

Define the error function

$$e(p,x) = |f(x) - a(p,x)|. \tag{1.2}$$

If $e(p_o,x)$ has maxima in exactly $r = n + 1$ points $x_o^j \in B$, $j = 1, \ldots, r$, Newton's method is a wellknown and efficient means of computing p_o (cf. [6,8]). If $B = [a,b] \subset R$, $x_o^1 = a$, $x_o^{n+1} = b$, it may be formulated as follows:

p_o, $d_o = \|e(p_o,.)\|$ and x_o^j, $j = 1, \ldots, r$ $(r=n+1!)$, satisfy

$$\begin{aligned} e(p,x^j) - d &= 0, \quad j = 1, \ldots, r, \\ e_x(p,x^j) &= 0, \quad j = 2, \ldots, r-1, \end{aligned} \tag{1.3}$$

a system of $2n - 1$ equations for a same number of unknowns. Given initial values

p_1, d_1, x_1^j, (1.3) is solved by Newton's method. Sufficient conditions for local convergence are given in [6].

If $r < n + 1$, there are more unknowns than equations (1.3) and the method is not applicable. In nonlinear programming the same difficulty is encountered if the number of active constraints is less than n. To overcome this difficulty, the equations required by the Kuhn-Tucker condition are added and the dual parameters occuring in this condition are regarded as unknowns too (see e.g. [10]). In Section 4 the same idea is used to formulate a method which is also applicable if $r < n + 1$. In Section 5 conditions for local convergence of the method are given in terms of sufficient conditions of the second order for locally best approximations (cf. Section 4).

Two versions of the method are presented. For linear approximation, the first one is identical to that given (without proof of convergence) in [1] for semi-infinite, linearly constraint programming. We remark that the methods in Section 4 easily can be extended to nonlinear, semi-infinite programming. To prove convergence, conditions given in [9] or [2] may be used instead of those given in Section 3. Using results from [4], approximation subject to constraints (e.g. restricted range approximation) can be treated too.

2. Notation. To facilitate reading we give a list of some essential symbols together with a short description or a reference to the place where they are introduced. We denote by

x	a point in R^m, generally $x \in B$
x^j, x_o^j	maxima of the error function $e(p,x)$, $e(p_o,x)$; $x_o^j \in E_o$
$x^j(t)$	local maxima of $e(p(t),x)$, $x^j(0) = x_o^j$; Theorem 3.1
μ_j	the derivative of $x^j(t)$ in $t = 0$: $\mu_j = x_t^j(0)$
p, p_o	points in $P \subseteq R^n$
$p(t)$	an arc in P, $p(0) = p_o$
ξ	the derivative of $p(t)$ in $t = 0$: $\xi = p_t(0)$
w^{ij}, w_o^{ij}	Lagrangean parameters for extrema of $e(p,x)$, $e(p_o,x)$; (3.11)
w^j, w_o^j	vectors in R^{m_j} with components w^{ij}, w_o^{ij}
$w^j(t)$	parameter vectors for extrema of $e(p(t),x)$, $w^j(0) = w_o^j$; Theorem 3.1
u^j, u_o^j	parameters in first order necessary condition; (3.20), (3.21)
$q(u,\xi)$	for given u a quadratic form in ξ; (3.22)

$f(x)$ the function to be approximated

$a(p,x)$ the approximating function

$e(p,x)$ the error function, $e(p,x) = |f(x) - a(p,x)|$

$g^i(x)$ functions defining B, (3.1)

B a compact region, where f is to be approximated, $B \subset R^m$

P an open set of parameters, $P \subseteq R^n$

E_o the set of x_o^j; (3.3), (3.7)

K a cone; (3.19)

$I, I(x)$ sets of indices; (3.1), (3.2)

G_j, M_j matrices; (3.8), (3.9)

$F_z(z)$ matrix for Newton's method; (4.9)

Derivatives are indicated by lower indices. For instance $p_t(t) = \frac{d}{dt}p(t)$ or

$$e_p(p,x) = \left(\frac{\partial e(p,x)}{\partial p_1}, \dots, \frac{\partial e(p,x)}{\partial p_n}\right)^T.$$

Lower indices xx, xp, px, pp denote resp. $m{\times}m$-, $m{\times}n$-, $n{\times}m$-, $n{\times}n$-matrices of second order derivatives. For instance

$$e_{xp}(p,x) = (e_{px}(p,x))^T = \begin{bmatrix} \frac{\partial^2 e(p,x)}{\partial x_1 \partial p_1} & \cdots\cdots & \frac{\partial^2 e(p,x)}{\partial x_1 \partial p_n} \\ \cdots & \cdots & \cdots \\ \frac{\partial^2 e(p,x)}{\partial x_m \partial p_1} & & \frac{\partial^2 e(p,x)}{\partial x_m \partial p_n} \end{bmatrix}.$$

An upper index j indicates that a function is evaluated for the arguments $p = p_o$, $x = x_o^j$. For instance $e_{xp}^j = e_{xp}(p_o, x_o^j)$ or $g^{ij} = g^i(x_o^j)$.

Finally $C^\nu(A,B)$, $A \subseteq R^k$, $B \subseteq R^\ell$, denotes the set of all functions defined on A with values in B, having continuous derivatives up to order ν. By assumption $f \in C^2(B,R)$, $a \in C^2(P{\times}B,R)$.

3. Conditions for locally best approximations. In this section some second order conditions for locally best approximations are stated without proof (cf. [3]).

From now the following assumptions are assumed to hold.

Assumption 3.1. The (compact) set B is given by

$$B = \{x \in R^m \mid g^i(x) \le 0,\ i \in I\}, \tag{3.1}$$

with I a finite set of indices and $g^i \in C^2(R^m,R)$. For $x \in B$ define

$$I(x) = \{i \in I \mid g^i(x) = 0\}. \tag{3.2}$$

Then, for every $x \in B$, $g^i_x(x)$, $i \in I(x)$, are linearly independent.

In the following, $p_o \in P$ is a fixed point such that $\|e(p_o,\cdot)\| > 0$. Define

$$E_o = \{x \in B \mid e(p_o,x) = \|e(p_o,\cdot)\|\}. \tag{3.3}$$

Assumption 3.2. For every $\bar{x} \in E_o$ there are $\bar{w}^i > 0$, $i \in I(\bar{x})$, such that the properties (i) and (ii) hold:

(i) $$e_x(p_o,\bar{x}) - \sum_{i\in I(\bar{x})} \bar{w}^i g^i_x(\bar{x}) = 0. \tag{3.4}$$

(ii) The quadratic form $\mu^T\bar{M}\mu$,

$$\bar{M} = e_{xx}(p_o,\bar{x}) - \sum_{i\in I(\bar{x})} \bar{w}^i g^i_{xx}(p_o,\bar{x}) \tag{3.5}$$

is negative definite on the subspace

$$\bar{T} = \{\mu \in R^m \mid \mu^T g^i_x(p_o,\bar{x}) = 0,\ i \in I(\bar{x})\}. \tag{3.6}$$

Assumption 3.2 implies that E_o is a finite set

$$E_o = \{x^1_o, \dots, x^r_o\}. \tag{3.7}$$

Thus, for $j = 1, \dots, r$, there are $w^{ij}_o > 0$, $i \in I(x^j_o)$, such that (i) and (ii) in Assumption 3.2 hold for $\bar{x} = x^j_o$, $\bar{w}^i = w^{ij}_o$.

Let $m_j = \mathrm{card}(I(x^j_o))$ and $w^j_o \in R^{m_j}$ be the vector with components w^{ij}_o, $i \in I(x^j_o)$. Define $m\times m_j$-matrices

$$G_j = (-g^{ij}_x),\quad i \in I(x^j_o). \tag{3.8}$$

Let further

$$M_j = e^j_{xx} - \sum_{i\in I(x^j_o)} w^{ij}_o g^{ij}_{xx} \tag{3.9}$$

and

$$T_j = \{\mu \in R^m \mid \mu^T G_j = 0\}. \tag{3.10}$$

Then, for $j = 1, \ldots, r$, we have

$$e^j_x + G_j w^j_o = 0, \tag{3.11}$$

$$\mu^T M_j \mu < 0 \text{ for } \mu \in T_j,\ \mu \neq 0, \tag{3.12}$$

and

$$w^j_o > 0. \tag{3.13}$$

P is an open set. Therefore, for every $\xi \in R^n$ we can find $t^* > 0$ and $p \in C^2([0,t^*],P)$ such that

$$p(0) = p_o, \quad p_t(0) = \xi. \tag{3.14}$$

For every $x^j_o \in E_o$ define $\Phi^j : R^m \times R^{m_j} \times R \to R^{m+m_j}$ by

$$\Phi^j(x,w,t) = \begin{bmatrix} e_x(p(t),x) - \sum_{k\in I(x^j_o)} w_k g^k_x(x) \\ \vdots \\ -g^i(x) \\ \vdots \end{bmatrix}, \quad i \in I(x^j_o). \tag{3.15}$$

(3.1) and (3.11) imply

$$\Phi^j(x^j_o,w^j_o,0) = 0, \quad j = 1, \ldots, r. \tag{3.16}$$

The following theorem is proved in [3].

<u>Theorem 3.1</u>. There are neighbourhoods $U(x^j_o)$, $U(w^j_o)$ of x^j_o, w^j_o, a $t_o > 0$, and functions $x^j \in C^2([0,t_o],U(x^j_o))$, $w^j \in C^2([0,t_o],U(w^j_o))$ such that $x^j(0) = x^j_o$, $w^j(0) = w^j_o$, and

(i) For $(t,x,w) \in [0,t_o] \times U(x^j_o) \times U(w^j_o)$ we have $\Phi^j(x,w,t) = 0$ if and only if $x = x^j(t)$, $w = w^j(t)$.

(ii) For $t \in [0,t_o]$, $e(p(t),x)$ has local maxima in $\bigcup_{j=1}^{r} U(x_o^j)$ in exactly the points $x^1(t), \ldots, x^r(t)$.

(iii) Let G_j and M_j be given by (3.8), (3.9). The derivatives $x_t^j(0)$, $w_t^j(0)$ are uniquely determined by

$$\begin{bmatrix} M_j & G_j \\ G_j^T & 0 \end{bmatrix} \begin{bmatrix} x_t^j(0) \\ w_t^j(0) \end{bmatrix} = \begin{bmatrix} -e_{xp}^j \xi \\ 0 \end{bmatrix}. \tag{3.17}$$

Let

$$\mu_j = x_t^j(0). \tag{3.18}$$

The following conditions are established in [3].

<u>Theorem 3.2</u>. If p_o is a locally best approximation, then, for every $\xi \in K$,

$$K = \{\xi \mid \xi^T e_p^j \le 0,\ j = 1, \ldots, r\}, \tag{3.19}$$

there are real numbers $u_o^j \ge 0$, such that

$$\sum_{j=1}^{r} u_o^j = 1, \tag{3.20}$$

$$\sum_{j=1}^{r} u_o^j e_p^j = 0, \tag{3.21}$$

and

$$q(u_o,\xi) = \xi^T\{\sum_{j=1}^{r} u_o^j e_{pp}^j\}\xi - \sum_{j=1}^{r} u_o^j \mu_j^T M_j \mu_j \ge 0. \tag{3.22}$$

Observe that, by (3.17), we have

$$\mu_j^T M_j \mu_j = \xi^T\{[e_{px}^j \ 0]\begin{bmatrix} M_j & G_j \\ G_j^T & 0 \end{bmatrix}^{-1}\begin{bmatrix} e_{xp}^j \\ 0 \end{bmatrix}\}\xi. \tag{3.23}$$

Thus, given $u_o = (u_o^1,\ldots,u_o^r)^T$, $q(u_o,\xi)$ is a quadratic form in ξ.

If $q(u_o,\xi) \ge 0$ is replaced by $q(u_o,\xi) > 0$, the condition proves to be sufficient.

Theorem 3.3. If, for every $\xi \in K$, there are $u_o^j \geq 0$, such that (3.20), (3.21) hold and, if $\xi \neq 0$, $q(u_o,\xi) > 0$, then p_o is a locally best approximation in the strict sense, i.e. there is a neighbourhood $U_o \subset P$ of p_o such that $\|e(p,\cdot)\| > \|e(p_o,\cdot)\|$ for $p \in U_o \sim \{p_o\}$.

Obviously, Theorem 3.3 implies:

Theorem 3.4. If (3.20), (3.21) has a unique solution $u_o^j > 0$, $j = 1, \ldots, r$, such that $q(u_o,\xi) > 0$ for $\xi \in K \sim \{0\}$, then p_o is a strict, locally best approximation.

Note that the assumptions in Theorem 3.4 imply that K is a linear subspace: $K = \{\xi \mid \xi^T e_p^j = 0,\ j = 1, \ldots, r\}$. Therefore, the positive definiteness of $q(u_o,\xi)$ on K is sufficient (cf. [9]).

4. The method. Consider the following system of $N = n + r + 1 + mr + \sum_{j=1}^{r} m_j$ equations for the N unknowns $p \in R^n$, $u \in R^r$, $d \in R$, $x^j \in R^m$, $w^j \in R^{m_j}$, $j = 1, \ldots, r$:

$$\sum_{j=1}^{r} u^j e_p(p,x^j) = 0 \tag{4.1}$$

$$e(p,x^j) - d = 0, \quad j = 1, \ldots, r \tag{4.2}$$

$$-\sum_{j=1}^{r} u^j = -1 \tag{4.3}$$

$$e_x(p,x^j) - \sum_{i \in I(x^j)} w^{ij} g_x^i(x^j) = 0, \; j = 1, \ldots, r \tag{4.4}$$

$$g^i(x^j) = 0, \quad i \in I(x^j), \quad j = 1, \ldots, r. \tag{4.5}$$

If p_o is a locally best approximation, the relations (3.21), (3.3), (3.20), (3.4), and (3.2) show that p_o, u_o, $d_o = \|e(p_o,\cdot)\|$, x_o^j, w_o^j solve (4.1) - (4.5).

Method I. Given some approximation $z_1 = (p_1^T, u_1^T, d_1, x_1^{jT}, w_1^{jT})^T \in R^N$ of $z_o = (p_o^T, u_o^T, d_o, x_o^{jT}, w_o^{jT})^T$, system (4.1) - (4.5) is solved by Newton's method. Naturally, convergence is not secure in general (cf. Section 5).

If the system is briefly denoted by

$$F(z) = b, \tag{4.6}$$

approximations z_i, $i = 2, 3, \ldots$, are computed according to

$$z_{i+1} = z_i + \Delta z_i, \tag{4.7}$$

where Δz_i is the solution of the linear system

$$F_z(z_i)\Delta z_i = b - F(z_i). \tag{4.8}$$

If $F_z(z_o)$ is nonsingular and z_1 sufficiently close to z_o, then the z_i are known to converge to z_o, the convergence being superlinear and even quadratic if some additional assumptions hold (cf. [7]). Sufficient conditions for $F_z(z_o)$ to be nonsingular are given in Section 5. $F_z(z)$ may be written as follows

$$F_z(z) = \left[\begin{array}{c|c|c|c|c} A(z) & B(z) & 0 & D(z) & 0 \\ \hline (B(z))^T & 0 & C & S(z) & 0 \\ \hline 0 & C^T & 0 & 0 & 0 \\ \hline D'(z) & 0 & 0 & M(z) & G(z) \\ \hline 0 & 0 & 0 & (G(z))^T & 0 \end{array}\right]. \tag{4.9}$$

Here

$$A(z) = \sum_{j=1}^{r} u^j e_{pp}(p,x^j) \qquad \text{(n×n-matrix)} \tag{4.10}$$

$$B(z) = (e_p(p,x^1),\ldots,e_p(p,x^r)) \qquad \text{(n×r-matrix)} \tag{4.11}$$

$$C^T = (-1,\ldots,-1) \qquad \text{(1×n-matrix)} \tag{4.12}$$

$$D(z) = (u^1 e_{px}(p,x^1)|\ldots|u^r e_{px}(p,x^r)) \qquad \text{(n×rm-matrix)} \tag{4.13}$$

$$(D'(z))^T = (e_{px}(p,x^1)|\ldots|e_{px}(p,x^r)) \qquad \text{(n×rm-matrix)} \tag{4.14}$$

$$S(z) = \begin{bmatrix} (e_x(p,x^1))^T & & 0 \\ & \ddots & \\ 0 & & (e_x(p,x^r))^T \end{bmatrix} \qquad \text{(r×rm-matrix)} \tag{4.15}$$

$$M(z) = \begin{bmatrix} M_1(z) & & 0 \\ & \ddots & \\ 0 & & M_r(z) \end{bmatrix} \qquad \text{(rm×rm-matrix)} \tag{4.16}$$

$$M_j(z) = e_{xx}(p,x^j) - \sum_{i\in I(x^j)} w^{ij} g^i_{xx}(x^j), \quad j = 1, \dots, r \tag{4.17}$$

$$G(z) = \begin{bmatrix} G_1(z) & & 0 \\ & \ddots & \\ 0 & & G_r(z) \end{bmatrix} \qquad (rm\times(\sum_{j=1}^{r} m_j)\text{-matrix}) \tag{4.18}$$

$$G_j(z) = (-g^i_x(x^j)), \qquad i \in I(x^j) \qquad (m\times m_j\text{-matrix}). \tag{4.19}$$

Apart from coefficients u^j in $D(z)$ (recall $D'(z)$) and the submatrix $S(z)$, $F_z(z)$ is symmetric. Assuming $u^j > 0$, $j = 1, \dots, r$ (we need this assumption for our proof of convergence too) a fully symmetric matrix is obtained if in (4.4), (4.5) the resp. j-th group of equations is multiplied by u^j and (4.2) is replaced by $e(p,x^j) - \sum_{i\in I(x^j)} w^{ij} g^i(x^j) - d = 0$. Then, the submatrix $S(z)$ is replaced by a matrix $S^*(z)$ with $S^*(z_o) = 0$ (cf. (3.11)), so that, in practice, this submatrix may be neglected.

A second method based on Theorem 3.1 can be formulated, which, however, will be shown to be essentially equivalent to the first one.

Let z_1 be an approximation of the solution with the property that x^j_1 are exact local maxima of $e(p_1,x)$ and such that (i) and (ii) in Assumption 3.2 hold for $\bar{x} = x^j_1$, $\bar{w} = w^j_1$. Thus, (4.4) and (4.5) hold. Note that Theorem 3.1 is applicable and gives information about the dependence of x^j_1, w^j_1 on p.

<u>Method II</u>. In (4.1), (4.2) x^j are regarded as functions $x^j(p)$ of p. Then (4.1) - (4.3) is a system of n + r + 1 equations for the unknowns p, u, d. Compute p_2, u_2, d_2 by performing one step of Newton's method. Compute x^j_2, w^j_2 such that (4.4), (4.5) are satisfied and start again.

We show that Theorem 3.1 gives us all information about $x^j(p)$ needed to compute p_2, u_2, d_2. We have

$$p_2 = p_1 + \Delta p_1, \qquad u_2 = u_1 + \Delta u_1, \qquad d_2 = d_1 + \Delta d_1, \tag{4.20}$$

where Δp_1, Δu_1, Δd_1 solve the linear system

$$\begin{bmatrix} A'(z_1) & B(z_1) & 0 \\ (B(z_1))^T & 0 & C \\ 0 & C^T & 0 \end{bmatrix} \begin{bmatrix} \Delta p_1 \\ \Delta u_1 \\ \Delta d_1 \end{bmatrix} = b' \tag{4.21}$$

with C given by (4.12), B(z) given by (4.11), b' defined according to (4.6), (4.8), and

$$A'(z_1) = \sum_{j=1}^{r} u_1^j [e_{pp}(p_1,x_1^j) + e_{px}(p_1,x_1^j)x_p^j(p_1)]. \tag{4.22}$$

For this, we have used the relation

$$e_x(p_1,x_1^j)x_p^j(p_1) = 0, \qquad j = 1, \ldots, r, \tag{4.23}$$

which will be proved immediately.

Thus, Method II is fully defined, if $e_{px}(p_1,x_1^j)x_p^j(p_1)$ in (4.22) can be computed. This is possible by means of (3.17) : With $M_j(z_1)$, $G_j(z_1)$ and $e_{xp}(p_1,x_1^j)$ instead of M_j, G_j and e_{xp}^j, for $\xi = e_k \in R^n$, the k-th unit vector, the solution of (3.17) is just the k-th column of the $(m+m_j)\times n$-matrix $\begin{bmatrix} x_p^j(p_1) \\ w_p^j(p_1) \end{bmatrix}$. Thus we have

$$\begin{bmatrix} M_j(z_1) & G_j(z_1) \\ (G_j(z_1))^T & 0 \end{bmatrix} \begin{bmatrix} x_p^j(p_1) \\ w_p^j(p_1) \end{bmatrix} = \begin{bmatrix} -e_{xp}(p_1,x_1^j) \\ 0 \end{bmatrix}. \tag{4.24}$$

Especially $(G_j(z_1))^T x_p^j(p_1) = 0$. Taking account of (3.4), this shows (4.23).

In [3] it is shown that for z_1 in a certain neighbourhood of z_o the matrix on the left of (4.24) is nonsingular. Therefore, we get from (4.24)

$$e_{px}(p_1,x_1^j)x_p^j(p_1) = [e_{px}(p_1,x_1^j),0] \begin{bmatrix} M_j(z_1) & G_j(z_1) \\ (G_j(z_1))^T & 0 \end{bmatrix}^{-1} \begin{bmatrix} -e_{xp}(p_1,x_1^j) \\ 0 \end{bmatrix}. \tag{4.25}$$

By means of (4.25) we can further show that p_2, u_2, d_2 computed according to (4.20) are the same as in Method I for p_1, u_1, d_1, x_1^j, w_1^j satisfying (4.4), (4.5): in this case, Δp, Δx, and Δw satisfy

$$\begin{bmatrix} D'(z_1) \\ 0 \end{bmatrix} \Delta p + \begin{bmatrix} M(z_1) & G(z_1) \\ (G(z_1))^T & 0 \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta w \end{bmatrix} = 0. \tag{4.26}$$

From (4.26) we get

$$[D(z_1),0]\begin{bmatrix}\Delta x\\ \Delta w\end{bmatrix} = [D(z_1),0]\begin{bmatrix}M(z_1) & G(z_1)\\ (G(z_1))^T & 0\end{bmatrix}^{-1}\begin{bmatrix}D'(z_1)\\ 0\end{bmatrix}\Delta p.$$

Substitution in (4.8) shows that the first $n + r + 1$ equations in (4.8) are identical with (4.21).

Concerning the corrections on x^j, w^j, the methods are different. By Method II x_2^j, w_2^j are computed such that (4.4), (4.5) hold exactly (at least theoretically). Again, this can be done using Newton's method, which is locally convergent due to the nonsingularity of $\begin{bmatrix}M(z_1) & G(z_1)\\ (G(z_1))^T & 0\end{bmatrix}$. By Method I, however, only one step of Newton's method is executed. Taking account of this relationship it is easily shown that local (quadratic) convergence of Method I implies the same for Method II.

If $B = [\alpha,\beta]$(approximation on an interval) $A'(z)$ can easily be computed explicitely. For $x^j \in \{\alpha,\beta\}$ we get $e_{px}(p,x^j)x_p^j(p) = 0$ and for $x^j \in (\alpha,\beta)$

$$e_{px}(p,x^j)x_p^j(p) = -\frac{e_{px}(p,x^j)e_{xp}(p,x^j)}{e_{xx}(p,x^j)}\,. \tag{4.27}$$

Method II may give rise to the question if we could not proceed more simply as follows: Solve alternately (4.1) - (4.3) with x^j fixed and (4.4), (4.5) with p, u, d fixed. The following example shows that convergence may be destroyed:

Approximate $f(x) = 1 - x^2$ in $[-1,1]$ by $a(p,x) = \frac{1}{2}p^2 - 2px$, $p \in R$. It is easily shown that the sufficient condition given in Theorem 3.4 holds for $p_o = 0$. We have $e(p,x) = |1 - x^2 - \frac{1}{2}p^2 + 2px|$ and $E_o = \{x^1\} = \{0\}$. For $|p| < 1$, the only global maximum of $e(p,x)$ in $[-1,1]$ is $x = p$. (4.1) - (4.5) become:
$u^1(-p+2x^1) = 0$, $1 - (x^1)^2 - \frac{1}{2}p^2 + 2px^1 - d = 0$, $u^1 = 1$, $-2x^1 + 2p = 0$.

Given p_1, $|p_1| < 1$, from the last equation we get $x_1^1 = p_1$ and, therefore, from the first one $p_2 = 2p_1$. Thus for every $p_1 \neq 0$, the method diverges.

On the other hand Method I and II converge: From the last equation we get $x^1 = p$ and, therefore, from the first one $-p + 2p = p = 0$.

5. Convergence. As we have pointed out in the last section, Method I and II converge for a starting point z_1, sufficiently close to a solution z_o of $F(z) = 0$, if $F_z(z_o)$ (see (4.9)) is nonsingular.

Theorem 5.1. If Assumptions 2.1 and 2.2 hold and if the sufficient condition for locally best approximations given in Theorem 3.4 holds in $p = p_o$, then $F_z(z_o)$ is nonsingular.

Proof. Let G_j and M_j be given by (3.8), (3.9). Then we have

$$F_z(z_o) = \left[\begin{array}{c|c|c|c|c} \sum\limits_{j=1}^{r} u_o^j e_{pp}^j & e_p^1 \;\cdots\; e_p^r & 0 & u_o^1 e_{px}^1 \;\cdots\; u_o^r e_{px}^r & 0 \\ \hline \begin{matrix} (e_p^1)^T \\ \vdots \\ (e_p^r)^T \end{matrix} & 0 & \begin{matrix} -1 \\ \vdots \\ -1 \end{matrix} & 0 & 0 \\ \hline 0 & -1 \;\cdots\; -1 & 0 & 0 & 0 \\ \hline \begin{matrix} e_{xp}^1 \\ \vdots \\ e_{xp}^r \end{matrix} & 0 & 0 & \begin{matrix} M_1 & & 0 \\ & \ddots & \\ 0 & & M_r \end{matrix} & \begin{matrix} G_1 & & 0 \\ & \ddots & \\ 0 & & G_r \end{matrix} \\ \hline 0 & 0 & 0 & \begin{matrix} G_1^T & & 0 \\ & \ddots & \\ 0 & & G_r^T \end{matrix} & 0 \end{array}\right] \tag{5.1}$$

We show that the assumption that $F_z(z_o)$ is singular leads to a contradiction.

If $F_z(z_o)$ is singular, then there are $\xi \in R^n$, $\eta \in R^r$, $\delta \in R$, $\xi^1, \ldots, \xi^r \in R^m$, and $\omega^j \in R^{m_j}$, $j = 1, \ldots, r$, not all equal to zero, such that

$$F_z(z_o)(\xi^T, \eta^T, \delta, \xi^{1^T}, \ldots, \xi^{r^T}, \omega^{1^T}, \ldots, \omega^{r^T})^T = 0. \tag{5.2}$$

We may assume that

$$\delta \leq 0. \tag{5.3}$$

First, we show that $\xi = 0$ implies that all other vectors η, δ, ξ^j, ω^j are zero, contrary to our assumption.

Let $\xi = 0$. Then, for $j = 1, \ldots, r$, $G_j^T \xi^j = 0$. Thus $\xi^j \in T_j$ (see (3.10)) and

$$M_j \xi^j + G_j \omega^j = 0. \tag{5.4}$$

Multiplication by $(\xi^j)^T$ shows

$$(\xi^j)^T M_j \xi^j = 0.$$

Therefore, by (3.12), $\xi^j = 0$, $j = 1, \ldots, r$. Hence, by (5.4), we have $G_j\omega^j = 0$. Since the $m\times m_j$-matrix G_j has full rank $m_j \leq m$ (see Assumption 3.1), $G_j\omega^j = 0$ is possible if and only if $\omega^j = 0$, $j = 1, \ldots, r$.

Moreover, since $\xi^j = 0$, we have

$$\sum_{j=1}^{r} \eta_j e_p^j = 0, \qquad \sum_{j=1}^{r} \eta_j = 0 \qquad (\eta=(\eta_1,\ldots,\eta_r)^T).$$

Hence, by the uniqueness of u_o^j assumed in Theorem 3.4, $\eta = 0$.

Finally, $\xi = 0$ implies $\delta = 0$.

Now assume that $\xi \neq 0$. By (5.3), $\xi^T e_p^j \leq 0$, $j = 1, \ldots, r$. Hence, $\xi \in K$ (see (3.19)). From

$$\sum_{j=1}^{r} u_o^j e_{pp}^j \xi + \sum_{j=1}^{r} \eta_j e_p^j + \sum_{j=1}^{r} u_o^j e_{px}^j \xi^j = 0$$

we obtain

$$\xi^T(\sum_{j=1}^{r} u_o^j e_{pp}^j)\xi + \sum_{j=1}^{r} u_o^j \xi^T e_{px}^j \xi^j + \sum_{j=1}^{r} \eta_j \xi^T e_p^j = 0. \tag{5.5}$$

Moreover, the relations

$$e_{xp}^j \xi + M_j \xi^j + G_j \omega^j = 0, \qquad G_j^T \xi = 0 \tag{5.6}$$

hold and show that ξ^j and ω^j are the unique solutions of (3.17) pertinent to ξ. Especially, $\xi^j = \mu_j$, $j = 1, \ldots, r$ (see (3.18)). Finally,

$$\sum_{j=1}^{r} \eta_j \xi^T e_p^j = \delta \sum_{j=1}^{r} \eta_j = \delta 0 = 0. \tag{5.7}$$

Hence, by (5.5), (5.6), (5.7), and $\xi^j = \mu_j$, we obtain

$$q(u_o,\xi) = \xi^T(\sum_{j=1}^{r} u_o^j e_{pp}^j)\xi - \sum_{j=1}^{r} u_o^j \mu_j^T M_j \mu_j = 0,$$

contrary to $q(u_o,\xi) > 0$ for $\xi \in K \sim \{0\}$. This completes the proof.

<u>Numerical example</u>. Consider the problem of approximating $f(x) = \sqrt{x}$ on $[0.25, 1]$ by the H-polynomial (cf [5]) $a(p,x) = \sigma((p_o x+p_1)x+p_2)^2 + p_3$, $\sigma = \pm 1$. In a first step, the

discrete problem on $\{x_o, \ldots, x_4\} \subset [0.25, 1]$, $x_i = 0.25 + i(1-0.25)/4$, $i = 0, \ldots, 4$, has been solved by the method described in [5]. The solution is approximately

$$a(p^1,x) = -((0.0874x-0.4956)x+1.118)^2 + 1.501.$$

$e(p^1,x)$ has local maxima in $x = 0.25, 0.387, 0.764, 1.0$ with resp. the values 0.00246, 0.00282, 0.00265, 0.00246.

Carrying out three steps of Method I, we find p^*,

$$p_o^* = 0.088\ 090\ 539\ 351\ , \qquad p_1^* = -\ 0.495\ 240\ 777\ 440$$

$$p_2^* = 1.119\ 066\ 635\ 328\ , \qquad p_3^* = 1.504\ 174\ 868\ 404$$

The values of $e(p^*,x)$ in the maxima $x = 0.25, 0.388, 0.760, 1.0$ differ less than 10^{-8}. The maximal error is about 0.00265. It is easily shown that the sufficient condition given in Theorem 3.4 holds.

Approximation by H-polynomials often results in locally best approximations with maximal error in less than $n + 1$ points. For other types of approximation on an interval $B = [\alpha,\beta]$, e.g. rational approximation, this case may be considered degenerate.

The results presented by Wetterling ([11]) indicate that, for $B \subset R^m$, $m > 1$, the case that the number of maxima is less than $n + 1$ is rather a rule than an exception.

References.

1. S.A. Gustafson and K.O. Kortanek, Numerical treatment of a class of semi-infinite programming problems, Nav. Res. Log. Quart., 20 (1973), 477-504.
2. R. Hettich, Extremalkriterien für Optimierungs- und Approximationsaufgaben, Technische Hogeschool Twente, Enschede, Dissertation, 1973.
3. R. Hettich, Kriterien zweiter Ordnung für lokal beste Approximationen, Numer. Math., 22 (1974), 409-417.
4. R. Hettich, Kriterien erster und zweiter Ordnung für lokal beste Approximationen bei Problemen mit Nebenbedingungen, Numer. Math., 25 (1975), 109-122.
5. R. Hettich, Chebyshev approximation by H-polynomials: a numerical method, J. Approximation Theory, 17 (1976).
6. G. Meinardus und D. Schwedt, Nicht-lineare Approximationen, Arch. Rat. Mech. Anal., 17 (1964), 297-326.
7. J.M. Ortega and W.C. Rheinboldt, Iterative solution of nonlinear equations in several variables, Academic Press, New York, 1970.
8. W. Wetterling, Anwendung des Newtonschen Iterationsverfahrens bei der Tschebyscheff-Approximation, insbesondere mit nichtlinear auftretenden Parametern, MTW, 10 (1963), Teil I: 61-63, Teil II: 112-115.
9. W. Wetterling, Definitheitsbedingungen für relative Extrema bei Optimierungs- und Approximationsaufgaben, Numer. Math., 15 (1970), 122-136.

10. W. Wetterling, Über Minimalbedingungen und Newton-Iteration bei nichtlinearen Optimierungsaufgaben, in: Iterationsverfahren, Numerische Mathematik, Approximationstheorie, ISNM, 15 (1970), 93-99, Birkhäuser, Basel-Stuttgart.

11. W. Wetterling, Numerische Anwendung von Kriterien zweiter Ordnung für lokal beste Approximationen, presented at the colloquium underlying these Proceedings.