JOURNAL OF APPROXIMATION THEORY **36**, 237–256 (1982)

On the Convergence of a Class of Nonlinear Approximation Methods

REMBERT REEMTSEN

Fachbereich Mathematik, Technische Hochschule Darmstadt, 6100 Darmstadt, West Germany

Communicated by Oved Shisha

Received May 19, 1980; revised July 28, 1982

A family of algorithms for nonlinear approximation is defined by point-to-set-maps. Then Zangwill's general convergence theorem is used for the convergence proof of a class of methods for the nonlinear approximation problem with infinitely many linear constraints. Numerous well-known techniques are included and generalized in this way. The convergence proof presented here is kept so general that convergence could be shown by it for a variety of similar methods.

1. INTRODUCTION

DEFINITION 1.1. Let A and B be sets, and let $P(B)$ be the set of all subsets of B. Then a map $Z: A \to P(B)$ which assigns to each $a \in A$ exactly one subset of B is called a point-to-set map.

Zangwill [39, 40] seems to have been the first who fully used point-to-set maps in the field of mathematical programming. He recognized, in particular, that an algorithm can be defined by a point-to-set map $Z: A \to P(B)$ where for $a_0 \in A$ given, the iteration of the algorithm generates a sequence (a_k) such that $a_k \in Z(a_{k-1})$. Zangwill states then a general convergence theorem, the main assumption of which is the closedness of the algorithmic map Z. In the German literature recently, Zangwill's theorem has been systematically developed and applied to the proof of the convergence of numerous nonlinear programming algorithms in the book of Horst [15]. By means of simple examples it has been shown there that the assumptions of the convergence theorem can be only slightly relaxed. Extensions of Zangwill's theorem can, for instance, be found in the papers of Huard [16] and Tishyadhigama *et al.* [35]. For further references and information on point-to-set maps, we refer, in particular, to [23] and its introduction by Huard.

While the use of point-to-set maps and their topological concepts requires

0021-9045/82/110237-20$02.00/0

a certain inconvenient terminology, it offers on the other hand a clear insight into the conditions for the convergence of algorithms as well as enables the treatment of a broad class of methods in a unified way. Further, parts of the algorithms can be altered without touching the others and so without making a completely new convergence proof necessary. In this paper we want to make use of Zangwill's theorem for the convergence proof of a class of algorithms that seek a stationary point, i.e., a saddle point or a local optimum, of the following nonlinear approximation problem:

$$\text{Find } \hat{a} \in E \text{ such that } z(\hat{a}) \leqslant z(a) \text{ for all } a \in E \text{ where } z(a) = \|f - F(a)\|. \tag{1.1}$$

Thereby $A \subseteq \mathbb{R}^p$ is an open set and $E \subseteq A$ is a nonempty set of feasible points; further $F: \mathbb{R}^p \to C(T)$ is a given mapping with $C(T)$ being the space of all continuous functions on a compact set $T \subset \mathbb{R}^n$ and $f \in C(T)$ is a given function which shall be approximated. The norm $\|\cdot\|$ on $C(T)$ can be chosen here arbitrarily.

DEFINITION 1.2. We say $F(\hat{a})$ for $\hat{a} \in E$ is a locally best approximation to f on T with respect to E if there exists an $\varepsilon > 0$ such that

$$\|f - F(\hat{a})\| \leqslant \|f - F(a)\| \qquad \forall a \in E \cap U_{\hat{a}}^{\varepsilon}$$

where $U_{\hat{a}}^{\varepsilon}$ is an ε-neighborhood of $\hat{a}$.

The existence of best approximations is usually difficult to verify and not investigated here. We refer the reader to the relevant literature.

In Section 2 some topological concepts of point-to-set maps as well as Zangwill's general convergence teorem are provided. A family of algorithms for problem (1.1) is then defined in Section 3 wherein one has a certain freedom in establishing the set $\Omega(a)$ of feasible directions at $a \in E$. For some choices of Ω theorems on the convergence of the corresponding algorithms are stated. The main part of their proofs, which are summarized in Appendix 2, consists of verifying the closedness of the algorithmic map. It results here in the proof of the continuity of a mapping $a \to m(a)$ which assigns to each parameter vector $a \in E$ the minimal value $m(a)$ of a certain linear optimization problem. To show the latter, results on the continuous dependence of the feasible set and the optimal value in an optimization problem on the parameters are used. The corresponding statements which we need here can be found in Krabs [18] and are cited in Appendix 1. They are also naturally expressed in terms of point-to-set maps. Finally we present in Section 4 an exemplary list of references to algorithms pertaining to the class considered here and to further information on them. Thereby emphasis is put on the maximum norm case because of its importance.

2. A General Convergence Theorem

We begin with the formulation of some definitions as they are given in Krabs [18].

Definition 2.1. Let $Z: A \to P(B)$ be a point-to-set map from $A \subseteq \mathbb{R}^p$ into $B \subseteq \mathbb{R}^m$ with $Z(a) \neq \emptyset$ for all $a \in A$.

(a) Z is closed at $a \in A$ if for each sequence (a_k) in A with $a_k \to a$ and for each sequence (b_k) in B with $b_k \in Z(a_k)$ for almost all k and $b_k \to b \in B$, it follows that $b \in Z(a)$.

(b) Z is open at $a \in A$ if for each sequence (a_k) in A with $a_k \to a$ the following holds true: to each $b \in Z(a)$ there exists a sequence (b_k) in B with $b_k \in Z(a_k)$ for almost all k and $b_k \to b$.

(c) Z is continuous at $a \in A$ if Z is open as well as closed at $a \in A$.

Reasonably an algorithm can be defined only if an element is characterized for which the algorithm shall look. This element will usually belong to a set Λ which we shall call the set of solution points. In the case of problem (1.1), one will choose Λ to be the set of stationary points, a definition of which is given later.

From Satz 2.21 in [15] or from convergence Theorem A in [40], the following theorem on the convergence of an algorithm for problem (1.1) can now be derived.

Theorem 2.1. *Let z be defined by* (1.1). *Further, let a set of feasible points $E \subseteq A \subseteq \mathbb{R}^p$ and a set Λ of solution points be known. Finally, let the point-to-set map $Z: E \to P(E)$ determine an algorithm that, given a point $a_0 \in E$, generates a sequence (a_k) in E with $a_k \in Z(a_{k-1})$. Suppose*

(i) *all a_k of the sequence (a_k) lie in a compact set $R \subseteq E$;*

(ii) *if $a_k \in \Lambda$, then the algorithm terminates;*

(iii) *if $a_k \notin \Lambda$, then for any $a_{k+1} \in Z(a_k)$, $z(a_{k+1}) < z(a_k)$;*

(iv) *the map Z is closed at each $a \in R \backslash \Lambda$.*

Then either the algorithm stops after finitely many steps at an $\hat{a} \in \Lambda$ or it generates an infinite sequence (a_k) which possesses accumulation points and each accumulation point of it is an element of Λ.

Remark 2.1. We assume throughout this paper that the algorithms in question are constructed in such a way that they recognize if $a_k \in \Lambda$ and stop, i.e., that assumption (ii) of Theorem 2.1 is always satisfied.

Remark 2.2. The assumption $C \cap \Lambda \neq \emptyset$ in Satz 2.21 of [15], which is equivalent to $R \cap A \neq \emptyset$ here, is not needed for the proof, but is a consequence of the other assumptions. It is clear that a sequence (a_k) as characterized in Theorem 2.1 cannot exist if $R \cap A = \emptyset$.

Remark 2.3. Theorem 2.1 can be proved analogously to the quoted theorems under the following relaxation of assumption (i):

(i′) Each sequence (a_k) in E with $a_k \in Z(a_{k-1})$ has an accumulation point in E.

The point-to-set map Z that defines an algorithm is often composed of several mappings. Hence for the proof of the closedness of Z, the following theorem is useful (cf. [15]).

THEOREM 2.2. *Let $E \subseteq \mathbb{R}^p$, $B \subseteq \mathbb{R}^m$, and $C \subseteq \mathbb{R}^l$. Further, let $D: E \to P(B)$ and $S: B \to P(C)$ be point-to-set maps which satisfy the conditions*:

(i) *$D(a) \neq \emptyset$ for all $a \in E$ and $S(b) \neq \emptyset$ for all $b \in B$*;

(ii) *D is closed at $a \in E$*;

(iii) *S is closed at each $b \in D(a)$*;

(iv) *For each sequence (a_k) in E with $a_k \to a$, each sequence (b_k) in B with $b_k \in D(a_k)$ possesses an accumulation point.*

Then also the composition $Z = SD$ of D and S is closed at $a \in E$.

3. A CLASS OF ALGORITHMS FOR THE APPROXIMATION PROBLEM

The proofs of the lemmas and theorems of this section can be found in Appendix 2. Throughout this paper we require

ASSUMPTION A.1. *The mapping $F: A \to C(T)$ is once continuously Fréchet differentiable on A.*

In the following for some members of a class of algorithms, it will be shown that assumptions (iii) and (iv) of Theorem 2.1 are satisfied. Hence to get convergence of the corresponding algorithms, one will have generally to provide

ASSUMPTION A.2. *All elements a_k of a sequence (a_k) in $E \subseteq \mathbb{R}^p$ generated by the algorithm under consideration lie in a compact set $R \subseteq E$.*

Remark 3.1. Schaback [33] calls a parametrization $F: A \to C(T)$ inversely compact with respect to $f \in C(T)$ if each sequence (a_k) in A satisfying $z(a_{k+1}) \leqslant z(a_k)$ has an accumulation point in A. Inverse compactness is a global property of a family of functions which is independent of any algorithm. To verify Assumption A.2 for a given parametrization is by no means a trivial task.

We define now a class of algorithms for the numerical solution of problem (1.1) through a point-to-set map $Z: E \to P(E)$ where $Z = SD$ is a composition of the two maps D and S which are characterized in the following.

$$D: E \to P(E \times \mathbb{R}^p) \Leftrightarrow D(a) \neq \varnothing \quad \text{for all} \quad a \in E \quad \text{and}$$
$$D(a) = \{(b, d) \in E \times \mathbb{R}^p \mid b = a, \| f - F(a) - F'(a)d \| = \inf_{h \in \Omega(a)} \| f - F(a) - F'(a)h \| \}. \tag{3.1}$$

Therein $\Omega(a)$ is a set of the form

$$\Omega(a) = \{h \in W(a) \mid g(a, h) \in Q\} \tag{3.2}$$

where

$$Q = \{r \in \mathbb{R} \mid r \geqslant 0\} \tag{3.3}$$

and $W: E \to P(\mathbb{R}^p)$ and $g: E \times \mathbb{R}^p \to \mathbb{R}$ are certain maps which will be defined in different ways below. Further,

$$S: E \times \mathbb{R}^p \to P(E) \Leftrightarrow S(b, d) \neq \varnothing \quad \text{for all} \quad (b, d) \in D(a) \text{ and}$$
$$S(b, d) = \{c \in E \mid c = b + \alpha d, z(c) = \min_{0 \leqslant \lambda \leqslant 1} z(b + \lambda d)\} \tag{3.4}$$

with z as in (1.1) where E and $\Omega(a)$ have to be such that for each $(b, d) \in D(a)$ all $c \in S(b, d)$ lie again in E. The definition of a solution set Λ is closely related to the definition of Ω or W and g, respectively.

Approximation problems with different kinds of constraints as well as a variety of methods for their solution can be treated with the algorithmic map $Z = SD$, (3.1)–(3.4), by a suitable choice of the point-to-set map $\Omega: E \in \mathbb{R}^p \to P(\mathbb{R}^p)$. For a better understanding let us give a simple example. If $E = \mathbb{R}^p$ in (1.1), we can define

$$g(a, h) = \| h \| \qquad \forall (a, h) \in A \times \mathbb{R}^p$$

and

$$W(a) = \{h \in \mathbb{R}^p \mid \| a + h \| \geqslant 0\} = \mathbb{R}^p \qquad \forall a \in E$$

for any vector norm $\| \cdot \|$ such that $\Omega(a)$ becomes equal to $\mathbb{R}^p$. In this case (3.1) means numerically the solution of an unconstrained linear approximation problem, a solution of which should yield together with (3.4) a direction of descent. If E in (1.1) is defined as the set of solutions of linear or nonlinear constraints, these have to be regarded in an algorithm. In our model of an algorithm such constraints may be handled by an appropriate choice of W (cf. (3.5), (3.6)). Finally, one may want to avoid the one-dimensional minimization in (3.4). In this case, the linear approximation problem in (3.1) has to be solved under certain constraints on h to guarantee that each d is a downward direction. Such controls of the magnitude of d can be described through the function g (cf. (3.8), (3.9), and Remark 3.3). Roughly speaking, $\Omega(a)$ should be of such a form that $D(a) \neq \varnothing$ for all $a \in E$ and each $d \in \Omega(a)$ with $(b, d) \in D(a)$ is (possibly in connection with (3.4)) a direction of descent in $a \in E$. In particular, Ω has to satisfy the assumptions of Theorem 5.2. Finally, Λ should be the set of stationary points for the problem under consideration.

In this paper we discuss the nonlinear approximation problem where linear constraints are imposed on the choice of the parameters. For this purpose let I and J be index sets. If not emphasized otherwise, I and J can have infinitely many elements. Let further $u_i(a) = s_i^T a + r_i$, $i \in I$, and $v_j(a) = s_j^T a + r_j$, $j \in J$, be functionals on A with $s_k \in \mathbb{R}^p$ and $r_k \in \mathbb{R}$ for all k of I and J. Then the set of feasible points is here defined by

$$E = \{a \in A \mid u_i(a) \geqslant 0, i \in I, v_j(a) = 0, j \in J\}, \tag{3.5}$$

where E is assumed to be nonempty. Further, we set

$$W(a) = \{h \in \mathbb{R}^p \mid u_i(a+h) \geqslant 0, i \in I, \text{and } v_j(a+h) = 0, j \in J\}, a \in E. \tag{3.6}$$

The set Λ of solution points is then given by

$$\Lambda = \{a \in E \mid \inf_{h \in W(a)} \| f - F(a) - F'(a)h \| = \| f - F(a) \| \}. \tag{3.7}$$

The elements of Λ are also referred to as stationary points. As a first result we now obtain

LEMMA 3.1. *If for $a \in E$ and $h \in W(a)$ the inequality $\| f - F(a) - F'(a)h \| < \| f - F(a) \|$ is valid, then for all sufficiently small $\lambda > 0$, $a + \lambda h \in E$ as well as $\| f - F(a + \lambda h) \| < \| f - F(a) \|$ holds true.*

COROLLARY 3.1. *If $F(\hat{a})$, $\hat{a} \in E$, is a locally best approximation to f on T with respect to E, then $\hat{a}$ is an element of Λ.*

Remark 3.2. In general only very restrictive a priori assumptions or local considerations, respectively, will guarantee that conversely $F(\hat{a})$ is a locally best approximation if $\hat{a}$ is a stationary point.

To prove the convergence of an algorithm of the class (3.1)–(3.7) for the approximation problem (1.1), (3.5), we shall have to ensure $D(a) \neq \varnothing$ for all $a \in E$ and, in particular, to satisfy conditions (iv) of Theorems 2.2 and 5.1, respectively. This suggests first to imbed the directions d with $(b, d) \in D(a)$ into a compact set. Therefore, we define g as

$$g(a, h) = l(a) - n(h) \qquad \forall\, (a, h) \in E \times \mathbb{R}^p, \tag{3.8}$$

where $n(h)$ is an arbitrary vector norm of h and $l: E \to \mathbb{R}$ is a continuous functional on E with the property

$$0 < l(a) \qquad \forall\, a \in E \backslash \Lambda \qquad \text{and} \qquad 0 \leqslant l(a) \qquad \forall\, a \in \Lambda. \tag{3.9}$$

LEMMA 3.2. *Let $\Omega(a)$, $a \in E$, be defined through* (3.2), (3.3), (3.5), (3.6), (3.8), *and* (3.9). *Then the infimum in* (3.1) *is achieved for a $\hat{h} \in \Omega(a)$.*

From now on we assume further that either for each $a \in E$ $l(a)$ is small enough such that $a + \lambda h \in A$ for all $h \in \Omega(a)$ and $\lambda \in [0, 1]$ or that $A = \mathbb{R}^p$; hence in any case $a + \lambda h \in E$, $0 \leqslant \lambda \leqslant 1$, for E (3.5). So under Assumptions A.1 and A.2 the algorithm as described in the following theorem converges in the sense of Theorem 2.1 to a stationary point of the approximation problem (1.1), (3.5).

THEOREM 3.1. *For the algorithmic map $Z = SD$ and the solution set Λ which are defined through* (3.1)–(3.9), *assumptions* (iii) *and* (iv) *of Theorem* 2.1 *are satisfied.*

Lemmas 3.1 and 3.2, Corollary 3.1, and Theorem 3.1 include obviously the case of the unconstrained approximation problem with

$$E = A \qquad \text{and} \qquad W(a) = \mathbb{R}^p \qquad \text{for all} \quad a \in E.$$

Now it is worth considering the case where $\Omega(a) = W(a)$ for all $a \in E$, i.e., the case where $\Omega(a)$ is possibly unbounded. To maintain the definition (3.2), (3.3) of Ω, we set for this matter

$$g(a, h) = n(h) \qquad \forall\, (a, h) \in E \times \mathbb{R}^p. \tag{3.10}$$

LEMMA 3.3. *If I and J have finitely many elements, then for $\Omega(a)$, $a \in E$, given by* (3.2), (3.3), (3.5), (3.6), *and* (3.10) *the infimum in* (3.1) *is attained for a $\hat{h} \in \Omega(a)$. If, further, $[F'(a)]^{-1}$ exists, this statement is also valid for arbitrary index sets I and J.*

Then evidently the following theorem is true.

THEOREM 3.2. *Let an algorithm be defined by* (3.1)–(3.7) *and* (3.10), *where I and J are sets with finitely many members. Provided that for each* *sequence* (a_k) *in* R *converging to an* $a \in R\backslash A$ *each sequence* (d_k) *with* $(b_k, d_k) \in D(a_k)$ *possesses an accumulation point in* $\mathbb{R}^p$, *conditions* (iii) *and* (iv) *of Theorem* 2.1 *are fulfilled.*

For the existence proof of accumulation points of such sequences (d_k), the next lemma is helpful.

LEMMA 3.4. *Let* $h \in \mathbb{R}^p$ *satisfy the inequality* $\|f - F(a) - F'(a)h\| \leqslant \|f - F(a)\|$ *for an* $a \in A$. *If* $[F'(a)]^{-1}$ *exists, then* $\|h\| \leqslant 2\|f - F(a)\| \|[F'(a)]^{-1}\|$.

Hence if for all $a \in R$ the inverse $[F'(a)]^{-1}$ exists and if in addition the mapping $a \to \|[F'(a)]^{-1}\|$ from $\mathbb{R}^p$ into $\mathbb{R}$ is continuous, then the algorithm of Theorem 3.2 becomes a special case of that one in Theorem 3.1 with

$$l(a) = 2\|f - F(a)\| \|[F'(a)]^{-1}\|.$$

Instead of verifying the continuity of the map $a \to \|[F'(a)]^{-1}\|$, it would be sufficient to show that there exists a constant N with

$$\|[F'(a)]^{-1}\| \leqslant N \qquad \text{for all} \quad a \in R \tag{3.11}$$

which would suggest setting $l(a) = 2N\|f - F(a)\|$.

LEMMA 3.5. *Provided that* $F: A \subseteq \mathbb{R}^p \to \mathbb{R}^n$ *with* $n \geqslant p$ *possesses a continuous Fréchet derivative* $F'(a)$ *with* $\operatorname{rank}(F'(a)) = p$ *for all elements* $a \in A$, *a constant* N *exists such that* (3.11) *holds true.*

Consequently, with the assumptions of Lemma 3.5 we can derive from Theorem 3.2 and Lemma 3.3.

COROLLARY 3.2. *Let* $F: A \subseteq \mathbb{R}^p \to \mathbb{R}^n$ *with* $n \geqslant p$ *have a continuous Fréchet derivative* $F'(a)$ *for all* $a \in A$ *and let* $\operatorname{rank}(F'(a)) = p$ *on* A. *Then for the algorithm determined through* (3.1)–(3.7) *and* (3.10), *assumptions* (iii) *and* (iv) *of Theorem* 2.1 *are satisfied.*

We finish this section with the following

Remark 3.3. From Lemma 3.1 it can be comprehended that the mapping S (3.4) is needed to guarantee assumption (iii) of Theorem 2.1, namely, that $a_k \notin A$ implies $z(a_{k+1}) < z(a_k)$ for $a_{k+1} \in S(b_k, d_k)$. Other choices of S are possible, but are not discussed here. It is also clear that S can be omitted completely and that Z can be defined by

$$\begin{aligned} Z(a) = G(a) = \{c \in \mathbb{R}^p \mid c = a + d, \|f - F(a) - F'(a)d\| \\ = \inf_{h \in \Omega(a)} \|f - F(a) - F'(a)h\|\} \end{aligned}$$

if $\Omega(a_k)$, $a_k \notin A$, can be controlled in such a way that already for $a_{k+1} \in G(a_k)$ the condition $z(a_{k+1}) < z(a_k)$ holds true. In this case we speak of the full-step method.

4. General Discussion

In this section we are going to refer to a variety of well-known algorithms which belong to the class of methods considered here or are basically of the same kind. So we shall not mention, in particular, if in a specific method the determination of an appropriate step length happens in another way than here in (3.4). Further, in practice a continuous functional $l: E \to \mathbb{R}$ with (3.9) will usually not be given explicitly, but upper bounds $\mu_k = l(a_k)$ for the length of the directions of descent will be computed by a suitable rule only at the discrete points $a_k \in E$ of the iteration. However, then a continuous functional l on E passing through the points $(a_k, \mu_k) \in \mathbb{R}^p \times \mathbb{R}$ and satisfying (3.9) could be constructed a posteriori provided that a constant N exists with $0 < \mu_k = l(a_k) \leqslant N$ for $a_k \notin A$ and $0 \leqslant \mu_k = l(a_k) \leqslant N$ if $a_k \in A$. Moreover, it is clear that the global convergence of the algorithms in question is independent of the sizes of such μ_k's as long as the a_k lie in E again.

The algorithms of Theorems 3. 1 and 3.2 generalize in different ways those we are referring to below: they are defined for all norms, they are valid for the discrete as well as the uniform nonlinear approximation problem with infinitely many linear constraints, and finally they permit a variety of strategies to control the length of the directions of descent in each iteration. We are furthermore convinced that the general algorithm characterized by (3.1)–(3.4) and the concept of its proof can be exploited for a variety of other problems with different kinds of constraints including nonlinear ones as they are, for example, treated in Cromme [9]. The list of references given here is by no means believed to be complete. For practical purposes, some additional information is given. Thereby the discussion of the maximum norm case is somewhat emphasized because the corresponding class of algorithms has, in particular, there turned out to be very successful.

To our knowledge Schaback [33] is the only one who considered the global convergence of an algorithm of the class (3.1)–(3.4) independently of the norm. The algorithm in [33] can be obtained by setting

$$W(a) = \mathbb{R}^p \qquad \text{and} \qquad l(a) = K$$

for a constant K.

Remark 4.1. Satz 1 in [33] just proves assumption (γ_3) of Corollary 3 in Huard [16] which is an extension of Zangwill's convergence theorem. Hence convergence of Schaback's algorithm can be concluded from Corollary 3 in [16] together with Satz 1 in [33].

If not mentioned otherwise, the convergence results of the algorithms we are referring to below concern only the discrete case of the unconstrained approximation problem. Let us now first discuss the case of the L_∞-norm. The algorithm (3.1)–(3.4) with $\Omega(a) = \mathbb{R}^p$ for all $a \in \mathbb{R}^p$ is widely known as the Osborne–Watson algorithm [24]. A slightly different version of this algorithm was already suggested before by Ishizaki and Watanabe [17]; however, without any results on its convergence. In [24] a proof of convergence is presented; however, the assumptions for the convergence as, for instance, the existence of constants m and K in Lemma 2.4 there, are not clearly indicated. Further, it can be shown easily by an example that in contrast to Lemma 2.2 there, even the condition $\| F_i'(\hat{a}) \| > 0$ for all i at a stationary point $\hat{a} \in \mathbb{R}^p$ is not sufficient for the nonexistence of directions of descent at $\hat{a}$. Another convergence proof of the Osborne–Watson algorithm can be found in Anderson [1]. However, the existence of a constant "reference" which is needed for the verification of Lemma 5.1.5 cannot be concluded from Lemma 5.1.4 as is claimed there, but has to be provided as an additional assumption. This is done in Anderson and Osborne [2] where the convergence proof of [1] is extended to so-called polyhedral norms. Under assumptions which are stronger than those of Section 3, it is shown there that each limit point of a sequence (a_k) generated by the algorithm is a stationary point. Under the additional assumption of a so-called "multiplier rule," local quadratic convergence of the full-step method is proved where the corresponding proofs of Osborne [26] and Anderson [1] are extended to the polyhedral norm case. The results of Anderson, Osborne, and Watson are thoroughly discussed in Osborne and Watson [30] where also convergence results are obtained for smooth, strictly convex, and monotonic norms which include the L_p-norms for $1 < p < \infty$. Closely related to the papers just mentioned is the work of Watson [36], where an alternative procedure is suggested for the case that already a reasonably good approximation is available.

Another proof of local quadratic convergence of the Osborne–Watson algorithm for the uniform norm is given by Cromme [7]. The assumptions in [7] are that a smoothness condition is satisfied, that $F(\hat{a})$ is a strongly unique locally best approximation to f on T with respect to A, and that $F'(\hat{a})$ satisfies a regularity condition. In [8] and [10] Cromme shows that the assumption of strong uniqueness is crucial for a locally good behavior of certain iterative procedures.

Another group of papers deals with modifications of the Osborne–Watson algorithm, where the set of feasible points of the linear minimization problem in (3.1) is not the full space $\mathbb{R}^p$, but a bounded region. In our terminology

Madsen [20] considers the case $W(a) = \mathbb{R}^p$ and $n(h) = \| h \|_\infty$, where in each iteration the $\mu_k = l(a_k)$ are altered in such a way that for the full-step method a convergence result that is in character of the same kind as Theorem 3.1 here can be obtained. Madsen and Schjaer–Jacobsen [22] generalized the algorithm of [20] to the case $W(a)$ (3.6) here where I and J have finitely many indices. In addition, the authors of [22] show local quadratic convergence of their algorithm provided that a certain system of functions satisfies Haar's condition at a limit point of a sequence (a_k) generated by the algorithm. For practical purposes, let us mention that there exist also modifications of the algorithms in [20] and [22] by Madsen [21] and Hald and Schjaer–Jacobsen [12], respectively, in which the calculation of derivatives is avoided. Further Hald and Madsen [13] combine the method in [20] with a method using second order information; in this connection, see also the survey article of Hettich [14].

A member of the group of algorithms discussed at last is basically also the Levenberg-like method of Anderson [1] which was generalized to polyhedral norms by Anderson and Osborne [3]. A modification of the algorithm in [3] was presented by Watson [37] where second derivatives are taken into account. In all these algorithms the sets of feasible directions in the linear subproblems are bounded in a way which was suggested by Levenberg for a least squares algorithm.

Finally, we want to mention that there exists a class of algorithms for the nonlinear Chebyshev problem which uses the so-called local Kolmogoroff criterion for the computation of the directions of descent.

See, for instance, Schultz [34], where more references can be found, and the comments in Hettich [14]. It is known that for $\hat{a} \in A$ the condition

$$\| f - F(\hat{a}) \| \leqslant \| f - F(\hat{a}) - F'(\hat{a}) h \| \qquad \forall h \in \mathbb{R}^p$$

and the local Kolmogoroff criterion

$$\min_{x \in I(\hat{a})} (f(x) - F(\hat{a}, x)) F'(\hat{a}, x) h \leqslant 0 \qquad \forall h \in \mathbb{R}^p,$$

where

$$I(a) = \{x \in B \mid |f(x) - F(a, x)| = \| f - F(a) \| \}, \qquad a \in A,$$

are equivalent and necessary for $F(\hat{a})$ to be a locally best approximation to f on T with respect to A (see, e.g. Reemtsen [31].) If now

$$(f(x) - F(a, x)) F'(a, x) d > 0 \qquad \forall x \in I(a) \tag{4.1}$$

is valid for a $d \in \mathbb{R}^p$, it can be easily verified that d is a direction of descent at $a \in A$. And it can further be shown that if $\| d \|$ is smaller than a certain number, (4.1) implies

$$\| f - F(a) - F'(a)d \| < \| f - F(a) \|$$

(see [31].) So it might be possible to fit some of those algorithms in our model by choosing Ω in a proper way.

Let us close with some remarks on other norms. For the discrete unconstrained L_1-approximation problem, Osborne and Watsen [25], Osborne [26], Anderson and Osborne [2], and Osborne and Watson [30] derive results on the convergence of algorithm (3.1)–(3.4) with $\Omega(a) = \mathbb{R}^p$, analogously to the L_∞ case. A Levenberg-like method where the directions of descent are bounded in some sense is presented in Anderson and Osborne [3].

In the case of $\| \cdot \|$ being the L_2-norm, a variety of well-known techniques is summarized and extended by algorithm (3.1)–(3.4). The literature especially on modifications of Newton's method is so extensive that we can confine ourselves to only the most elementary instances. For example, if F is a mapping from $\mathbb{R}^p$ into $\mathbb{R}^n$ $(p \leqslant n)$, the problem of finding

$$\min_{h \in \mathbb{R}^p} \| f - F(a) - F'(a)h \|$$

is equivalent to the determination of $h = [F'(a)]^+ (f - F(a))$, where $[F'(a)]^+$ is the pseudo-inverse of $F'(a)$ (see Luenberger [19]); hence in the case $\Omega(a) = \mathbb{R}^p$ the full-step method (3.1)–(3.4) just becomes Newton's method for $p = n$ (e.g., in [19]) and the Newton–Raphson method for $p < n$ (Ben–Israel [6].) For more information and results on the respective convergence behavior, we refer, in particular, to Osborne [27]; but compare also Osborne [26] and Osborne and Watson [30]. Generalizations of Newton's method are the Levenberg-like algorithms in Osborne [28, 29] which include the methods of Levenberg, Marquardt, and Morrison as special cases (for references, see [28].) Therein again the magnitude of the directions of descent is controlled in a certain way.

The convergence behavior of algorithm (3.1)–(3.4) with $\Omega(a) = \mathbb{R}^p$ for the L_p-norms, $1 < p < \infty$, is investigated in Osborne and Watson [30].

Finally, we give some references to algorithms solving the linear subproblem in (3.1) in the discrete case. For the linear unconstrained L_∞-problem, Barrodale and Young [4] supply a modified simplex algorithm; linear constraints can be added there easily. By the authors of [4] in addition an algorithm for linear L_1-approximation was developed which was improved and extended to problems with linear constraints by Barrodale and Roberts [5]. Wolfe [38] analyzes the convergence of an algorithm for the unconstrained L_p-approximation, $1 \leqslant p < 2$, that had been studied by other authors before. Finally, Fletcher *et al.* [11] derive a method for the solution of the L_p-problem without constraints in the case $2 \leqslant p < \infty$ which reduces to the solution of the normal equations for $p = 2$.

Since this paper was written, several related papers and books have been published. We want to mention here merely the books of G. A. Watson "Approximation theory and numerical methods" (Chichester–New York–Brisbane–Toronto, 1980) and of R. Hettich and P. Zencke "Numerische Methoden der Approximation und semi-infiniten Optimierung" (Stuttgart, 1982); both have long chapters on algorithms and offer an extensive list of references. They are in particular recommended for L_∞-approximation and its numerical aspects.

APPENDIX 1

For the proof of the closedness of the mapping $Z: E \to P(E)$ in the convergence Theorem 2.1, the following results out of Krabs [18] are helpful.

Let X and Y be metric spaces, U a normed vector space, and $\Phi: X \times Y \to \mathbb{R}$, $g: X \times Y \to U$ given mappings. Further, let $W: X \to P(Y)$ be a point-to-set map from X into $P(Y)$. Finally, Q shall be a nonempty subset of U. Then for each $x \in X$, we define

$$\Omega(x) = \{y \in W(x) \mid g(x, y) \in Q\} \subseteq Y$$

and assume $\Omega(x) \neq \varnothing$ for all $x \in X$. With these preliminaries we consider now the problem to minimize the function $\Phi(x, \cdot)$ on $\Omega(x)$ and define the optimal value as

$$m(x) = \inf\{\Phi(x, y) \mid y \in \Omega(x)\}$$

and the set of optimal solutions as

$$O(x) = \{y \in \Omega(x) \mid \Phi(x, y) = m(x)\}.$$

Beside the notion of the continuity of a point-to-set map, which was defined by Definition 2.1, we shall need for the formulation of the next theorems the following

DEFINITION 5.1. (a) $\Phi: X \times Y \to \mathbb{R}$ is said to be continuous with respect to $\{\hat{x}\} \times \Omega(\hat{x})$ if for all sequences (x_k) in X with $x_k \to \hat{x}$ and all sequences (y_k) in Y with $y_k \in \Omega(x_k)$ for almost all k and $y_k \to \hat{y}$ for a $\hat{y} \in \Omega(\hat{x})$, $\lim_{k \to \infty} \Phi(x_k, y_k) = \Phi(\hat{x}, \hat{y})$ holds true.

(b) $g: X \times Y \to U$ is said to be continuous with respect to $\{\hat{x}\} \times W(\hat{x})$ if for all sequences (x_k) in X with $x_k \to \hat{x}$ and all sequences (y_k) in Y with $y_k \to \hat{y} \in W(\hat{x})$, $\lim_{k \to \infty} g(x_k, y_k) = g(\hat{x}, \hat{y})$ holds true.

Then Satz 2.4 in [18] says:

THEOREM 5.1. *If*

(i) Ω *is continuous at* $\hat{x}$,

(ii) Φ *is continuous with respect to* $\{\hat{x}\} \times \Omega(\hat{x})$,

(iii) *for each sequence* (x_k) *in* X *with* $x_k \to \hat{x}$ *the corresponding sets of optimal solutions* $O(x_k)$ *are nonempty, and*

(iv) *each sequence* (y_k) *in* Y *with* $y_k \in O(x_k)$ *for almost all* k *possesses an accumulation point,*

then the function $x \to m(x)$ *is continuous at* $\hat{x}$.

For the continuity of Ω at $\hat{x}$, the assumptions of the next theorem are sufficient (Satz 4.1 and Satz 4.2 in [18]):

THEOREM 5.2. *If*

(i) $W: X \to P(Y)$ *is continuous at* $\hat{x} \in X$,

(ii) Q *has a nonempty interior* $\mathring{Q}$ *and*

$$\Omega(\hat{x}) = \overline{\{y \in W(\hat{x}) \mid g(\hat{x}, y) \in \mathring{Q}\}},$$

where $\bar{B}$ *is the closure of* B,

(iii) Q *is a closed set,*

(iv) $g: X \times Y \to U$ *is continuous with respect to* $\{\hat{x}\} \times W(\hat{x})$,

then the mapping $x \to \Omega(x)$ *is continuous at* $\hat{x}$.

APPENDIX 2

Proof of Lemma 3.1. First we observe that $W(a)$ (3.6) is a convex set that encloses the origin. Therefore, if $h \in W(a)$ then also λh is contained in $W(a)$ for all $\lambda \in [0, 1]$. Hence for all sufficiently small $\lambda > 0$ with $a + \lambda h \in A$, $a + \lambda h$ belongs to E. The remainder of the proof is then a consequence of the estimation

$$\begin{aligned} \| f - F(a + \lambda h) \| &= \|(1 - \lambda)(f - F(a)) + \lambda(f - F(a) - F'(a)h) \\ &\quad + (F(a) - F(a + \lambda h) + F'(a)\lambda h)\| \\ &\leqslant (1 - \lambda) \| f - F(a) \| + \lambda \| f - F(a) - F'(a)h \| + o(\lambda \| h \|) \\ &= \| f - F(a) \| - \lambda C + o(\lambda \| h \|) < \| f - F(a) \| \end{aligned}$$

for all sufficiently small $\lambda \in (0, 1]$, where $C = |f - F(a)| - |f - F(a) - F'(a)h|$.

Proof of Lemma 3.2. $W(a)$ is nonempty since $0 \in W(a)$ and in any case a closed set. Consequently, $\Omega(a)$ is here a nonempty compact set. Since further the map $h \to |f - F(a) - F'(a)h|$ is continuous, the statement of Lemma 3.2 can be concluded from the theorem of Weierstrass.

Proof of Lemma 3.3. Let

$$H = \{F'(a)h \mid h \in \Omega(a)\}.$$

If $\Omega(a) = \mathbb{R}^p$, H is obviously closed. If otherwise I and J have finitely many elements, then $\Omega(a)$ is the set of solutions of a finite system of linear inequalities, i.e., a so-called "polyhedral" convex set in $\mathbb{R}^p$. By Theorem 19.1 in Rockafellar [32], $\Omega(a)$ can, therefore, be "finitely generated" which means that there exist vectors $e_1,..., e_k, e_{k+1},..., e_m$ in $\mathbb{R}^p$ and a fixed integer k, $0 \leqslant k \leqslant m$, such that

$$\Omega(a) = \left\{ \sum_{i=1}^{m} \gamma_i e_i \mid \gamma_1 + \cdots + \gamma_k = 1, \gamma_i \geqslant 0 \text{ for } i = 1,..., m \right\}.$$

Consequently,

$$H = \left\{ \sum_{i=1}^{m} \gamma_i (F'(a)\, e_i) \mid \gamma_1 + \cdots + \gamma_k = 1, \gamma_i \geqslant 0 \text{ for } i = 1,..., m \right\}.$$

The $F'(a)e_i$, $i = 1,..., m$, generate a linear subspace of $C(T)$ which is isomorphic to a space $\mathbb{R}^s$. The corresponding isomorphic image of H in $\mathbb{R}^s$ is then again a finitely generated and hence a closed set (see [32]). Consequently, H has to be closed, too. One can now further easily verify that with $\rho = F'(a)h$

$$\inf_{\rho \in H} |f - F(a) - F'(a)h| = \inf_{\rho \in V_1} |f - F(a) - F'(a)h|$$

where

$$V_1 = \{\rho \in H \mid |F'(a)h| \leqslant |f - F(a) - F'(a)h^*| + |f - F(a)|\}$$

for $h^* \in \Omega(a)$ arbitrary, but fixed. Since V_1 is compact and the mapping $F'(a)h \to |f - F(a) - F'(a)h|$ is continuous, we can apply Weierstrass' theorem. If now I and J are arbitrary and if $[F'(a)]^{-1}$ exists, we can infer as follows: Since $0 \in W(a)$,

$$\inf_{h \in W(a)} |f - F(a) - F'(a)h| = \inf_{h \in V_2} |f - F(a) - F'(a)h|$$

with

$$V_2 = \{h \in W(a) \mid |f - F(a) - F'(a)h| \leqslant |f - F(a)|\}.$$

V_2 is closed and by virtue of Lemma 3.4 also bounded.

Proof of Lemma 3.4.

$$\begin{aligned} |h| &= |[F'(a)]^{-1}F'(a)h| \leqslant |[F'(a)]^{-1}|\,|F'(a)h| \\ &\leqslant |[F'(a)]^{-1}|\,|(f - F(a) - F'(a)h) - (f - F(a))| \\ &\leqslant 2\,|f - F(a)|\,|[F'(a)]^{-1}|. \end{aligned}$$

Proof of Lemma 3.5. If $L(a) = [F'(a)]^T[F'(a)]$, then $\operatorname{rank}(L(a)) = p$ for all $a \in A$. Henceforth, for $a_j \in R$ fixed $L^{-1}(a_j)$ exists. By assumption $L^{-1}(a)$ exists also for all $a \in B_j$ of an open ball $B_j \subseteq A$ centered at a_j so that further

$$K_j = \sup_{a \in B_j} |L^{-1}(a)|$$

is a finite number. By the theorem of Heine–Borel, there are now finitely many B_j's which cover R. Hence a constant K exists such that $|L^{-1}(a)| \leqslant K$ for all $a \in R$. Since the map $a \to [F'(a)]^T$ is continuous on R, we can finally conclude

$$\begin{aligned} |[F'(a)]^+| &= |[(F'(a))^T(F'(a))]^{-1}[F'(a)]^T| \\ &\leqslant |L^{-1}(a)|\,|[F'(a)]^T| \leqslant N \qquad \forall a \in R. \end{aligned}$$

where $[F'(a)]^+$ denotes the pseudoinverse of $F'(a)$.

Proof of Theorems 3.1 *and* 3.2. In the following, 2.1(iii) refers to assumption (iii) of Theorem 2.1, etc.

2.1(iii). Let $a_k \in R \backslash A$ be fixed.

(a) If g is defined by (3.10) as in Theorem 3.2, assumption 2.1(iii) is a consequence of Lemmas 3.1 and 3.3.

(b) Let now g (3.8), (3.9) be as in Theorem 3.1. Since $a_k \notin A$, there is an $h \in W(a_k)$ such that

$$|f - F(a_k) - F'(a_k)h| < |f - F(a_k)|; \tag{6.1}$$

(6.1) yields further

$$\begin{aligned} &|f - F(a_k) - F'(a_k)\,\lambda h| \\ &\quad\leqslant \lambda\,|f - F(a_k) - F'(a_k)h| + (1-\lambda)\,|f - F(a_k)| \\ &\quad< \lambda\,|f - F(a_k)| + (1-\lambda)\,|f - F(a_k)| = |f - F(a_k)| \end{aligned} \tag{6.2}$$

for all $\lambda \in (0, 1]$. Since λh, $h \in W(a_k)$, is element of $\Omega(a_k)$ for all sufficiently small $\lambda > 0$, we can conclude from (6.2) that

$$\inf_{h \in \Omega(a_k)} |f - F(a_k) - F'(a_k)h| < |f - F(a_k)|; \tag{6.3}$$

(6.3) together with Lemmas 3.1 and 3.2 then guarantees

$$\min_{0 \leqslant \lambda \leqslant 1} |f - F(b_k + \lambda d_k)| < |f - F(a_k)| \qquad \forall\, (b_k, d_k) \in D(a_k).$$

2.1(iv). For the proof of the closedness of the point-to-set map Z at each $a \in R \backslash \Lambda$, we verify the assumptions of Theorem 2.2.

2.2(i). E and $\Omega(a)$, $a \in E$, are nonempty, so that $D(a) \neq \varnothing$ for all $a \in E$ is a consequence of Lemmas 3.2 and 3.3, respectively. $S(b, d) \neq \varnothing$ for all $(b, d) \in E \times \mathbb{R}^p$ is then obvious under our assumptions.

2.2(iii). To prove the closedness of S at each $(b, d) \in D(a)$, $a \in R \backslash \Lambda$, one can follow the proof of Lemma 2.2 in [15]. Note that here $d \neq 0$ and S is a mapping from $E \times \mathbb{R}^p$ into $P(E)$.

2.2(iv). (a) l (3.8), (3.9) is continuous on R and, therefore, achieves its maximum there. Hence for all $(b, d) \in D(a)$ and all $a \in R$, we have

$$n(d) \leqslant \max_{a \in R} l(a),$$

i.e., all possible directions $d \in \Omega(a)$, $a \in R$, lie in a compact set.

(b) In Theorem 3.2 this condition is taken into the formulation of the theorem as an assumption.

2.2(ii). D is closed at each $a \in R \backslash \Lambda$, if for each sequence (a_k) in R, $a_k \to a$, and each sequence $(b_k, d_k) \in D(a_k)$ with $d_k \to d$, it follows that $(b, d) \in D(a)$. Consequently, D is closed at $a \in R \backslash \Lambda$ if $a_k \to a$ implies that

$$\begin{aligned} \lim_{k \to \infty} m(a_k) &= \lim_{k \to \infty} |f - F(a_k) - F'(a_k)d_k| \\ &= \lim_{k \to \infty} \inf_{h \in \Omega(a_k)} |f - F(a_k) - F'(a_k)h| \\ &= \inf_{h \in \Omega(a)} |f - F(a) - F'(a)h| \\ &= |f - F(a) - F'(a)d| = m(a). \end{aligned}$$

For the proof of this implication we verify the assumptions of Theorem 5.1. Therefore, we set $X = R$, $Y = \mathbb{R}^p$ and $U = \mathbb{R}$ and define

$$\Phi(a, h) = | f - F(a) - F'(a)h |$$

and

$$O(a) = \{d \in \Omega(a) \mid | f - F(a) - F'(a)d | = m(a)\}.$$

$\Omega(a) \neq \varnothing$ for all $a \in R$ is obviously satisfied for Ω as in Theorems 3.1 and 3.2.

5.1(ii).

$$\begin{aligned} &\big| | f - F(a) - F'(a)h | - | f - F(a_k) - F'(a_k)h_k | \big| \\ &\quad \leqslant | (F(a) - F(a_k)) - (F'(a)\, h - F'(a_k)h_k)| \\ &\quad \leqslant | F(a) - F(a_k)| + | F'(a)(h_k - h)| + | (F'(a_k) - F'(a))h_k | \\ &\quad \leqslant | F(a) - F(a_k)| + | F'(a)| \, |h_k - h| + |F'(a_k) - F'(a)| \, (| h | + \varepsilon). \end{aligned}$$

5.1(iii) was proved with Lemmas 3.2 and 3.3.

5.1(iv). See 2.2(iv) above.

5.1(i). The continuity of the mapping $\Omega: E \to P(\mathbb{R}^p)$ at each $a \in R \backslash \Lambda$ will be proved by checking the assumptions of Theorem 5.2.

5.2(i). Let $W: E \to P(\mathbb{R}^p)$ be defined by (3.5), (3.6).

(a) The closedness of $W(a)$ at each $a \in R \backslash \Lambda$ results directly from the continuity of the functionals u_i and v_j for all $i \in I$ and $j \in J$, respectively.

(b) W is open at each $a \in R \backslash \Lambda$ since for each sequence (a_k) in R with $a_k \to a$ and each $h \in W(a)$, the sequence (h_k) where $h_k = h + (a - a_k)$ is such that $h_k \in W(a_k)$ and $h_k \to h$.

5.2(ii). For $a \in R \backslash \Lambda$, $W(a)$ has to contain an element $h \neq 0$ beside the zero element and $l(a)$ has to be a positive number. Therefore, both sets

$$B_1 = \{h \in W(a) \mid n(h) < l(a)\}, \qquad B_2 = \{h \in W(a) \mid n(h) > 0\}$$

are nonempty. That according to its definition, $\Omega(a)$ equals the closure of B_1 and B_2, respectively, can now be comprehended easily.

5.2(iii) is obvious for Q (3.3).

5.2(iv). Let (a_k) be a sequence in R with $a_k \to a$, $a \in R \backslash \Lambda$, and (h_k) be a sequence in $\mathbb{R}^p$ such that $h_k \to h \in W(a)$.

(a) Let g be defined through (3.8), (3.9) as in Theorem 3.1. Then because of the continuity of the maps $l: E \to \mathbb{R}$ and $n: \mathbb{R}^p \to \mathbb{R}$, it results that

$$\lim_{k \to \infty} g(a_k, h_k) = \lim_{k \to \infty} (l(a_k) - n(h_k)) = l(a) - n(h).$$

(b) For g (3.10) (Theorem 3.2) we have analogously

$$\lim_{k \to \infty} g(a_k, h_k) = \lim_{k \to \infty} n(h_k) = n(h).$$

References

1. D. H. Anderson, "Linear Programming and the Calculation of Maximum Norm Approximations," Ph. D. Thesis, Canberra, 1975.
2. D. H. Anderson and M. R. Osborne, Discrete, nonlinear approximation problems in polyhedral norms, *Numer. Math.* **28** (1977), 143–156.
3. D. H. Anderson and M. R. Osborne, Discrete, nonlinear approximation problems—A Levenberg-like algorithm, *Numer. Math.* **28** (1977), 157–170.
4. I. Barrodale and A. Young, Algorithms for best L_1 and L_∞ linear approximations on a discrete set, *Numer. Math.* **8** (1966), 295–306.
5. I. Barrodale and F. D. K. Roberts, An efficient algorithm for discrete l_1 linear approximation with linear constraints, *SIAM. J. Numer. Anal.* **15** (1978), 603–611.
6. A. Ben-Israel, A Newton–Raphson method for the solution of systems of equations, *J. Math. Anal. Appl.* **15** (1966), 243–252.
7. L. Cromme, Eine Klasse von Verfahren zur Ermittlung bester nichtlinearer Tschebyscheff-Approximationen, *Numer. Math.* **25** (1976), 447–459.
8. L. Cromme, Zur Tschebyscheff-Approximation bei Ungleichungsnebenbedingungen im Funktionenraum, *in* "Approximation Theory" (R. Schaback and K. Scherer, Eds.), pp. 144–153, Lecture Notes in Mathematics No. 556, Springer-Verlag, Berlin/New York, 1976.
9. L. Cromme, Numerische Methoden zur Behandlung einiger Problemklassen der nichtlinearen Tschebyscheff-Approximation mit Nebenbedingungen, *Numer. Math.* **28** (1977), 101–117.
10. L. Cromme, Strong uniqueness—A far-reaching criterion for the convergence analysis of iterative procedures, *Numer. Math.* **29** (1978), 179–193.
11. R. Fletcher, J. A. Grant, and M. D. Hebden, The calculation of linear best L_p approximations, *Computer J.* **14** (1971), 276–279.
12. J. Hald and H. Schjaer-Jacobsen, Linearly constrained minimax optimization without calculating derivatives, *in* "Operations Research Verfahren 31" (F. Steffens, Ed.), pp. 289–301. Verlag Anton Hain, Meisenheim, 1979.
13. J. Hald and K. Madsen, A 2-stage algorithm for minimax optimization, Rep. No. NI-78-11, Tekniske Højskole Lyngby, Denmark, 1978.
14. R. Hettich, Numerical methods for nonlinear Chebyshev approximation, *in* "Approximation in Theorie und Praxis" (G. Meinardus, Ed.), pp. 139–156. Bibliographisches Institut, Mannheim/Wien/Zürich, 1979.
15. R. Horst, "Nichtlineare Optimierung," Carl Hanser Verlag, Munich/Vienna, 1979.
16. P. Huard, Extensions of Zangwill's theorem, *Math. Progr. Study* **10** (1979), 98–103.
17. Y. Ishizaki and H. Watanabe, An iterative Chebyshev approximation method for network design, *IEEE Trans. Circuit Theory* **CT-15** (1968), 326–336.

18. W. KRABS, Stetige Abänderung der Daten bei nichtlinearer Optimierung und ihre Konsequenzen, *in* "Operations Research Verfahren 25" (A. Angermann, Ed.), pp. 93–113.Verlag Anton Hain, Meisenheim, 1976.
19. D. LUENBERGER, "Optimization by Vector Space Methods," Wiley, New York, 1969.
20. K. MADSEN, An algorithm for minimax solution of overdetermined systems of non-linear equations, *J. Inst. Math. Appl.* **16** (1975), 321–328.
21. K. MADSEN, Minimax solution of non-linear equations without calculating derivatives, *Math. Progr. Study* **3** (1975), 110–126.
22. K. MADSEN AND H. SCHJAER–JACOBSEN, Linearly constrained minimax optimization, *Math. Progr.* **14** (1978), 208–223.
23. Mathematical Programming Study 10 (1979).
24. M. R. OSBORNE AND G. A. WATSON, An algorithm for minimax approximation in the nonlinear case, *Computer J.* **12** (1969), 64–69.
25. M. R. OSBORNE AND G. A. WATSON, On an algorithm for discrete nonlinear L_1 approximation *Computer J.* **14** (1971), 184–188.
26. M. R. OSBORNE, An algorithm for discrete, nonlinear, best approximation problems, *in* "Numerische Methoden der Approximationstheorie" (L. Collatz and G. Meinardus, Eds.), pp. 117–126. ISNM 16, Birkhäuser-Verlag, Basel, 1972.
27. M. R. OSBORNE, A class of methods for minimising a sum of squares, *Austral. Computer J.* **4** (1972), 164–169.
28. M. R. OSBORNE, Some aspects of non-linear least squares calculations, *in* "Numerical Methods for Nonlinear Optimization" (F. A. Lootsma, Ed.), pp. 171–189. Academic Press, London/New York, 1972.
29. M. R. OSBORNE, Nonlinear least squares—The Levenberg algorithm revisited. *J. Austral. Math. Soc.* **19** (1976), 343–357.
30. M. R. OSBORNE AND G. A. WATSON, Nonlinear approximation problems in vector norms, *in* "Numerical Analysis" (G. A. Watson, Ed.), pp. 117–132, Lecture Notes in Mathematics No.630, Springer-Verlag, Berlin/New York, 1978.
31. R. REEMTSEN, "Charakterisierungen und Optimalitätskriterien für lokale Minima bei der Tschebyscheff-Approximation," Ph. D. Thesis, Darmstadt, 1978.
32. R. T. ROCKAFELLAR, "Convex Analysis," Princeton Univ. Press, Princeton, N. J., 1972.
33. R. SCHABACK, Globale Konvergenz von Verfahren zur nichtlinearen Approximation, *in* "Approximation Theory" (R. Schaback and K. Scherer, Eds.), pp. 352–363, Lecture Notes in Mathematics No. 556, Springer-Verlag, Berlin/New York, 1976.
34. R. SCHULTZ, "Ein Abstiegsverfahren für Approximationsaufgaben in normierten Räumen," Ph. D. Thesis, Hamburg, 1977.
35. S. TISHYADHIGAMA, E. POLAK, AND R. KLESSIG, A comparative study of several general convergence conditions for algorithms modeled by point-to-set maps, *Math. Progr. Study* **10** (1979), 172–190.
36. G. A. WATSON, Dual methods for nonlinear best approximation problems, *J. Approx. Theory* **26** (1979), 142–150.
37. G. A. WATSON, The minimax solution of an overdetermined system of nonlinear equations, *J. Inst. Math. Appl.* **23** (1979), 167–180.
38. J. M. WOLFE, On the convergence of an algorithm for discrete L_p approximation, *Numer. Math.* **32** (1979), 439–459.
39. W. I. ZANGWILL, Convergence conditions for nonlinear programming algorithms, *Management Sci.* **16** (1969), 1–13.
40. W. I. ZANGWILL, "Nonlinear Programming: A unified Approach," Prentice-Hall, Englewood-Cliffs, N. J., 1969.