CUCS-42-83

Optimal Solution of Nonlinear Equations Satisfying a Lipschitz Condition

K. Sikorski
Department of Computer Science
Columbia University
New York, New York
(On leave from the University of Warsaw)

AMS (MOS) Subject classifications (1970) Primary 65H10.

Key words and phrases: algorithm, complexity, information operator, Lipschitz condition, nonlinear equations.

Abstract.

For a given nonnegative ε we seek a point x^* that $|f(x^*)| \leq \varepsilon$ where f is a nonlinear transformation of the cube $B = [0,1]^m$ into $\mathbb{R}$ (or $\mathbb{R}^p$, $p > 1$) satisfying a Lipschitz condition with the constant K and having a zero in B.

The information operator on f consists of n values of arbitrary linear functionals which are computed adaptively. The point x^* is constructed by means of an algorithm which is a mapping depending on the information operator. We find an optimal algorithm, i.e., algorithm with the smallest error, which uses n function evaluations computed adaptively. We also exhibit nearly optimal information operators, i.e., the linear functionals for which the error of an optimal algorithm that uses them is almost minimal. Nearly optimal information operators consist of n nonadaptive function evaluations at equispaced points x_j in the cube B. This result exhibits the superiority of the T. Aird and J. Rice procedure ZSRCH (IMSL library [1]) over Sobol's approach [7] for solving nonlinear equations in our class of functions. We also prove that the simple search algorithm which yields a point $x^* = x_k$ such that $|f(x_k)| = \min_{1 \leq j \leq n} |f(x_j)|$ is nearly optimal. The complexity, i.e., the minimal cost of solving our problem is roughly equal to $(K/\varepsilon)^m$.

0. Introduction.

Let $\mathbb{R}$ denote the real line, B be the m-dimensional unit cube and let ε be a nonnegative number. Two basic error criteria are used for determining an approximate solution x^* of a nonlinear equation

$$(0.1) \qquad f(x) = 0,$$

where $f:B \to \mathbb{R}$. (If $f:B \to \mathbb{R}^p$ and $p > 1$, we show in Section 1 how to transform this to the case $p = 1$.) Assuming that $f(\alpha) = 0$ these two error criteria are defined as follows:

$$(0.2) \qquad \text{root criterion:} \qquad \|x^*-\alpha\| \le \varepsilon,$$

$$(0.3) \qquad \text{residual criterion:} \qquad |f(x^*)| \le \varepsilon.$$

We assume that f belongs to the class F of transformations satisfying a Lipschitz condition in the infinity norm with a constant K and having a zero in B.

The information operator N_n on f consists of n function values, or more generaly of n values of arbitrary linear functionals which are computed adaptively. The approximation x^* to α is constructed by means of an algorithm φ which is a mapping depending on the information operator.

It was shown in [6] that there exist functions in F such that for every $\varepsilon < 1/2$ it is impossible to find x^*

---

satisfying the root criterion (0.2) no matter how large is n and no matter what algorithm is used. Therefore in this paper we deal only with the residual criterion (0.3).

We define the radius of an information operator N_n which is the sharp lower bound on the error of every algorithm using N_n. For a given information operator N_n consisting of adaptive evaluations of function values we determine an algorithm φ which has the smallest error, i.e., which is optimal for the worst case model. We exhibit information operators $N_{n,i}$, $i = 1,2,\ldots,n+1$ which have almost minimal radius, i.e. are nearly optimal. We prove that the operators $N_{n,i}$ consist of n <u>nonadaptive</u> function evaluations at equispaced points x_j in the cube B. This result exhibits the superiority of T. Aird and J. Rice procedure ZSRCH (IMSL library [1]) over Sobol's approach [7] for solving (0.3).

We also consider the complexity (minimal cost) of solving (0.3). It is roughly equal to $(K/\varepsilon)^m$. Even for K near unity and moderate ε the complexity is large for the high dimensional case.

We develop two simple search algorithms φ^* and φ^{**} which use the nearly optimal information operators $N_{n,i}$. The algorithm φ^* requires the knowledge of the constant K. The algorithm φ^{**} which yields a point $x^* = x_k$, such that

$|f(x_k)| = \min_{1 \leq j \leq n} |f(x_j)|$, does not require the knowledge of K, but its cost can exceed the complexity by a factor of 2^m.

The cost of the algorithm φ, which also requires the knowledge of K, and is "strongly optimal" is not known.

Sukharev [9] considered the scalar case $m = 1$. This paper generalizes Sukharev's results to arbitrary m.

We use however, different notation and proof technique. Our notation is adopted from Traub and Woźniakowski [10] (see also [11]).

We briefly summarize the contents of the paper. In Section 1 we define information operator, algorithm and specify what we mean by optimal information operator and optimal algorithm. In Section 2 we find the radius of information operator consisting of adaptive evaluations of function values. In Section 3 we exhibit optimal information operators $N_{n,i}$ and the algorithms φ, φ^* and φ^{**}. In Section 4 we deal with the class of general information operators and in Section 5 we find the complexity of the problem (0.3). Finally in Section 6 we pose some open problems concerning optimal information operators and algorithms for different <u>classes</u> of functions.

1. Basic definitions and theorems - formulation of the problem.

Let $B = [0,1]^m$ be the unit m-dimensional cube of $\mathbb{E}^m$ and let $\|x\| = \max_{1\le i\le m} |x_i|$ be the infinity norm in $\mathbb{R}^m$. Define G as the class of functions $f:B \to \mathbb{R}$ satisfying a Lipschitz condition with constant K, i.e.,

$$(1.1) \qquad G = \{f:B \to \mathbb{R}: |f(x)-f(y)| \le K\|x-y\|, \quad x,y \in B\}.$$

Let F be a subclass of G defined by

$$(1.2) \qquad F = \{f \in G: \exists\, \alpha \in B: f(\alpha) = 0\}$$

For a given ε, $\varepsilon \ge 0$, define the set

$$(1.3) \qquad S(f,\varepsilon) = \{x \in B: |f(x)| \le \varepsilon\}, \ \forall f \in F.$$

This set is not empty since f has a zero in B. The problem is to find a point x^* satisfying the residual criterion (0.3), i.e.,

$$(1.4) \qquad x^* \in S(f,\varepsilon).$$

<u>Remark</u> 1.1: One may wish to solve (1.4) in the class of functiong $g:B \to \mathbb{R}^p$. $p > 1$, satisfying a Lipschitz condition and having a zero in B. This problem is, however, equivalent to the case $p = 1$.

Indeed, for a given $g:B \to \mathbb{R}^p$ define the function f, $f:B \to \mathbb{R}$. by $f(x) = \max_{1\le j\le p} |g_j(x)| = \|g(x)\|_\infty$. This f satisfies a Lipschitz condition with the same constant as g. Note that f has a zero in B. Moreover $S(f,\varepsilon) = \{x \in B: \|g(x)\| \le \varepsilon\}$.

Therefore we may assume without loss of generality that $p = 1$. □

To find x^* satisfying (1.4) we use an <u>information operator</u> N_n and an <u>algorithm</u> φ using N_n. These are defined as in Traub and Wozniakowski [10].

Let $f \in G$ and

$$(1.5) \qquad N_n(f) = [L_1(f), L_2(f;y_1), \ldots, L_n(f;y_1,\ldots,y_{n-1})]$$

where $y_i = L_i(f;y_1,\ldots,y_{i-1})$ and

$$(1.6) \qquad L_{i,f}(\cdot) \overset{df}{=} L_i(\cdot;y_1,\ldots,y_{i-1}) : G \to \mathbb{R}$$

is a linear functional, $i = 1,2,\ldots,n$.

If $L_{i,f}(\cdot) = L_i(\cdot)$, $\forall$ i, i.e., $L_{i,f}$ does not depend on the previously computed values $y_1,\ldots,y_{i-1}$ the information operator is called <u>nonadaptive</u>; otherwise it is called <u>adaptive</u>.

The total number of functional evaluations n is called the <u>cardinality</u> of N_n.

Knowing $N_n(f)$ we approximate x^* by an <u>algorithm</u> φ which is a mapping

$$(1.7) \qquad \varphi : N_n(F) \to B .$$

<u>The error of the algorithm</u> φ is defined by

$$(1.8) \qquad e(\varphi) = \sup_{f \in F} |f(\varphi(N_n(f)))| .$$

Thus $x^* = \varphi(N_n(f))$ satisfies (1.4) for every f in F iff

$e(\varphi) \le \varepsilon$.

Note that if two functions f and $\tilde{f}$ from F have the same information, $N_n(f) = N_n(\tilde{f})$, then the value of the algorithm φ is the same for f and $\tilde{f}$, $\varphi(N_n(f)) = \varphi(N_n(\tilde{f}))$. Thus (1.8) can be restated as

$$(1.9) \qquad e(\varphi) = \sup_{f \in F} e(\varphi, f)$$

where the <u>local error</u> $e(\varphi, v)$ is defined by

$$(1.10) \qquad e(\varphi, f) = \sup\{|\tilde{f}(\varphi(N_n(f))| : \tilde{f} \in F, N_n(\tilde{f}) = N_n(f)\}.$$

Define <u>the radius of the information operator</u> N_n (briefly <u>radius of information</u>) by

$$(1.11) \qquad r(N_n) = \sup_{f \in F} r(N_n, f)$$

where the <u>local radius</u> $r(N_n, f)$ is given by

$$(1.12) \qquad r(N_n, f) = \inf_{x \in B} \sup\{|\tilde{f}(x)| : \tilde{f} \in F, N_n(\tilde{f}) = N_n(f)\}.$$

Let $\Phi = \Phi(N_n)$ be the class of all algorithms using the information operator N_n. It is obvious that

$$(1.13) \qquad \sup_{\varphi \in \Phi(N_n)} e(\varphi, f) = r(N_n, f), \quad \forall \; f \in F$$

and

$$(1.14) \qquad \inf_{\varphi \in \Phi(N_n)} e(\varphi) = r(N_n).$$

We are interested in algorithms for which $e(\varphi, f)$ and $e(\varphi)$ are minimal. An algorithm φ^{so}, $\varphi^{so} \in \Phi$, is _strongly optimal_ iff

$$(1.15) \qquad e(\varphi^{so}, f) = r(N_n, f), \ \forall \ f \in F.$$

An algorithm φ^{o}, $\varphi^{o} \in \Phi$, is _optimal_ iff

$$(1.16) \qquad e(\varphi^{o}) = r(N_n).$$

It is obvious that every strongly optimal algorithm is optimal, but the converse is in general not true. It may happen that due to some special properties of f, $r(N_n(f)) << r(N_n)$. A strongly optimal algorithm φ^{so} takes advantage of this favorable f since $e(\varphi^{so}, f) = r(N_n(f))$. For some optimal algorithm φ^{o} it may happen that $e(\varphi^{o}, f) = e(\varphi^{o}) >> e(\varphi^{so}, f)$.

We are also interested in algorithms for which the errors $e(\varphi)$ are close to minimal. An algorithm φ^{ao}, $\varphi^{ao} \in \Phi$, is _almost optimal_ iff

$$(1.17) \qquad e(\varphi^{ao}) = c_n r(N_n)(1 + o(1)) \text{ as } n \to \infty$$

where the constants c_n are in the range $1 \le c_n \le 2$.

The radius of information measures the strength of an information operator. We can solve the problem (1.4) iff $r(N_n) \le \epsilon$.

For a given n we want to find the functionals in (1.5) such that the radius of information is minimized. More precisely let $\mathcal{N}_n$ be the class of all adaptive or nonadaptive, information operators with cardinality at most n. Then the information operator N_n^o, $N_n^o \in \mathcal{N}_n$, is <u>optimal</u> iff

$$(1.18) \qquad r(N_n^o) = \inf_{N\in\mathcal{N}_n} r(N).$$

The information operator N_n^{ao}, $N_n^{ao} \in \mathcal{N}_n$, is <u>almost optimal</u> iff

$$(1.19) \qquad r(N_n^{ao}) = b_n \inf_{N\in\mathcal{N}_n} r(N)\ (1 + o(1)) \text{ as } n \to \infty$$

where the constant b_n are in the range $1 \le b_n \le 2$.

We are now in a position to state the main problems of this paper.

(1.20) What is the optimal information N_n^o?

(1.21) What is the minimal cardinality of the optimal information N_n^o, such that $r(N_n^o) \le \varepsilon$?

(1.22) What is a strongly optimal, optimal or almost optimal algorithm using the optimal information N_n^o?

In Sections 2 and 3 we deal with the information operator consisting of adaptive evaluations of function values, i.e.,

$$(1.23) \qquad N_n(f) = [f(x_1),\dots,f(x_n)]$$

where x_1 is some point chosen a priori, $x_1 \in B$, and

$$x_i = \tilde{x}_i(f(x_1),\dots,f(x_{i-1})), \quad i = 2,3,\dots,n$$

where $\tilde{x}_i$ is a transformation $\tilde{x}_i : \mathbb{R}^{i-1} \to B$.

In Section 4 we consider the general information operator given by (1.5) and in Section 5 we deal with the problem of complexity (minimal cost) of solving (1.4).

2. Local Radius of Information

In this section we show how to calculate the local radius $r(N_n,f)$ see (1.12), for the information operator consisting of adaptive evaluations of function values (1.23).

Let $y = N_n(f)$, i.e., $y_j = f(x_j)$, $j = 1,2,\dots,n$. Define the set

$$(2.1) \qquad Z = Z(N_n(f)) = \{z \in B : \exists \tilde{f} \in F : N_n(\tilde{f}) = N_n(f) \text{ and } \tilde{f}(z) = 0\}.$$

Thus Z is the set of zeros of all functions $\tilde{f}$ in F which share the same information operator value with f. From the definition of the class we have

$$(2.2) \qquad y_j - K\|x-x_j\| \le \tilde{f}(x) \le y_j + K\|x-x_j\|$$

for all j, $x \in B$ and $\tilde{f} \in F$ such that $N_n(\tilde{f}) = N_n(f)$.

Define the functions

$$
(2.3) \qquad
\begin{aligned}
g_n^-(x) &= \max_{1 \le j \le n} (y_j - K\|x-x_j\|), \\
g_n^+(x) &= \min_{1 \le j \le n} (y_j - K\|x-x_j\|).
\end{aligned}
$$

Thus, in view of (2.2), (2.3) implies that

$$(2.4) \qquad g_n^-(x) \le f(x) \le g_n^+(x), \quad \forall \ x \in B$$

for all $\tilde{f} \in F$ such that $N_n(\tilde{f}) = N_n(f)$.

Let $B(x,r) = \{y \in \mathbb{R}^m : \|y-x\| \le r\}$. Then it is obvious that

$$(2.5) \qquad Z \subset \tilde{Z} = B - \cup_{j=1}^{n} \operatorname{Int} B(x_j, |y_j|/K).$$

Take any $z \in \tilde{Z}$ and define the function $\tilde{f}$ by

$$(2.6) \qquad \tilde{f}(x) = \max(g_n^-(x), -K\|z-x\|).$$

This f satisfies a Lipschitz condition with the constant K. Moreover from (2.5) we have $K\|z-x_j\| \ge |y_j|$, for all j, which implies by (2.3) and (2.6) that $g_n^-(z) \le 0$ and $\tilde{f}(z) = 0$. Similarly $N_n(\tilde{f}) = N_n(f)$ which means that $\tilde{f} \in F$ and $z \in Z$. From (2.5) we conclude that

$$(2.7) \qquad Z = Z(N_n(f)) = B - \cup_{j=1}^{n} \operatorname{Int} B(x_j, |y_j|/K).$$

Define $B(c_n, r_n)$ as a cube of the minimal radius containing the set Z. Thus c_n is a center of Z and r_n is the radius of Z. Denote

$$(2.8) \qquad D_n = Kr_n.$$

Let $\tilde{f} \in F$ and $N_n(\tilde{f}) = N_n(f)$. Then there exists a zero $\tilde{z}$, $\tilde{z} \in Z$, of f such that $\|c_n - \tilde{z}\| \le r_n$. Hence

$$(2.9) \qquad |\tilde{f}(c_n)| = |\tilde{f}(c_n) - \tilde{f}(\tilde{z})| \le K\|c_n - \tilde{z}\| \le Kr_n = D_n$$

for all $\tilde{f}$.

Observe also that

$$(2.10) \qquad |\tilde{f}(x)| \le K\|x - c_n\| + |\tilde{f}(c_n)| \le D_n + K\|x - c_n\|.$$

Define the functions f_n^- and f_n^+ by

$$(2.11) \qquad f_n^-(x) = \max(g_n^-(x), -D_n - K\|x - c_n\|),$$

$$(2.12) \qquad f_n^+(x) = \min(g_n^+(x), D_n + K\|x - c_n\|).$$

From (2.4) and (2.10) it is obvious that for all $\tilde{f} \in F$ such that $N_n(\tilde{f}) = N_n(f)$ we have

$$(2.13) \qquad f_n^-(x) \le \tilde{f}(x) \le f_n^+(x), \quad \forall \; x \in B.$$

This shows that f_n^- and f_n^+ are the envelopes of the functions $\tilde{f}$.

We are now ready to prove:

Lemma 2.1: Let $I = \{i : y_i > 0\}$ and $J = \{j : y_j < 0\}$. Then

$$(2.14) \qquad r(N_n, f) = \min\{|y_1|, \ldots, |y_n|, D_n, d_n\}, \quad \forall \; f \in F,$$

where $d_n = 1/2\ K \min\limits_{\substack{i\in I\\ j\in J}}\{\|x_i - x_j\| - y_i/K + y_j/K\}$.

(Assuming that $\min \emptyset = +\infty$.) $\square$

Proof: If $y_p = 0$ for some p then obviously $r(N_n,f) = 0$. Thus we can assume that $y_p = 0$ for all p. Note that $\|x_i - x_j\| \geq y_i/K - y_j/K\ \forall\ i \in I$ and $j \in J$, which implies that $d_n \geq 0$.

Denote $D = \min\{|y_1|,\ldots,|y_n|,D_n,d_n\}$. We first prove that

$$(2.15) \qquad r(N_n,f) \leq D,\ \forall\ f \in F.$$

Setting $x = x_p$ in (1.12) we observe that $r(N_n,f) \leq |y_p| = |f(x_p)|$. Taking $x = c_n$ in (1.12), (2.9) yields $r(N_n,f) \leq D_n$. Thus it is enough to prove that

$$(2.16) \qquad r(N_n,f) \leq d_n,\ \forall\ f \in F.$$

for nonempty I and J choose $i_0 \in I$ and $j_0 \in J$ such that

$$d_n = (\|x_{i_0} - x_{j_0}\| - y_{i_0}/K + y_{j_0}/K)K/2.$$

Define

$$(2.17) \qquad p = (p_1 + p_2)/2$$

where

$$\{p_1\} = \overline{x_{i_0}x_{j_0}} \cap \partial B(x_{i_0}, y_{i_0}/K),$$

$$\{p_2\} = \overline{x_{i_0}x_{j_0}} \cap \partial B(x_{j_0}, -y_{j_0}/K),$$

and $\partial B(x,y)$ denotes the boundary of $B(x,y)$.

We now prove

$$(2.18) \qquad f_n^+(p) \le d_n.$$

From (2.17) we get $d_n = 1/2\, K(\|x_{i_0} - x_{j_0}\| - \|x_{i_0} - p_1\| - \|x_{j_0} - p_2\|)$ $= K\|p_2 - p\|$. Thus

$$\begin{aligned} y_{j_0} + K\|p-x_{j_0}\| &= y_{j_0} + K(\|p-p_2\| + \|p_2-x_{j_0}\|) \\ &= y_{j_0} + d_n - y_{j_0} = d_n. \end{aligned}$$

The definition of f_n^+ implies (2.18).

Similarly we can show that $f_n^-(p) \ge -d_n$. Thus (2.13) yields $|\tilde{f}(p)| \le d_n$ for every $\tilde{f} \in F$ such that $N_n(\tilde{f}) = N_n(f)$. Hence (1.12) implies (2.16) and (2.15).

We now prove that $r(N_n, f) \ge D$. For an arbitrary $x_0 \in B$ we construct a function $\tilde{f}$ in F such that $N_n(\tilde{f}) = N_n(f)$ and

$$(2.19) \qquad |\tilde{f}(x_0)| \ge D.$$

Define $\tilde{f}$ by

$$(2.20) \qquad \tilde{f}(x) = \begin{cases} \max(f_n^-(x), D-K\|x_0-x\|) & \text{if } x_0 \notin \cup_{j\in J} B(x_j, (D-y_j)/K), \\ \min(f_n^+(x), -D+K\|x_0-x\|) & \text{otherwise.} \end{cases}$$

This f satisfies a Lipschitz condition with the constant K. Suppose that $x_0 \notin \cup_{j \in J} B(x_j, (D - y_j)/K)$. If $j \in J$ then $\|x_0 - x_j\| > (D - y_j)/K$. Thus $D - K\|x_j - x_0\| < y_j$. This implies $\tilde{f}(x_j) = f(x_j)$. If $i \in I$ then $D - K\|x_i - x_0\| \leq y_i$, so $\tilde{f}(x_i) = f(x_i)$. Thus $N_n(\tilde{f}) = N_n(f)$. From the definition of r_n there exists z in $Z(N_n(f))$ such that $Kr_n - K\|z - x_0\| \leq 0$. Thus $D - K\|z - x_0\| \leq 0$. Of course $f_n^-(z) \leq 0$ which yields $\tilde{f}(z) \leq 0$. Thus $\tilde{f}$ has a zero in B since $\tilde{f}(x_0) \geq D \geq 0$. Therefore $\tilde{f} \in F$. Hence (2.19) holds.

Similarly we can show (2.19) if $x_0 \in \cup_{j \in J} B(x_j, (D - y_j)/K)$. Note that (2.19) yields $\sup\{|\tilde{f}(x_0)|: \tilde{f} \in F: N_n(\tilde{f}) = N_n(f)\} \geq D$. Since x_0 is arbitrary, (1.12) implies that $r(N_n, f) \geq D$. Combining this with (2.15) we get (2.14). □

3. Optimality Results.

In this section we find the optimal information operator of the form (1.23), the minimal cardinality $n(\varepsilon)$ of the information operator N_n such that $r(N_n) \leq \varepsilon$, and optimal algorithms.

We first assume that the cardinality of N_n is of the form $n = M^m - 1$ for some integer M, $M > 1$. The case of general n will be discussed later. Let

$$R = 1/(2M). \tag{3.1}$$

Define the set X^* by

$$X^* = \{z \in B: z = [(2j_1-1)R, \ldots, (2j_m-1)R] \; j_k=1,\ldots,M, \; k=1,\ldots,m\}.$$

The set X^* has $M^m = n + 1$ elements. Let $x_1^*, x_2^*, \ldots, x_{n+1}^*$ be distinct elements of X^*, i.e., $X^* = \{x_1^*, \ldots, x_{n+1}^*\}$. Note that X^* is the set of centers of the cubes $B(x_i^*, R)$ which form the optimal covering of B. Here optimal covering means that $B \subset \cup_{j=1}^{n+1} B(x_j^*, R)$ and for every points z_j such that $B \subset \cup_{j=1}^{n+1} B(z_j, r)$ it may be shown that $r \geq R$, see Sukharev [8].

Let us fix $i \in \{1,2,\ldots,n+1\}$ and define a <u>nonadaptive</u> $N_{n,i}$ by

$$(3.2) \qquad N_{n,i}(f) = [f(x_1^*), \ldots, f(x_{i-1}^*), f(x_{i+1}^*), \ldots, f(x_{n+1}^*)].$$

Note that we do not compute $f(x_i^*)$ and therefore the cardinality of $N_{n,i}$ is equal to n. We are now ready to prove optimality of the information operator $N_{n,i}$.

<u>Theorem 3.1</u>: For every $i \in \{1,2,\ldots,n+1\}$ the information operator $N_{n,i}$ is optimal and $r(N_{n,i}) = K/(2M)$, where $n = M^m - 1$. □

<u>Proof:</u> Let $v = KR = K/(2M)$. We first show that

$$(3.3) \qquad r(N_{n,i}) \leq v.$$

Let f be an arbitrary element of F. Let $y_j = f(x_j^*)$ for all j. Suppose that there exists an index j such that $|y_j| \leq v$.

Then (2.14) yields $r(N_{n,i},f) \leq v$. We can assume now that $|y_j| > v$ for all j. Let $z \in Z = Z(N_{n,i}(f))$. From (2.7) $z \notin \mathrm{Int}\, B(x_j^*, |y_j|/K)$. Thus $\|z-x_j^*\| > v/K = R$. Thus $z \in B(x_i^*,R)$ and consequently $Z \subset B(x_i^*,R)$. From (2.8) we conclude that $D_n \leq v$ and (2.14) implies $r(N_{n,i},f) \leq v$. Hence $r(N_{n,i},f) \leq v$, $\forall\, f \in F$. Taking the supremum we get (3.3).

We now show that for every information operator N_n in $\mathcal{T}_n$ there exists a function g in F such that

$$(3.4) \qquad r(N_n,g) = v.$$

Recall now that the information operator N_n is of the form $N_n(f) = [f(x_1),\dots,f(x_n)]$, where x_1 is given a priori, $x_1 \in B$, and $x_i = \tilde{x}_i(f(x_1),\dots,f(x_{i-1}))$.

Define the function g by

$$(3.5) \qquad g(x) = \max_{1\leq j\leq n} (v - K\|x - z_j\|), \ \forall\ x \in B$$

where $z_1 = x_1$ and $z_1 = \tilde{x}_i(\underbrace{v,v,\dots,v}_{i-1\text{ times}})$. Then

$$(3.6) \qquad N_n(g) = [v,v,\dots,v].$$

Of course g satisfies a Lipschitz condition with the constant K. To guarantee that $g \in F$ it is enough to show that the set $A = \{z \in B\colon g(z) = 0\}$ is not empty. From (3.5) we have

$$(3.7) \qquad A = \partial \bigcup_{j=1}^{n} B(z_j,R) \cap B.$$

Suppose that $A = \emptyset$. This implies that $B \subset \bigcup_{j=1}^{n} \operatorname{Int} B(z_j, R)$ and due to (2.7) $Z = \emptyset$. Thus it is enough to prove that $Z = Z(N_n(g)) \neq \emptyset$. We shall show more by proving that

$$(3.8) \qquad \operatorname{Vol}(Z) \geq (2R)^m$$

where Vol denotes the m dimensional volume.

To obtain (3.8) observe that

$$\operatorname{Vol}(Z) = \operatorname{Vol}(B - \cup_{j=1}^{n} B(z_j, R)) \geq \operatorname{Vol}(B) - \Sigma_{j=1}^{n} \operatorname{Vol}(B(z_j, R))$$

$$= 1 - n(2R)^m = (2R)^m.$$

This yields that g has a zero and belongs to F. From (3.8) we conclude that the radius of Z is at least R. Thus from (2.8) $D_n \geq v$. From (2.14) we finally conclude that $r(N_n, g) = v$. This proves that $r(N_n) \geq v$ for any information operator N_n. Theorem 6.1 now follows from (3.3). $\square$

Theorem 6.1 says that the nonadaptive information operator is optimal. Thus even if adaptive information operators are permitted it <u>does not</u> help. The nodes of the optimal information operator are given a priori.

There are a number of problems for which the same result holds. For instance it is known that for the linear problems adaptive information operators do not help, see [3] and [10].

There are known cases of nonlinear problems for which adaptive information operators do not help. See for example [2], [4], [6], [8], [9], and [12].

For some nonlinear problems it may happen that adaptive information operators are significantly better than nonadaptive, see [5], [9] and Chapter 8 of [10]. It may be noted that for the class $F' = \{f:[a,b]\to\mathbb{R} : f(a)\le 0, f(b)\ge 0$ and $|f(x)-f(y)|\le K|x-y|\ x,y\in[a,b$ which is similar to our class F for $m = 1$, Sukharev proved in [9] that adaptive information operators are much more powerful than nonadaptive. This means that the assumption of opposite signs at the endpoints is much stronger than the assumption about existence of a zero.

We now want to find the minimal cardinality of N_n such that $r(N_n) \le \varepsilon$. Note that Theorem 3.1 states that $N_{n,i}$ is optimal if n is of the special form $n = M^n - 1$. We are unable to find the exact radius for an arbitrary n. We can however prove:

Theorem 3.2: Let $n(\varepsilon)$ be the minimal cardinality of the information operator N_n such that $r(N_n) \le \varepsilon$. Then

$$n(\varepsilon) = (\lceil K/(2\varepsilon)\rceil - a)^m - 1 \tag{3.9}$$

where $a \in (-1,0]$. □

Proof: Theorem 3.1 states that $r(N_{n,i}) = K/(2M)$ with $n = M^m - 1$. To guarantee that $r(N_{n,i}) \leq \varepsilon$ we choose the minimal n^* such that

$$(3.10) \qquad K/(2M^*) \leq \varepsilon, \; n^* = M^{*m} - 1.$$

This yields $M^* = \lceil K/(2\varepsilon) \rceil$. Suppose that $n \leq q = (M^* - 1)^m - 1$. Then for arbitrary information operator N_n we have

$$(3.11) \qquad r(N_n) \geq r(N_{n,i}) = K/(2(M^* - 1)) > \varepsilon.$$

From (3.10) and (3.11) we conclude that $n(\varepsilon)$ satisfies

$$(3.12) \qquad (M^* - 1)^m - 1 < n(\varepsilon) \leq M^{*m} - 1.$$

This can be rewritten as $n(\varepsilon) = (M^* - a)^m - 1$ with $a \in (-1,0]$. Hence (3.9) is proven. □

Suppose that $K = 2$. Then (3.9) implies that $n(\varepsilon)$ is essentially equal to $(1/\varepsilon)^m$. Note that $n(\varepsilon)$ depends strongly on the dimension of the problem. Suppose we can solve the problem using $n = 10^6$ function evaluations. Then the accuracy ε which can be guaranteed is no better than $10^{-6/m}$. Thus for $m = 1$ we get $\varepsilon \geq 10^{-6}$, for $m = 3$, $\varepsilon \geq 10^{-2}$ and for $m = 6$, $\varepsilon \geq 10^{-1}$!

We now wish to find an optimal algorithm.

Let N_n be any information operator in η_n. Define the algorithm φ by

$$(3.13) \qquad \varphi(N_n(f)) = \begin{cases} x_j & \text{if} \quad |y_j| = D, \\ c_n & \text{if} \quad D_n = D, \\ p & \text{if} \quad d_n = d, \end{cases}$$

where D, c_n, D_n, d_n and p are defined as in Section 2.

Then (2.14) and (1.10) imply that $e(\varphi,f) = r(N_n,f)$, $\forall\ f \in F$. Thus φ is a <u>strongly</u> <u>optimal</u> algorithm. The combinatory complexity of φ, i.e., the cost of computing $\varphi(y)$ for a given $y = N_n(f)$ may be large since it requires the computation of a center c_n of the set $Z(y)$. It is an interesting combinatorial problem to find the complexity, i.e., minimal cost, of computing a center of the set $B - \bigcup_{j=1}^{n} \mathrm{Int}(B(x_j,b_j))$.

For the optimal information operator $N_{n,i}$ we propose an algorithm which has combinatory complexity linear in n. Recall that $v = KR$. Define φ^* by

$$(3.14) \qquad \varphi^*(N_{n,i}(f)) = \begin{cases} x_i^* & \text{if} \quad |y_j| > v \text{ for all } j, \\ \\ x^*_j & \text{otherwise, where} \\ & |y_j| = \min\{|y_1|,\dots,|y_{i-1}|\,|y_{i+1}|,\dots,|y_{n+1}| \end{cases}$$

Thus the computation of $w = \varphi^*(y)$ for a given $y = N_{n,i}(f)$ requires only n comparisons. Equations (3.14) and (2.9) imply that

$$|f(w)| \le \min\{|y_1|,\dots,|y_{i-1}|,|y_{i+1}|,\dots,|y_{n+1}|,D_n\} = D^*.$$

From the proof of (3.3) it follows that $D^* \le v$. Thus $e(\varphi^*) \le v$. By Theorem 3.1 $r(N_{n,i}) = v$ which yields $e(\varphi^*) = r(N_{n,i})$. Hence φ^* is optimal. We summarize these results as:

Theorem 3.3: The algorithm φ^* defined by (3.14) is optimal, but not strongly optimal. The combinatory complexity of φ^* is equal to the cost of n comparisons. □

We stress that to compute $\varphi^*(y)$ we have to know the constant K. The user may not know K. Thus we propose a third algorithm which is almost optimal, does not require a knowledge of K and has combinatory complexity linear in n.

Define φ^{**} by

$$\varphi^{**}(N_{n,i}(f)) = x_j^* \tag{3.15}$$

where $|y_j| = \min\{|y_1|,\dots,|y_{i-1}|,|y_{i+1}|,\dots,|y_{n+1}|\}$.

We first find the error of this algorithm. Let f be a function, $f: B \to \mathbb{R}$, satisfying a Lipschitz condition with the constant K. If $|f(x_j^*)| > 2v$ for all j then (2.2) shows that the set $Z = \{z \in B: f(z) = 0\}$ is empty. Thus $f \notin F$. This implies

$$\forall\ f \in F\ \exists j \text{ such that } |f(x_j^*)| \le 2v. \tag{3.16}$$

Hence

$$e(\varphi^{**}) \le 2v. \tag{3.17}$$

Let

$$(3.18) \qquad g_n^-(x) = \max_{\substack{1\le j\le n+1\\ j\ne i}} (2v - K\|x - x_j^*\|).$$

Then g_n^- satisfies a Lipschitz condition with the constant K. Furthermore $g_n^-(x_i^*) = 0$ since there exists j_0 such that $\|x_i^* - x_{j_0}^*\| = 2R$ and $\|x_j^* - x_i^*\| \geq 2R$ for all j. Thus $g_n^- \in F$. This and (3.17) yield

$$(3.19) \qquad e(\varphi^{**}) = 2v = 2r(N_{n,i}).$$

This inequality says that φ^{**} is almost optimal, see (1.17). The combinatory complexity of φ^{**} is equal to the cost of $n - 1$ comparisons. We summarize these results as:

<u>Theorem 3.4:</u> (i) The algorithm φ^{**} is almost optimal, and

$$e(\varphi^{**}) = 2e(\varphi^*).$$

(It is not strongly optimal.)

(ii) The computation of $\varphi^{**}(y)$, for a given $y = N_{n,i}(f)$, does not require the knowledge of the constant K.

(iii) The combinatory complexity of φ^{**} is equal to the cost of $n - 1$ comparisons. □

4. General Information Operators

In Sections 2 and 3 we were dealing with the information operators (1.23) comprised of n function evaluations at some adaptively chosen points in the cube B. In this section we shall study the class $\mathcal{N}_n$ of general information operators (1.5) consisting of adaptive evaluations of arbitrary linear functionals. It is surprising that even in the class $\mathcal{N}_n$ the nonadaptive information operator $N_{n,i}$, see (3.2), is almost optimal. This is proven in Theorem 4.1.

Theorem 4.1: Let $n = M^m - 2$ and $n' = (M-1)^m - 1$ for some integer $M > 1$. Then

$$(4.1) \qquad K/(2M) \le \inf_{N_n \in \mathcal{N}_n} r(N_n) \le K/(2(M-1)) = r(N_{n',j}), \quad j = 1,2,\ldots,(M-1)^m. \qquad \square$$

Proof: Since $n' \le n$ we have that

$$\inf_{N_n \in \mathcal{N}_n} r(N_n) \le \inf_{N_{n'} \in \mathcal{N}_{n'}} r(N_{n'}) \le r(N_{n',j}) = K/(2(M-1)), \quad j = 1,2,\ldots,(M-1)^m.$$

This establishes the two right-nearest relations in (4.1).

Therefore it is enough to show that for every N_n from $\mathcal{N}_n$

$$K/(2M) \le r(N_n).$$

Let $R(x) = R = K/(2M)$, $x \in B$. Applying the information operator N_n to the function R we get the nonadaptive information operator, see (1.6),

$$N_{n,R}(\cdot) = [L_{1,R}(\cdot),\dots,L_{n,R}(\cdot)].$$

Let

$$h_i(x) = \begin{cases} 0 & \text{if } x \notin B(x_i^*, 1/(2M)) \\ R - K\|x-x_i\| & \text{otherwise,} \end{cases}$$

where x_i^* is defined in section 3, (3.2), and $i = 1,2,\dots,n+2$. Let $\vec{c} = (c_1,c_2,\dots,c_{n+1})$ be a nonzero solution of the homogeneous system of n linear equations with $n + 1$ unknowns:

$$\Sigma_{i=1}^{n+1} c_i L_{j,R}(h_i) = 0, \quad j = 1,2,\dots,n.$$

Let $|c_k| = \max_{1\le i\le n+1} |c_i|$. Define the functions H and f by

$$H(x) = \Sigma_{i=1}^{n+1} c_i h_i(x)/|c_k|,$$

$$f(x) = \begin{cases} R + H(x) & \text{if } c_k < 0 \\ R - H(x) & \text{otherwise.} \end{cases}$$

The function f satisfies a Lipschitz condition with the constant K and has a zero in B, since $f(x_k) = 0$. Therefore f belongs to the class F. Note that $f(x) = R$ for $x \in B(x_{n+2}^*, 1/(2M)$

Choose an arbitrary point x_0 from B. Thus $x_0 \in B(x_{i_0}^*, 1/(2M))$ for some $i_0 \in \{1,2,\dots,n+2\}$.

As before, let $\vec{c} = (c_1,\dots,c_{i_0-1},c_{i_0+1},\dots,c_{n+2})$ be a nonzero solution of the system

$$\Sigma_{\substack{i=1\\ i\neq i_0}}^{n+2} c_i L_{j,R}(h_i) = 0, \quad j = 1,2,\ldots.n.$$

Let $|c_k| = \max\{|c_i| : i \neq i_0\}$. Define the functions $\tilde{H}$ and $\tilde{f}$ by

$$\tilde{H}(x) = \Sigma_{\substack{i=1\\ i\neq i_0}}^{n+2} c_i h_i(x)/|c_k|,$$

$$\tilde{f}(x) = \begin{cases} R + H(x) & \text{if } c_k < 0 \\ R - H(x) & \text{otherwise.} \end{cases}$$

Note that $\tilde{f}(x) = R$ for $x \in B(x^*_{i_0}, 1/(2M))$ and $\tilde{f}$ belongs to F. It is crucial to notice that $N_n(\tilde{f}) = N_n(f)$.

Thus for every information operator N_n we constructed a function $f \in F$ and for every $x_0 \in B$ we constructed a function $\tilde{f} \in F$ such that $N_n(f) = N_n(\tilde{f})$ and $\tilde{f}(x_0) = R$. Due to (1.12) and (1.10) it follows that $r(N_n) \geq R$ which proves the left-most relation of (4.1).

Hence the proof of Theorem 4.1 is completed. □

<u>Corollary 4.1</u>: Let $n(\varepsilon)$ be the minimal cardinality of the information operator N_n in $\mathcal{N}_n$ such that $r(N_n) \leq \varepsilon$. Then

$$n(\varepsilon) = (K/(2\varepsilon))^m(1 + o(1)) \text{ as } \varepsilon \to 0.$$ □

<u>Proof:</u> Theorem 3.2 implies that

(4.2) $$n(\varepsilon) \leq (K/(2\varepsilon) + 2)^m - 1.$$

Choose the maximal M such that $K/(2M) > \varepsilon$, i.e., $M = K/(2\varepsilon) - b$ for some $b \in [0,1]$. Theorem 4.1 yields that $r(N_n) \geq K/(2M)$ if $n = M^m - 2$. Thus $n(\varepsilon)$ has to satisfy

$$(4.3) \qquad n(\varepsilon) > M^m - 2.$$

The inequalities (4.2) and (4.3) imply that $n(\varepsilon) = (K/(2\varepsilon))^m(1 + o(1))$. This completes the proof. $\square$

5. Complexity of the Problem

As in [10] by the complexity $\mathrm{comp}(\varepsilon)$ of the problem we mean the minimal cost of solving (1.4). Thus $\mathrm{comp}(\varepsilon)$ is the sum of the computational cost of evaluation an information operator N_n and the minimal combinatory cost (combinatory complexity) of an optimal algorithm using N_n, where n is the minimal cardinality such that $r(N_n) \leq \varepsilon$.

The results of Section 3 and 4 enable us to find the complexity $\mathrm{comp}(\varepsilon)$. Assume that c_1 is the cost of one functional evaluation and that arithmetic operations and comparisons cost unity. Moreover, assume that any optimal algorithm has to use each y_j at least once. This implies that its combinatory complexity has to be at least equal to $n - 1$. Thus the algorithm φ^* has an almost minimal combinatory complexity. We summarize these results in

Theorem 4.2: (i) The complexity of the problem (1.4) is

$$\mathrm{comp}(\varepsilon) = n(\varepsilon)(c_1 + b)$$

where $b \in [1 - 1/n(\varepsilon), 1]$.

(ii) The complexity $\mathrm{comp}(\varphi^*, \varepsilon)$ of the algorithm φ^*, i.e., the sum of the computational cost of the information operator $N_{n,i}$ and the combinatory complexity of φ^*, is almost minimal since

$$\mathrm{comp}(\varphi^*, \varepsilon) = \mathrm{comp}(\varepsilon)(1 + u)$$

where $u = (1 - b)/(c_1 + b)$ and $u \leq 1/(c_1 n(\varepsilon))$.

(iii) The complexity $\mathrm{comp}(\varphi^{**}, \varepsilon)$ of the algorithm φ^{**}, which does not require the knowledge of K, is

$$\mathrm{comp}(\varphi^{**}, \varepsilon) = \mathrm{comp}(\varepsilon)(2 + w_1)^m (1 + w_2)$$

where

$$|w_1| \leq 3/(n(\varepsilon) + 1)^{1/m}$$

and

$$|w_2| \leq 1/n(\varepsilon)(1 + 2/c_1).$$

Thus asymptotically

$$\mathrm{comp}(\varphi^{**}, \varepsilon) = 2^m \mathrm{comp}(\varepsilon)(1 + o(1)) \text{ as } \varepsilon \to 0.$$

6. Final Remarks.

It is important to note that our negative complexity result depends significantly on the class of functions F. Indeed, define the class F_1 by

$$F_1 = \{f:[0,1] \to \mathbb{R} : K_1|x-y| \leq |f(x)-f(y)| \leq K_2|x-y|$$
$$\text{for all } x,y \in [0,1] \text{ and } \exists\, z \in [0,1] : f(z) = 0\}.$$

Thus it is a class of functions satisfying a two-way Lipschitz condition with the constants K_1 and K_2, $0 < K_1 \leq K_2$, and having a zero in $[0,1]$. As in Sections 2 and 3 we can prove that

$$r(N_{n,i}) = (K_2 - K_1)/(2(n+1))$$

where $n \geq 2$ and $i \in \{1,2,...,n+1\}$.

Thus $n(\varepsilon)$ defined as in Section 3, Theorem 3.2, is no greater than $M^* = \max(\lceil (K_2 - K_1)/(2\varepsilon)\rceil - 1, 2)$. We can prove that there exists an optimal algorithm using $N_{n,i}$ with combinatory complexity no greater than cn, where c is a constant. Therefore the complexity $\mathrm{comp}(\varepsilon)$ is no greater than $M^*(c_1 + c)$.

Note that if K_1 is close enough to K_2 then the complexity $\mathrm{comp}(\varepsilon)$ is essentially equal $2\,c_1$. This is intuitively obvious since for K_1 tending to K_2 the class F_1 shrinks to the class consisting of linear functions. A linear function f is uniquely determined by the formula $|f(x) - f(y)| = K_2|x-y|$

and the values at two different points.

It is an open problem to generalize the above result for the class F_1 to the m dimensional case. We conjecture that the complexity is roughly $(\lceil (K_2 - K_1)/(2\varepsilon)\rceil - 1)^m (c_1 + 1)$.

Acknowledgements.

I am extremely indebted to H. Woźniakowski for his help and comments during the preparation of this paper. He has convinced me that the problem of writing a paper is an adaptive procedure. I am also indebted to J. Traub and G. Wasilkowski for reading the manuscript and valuable remarks.

References

[1] Aird, T.J., Rice, J.R., Systematic Search in High Dimensional Sets, SIAM J. Num. Anal., vol. 14. 1977. 296-312.

[2] Bakhvalov, N.S., On the Optimality of Linear Methods for Operator Approximation in Convex Classes of Functions, USSR Comp. Math. and Math. Phys., 11, no. 4, 1971, 244-249.

[3] Gal, S., Micchelli, A.C., Optimal Sequential and Non Sequential Procedures for Evaluating of Functional, Appl. Anal., 10, 1980, 105-120.

[4] Kiefer, J., Optimum Sequential Search and Approximation Methods Under Minimum Regularity Assumptions, J. Soc. Indust. Appl. Math., 5, 1957, 105-136.

[5] Sikorski, K., Bisection is Optimal, 1981, to appear in Num. Math.

[6] Sikorski, K., Woźniakowski, H., For What Error Criteria We Can Solve Nonlinear Equations, 1981, submitted to BIT.

[7] Sobol, I.M., On the Systematic Search in a Hypercube, SIAM J. Num. Anal., vol., 16, 1979, 790-793.

[8] Sukharev, A.G., Optimal Strategies of the Search for an Extremum, USSR, Comp. Math. and Math. Phys., 11, no. 4, 1971, 119-137.

[9] Sukharev, A.G., Optimal Search for a Zero of Function Satisfying a Lipschitz Condition, USSR Comp. Math. and Math. Phys., 16, no. 1, 1976, 17-26.

[10] Traub, J.F., Woźniakowski, H., A General Theory of Optimal Algorithms, Academic Press, New York 1980.

[11] Traub, J.F., Wasilkowski, G., Woźniakowski, H., Information, Uncertainty, Complexity, Addison-Wesley, 1983.

[12] Zaliznyak, N.F., Ligun, A.A., Optimal Strategies for Seeking the Global Maximum of a Function, USSR Comp. Math. and Math. Phys. 18, no. 2, 1976, 31-38.