CUCS-58-83

Some Nonlinear Operators Are As Easy To Approximate As The Identity Operator

G.W. Wasilkowski

Department of Computer Science
Columbia University
New York, New York

May 1982

Revised May 1983

This research was supported in part by the National Science Foundation under Grant MCS 82-14322.

Abstract

In this paper we study the following problem. Given an operator S and a subset F_0 of some linear space, approximate $S(f)$ for any $f \in F_0$ possessing only partial information on f. Although all operators S considered here are nonlinear (e.g. $\min f(x)$, $\min |f(x)|$, $\frac{1}{f}$ or $\|f\|$), we prove that these problems are "equivalent" to the problem of approximating $S(f) = f$, i.e., $S = I$. This equivalence provides optimal (or nearly optimal) information and algorithms.

1. Introduction

There are many papers dealing with the following problem: approximate an element f which belongs to a subclass F_0 of a linear normed space possessing only partial information on f. For many subclasses F_0 we know optimal information, optimal algorithms and we know that adaptive information is not more powerful than nonadaptive information. (see e.g. [1], [4]). This is an example of a linear problem; that is one wants to approximate S(f) for a linear operator S.

The situation is quite different for nonlinear problems; that is one wants to approximate S(f) where S is a nonlinear operator. Nonlinearity of S usually makes the problem of finding optimal information and optimal algorithms more difficult.

In this paper we give sufficient conditions for a nonlinear problem to be equivalent to the problem of approximating S(f) = f. This equivalence leads to optimal (or nearly optimal) information and algorithms for this nonlinear problem. We will present some nonlinear problems for which these sufficient conditions hold.

We summarize the contents of this paper. In Section 2 we present the basic definitions and results which will be needed in this paper. We define what we mean by a problem, information and an algorithm. We recall the concept of the error of an algorithm and of the radius of information. We show how these concepts become simpler for the problem with $S = I$ i.e., $S(f) = f$. In Section 3 we prove two simple lemmas which give sufficient conditions for a nonlinear problem to be equivalent to the problem with $S = I$. We illustrate these lemmas by such problems as approximation of $S(f) = \frac{1}{f}$ or $S(f) = \sqrt{f}$ where f is a function. In Section 4 we consider the problem of estimating $S(f) = \|f\|$. In the last section we study three problems which are related to the problem of finding the minimum of a given function f.

For all these problems we exhibit nearly optimal information and nearly optimal algorithms. We also prove that adaption does not help.

2. Basic concepts.

In this section we present the basic definitions and results which will be needed in this paper. A more detailed discussion can be found in [3] and [4].

Let F_1, F_2 be linear spaces and let F_0 be a subset of F_1. Let $\bar{S}$ be an operator

$$(2.1)\qquad \bar{S} : F_0 \times \mathbb{R}_+ \to 2^{F_2}$$

such that for every $f \in F_0$ and $\delta \geq 0$

$$(2.2)\qquad \bar{S}(f,\delta) \neq \emptyset$$

$$\bar{S}(f,\delta_1) \subset \bar{S}(f,\delta_2) \quad \text{whenever} \quad \delta_1 \leq \delta_2.$$

By the $(\bar{S},F_0)$-problem we mean the problem of constructing an element $g = g(f) \in F_2$ such that

$$(2.3)\qquad g(f) \in \bar{S}(f,\delta), \qquad \forall f \in F_0,$$

for a possibly small number δ.

To solve this problem we use an adaptive linear information operator N (briefly information operator or information) which is defined by

$$(2.4)\qquad N(f) = [L_1(f), L_2(f;y_1), \ldots, L_n(f;y_1,y_2,\ldots,y_{n-1})]$$

where $y_1 = y_1(f) = L_1(f)$, $y_i = y_i(f) = L_i(f;y_1,\ldots,y_{i-1})$ and

(2.5) $$L_{i,f}(\cdot) \stackrel{df}{=} L_i(\cdot;y_1,\ldots,y_{i-1}) : F_1 \to \mathbb{R}$$

is a linear functional, $i = 1,2,\ldots,n$. If $L_{i,f}$ does not depend on f, i.e., $L_{i,f} \equiv L_i$, for $i = 1,2,\ldots,n$, then N is called nonadaptive. By the cardinality of N we mean the total number n of functional evaluations, $\mathrm{card}(N) = n$.

Knowing $N(f)$ we construct $g(f)$ by an algorithm φ, i.e., $g(f) = \varphi(N(f))$. Here by an algorithm φ using N we mean any mapping

(2.6) $$\varphi : N(F_0) \to F_2.$$

The error of φ is defined as

(2.7) $$e(\varphi,N;\bar{S},F_0) = \inf\{\delta \geq 0 : \forall f \in F_0, \varphi(N(f)) \in \bar{S}(f,\delta)\}.$$

Let $\Phi(N)$ be the class of all algorithms using N,

$$\Phi(N) = \{\varphi : N(F_0) \to F_2\}.$$

By the radius of N we mean

(2.8) $$r(N;\bar{S},F_0) = \inf_{\varphi\in\Phi(N)} e(\varphi,N;\bar{S},F_0).$$

Thus the radius $r(N;\bar{S},F_0)$ is the sharp lower bound on errors of algorithms using N. An algorithm φ^*, $\varphi^* \in \Phi(N)$, is optimal iff

$$(2.9)\qquad e(\varphi^*,N;\bar{S},F_0) = r(N;\bar{S},F_0).$$

Let Ψ_n^a be the class of all adaptive linear information operators of cardinality not greater than n and let Ψ_n^{non} be the subclass of Ψ_n^a consisting of all nonadaptive linear information operators. The <u>nth adaptive radius for</u> the $(\bar{S},F_0)$ - <u>problem</u> is defined as

$$(2.10)\qquad r^a(n;\bar{S},F_0) = \inf_{N\in\Psi_n^a} r(N;\bar{S},F_0)$$

and the <u>nth nonadaptive radius for</u> the $(\bar{S},F_0)$- <u>problem</u> as

$$(2.11)\qquad r^{non}(n;\bar{S},F_0) = \inf_{N\in\Psi_n^{non}} r(N;\bar{S},F_0).$$

Of course, $r^{non}(n;\bar{S},F_0) \geq r^a(n;\bar{S},F_0)$.

We shall say that N^* is an <u>nth adaptive</u> (or <u>nonadaptive</u>) <u>optimal information operator for</u> the $(\bar{S},F_0)$- <u>problem</u> if

$$N^* \in \Psi_n^a \quad (\text{or} \quad N^* \in \Psi_n^{non})$$

$$r(N^*;\bar{S},F_0) = r^a(n;\bar{S},F_0) \quad (\text{or} = r^{non}(n;\bar{S},F_0)).$$

Roughly speaking, the error of any algorithm using an arbitrary information operator of cardinality at most n is not smaller than the nth radius. The error of an optimal algorithm using optimal information is equal to $r(n;\bar{S},F_0)$.

That is why we want to find optimal algorithms and optimal information operators.

Suppose now that the space F_2 is equipped with the norm $\|\cdot\|_{F_2}$ and that there exists an operator S (in general nonlinear),

$$(2.12)\qquad S : F_0 \to F_2,$$

such that

$$(2.13)\qquad \bar{S}(f,\delta) = \{g \in F_2 : \|S(f)-g\|_{F_2} \le \delta\},\quad \forall f \in F_0,\quad \forall \delta \ge 0.$$

In the $(\bar{S},F_0)$-problem we approximate $S(f)$, where the error is measured by $\|S(f) - g\|_{F_2}$. Such a problem is called a nonlinear problem. To stress the special form of this problem we drop the bar over S and denote it by the (S,F_0)-problem. For every algorithm φ we have

$$(2.14)\qquad e(\varphi,N;S,F_0) = \sup_{f\in F_0}\|S(f)-\varphi(N(f))\|_{F_2}.$$

For a nonlinear problem we can estimate the radius of information as follows. Let $N \in \Psi_n^a$ and

$$(2.15)\qquad d(N;S,F_0) = \sup_{f\in F_0}\ \sup\{\|S(f)-S(\tilde{f})\|_{F_2},\ \ N(\tilde{f}) = N(f),\ \tilde{f} \in F_0\}$$

be <u>the diameter of</u> N. Then

$$(2.16) \qquad \tfrac{1}{2}d(N;S,F_0) \le r(N;S,F_0) \le d(N;S,F_0).$$

In many cases we have the left equality in (2.16). This holds for instance, if S is a functional, i.e., $F_2 = \mathbb{R}$ and $\|\cdot\|_{F_2} = |\cdot|$. Note that $d(N;S,F_0)$ has a relatively simple form and provides a rather sharp estimate of $r(N;S,F_0)$. We also know that any interpolatory algorithm is nearly optimal. By an <u>interpolatory algorithm</u> we mean any algorithm $\varphi^I \in \Phi(N)$ such that

$$\varphi^I(N(f)) = S(\tilde{f}) \quad \text{for some } \tilde{f} \in F_0 \quad \text{and}$$

$$N(\tilde{f}) = N(f).$$

Then

$$r(N;S,F_0) \le e(\varphi^I,N;S,F_0) \le d(N;S,F_0) \le 2r(N;S,F_0).$$

Hence the error of φ^I differs at most by a factor of two from the error of an optimal algorithm.

We now consider a very special problem defined as follows. Let F_1 be equipped with the norm $\|\cdot\|_{F_1}$, let $S = I$ be the identity operator and let F_0 be balanced and convex (i.e., $f \in F_0$ implies $-f \in F_0$, $f_1,f_2 \in F_0$ implies $tf_1 + (1-t)f_2 \in F_0$, $\forall\, t \in [0,1]$). Then the (I,F_0)-problem is called the <u>approximation</u> <u>(I,F_0)-problem</u> or briefly the <u>approximation problem</u>. For the approximation problem it is easy to find the diameter of N. Indeed, let N^{non} be an nonadaptive information operator. Then

$$(2.17) \qquad d(N^{non};I,F_0) = 2 \sup_{h\in F_0 \cap \ker N} \|h\|_{F_1} .$$

For an adaptive information operator N^a of the form (2.4) we have

$$(2.18) \qquad d(N_0;I,F_0) \le d(N^a;I,F_0) \le \sup_{f\in F_0} d(N_f;I,F_0)$$

where, as in (2.5), $N_f = [L_{1,f},\dots,L_{n,f}]$ is a nonadaptive information operator.

From (2.16) and (2.18) it follows that

$$(2.19) \qquad r^a(n;I,F_0) \le r^{non}(n;I,F_0) \le \frac{1}{2}\, r^a(n;I,F_0).$$

Thus adaption does not essentially help for the approximation problem.

3. Two Lemmas.

In this section we prove two lemmas which will be used in the next sections. These lemmas provide lower and upper bounds on the diameter of information for a nonlinear (S,F_0)-problem. We estimate the diameter of N for (S,F_0)-problem by the diameter of N for the approximation $(I,\tilde{F}_0)$-problem for some $\tilde{F}_0$ which depends on F_0.

Lemma 3.1: Suppose there exist

(i) an element $f^* \in F_0$,

(ii) a balanced and convex subset $\tilde{F}_0 \subset F_1$,

(iii) a positive constant m such that

$$f^* + h \in F_0, \quad \forall h \in \tilde{F}_0$$

and

$$\max\{\|S(f^*)-S(f^*-h)\|_{F_2}, \|S(f^*)-S(f^*+h)\|_{F_2}, \|S(f^*-h)-S(f^*+h)\|_{F_2}\} \geq m\|h\|_{F_1}, \quad \forall h \in \tilde{F}_0.$$

Then for every information operator N, $N \in \Psi_n^a$,

$$(3.1) \qquad d(N;S,F_0) \geq \frac{1}{2}\, m d(N_{f^*};I,\tilde{F}_0)$$

and

$$(3.2) \qquad r^a(n;S,F_0) \geq \frac{1}{4}\, m r^{non}(n;I,\tilde{F}_0). \qquad \square$$

Proof: Let $h \in \tilde{F}_0 \cap \ker N_{f^*}$. Then $f^* \pm h \in F_0$ and

$N(f^*) = N(f^* \pm h)$. Due to (2.15) and (2.17),

$$d(N;S,F_0) \geq \sup_{h\in\tilde{F}_0\cap \ker N_{f^*}} \max\{\|S(f^*)-S(f^*-h)\|_{F_2},$$

$$\|S(f^*)-S(f^*+h)\|_{F_2}, \|S(f^*-h)-S(f^*+h)\|_{F_2}\}$$

$$\geq m \sup_{h\in\tilde{F}_0\cap \ker N_{f^*}} \|h\|_{F_1} = \frac{1}{2}\, m d(N_{f^*};I,\tilde{F}_0).$$

This proves (3.1). Since N_{f^*} is nonadaptive, $d(N_{f^*};I,\tilde{F}_0) \geq r^{non}(n;I,\tilde{F}_0)$ and (2.16) yields (3.2). □

We need the following definition. Let conv(A) denote the convex hull of a set A, $A \subset F_1$. For a given subset F_0, $F_0 \subset F_1$, let $BC(F_0)$ be the balanced convex hull of F_0 defined by

(3.3) $$BC(F_0) = \frac{1}{2}\,\mathrm{conv}(F_0+(-F_0)) \quad (= \mathrm{conv}\{\tfrac{1}{2}(f_1-f_2): f_1,f_2 \in F_0\}).$$

Of course, $BC(F_0)$ is balanced and convex. Furthermore $BC(F_0) = F_0$ iff F_0 is balanced and convex.

<u>Lemma 3.2</u>: Suppose there exists a constant M such that

(3.4) $$\|S(f_1)-S(f_2)\|_{F_2} \leq M\|f_1-f_2\|_{F_1}, \quad \forall f_1,f_2 \in F_0.$$

Then for every information operator N, $N \in \Psi_n^a$, (3.5) $$d(N;S,F_0) \leq M \sup_{f\in F_0} d(N_f;I,BC(F_0))$$

and

(3.6) $$r^a(n;S,F_0) \leq 2M\, r^a(n;I,BC(F_0)). \qquad \square$$

Proof: Let $f,\tilde{f} \in F_0$; $N(f) = N(\tilde{f})$. Define $h^* = \frac{1}{2}(f-\tilde{f})$. Then $h^* \in BC(F_0) \cap \ker N_f$. Therefore

$$\|S(f)-S(\tilde{f})\|_{F_2} \leq \frac{M}{2}\|h^*\|_{F_1} \leq \frac{M}{2} \sup_{h\in BC(F_0)\cap\ker N_f} \|h\|_{F_1}$$

$$= Md(N_f;I,BC(F_0)).$$

Since $d(N;S,F_0) = \sup_{f\in F_0} \sup\{\|S(f)-S(\tilde{f})\|_{F_2} : f \in F_0,\ N(\tilde{f}) = N(f)\}$ then

$$d(N;S,F_0) \leq M \sup_{f\in F_0} d(N_f;I,BC(F_0)).$$

This proves (3.5). Since (3.6) easily follows from (3.5) and (2.16) the proof is completed. $\square$

We illustrate Lemmas 3.1 and 3.2 by the following problem. Let $F_1 = C[0,1]$ be the space of continuous functions with the sup norm

$$\|f\|_{F_1} = \|f\| = \sup_{x\in[0,1]} |f(x)|.$$

Let $F_0 = \{f \in F_1 : f(x) \in [1,3], |f'(x)| \leq 1,\ \forall x \in [0,1] \text{ a.e.}\}$ and let $\bar{g}$ be a function, (3.7) $g : [1,3] \to \mathbb{R}$

such that $g'(x) \in [m_1, M_1]$. Define $S : F_0 \to F_2 = F_1$ as

(3.8) $S(f)(x) = g(f(x))$.

We now apply Lemma 3.1. Take $f^*(x) = 2$ and $\tilde{F}_0 = \{h \in F_1 : \|h\| \le 1, |h'(x)| \le 1, \forall x \in [0,1] \text{ a.e.}\}$. Then $f^* + h \in F_0$ for every $h \in \tilde{F}_0$. Furthermore for every $h \in \tilde{F}_0$ we have

$$\|S(f^*)-S(f^*+h)\| = \sup_{x\in[0,1]} |g(f^*(x))-g(f^*(x)+h(x))|$$

and due to the Taylor expansion of g we get

$$\|S(f^*)-S(f^*+h)\| \ge m_1 \sup_{x\in[0,1]} |h(x)| = m_1\|h\|.$$

Hence Lemma 3.1 holds with $m = m_1$ and

(3.9) $d(N;S,F_0) \ge \frac{1}{2} m_1 d(N_{f^*};I,\tilde{F}_0)$.

We now apply Lemma 3.2. Using once more the Taylor expansion of g we easily conclude that $\|S(f_1)-S(f_2)\| \le M_1\|f_1-f_2\|$, $\forall f_1, f_2 \in F_0$. Hence,

(3.10) $d(N;S,F_0) \le M_1 \sup_{f\in F_0} d(N_f;I,BC(F_0))$

where $BC(F_0) = \{f \in F_1 : \|f\| \le 1, |f'(x)| \le 1, \forall x \in [0,1] \text{ a.e.}\}$. It can be proven that for every information N

$d(N;S,F_0) = 2r(N;S,F_0)$, $d(N;I,\tilde{F}_0) = 2r(N;I,\tilde{F}_0)$ and $d(N;I,BC(F_0)) = 2r(N;I,BC(F_0))$. Hence,

$$(3.11) \qquad \frac{1}{2} m_1 r(N;I,\tilde{F}_0) \le r(N;S,F_0) \le M_1 \sup_{f\in F_0} r(N_f,I,BC(F_0)).$$

It is easy to prove that

$$(3.12) \qquad N_n^*(f) = [f(x_1),f(x_2),\ldots,f(x_n)], \qquad x_i = \frac{2i-1}{2n},$$

is an nth optimal information for both $(I,\tilde{F}_0)$ and $(I,BC(F_0))$ problems and

$$r(N_n^*;I,BC(F_0)) = r(N_n^*;I,\tilde{F}_0) = r(n;I,BC(F_0))$$
$$= r(n;I,\tilde{F}_0) = \frac{1}{n}.$$

This means that N_n^* is an almost nth optimal information for the (S,F_0)-problem and

$$\frac{1}{2n} m_1 \le r^a(n;S,F_0) \le r(N_n^*;S,F_0) \le \frac{1}{n} M_1.$$

For example, if $g(x) = \frac{1}{x}$ (i.e., $S(f) = \frac{1}{f}$) then $m_1 = \frac{1}{9}$ and $M_1 = 1$, and if $g(x) = \sqrt{x}$ (i.e., $S(f) = \sqrt{f}$) then $m_1 = \frac{1}{2\sqrt{3}}$ and $M_1 = \frac{1}{2}$.

4. Optimal estimation of $\|f\|_{F_1}$.

In this section, we solve the following simple problem. Let F_0 be balanced and convex. Let $F_2 = \mathbb{R}$ with $\|\cdot\|_{F_2} = |\cdot|$ and let

$$(4.1) \qquad S(f) = \|f\|_{F_1}.$$

Thus our problem is to approximate the value $\|f\|_{F_1}$ for every $f \in F_0$.

Theorem 4.1: For every information operator N, $N \in \Psi_n^a$,

$$(4.2) \qquad \frac{1}{4}\, r(N_0;T,F_0) \le r(N;S,F_0) \le \sup_{f\in F_0} r(N_f;I,F_0)$$

and

$$(4.3) \qquad \frac{1}{4}\, r^{non}(n;I,F_0) \le r^a(n;S,F_0) \le r^{non}(n;I,F_0). \qquad \square$$

Proof: Let $f^* = 0$ and $h \in F_0$. Then $f^* + h \in F_0$ and $|\|f^*-h\|_{F_1} - \|f^*\|_{F_1}| = \|h\|_{F_1}$. This means that Lemma 3.1 holds with $f^* = 0$, $\tilde{F}_0 = F_0$ and $m = 1$. Hence,

$$\frac{1}{2}\, d(N_0;I,F_0) \le d(N;S,F_0), \qquad \forall N \in \Psi_n^a.$$

Observe that $|\|f_1\|_{F_1} - \|f_2\|_{F_1}| \le \|f_1-f_2\|_{F_1}$, $\forall f_1,f_2 \in F_0$. Thus Lemma 3.2 holds with $BC(F_0) = F_0$ and $M = 1$. Hence

$$d(N;S,F_0) \le \sup_{f\in F_0} d(N_f;I,F_0), \qquad \forall N \in \Psi_n^a.$$

Thus,

$$\frac{1}{2}\, d(N_0;I,F_0) \le d(N;S,F_0) \le \sup_{f\in F_0} d(N_f;I,F_0),$$

$$\forall N \in \Psi_n^a.$$

Since $2r(N;S,F_0) = d(N;S,F_0)$, $\forall N$, then

$$\frac{1}{4}\, r(N_0;I,F_0) \le r(N;S,F_0) \le \sup_{f\in F_0} r(N_f;I,F_0) \tag{4.4}$$

which proves (4.2). Since N_f is nonadaptive then (4.3) easily follows from (4.2). This completes the proof. $\square$

This theorem states that the problem of estimating the value of $S(f) = \|f\|_{F_1}$ is equivalent to the approximation (I,F_0)-problem. Hence every nth optimal information operator N^* for (I,F_0)-problem is also nearly optimal for the $(S.F_0)$-problem. Since this problem is an example of the nonlinear problem we know that every interpolatory algorithm φ^I using N is almost optimal. We illustrate this by the following example.

<u>Example 4.1</u>: Let F_1 be a separable Hilbert space and $F_0 = \{f \in F_1 : \|Tf\|_{F_1} \le 1\}$ where $T : F_1 \to F_1$ is a one-to-one linear operator. Let $K_1 = (T^{-1})^*(T^{-1})$. We assume that K_1 is compact. Then there exists an orthonormal basis

$\zeta_1, \zeta_2, \ldots,$ such that $K_1\zeta_i = \lambda_i\zeta_i$ and $\lambda_1 \geq \lambda_2 \ldots \geq 0$.
Define

(4.5) $$N^*(f) = [(f,\zeta_1),(f,\zeta_2),\ldots,(f,\zeta_n)].$$

From [4] we know that N^* is an nth optimal information operator for the (I,F_0)-problem and

(4.6) $$r(N^*;I,F_0) = r^a(n;I,F_0) = \sqrt{\lambda_{n+1}}.$$

Due to Theorem 4.1 we get that N^* is nearly optimal for the (S,F_0)-problem and

(4.7) $$r(N^*;S,F_0) = c_1\sqrt{\lambda_{n+1}} = c_2 r^a\ (n;S,F_0)$$

where $c_1, c_2 \in [1/4,1]$. Let $\varphi^I \in \Phi\ (N^*)$ be defined as

(4.8) $$\varphi^I(N^*(f)) = \|\Sigma_{i=1}^n\ (f,\zeta_i)\zeta_i\|_{F_1} = \{\Sigma_{i=1}^n\ (f,\zeta_i)^2\}^{1/2}.$$

Since φ^I is interpolatory then φ^I is nearly optimal and

$$e(\varphi^I,N^*;S,F_0) = cr(N^*;S,F_0)$$

where $c \in [1,2]$.

5. Minimum function problems.

In this section, we solve some (S,F_0)-problems which are related to the estimation of the minimum of functions from a given set F_0. We prove the equivalence between these problems and the approximation (I,F_0)-problem. Since for many subclasses F_0 we know an nth optimal information N^* for (I,F_0)-problem, this provides a nearly optimal information for the (S,F_0)-problems.

Let $F_1 = C[0,1]$ be the space of continuous functions with the sup norm, i.e.,

$$\|f\|_{F_1} = \|f\| = \sup_{x\in[0,1]} |f(x)|.$$

Let F_0 be a balanced and convex subset of F_1. We consider three problems in the successive subsections.

(i) Minimum-value problem

Let

$$S_1(f) = \min_{x\in[0,1]} f(x), \qquad S_1 : F_0 \to \mathbb{R} = F_2. \tag{5.1}$$

Consider the (S_1,F_0)-problem, i.e., we want to approximate the minimal value of f for every $f \in F_0$. Of course, this is a nonlinear problem.

Theorem 5.1: For every information operator N, $N \in \Psi_n^a$

$$\frac{1}{2}\, r(N_0;I,F_0) \le r(N;S_1,F_0) \le \sup_{f\in F_0} r(N_f;I,F_0) \tag{5.2}$$

and

$$\frac{1}{2}\, r^{non}(n;I,F_0) \le r^a(n;S_1,F_0) \le r^{non}(n;I,F_0). \tag{5.3}$$ □

Proof: Take $f^* = 0$. For $h \in F_0$ define

$$h_-(x) = \min(h(x),0) \quad \text{and} \quad h_+(x) = \max(h(x),0).$$

Then

$$|S_1(f^*) - S_1(f^*+h)| = |S_1(h)| = |\min_{x\in[0,1]} h(x)| = \|h_-\|$$

$$|S_1(f^*) - S_1(f^*-h)| = |S_1(-h)| = |\max_{x\in[0,1]} h(x)| = \|h_+\|.$$

Since $\max\{\|h_-\|,\|h_+\|\} = \|h\|$ then Lemma 3.1 holds with $\tilde{F}_0 = F_0$ and $m = 1$.

Hence $\frac{1}{2}\, d(N_0,I,F_0) \le d(N;I,F_0)$. It is known that $2r(N;I,F_0) = d(N;I,F_0)$. Since S_1 is a functional then $2r(N;S_1,F_0) = d(N;S_1,F_0)$, $\forall N$. This proves the left hand side of (5.2).

To prove the right hand side we apply Lemma 3.2. Since F_0 is balanced and convex then $BC(F_0) = F_0$. Then there exist $\alpha_1,\alpha_2 \in [0,1]$ such that $S_1(f_i) = f(\alpha_i)$, $i = 1,2$. Without loss of generality we can assume that $f_1(\alpha_1) \ge f_2(\alpha_2)$. Then

$$|S_1(f_1) - S_1(f_2)| = f_1(\alpha_1) - f_2(\alpha_2) \le f_1(\alpha_2) - f_2(\alpha_2)$$

$$= |f_1(\alpha_2) - f_2(\alpha_2)| \le \|f_1 - f_2\|.$$

Thus Lemma 3.2 holds with M = 1. Hence

$$d(N;S_1,F_0) = 2r(N;S_1,F_0) \le \sup_{f\in F_0} d(N_f;I,F_0)$$

$$= 2 \sup_{f\in F_0} r(N_f;I,F_0).$$

This proves (5.2). Since (5.3) follows immediately from (5.2), the proof is completed. ⊐

We specify Theorem 5.1 by taking

(5.4) $$F_0 = \{f \in F_1 : f^{(r-1)} \text{ abs. continuous, } \|f^{(r)}\|_\infty \le 1\}.$$

From [2] we know that the information operator

(5.5) $$N^*(f) = [f(x_1), f(x_2), \ldots, f(x_n)], \quad x_i = \frac{2i-1}{2n}$$

is nearly optimal for (I,F_0)-problem and

$$r(N^*;I,F_0) = \Theta(n^{-r}) = r^a(n;I,F_0).$$

Due to Theorem 5.1 we get that N^* is nearly optimal for (S_1,F_0)-problem and

(5.6) $$r(N^*;S_1,F_0) = \Theta(n^{-r}) = r^a(n;S_1,F_0).$$

From Section 2 we also know that every interpolatory algorithm

$\varphi^I \in \Phi(N^*)$ has the error

(5.7) $$e(\varphi^I, N^*; S, F_0) = \Theta(n^{-r}).$$

(ii) Modulus minimum-value problem

Let

(5.8) $$S_2(f) = \min_{x\in[0,1]} |f(x)| \qquad (= S_1(|f|)),$$

and consider the (S_2, F_0)-problem. Thus, we now approximate the minimum of the absolute values of $f(x)$. It is easy to observe that Lemma 3.2 is satisfied with $M = 1$. Indeed, for $f_1, f_2 \in F_0$, $|S_2(f_1) - S_2(f_2)| = ||f_1(\beta_1)| - |f_2(\beta_2)|| \le \|f_1 - f_2\|$ where $S_2(f_i) = |f_i(\beta_i)|$. Hence

$$r(N; S_2, F_0) \le \sup_{f\in F_0} r(N_f; I, F_0), \quad \forall N \in \Psi_n^a.$$

Assume that there exists a positive constant c such that

(5.9) $$f_c(x) \equiv c \in F_0.$$

Define $F_0(c)$ as follows:

(5.10) $$F_0(c) = \{h \in F_1 : f_c + h \in F_0 \text{ and } \|h\| \le c\}.$$

Of course, $F_0(c)$ is balanced and convex. Furthermore for every $h \in F_0(c)$ we have $f_c + h \in F_0$, $|S_2(f_c) - S_2(f_c+h)| \ge \|h_-\|$

and $|S_2(f_c)-S_2(f_c-h)| \geq \|h_+\|$. Hence Lemma 3.1 holds with $\tilde{F}_0 = F_0(c)$, $m = 1$ and $f^* = f_c$. Since S_2 is a functional then

$$2r(N_{f_c};I,F_0(c)) \leq r(N;S_2,F_0).$$

We summarize this in the following theorem.

Theorem 5.2: For every information operator N, $N \in \Psi_n^a$,

$$(5.11) \qquad \frac{1}{2}\, r(N_{f_c};I,F_0(c)) \leq r(N;S_2,F_0) \leq \sup_{f\in F_0} r(N_f;I,F_0)$$

and

$$(5.12) \qquad \frac{1}{2}\, r^{non}(n;I,F_0(c)) \leq r^a(n;S_2,F_0) \leq r^{non}(n;I,F_0). \qquad \square$$

We specify Theorem 5.2 by taking F_0 defined by (5.4) with $r \geq 1$. Then for every positive c, $f_c \in F_0$ and $F_0(c) = \{h \in F_0 : \|h\| \leq c\}$. Hence

$$\frac{1}{2}\, r^{non}(n;I,F_0) \leq \frac{1}{2} \sup_{c>0} r(N_{f_c};I,F_0(c)) \leq r(N;S_2,F_0).$$

This means that for the (S_2,F_0)-problem we have

$$(5.13) \qquad \frac{1}{2}\, r^{non}(n;I,F_0) \leq r^a(n;S_2,F_0) \leq r^{non}(n;I,F_0).$$

Let N^* be defined by (5.5). Then N^* is nearly optimal also for this problem and

$$(5.14) \qquad r(N^*;S_2,F_0) = \Theta(n^{-r}) = r^a(n;S_2,F_0).$$

Every interpolatory algorithm $\varphi^I \in \Phi(N^*)$ is also nearly optimal.

(iii) Minimum point problems

We considered in (i) and (ii) the problems of approximating the minimal value of f and $|f|$ respectively without constructing points at which these values are attained. We now consider the problem of approximating a point $\alpha = \alpha(f)$ such that $f(\alpha) = S_1(f)$ $(= \min\{f(x) : x \in [0,1]\})$. (We do not consider the problem of approximating $\beta = \beta(f)$ where $f(\beta) = S_2(f)$ since they are similar.)

For $f \in F_0$ let

$$(5.15) \qquad P(f) = \{\alpha \in [0,1] : f(\alpha) = S_1(f)\}.$$

Thus, $P(f)$ is the set of all point α for which $f(\alpha)$ is minimal. Our problem is to construct $x = x(f)$ which approximates $P(f)$ in some sense.

Absolute error criterion. Let $\mathrm{dist}(P(f),x) = \inf\{\|x-\alpha\| : \alpha \in P(f)\}$. Suppose we want to construct $x = x(f)$ such that

$$(5.16) \qquad \mathrm{dist}(P(f),x) \text{ is small for every } f \in F_0.$$

In our terminology this is an $(\bar{S}_3,F_0)$-problem with $\bar{S}_3$ defined by

$$(5.17) \qquad \bar{S}_3(f,\delta) = \{x \in \mathbb{R} : \mathrm{dist}(P(f),x) \le \delta\}.$$

Note that this is not a nonlinear problem.

Theorem 5.3: Suppose that $C^{\infty}]0,1] \subset \text{lin}(F_0)$. Then for every information operator N, $N \in \Psi_n^a$,

$$(5.18) \qquad r(N;\bar{S}_3,F_0) = r^a(n;\bar{S}_3,F_0) = \frac{1}{2}. \qquad \square$$

Proof: Take $\varphi^* \in \Phi(N)$, $\varphi^*(N(f)) \equiv 1/2$. Since for every $f \in F_0$, dist $(P(f),1/2) \leq 1/2$ then

$$r(N;\bar{S}_3,F_0) \leq e(\varphi^*,N;\bar{S}_3,F_0) \leq \frac{1}{2}.$$

We now prove that $r(N;\bar{S}_3;F_0) \geq 1/2$. Take an arbitrary algorithm $\varphi \in \Phi(N)$ and $\delta > 0$. Since $C^{\infty}[0,1] \subset \text{lin}(F_0)$ then there exist $h_1,h_2 \in F_0 \cap \ker N_0$ such that $S_1(h_i) < 0$, supp $h_1 \subset [0,\delta]$ and supp $h_2 \subset [1-\delta,1]$. Let $x = \varphi(N(0)) = \varphi(N(h_i))$. Then

$$e(\varphi,N;\bar{S}_3,F_0) \geq \max\{\text{dist}(P(h_1),x),\text{dist}(P(h_2),x)\} \geq \frac{1}{2} - \delta.$$

Since φ and δ are arbitrary then

$$r(N;\bar{S}_3,F_0) \geq \frac{1}{2}.$$

This means that $r(N;\bar{S}_3,F_0) = 1/2$. Since N is arbitrary this completes the proof. $\square$

This theorem states that we cannot approximate any point α at which f is minimal with absolute error less than 1/2.

We now change the error criterion.

Residual error criterion. Suppose we want to construct $x = x(f)$ such that

(5.19) $\quad f(x) - S_1(f)$ is small for every $f \in F_0$.

In our terminology this is an $(\bar{S}_4, F_0)$-problem with $\bar{S}_4$ defined by

(5.20) $\quad \bar{S}_4(f,\delta) = \{x \in [0,1] : f(x) - S_1(f) \le \delta\}.$

This is not a nonlinear problem and we can not apply Lemmas 3.1 or 3.2. However we can give upper and lower bounds on $r(N;\bar{S}_4 F_0)$ using Theorem 5.1. For this purpose we need the following definition.

Let $N = [L_1, L_2, \ldots, L_n] \in \Psi_n^a$, $\delta > 0$ and $\varphi_\delta \in \Phi(N)$ be a δ-optimal algorithm, i.e.,

$$e(\varphi_\delta, N; \bar{S}_4, F_0) \le \delta + r(N; \bar{S}_4, F_0).$$

Let $\bar{N}_\delta \in \Psi_{n+1}^a$,

(5.21) $\quad \bar{N}_\delta(f) = [N(f); f(z)] \quad (= [L_1(f), \ldots, L_n(f; y_1, \ldots, y_{n-1});$
$f(z)])$

where $z = z(f,\delta) = \varphi_\delta(N(f))$.

Theorem 5.4: For every information operator N, $N \in \Psi_n^a$, and $\delta > 0$

$$(5.22) \qquad r(\bar{N}_\delta;S_1,F_0) - \delta \le r(N;\bar{S}_4,F_0) \le 2r(N;I,F_0)$$

and

$$(5.23) \qquad \frac{1}{2} r^{non}(n+1;I,F_0) \le r^a(n;\bar{S}_4,F_0) \le 2r^{non}(n;I,F_0). \quad \square$$

Proof: Let $\bar{\varphi}_\delta(\bar{N}_\delta(f)) \overset{df}{=} f(z)$. Of course, $\bar{\varphi}_\delta \in \Phi(\bar{N}_\delta)$ and

$$r(\bar{N}_\delta;S_1,F_0) \le e(\bar{\varphi}_\delta,\bar{N}_\delta;S_1,F_0) = e(\varphi_\delta,N;\bar{S}_4,F_0)$$

$$\le r(N;\bar{S}_4,F_0) + \delta$$

which proves the left-hand side of (5.22). We now prove the right-hand side. Without loss of generality we can assume that $r(N;I,F_0) < \infty$. For $f \in F_0$ let

$$V(N,f) = \{\tilde{f} \in F_0 : N(\tilde{f}) = N(f)\}.$$

For $x \in [0,1]$ let

$$(5.24) \qquad \underline{\sigma}(x) = \inf_{\tilde{f}\in V(N,f)} \tilde{f}(x), \qquad \bar{\sigma}(x) = \sup_{\tilde{f}\in V(N,f)} \tilde{f}(x).$$

Then $\underline{\sigma}$ and $\bar{\sigma}$ depend on N(f) and

$$(5.25) \qquad \sup_{f\in F_0} \sup_{x\in[0,1]} \frac{1}{2}(\bar{\sigma}(x)-\underline{\sigma}(x)) = r(N;I,f_0)$$

$$= \frac{1}{2} d(N;I,F_0) < +\infty, \qquad \forall x \in [0,1].$$

Hence $\underline{\sigma}(x)$ and $\bar{\sigma}(x)$ are finite for every $x \in [0,1]$. Furthermore

$$(5.26) \qquad \underline{\sigma}(x) \le \tilde{f}(x) \le \bar{\sigma}(x), \quad \forall \tilde{f} \in V(N,f), \quad \forall x \in [0,1].$$

For $\delta > 0$ let β be a point such that

$$(5.27) \qquad \underline{\sigma}(\beta) - \delta \le \inf_{x\in[0,1]} \underline{\sigma}(x).$$

Since β depends only on $N(f)$ and δ then the algorithm

$$(5.28) \qquad \varphi_\delta(N(f)) = \beta$$

is well defined and $\varphi_\delta \in \Phi(N)$. We now prove that

$$(5.29) \qquad e(\varphi_\delta, N; \bar{S}_4, F_0) \le 2r(N;I,F_0) + \delta.$$

Indeed, for $\tilde{f} \in V(N,f)$ let $\tilde{f}(\tilde{\alpha}) = S_1(\tilde{f})$. Then

$$\tilde{f}(\varphi_\delta(N(f))) - S_1(\tilde{f}) = \tilde{f}(\beta) - \tilde{f}(\tilde{\alpha}) \le \bar{\sigma}(\beta) - \underline{\sigma}(\tilde{\alpha})$$

and due to (5.27) and (5.25)

$$f(\varphi_\delta(N(f))) - S_1(\tilde{f}) \le \bar{\sigma}(\beta) - \underline{\sigma}(\beta) + \delta$$

$$\le 2r(N;I,F_0) + \delta.$$

Hence (5.29) is proven. Since δ is arbitrary we get

$$r(N;\bar{S}_4,F_0) \le 2r(N;I,F_0)$$

which proves (5.22). Note that (5.23) follows easily from Theorem 5.1 and (5.22). Hence the proof of Theorem 5.4 is completed.

Let F_0 be defined by (5.4). Then N^* defined by (5.5) is nearly optimal also for this problem and

$$r(N^*;\bar{S}_4,F_0) = \Theta(n^{-r}) = r^a(n;\bar{S}_4,F_0).$$

We end this section by

Remark 5.1: In this section we studied some problems with balanced and convex F_0. This was done only for simplicity. Similar results can be proven for other sets F_0 which are not necessary balanced and convex.

We also assumed that F_0 consists of real functions $f : [0,1] \to \mathbb{R}$. The similar theorems can be proven for a more general setting. For example, let A be a compact subset of a metric space and let F_3 be a linear space with the norm $\|\cdot\|_{F_3}$. Let F_1 be the space of continuous operators (not necessarily linear) $f: A \to F_3$ with the norm $\|f\|_{F_1} = \sup_{a\in A} \|f(a)\|_{F_3}$. Define

$$S_2(f) = \inf_{a\in A} \|f(a)\|_{F_3}.$$

Then the (S_2,F_0)-problem is equivalent to the approximation

(I, F_0)-problem, i.e.,

$$\frac{1}{2}\, r(N_{f_c}; I, F_0(c)) \le r(N; S_2, F_0) \le \sup_{f \in F_0} r(N_f; I, F_0)$$

(compare with Theorem 5.2).

Acknowledgements

I wish to thank J.F. Traub and H. Woźniakowski for their valuable suggestions concerning this paper.

References

[1] C.A. Micchelli and T.J. Rivlin, A survey of optimal recovery, in Optimal Estimation in Approximation Theory (C.A. Micchelli and T.J. Rivlin, eds.), 1-54, Plenum Press, New York, 1977.

[2] C.A. Micchelli, T.J. Rivlin and S. Winograd, The optimal recovery of smooth functions, Numer. Math. 26, 1976, 191-200.

[3] J.F. Traub, G.W. Wasilkowski and H. Woźniakowski, Information, Uncertainty, Complexity, Addison-Wesley, Reding Mass., 1983.

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