IBM Research

RC 5729
(#24739)
11/21/75
Mathematics

31 pages

OPTIMAL ESTIMATION OF LINEAR FUNCTIONALS

Charles A. Micchelli
IBM Thomas J. Watson Research Center
Mathematical Sciences Department
Yorktown Heights, New York 10598

RC 5729
(#24739)
11/21/75
Mathematics

20 pages

OPTIMAL ESTIMATION OF LINEAR FUNCTIONALS

Charles A. Micchelli
IBM Thomas J. Watson Research Center
Mathematical Sciences Department
Yorktown Heights, New York 10598

ABSTRACT: In this paper we deal with the problem of estimating a linear functional, Nx, when given limited information on x. We show, in a very general setting, that there always exists a linear algorithm for computing Nx which yields the least possible error relative to the given information on x. Thus the search for nonlinear algorithms will not yield smaller errors than those which are achievable by linear algorithms.

A number of special examples are also discussed.

Section 1. Introduction

Let X be a linear space and N a real-valued functional on X. In this paper we will be concerned with estimating Nx, given limited information on x.

Perhaps the most convenient model for the description of the problem we will study is the numerical computation of Nx. Thus we may view X as a function space and Nx as some functional of x which we wish to compute.

We desire an "algorithm" for computing Nx which uses only certain information on x. For instance, if Nx is the (unique) zero of the function x then the secant rule is an algorithm for computing Nx which uses only function values of x, while Newton's method requires the derivative of x as well. We define a mapping I from X onto some other linear space Y and interpret Ix as the admissible information obtained from x.

In practical computations, Ix is frequently known only with some error, either introduced in the computation of Ix or a result of errors in the tabulation of the data Ix. To measure this error we require that $Y(= (Y, ||\cdot||)$ be a normed linear space. Thus in the computation of Nx we actually know Iz where $||Ix-Iz|| \le \epsilon$, $z \in X$ and ϵ is some given nonnegative number.

We interpret an algorithm for computing Nx as a mapping A from Y into the real numbers R and AIz is the approximation to Nx obtained from this algorithm. An effective algorithm gives a good approximation to Nx when x is in some (hopefully large) subset of elements of X. We let K be a subset of X and search for an optimal algorithm for computing Nx given that $x \in K$.

The error in approximating Nx by A given the above information on x is

$$(1.1) \qquad E(A) = \sup\{|Nx - AIz| \;:\; x \in K,\; z \in X,\; ||Ix - Iz|| \le \epsilon\}$$

and the minimum error is

$$(1.2) \qquad E = \inf\{E(A) : A{:}Y \to R\}.$$

A_o is called an optimal estimator (algorithm) for N provided that

$$(1.3) \qquad E = E(A_o).$$

In this paper, we will prove that if N and I are linear, $\epsilon > 0$ and K is a convex balanced subset of X then there always exists an optimal estimator for N which is a continuous linear functional on Y. Thus in a very general setting the search for "nonlinear" algorithms will not yield a smaller error than linear algorithms.

Section 2. Optimal Linear Estimators

We make the following assumptions on N, I and K.

(2.1) I is a linear operator from X onto Y, K is a balanced convex subset of X and N a linear functional on X.

Also, to insure the "approximability" of N on K relative to I we demand that

(2.2) $E(A) < \infty$ for some $A \in Y^*$ (Y^* is the norm dual of Y).

Existence of optimal linear estimator

Theorem 2.1. If $\epsilon > 0$ then there exists a continuous linear functional L on Y such that L is an optimal estimator for N.

Proof. We base the proof of this theorem on the following version of the "minmax theorem", cf. Aubin, [1].

Theorem A. Let S, T be sets with S a closed subset of a topological space U. Let ϕ be a real-valued function defined on S×T such that

$$\inf_{s \in S} \sup_{t \in T} \phi(s,t) < \infty.$$

Assume that the following properties hold.

i) For any $t \in T$, the mapping $s \to \phi(s,t)$ is lower semi-continuous.

ii) ϕ is convex with respect to s and concave with respect to t.

iii) There exists a $t_o \in T$ such that for all $\lambda \in R$, the set

$$\{s \in S: \phi(s,t_o) \le \lambda\} \text{ is relatively compact.}$$

Then there exists an $\bar{s} \in S$ such that

$$\sup_{t\in T} \phi(\bar{s},t) = \sup_{t\in T} \inf_{s\in S} \phi(s,t) .$$

Proof of Theorem 2.1. We define for $L\in Y^*$ and $x\in X$

(2.3) $$\phi(L,x) = Nx - LIx + \epsilon||L||$$

where $||L|| = \sup\{|Ly| : ||y|| \le 1, y\in Y\}$ is the norm of the linear functional L on Y. Also, we define the following two quantities

(2.4) $$E_o = \sup\{|Nx| : ||Ix|| \le \epsilon , x\in K\}$$

and

(2.5) $$E_1 = \inf\{E(A): A\in Y^*\} .$$

We will prove that

$$E_o = E = E_1 .$$

Let us begin by showing that

(2.6) $$E_o \le E \le E_1 .$$

The second inequality follows from the definitions of E and E_1. To prove the first inequality let A be any mapping from Y into R, $x\in K$ and $||Ix|| \le \epsilon$. Then

$$|Nx - AIO| \le E(A)$$

$$|Nx + AIO| \le E(A) .$$

Here we used the fact that if $x\in K$ then $-x\in K$. Thus by the triangle inequality we obtain $E_o \le E$.

Since

$$\inf_{L\in Y^*} \phi(L,x) = \begin{cases} Nx, & \text{if } ||Ix|| \le \epsilon \\ -\infty, & \text{otherwise} \end{cases}$$

we conclude that $\sup_{x\in K} \inf_{L\in Y^*} \phi(L,x) = E_o$.

Furthermore, from the following inequality

$$|Nx - LIz| \le |Nx-LIx| + \epsilon \, ||L|| ,$$

valid for any $x\in K$, $z\in X$ with $||Ix-Iz|| \le \epsilon$ and $L\in Y^*$ we conclude that

(2.7) $\quad E(L) \le \sup_{x\in K} \phi(L,x)$.

Now, if we let $S = U = Y^*$, with the w^*-topology and $T = K$ then i) and ii) of Theorem A are valid for $\phi(L,x)$ while iii) is true with $x = 0$. Thus there exists an $L_o\in Y^*$ such that

$$\sup_{x\in K} \phi(L_o,x) = \sup_{x} \inf_{L} \phi(L,x) = E_o.$$

Hence we conclude from (2.6) and (2.7) that

$$E_o \le E \le E(L_o) \le E_o$$

and the theorem is proven.

Remark 2.1. If X is a set of real-valued functions on some set T and Nx is a linear operator from X into itself. Then, according to Theorem 2.1, $(Nx)(t)$ has an optimal linear estimator $(L_t x)(t)$ for every $t\in T$. Hence $t\to(L_t x)(t)$ is an optimal linear estimator for the operator $t\to(Nx)(t)$.

Let us observe that it was essential in the proof of Theorem 1 that $\epsilon > 0$. This may be seen from the following example.

Example 2.1. Let $K = \{f : f'' \in C[0,1]\ ,\ \max_{0 \le x \le 1} |f''(x)| \le 1\}$, $X = C[0,1]$ and $Nf = f'(0)$, $f \in K$. The linear functional Nf has an extension to a linear functional on X. If we choose $If = f$ then for any $\delta > 0$ we have

$$|f'(0) - \frac{1}{\delta}(f(\delta) - f(0))| \le \frac{\delta}{2} \max_{0 \le x \le 1} |f''(x)|.$$

Hence for $\epsilon = 0$ we have $E_1 = 0$. However, there does not exist any bounded linear functional L_o on X such that $L_o f = Nf$ for $f \in K$.

This example demonstrates that when $\epsilon = 0$ the existence assertion of Theorem 2.1 fails.

According to the proof of Theorem 2.1 any solution of the minimum problem

$$(2.8) \qquad \min_{L \in Y^*} (\sup_{x \in K} |Nx - LIx| + \epsilon \, ||L||)$$

is an optimal linear estimator for Nx.

We record below a useful criteria which characterizes an optimal linear estimator for Nx.

We define the set

$$K(\epsilon) = \{x : x \in K,\ ||Ix|| \le \epsilon\}.$$

Characterization of optimal linear estimator

Lemma 2.1. If $E_o = Nx_o$ where $x_o \in K(\epsilon)$ then L_o is an optimal linear estimator for N if and only if

$$(2.8) \qquad \phi(L_o, x) \leq \phi(L_o, x_o) \leq \phi(L, x_o)$$

for all $L \epsilon Y^*$ and $x \epsilon K$, that is, (L_o, x_o) is a saddle point for $\phi(L,x)$.

The inequalities (2.8) are equivalent to the conditions

$$(2.9) \qquad L_o I x_o = \epsilon ||L_o||$$

and

$$(2.10) \qquad \max_{x \epsilon K} (N - L_o I) x = (N - L_o I) x_o .$$

Furthermore, if L_o is an optimal linear estimator for N then $E_o = Nx_o$ where $x_o \epsilon K(\epsilon)$ if and only if (2.8) is valid.

The proof of this lemma is straightforward and we omit it.

Convolution of convex functions

Theorem 2.1 has a close relation to the following identity.

Here we assume that $|\cdot|_i$, $i = 1,2$ are two semi-norms defined on the linear space X. Thus they have the following properties.

$$|x+y|_i \leq |x|_i + |y|_i, \ i = 1,2,$$
$$|\alpha x|_i \leq |\alpha| \, |x|_i, \ i = 1,2 .$$

We define for any linear functional L on X its norm relative to $|\cdot|_i$, $i = 1,2$ by

$$||L||_i = \sup_{|x|_i \leq 1} |Lx|, \ i = 1,2 .$$

Then for any linear functional N on X

$$(2.11) \qquad \sup_{\substack{|x|_i \leq 1 \\ i=1,2}} |Nx| = \min_{L_1 + L_2 = N} \{||L_1||_1 + ||L_2||_2\}$$

where the minimum is taken over all pairs (L_1, L_2) of linear functionals such that $||L_1||_1 < \infty$, $||L_2||_2 < \infty$ (we assume that there is at least one such pair of linear functionals). The right hand side of (2.11) is the familiar notion of the convolution of two convex functions. If we define $\psi_i(L) = ||L||_i$, $i = 1,2$ then (2.11) becomes

$$\sup_{\substack{|x|_i \le 1 \\ i=1,2}} Nx = (\psi_1 + \psi_2)(N)$$

cf. [3], [4] .

The proof of (2.11) follows as a special case of Theorem 1. We define $K = \{x: |x|_2 \le 1\}$ and $M = \{x: |x|_1 = 0\}$. M is a linear subspace of X. Let Π be the canonical mapping from X onto X/M, defined by $\Pi x = x+M$. X/M becomes a normed linear space by defining $||\Pi x|| = |x|_1$. Now, if we choose $\epsilon = 1$ and $I = \Pi$ in Theorem 2.1 we obtain (2.11).

Equation (2.11) has the following immediate extension to more than two semi-norms,

$$\sup_{\substack{|x|_i \le 1 \\ i=1,\dots,m}} |Nx| = \min_{\sum_{j=1}^{m} L_j = N} \sum_{j=1}^{m} ||L_j||_j .$$

Remark 2.2. If we modify the hypothesis of Theorem 2.1 to allow I to be affine, that is, I preserves convex combinations, then a glance at the proof of Theorem 2.1 reveals that

$$\min_{L} \sup_{x \in K} (Nx - LIx) + \epsilon ||L||$$

$$= \sup\{Nx: x \in K(\epsilon)\},$$

provided that $K(\epsilon)\neq\emptyset$, and $\epsilon>0$. In particular, if I is linear and $x_o\in K$, then for $\epsilon>0$

$$\min_{L} \sup_{x\in K} \left(|Nx-LIx|+LIx_o+\epsilon||L||\right)$$
$$= \sup\{Nx:\ x\in K,\ ||Ix-Ix_o||\leq\epsilon\}\ .$$

Finite dimensional information vector

Perhaps, the most important special case of Theorem 2.1 in practice is the case that Ix is a finite collection of scalar information of x, $Ix = (I_1x,\ldots,I_nx)$. This case, however, excludes some important examples, see examples 2.3, 2.4. Our goal is to prove that Theorem 2.1 remains valid for $\epsilon = 0$ when Y is a finite dimensional linear space.

Corollary 2.1 Let X be a linear space, N a linear functional on X and K a convex balanced absorbing subset of X. Suppose I is a linear operator mapping X onto some finite dimensional normed linear space $(Y,||\cdot||)$. If there exists a linear functional L on Y such that $\sup_{x\in K} |Nx-LIx| < \infty$

then there exists a linear functional L_o on Y such that

$$\sup_{\substack{x\in K\\ Ix=0}} Nx = \min_{A} \{\sup |Nx-AIx| \ :\ A:Y\to R\}$$

$$= \sup_{x\in K} |Nx-L_oIx|\ .$$

The above result is due to Smolyak [9] and a proof of it may also be found in Bakhvalov [2]. We shall prove this result as a corollary of Theorem 2.1.

Proof. We assume without loss of generality that $Y = R^n$ and denote the inner product of vectors $x,y\in Y$ by $x\cdot y$. Let us also define

$$d(\epsilon) = \sup\{Nx : x\in K(\epsilon)\} ,$$

$$h_\epsilon(y) = \sup_{x\in K} Nx - y\cdot Ix + \epsilon||y||$$

and for $\epsilon>0$

$$h_\epsilon(y(\epsilon)) = \min\{h_\epsilon(y) : y\in Y\} .$$

For $\epsilon>0$, Theorem 2.1 implies that $d(\epsilon) = h_\epsilon(y(\epsilon))$.

We begin by proving that $\lim_{\epsilon\to 0^+} d(\epsilon) = d(0)$.

Let $|\cdot|$ be the Minkowski function of K. Thus $|x| = \inf\{c : x\in cK\}$ and $x\in K$ implies $|x|\le 1$. Since Y is finite dimensional there exists a bounded linear operator R from Y into X such that $I(x-RIx) = 0$, $x\in X$. Now, let $x\in K(\epsilon)$ then for $\delta>0$ there is a $c_\delta < |x-RIx|+\delta$ such that $c_\delta^{-1}(x-RIx)\in K$. Hence $N(x-RIx) \le c_\delta d(0)$. There exist constants a,b such that $|RIx|\le a||Ix||$, $|NRIx|\le b||Ix||$. Therefore

$$d(\epsilon) \le b\epsilon + (1+a\epsilon+\delta)\, d(0).$$

Letting $\epsilon \to 0^+$ and $\delta \to 0$ we obtain $\lim_{\epsilon\to 0^+} d(\epsilon) = d(0)$.

Now we show that $\lim_{\epsilon\to o} \min\{h_\epsilon(y):y\in Y\} = \min\{h_o(y):y\in Y\}$. To this end we observe that $\lim_{y\to\infty} h_o(y) = \infty$, since K is absorbing and I maps X onto Y. Thus there exists a $y_o\in Y$ such that $h_o(y_o) = \min\{h_o(y):y\in Y\}$. It is an easy matter to show that $||y(\epsilon)|| \ge ||y_o||$ for all $\epsilon>0$. Therefore there exists a subsequence $\{y(\epsilon')\}$ such that $\lim_{\epsilon'\to 0} y(\epsilon') = \bar{y}$. Now, it easily follows that $\lim_{\epsilon'\to 0} h_{\epsilon'}(y(\epsilon')) = h_o(\bar{y}) = \min\{h_o(y): y\epsilon Y\}$ and we conclude that $\lim_{\epsilon\to 0} \min\{h_\epsilon(y): y\epsilon Y\} = \min\{h_o(y): y\epsilon Y\} = h_o(y_o)$. This completes the proof of the corollary. (Note that $\bar{y}$ is a minimum of $h_o(y)$ which has least norm, that is, $||\bar{y}|| \le ||y^*||$ for any $y^*\epsilon Y$ for which $h_o(y^*) = \min\{h_o(y): y\epsilon Y\}$.

Remark 2.3. As in Remark 2.3 we also have under the hypothesis of corollary 2.1

$$\inf_y \sup |Nx - y\cdot Ix| + y\cdot Ix_o$$

$$= \sup\{Nx: x\epsilon K,\ Ix = Ix_o\}$$

provided that the infimum on the left hand side of the above equation is achieved by some $y\epsilon R^n$.

Examples

Let $Y = \ell_n^\infty$ and $Ix = (I_1x,\ldots,I_nx)$ where $I_ix, i=1,\ldots,n$, are linear functionals on X. Then Theorem 2.1 implies for $\epsilon > 0$

$$(2.12)\qquad \sup_{\substack{x\epsilon K \\ |I_ix|\le\epsilon \\ i=1,\ldots,n}} Nx = \min_{c_1,\ldots,c_n} \left(\sup_{x\epsilon K} |Nx - \sum_{i=1}^n c_iI_ix| + \epsilon \sum_{i=1}^n |c_i|\right)$$

Example 2.1 $X = K = \pi_n$, polynomials of degree $\le$ n. $I_ip = p(x_i)$, $i = 1,\ldots,n+1$, and $Np = \sum_{j=0}^n a_jd_j$ where $p(x) = \sum_{j=0}^n a_jx^j$. Then

$$\sup_{x\epsilon K} (Nx - \sum_{j=1}^{n+1} c_jI_jx) < \infty \quad \text{if and only if } Np = \sum_{j=1}^{n+1} c_jI_jp, \text{ for all } p\epsilon\pi_n.$$

These equations determine $c_1,\dots,c_{n+1}$ uniquely. If $\ell_i \in \pi_n$, $i=1,\dots,n+1$ and $\ell_i(x_j) = \delta_{ij}$ then $c_i = N\ell_i$. Thus (2.11) becomes, after simplification,

$$\max_{|p(x_i)|\le 1} Np = \sum_{i=1}^{n} |N\ell_i| .$$

<u>Example 2.2</u> Let Z be a normed linear space and $x_1,\dots,x_n \in Z$.

If in (2.12) we choose $X = Z^*$, $K = \{L : L\in X, ||L||\le 1\}$ and $I_i L = Lx_i$, $i = 1,\dots,n$, $NL = Lx$ we obtain, since $||c|| = \sum_{i=1}^{n} |c_j|$,

$$\max_{\substack{||L||\le 1 \\ |Lx_i|\le \epsilon \\ i=1,\dots,n}} Lx = \min_{c_1,\dots,c_n} \left(\sup_{||L||\le 1} |Lx - \sum_{j=1}^{n} c_j Lx_j| + \epsilon \sum_{j=1}^{n} |c_j|\right) .$$

The right hand side simplifies and we obtain

$$(2.13)\qquad \max_{\substack{||L||\le 1 \\ |Lx_i|\le \epsilon \\ i=1,\dots,n}} Lx = \min_{c_1,\dots,c_n}\left(||x - \sum_{j=1}^{n} c_j x_j|| + \epsilon \sum_{j=1}^{n} |c_j|\right).$$

Thus when $\epsilon = 0$, (2.13) reduces to the familiar duality formula

$$(2.14)\qquad \max_{\substack{||L||\le 1 \\ L\in M^{\perp}}} Lx = \text{dist}\,(x,M) = \inf_{c_i,\dots,c_n} ||x - \sum_{j=1}^{n} c_j x_j|| .$$

Note that when $\epsilon > 0$ (2.13) depends on a particular choice of basis in M. This distinction is unnecessary in (2.14).

<u>Example 2.3.</u> In this example we will examine Theorem 2.1 when both X and

Y are Hilbert spaces, $X = (X,|\cdot|)$, $Y = (Y,||\cdot||)$ and $K = \{x:|x| \le 1\}$.

Let $[x,x] = |x|^2$, $(y,y) = ||y||^2$, $Nx = [n,x]$, $n \in X$, $n \ne 0$, and suppose I is a bounded linear operator from X onto Y. According to Theorem 2.1

(2.15) $$\max_{\substack{||Ix|| \le \epsilon \\ |x| \le 1}} [n,x] = \min_{y \in Y} \Big(\max_{|x| \le 1} ([n,x] - (y,Ix)) + \epsilon ||y|| \Big) .$$

The adjoint $I^*: Y \to X$ of the mapping I is defined by the relation

(2.16) $$(y,Ix) = [I^* y, x], \quad x \in X,\ y \in Y.$$

Using (2.16), (2.15) becomes

(2.17) $$\max_{\substack{||Ix|| \le \epsilon \\ |x| \le |}} [n,x] = \min_{y \in Y} |n - I^* y| + \epsilon ||y||) .$$

This equation remains valid for $\epsilon = 0$. Moreover, there exists for $\epsilon \ge 0$, an $x(\epsilon) \in X$ $(y(\epsilon) \in Y)$ which solve the maximum (minimum) problem in (2.17). Clearly, $y(\epsilon)$ is unique since it is the minimum of a strictly convex function.

According to Lemma 2.1, $x(\epsilon)$ and $y(\epsilon)$ are determined by the following conditions

(2.18) $$(y(\epsilon), Ix(\epsilon)) = \epsilon ||y(\epsilon)||$$

(2.19) $$\max_{|x| \le 1} [n - I^* y(\epsilon), x] = [n - I^* y(\epsilon), x(\epsilon)]$$

and

(2.20) $$[x(\epsilon), x(\epsilon)] \le 1, \qquad (Ix(\epsilon), Ix(\epsilon)) \le \epsilon^2 .$$

Using the criteria for equality in the CBS inequality we see that (2.18) and (2.19) are equivalent to

(2.21) $$y(\epsilon) = 0 \quad \text{or} \quad Ix(\epsilon) = \frac{y(\epsilon)}{||y(\epsilon)||}$$

$$(2.22) \qquad e \equiv n - I^* y(\epsilon) = 0 \quad \text{or} \quad x(e) = \frac{e}{|e|}$$

It is convenient to analyze the above equations by considering several cases.

Case 1. $y(\epsilon) = 0$ if and only if

$$(2.23) \qquad \epsilon_1^2(n) \equiv \epsilon_1^2 \equiv \frac{(In, In)}{[n,n]} \leq \epsilon^2.$$

Furthermore, when $\epsilon \geq \epsilon_1$ then $x(\epsilon)$ is uniquely given by $x(\epsilon) = \frac{n}{|n|}$.

To prove this claim we first suppose $y(\epsilon) = 0$ then $e = n \neq 0$. Thus (2.22) implies $x(\epsilon) = \frac{n}{|n|}$. Hence (2.20) implies (2.23) is valid. Conversely, if (2.23) holds then $x(\epsilon) = \frac{n}{|n|}$ satisfies (2.20). Hence $E = |n|$ and $y(\epsilon) = 0$, since the minimum problem in (2.18) has a unique solution.

Case 2. $y(\epsilon) \neq 0$ and $n \notin R(I^*)$ (range of I^*).

In this case we have $\epsilon < \epsilon_1$ and $e \neq 0$. Thus we may solve (2.21) and (2.22) for $x(\epsilon)$ and $y(\epsilon)$.

To this end we define the linear operators

$$P_\lambda = (II^* + \lambda 1)^{-1} I$$

and

$$Q_\lambda = 1 - I^* P_\lambda .$$

Since I maps X onto Y the mappings P_λ, Q_λ are well-defined for $\lambda \geq 0$. A simple calculation shows that

$$y(\epsilon) = P_\mu n$$

and

$$x(\epsilon) = \frac{Q_\mu n}{|Q_\mu n|} .$$

The value μ is uniquely determined by the equation

$$g(\mu) = \epsilon$$

where
$$g(\lambda) = \lambda \frac{||P_\lambda n||}{|Q_\lambda n|} .$$

Again, in this case, $x(\epsilon)$ is unique.

Case 3. $n \epsilon R(I^*)$. We let $n = I^* \lambda$ for some $\lambda \epsilon Y$. Now, we claim that $e = 0$ if and only if

$$\epsilon^2 \le \epsilon_o^2(\epsilon) = \epsilon_o^2 \equiv \frac{(\lambda,\lambda)}{((II^*)^{-1}\lambda,\lambda)} \tag{2.24}$$

Clearly, if $e = 0$ then $y(\epsilon) = \lambda$ and $x(\epsilon)$ is any vector in X with $[x(\epsilon),x(\epsilon)] \le 1$ and $Ix(\epsilon) = \frac{\epsilon\lambda}{||\lambda||}$. Let γ be any element in Y then

$$(\gamma,\lambda) = (\gamma, Ix(\epsilon)) \frac{||\lambda||}{\epsilon} = [I^*\gamma, x(\epsilon)] \frac{||\lambda||}{\epsilon} .$$

Hence

$$\epsilon \le \frac{|I^*\gamma| \; ||\lambda||}{(\gamma\cdot\lambda)} . \tag{2.25}$$

If we choose $\gamma = (II^*)^{-1}\lambda$ we obtain the desired conclusion.

Conversely, if (2.24) is valid then (2.25) also holds for all $\gamma \epsilon Y$ since

$$\min_{\gamma \epsilon Y} \frac{|I^*\gamma|}{\lambda\cdot\gamma} = \frac{\epsilon_o}{||\lambda||} .$$

Since $||\lambda|| - ||\lambda-\gamma|| \le \frac{(\lambda,\gamma)}{||\lambda||}$ we conclude that

$$\epsilon||\lambda|| \le |I^*\gamma| + \epsilon||\lambda-\gamma|| \quad , \quad \gamma\epsilon Y .$$

Equivalently, $\min_y (|n-I^*y|| + \epsilon||y||) = \epsilon||\lambda||$.

Hence $y(\epsilon) = \lambda$ and therefore $e = 0$.

As a result of this analysis we also obtain the inequality

$$\epsilon_o(\lambda) < \epsilon_1(I^*\lambda)$$

since both $e = 0$ and $y(\epsilon) = 0$ cannot occur unless $n = 0$, $\lambda = 0$.

Thus, we see that when $n \epsilon R(I^*)$ three cases may occur.

<u>Case 3.1</u> $\epsilon \le \epsilon_o$. Then $y(\epsilon) = \lambda$ and $x(\epsilon)$ is <u>any</u> vector in X such that

$Ix(\epsilon) = \epsilon \dfrac{\lambda}{||\lambda||}$ and $[x(\epsilon), x(\epsilon)] \le 1$.

<u>Case 3.2</u> $\epsilon_o \le \epsilon \le \epsilon_1$. $(x(\epsilon), y(\epsilon))$ are given uniquely by case 2.

<u>Case 3.3</u> $\epsilon_1 \le \epsilon$ $(x(\epsilon), y(\epsilon))$ are given uniquely by case 1.

Let us also note that $\lim_{\lambda \to \infty} g(\lambda) = \epsilon_1$ while

$$\lim_{\lambda \to 0} g(\lambda) = \begin{cases} \epsilon_o(\lambda), & \text{if } n = I^*\lambda \\ 0, & \text{if } n \notin R(I^*). \end{cases}$$

Thus $g(\lambda)$ is a strictly increasing function for $\lambda \ge 0$.

As an example of the above analysis we consider the problem of predicting the value of a function f at time $T>0$, $f(T)$, from its values in the past, $\{f(t)\}_{t<0}$ which we can compute with error ϵ. The case $\epsilon = 0$ of this problem is implicit in the discussion of an example given by Wiener in [11], page 65.

We choose $X = H_1 = \{f: f, f' \epsilon L^2(R)\}$, $[f,f] = \int_{-\infty}^{+\infty} f^2(t)dt + \int_{-\infty}^{+\infty} (f'(t))^2 dt$ and $Y = L^2(-\infty, 0)$, $(f,f) = \int_{-\infty}^{0} f^2(t)dt$. The information operator I is the identity mapping taking X onto Y, given by, $(If)(t) = f(t)$, $t < 0$. Since the space X has a reproducing kernel $K(x,t) = \frac{1}{2}e^{-|x-t|}$ the adjoint of I is easily seen to be

$$(I^*g)\ (t) = \frac{1}{2} \int_{-\infty}^{0} e^{-|t-x|}\, g(x)dx, \qquad t\in R.$$

Also, we may represent the linear functional f(T) as [n,f] where $n(t) = \frac{1}{2}e^{-|T-t|}$, $t\in R$. Since f(T) is not in Y^* we conclude that $n\notin R(I^*)$. Thus the best estimator of f(T) is <u>zero</u> if $\epsilon \ge \epsilon_1 = \frac{1}{2}e^{-T}$, otherwise, we are led to examine the integral equation

$$(2.26) \qquad \frac{1}{2} \int_{-\infty}^{0} e^{-|t-x|}\, y(x)dx + \lambda y(t) = \frac{1}{2}e^{-|T-t|}, \quad t<0, \text{ where } y\in L^2(-\infty,0)$$

and $\lambda>0$. The solution of (2.26) is $y_\lambda(x) = (\sqrt{\lambda^{-1}+1}\ -1)e^{-T}e^{(\sqrt{\lambda^{-1}+1})x}$, $\lambda>0$ and

$$g(\lambda) = \frac{\lambda||y\lambda||}{|n-I^*y_\lambda|}$$

$$= \frac{1}{(\sqrt{\lambda^{-1}+1}\ +1)\sqrt{(\sqrt{\lambda^{-1}+1}\ -1)\ (e^{2T}-1) + e^{2T}}}.$$

Thus the best estimator of f(T) when $\epsilon < \frac{1}{2}e^{-T}$ using the information $\{f(t)\}_{t<0}$ is

$$(\beta-1) \int_{-\infty}^{0} e^{-T}\ e^{\beta x}\ f(x)\ dx$$

where β is uniquely determined by the equation

$$(\beta+1)^2[(\beta-1)\ (e^{2T}-1) + e^{2T}] = \epsilon^{-2}.$$

<u>Example 2.4</u>. In this example we are concerned with the best estimate for the derivative of a smooth function.

Specifically, let $W^{2m}(R) = \{f:\ f^{(2m-1)}$ absolutely continuous on every finite interval, $f^{(2m)}\in L^\infty(R)\}$. Given real numbers $0\le t_1\le\ldots\le t_m$, define

$$(Lf)(x) = \prod_{j=1}^{m} \left(\frac{d^2}{dx^2} - t_j^2\right) f(x).$$

We desire the best estimate for $f'(0)$ using only function values of f, known within error ϵ, and given that $||Lf||_\infty = \text{ess sup } \{|(Lf)(x)| : x \epsilon R\} \leq 1, f \epsilon W^{2m}(R)$. According to Theorem 2.1 the minimum error is

$$E = \sup_{\substack{||f||_\infty \leq \epsilon \\ ||Lf||_\infty \leq 1}} |f'(0)| \tag{2.27}$$

A detailed discussion of this extremal problem is given in [6].

The dual version of (2.27) which is important in determining the best estimator for $f'(0)$ is obtained in the following way. For simplicity, we assume $t_1 > 0$ then

$$f(x) = \int_{-\infty}^{+\infty} \Omega(x-t)(Lf)(t)dt$$

where

$$\Omega(t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \frac{e^{itx}}{p_{2m}(ix)} \, dx$$

$$p_{2m}(x) = \prod_{j=1}^{m} (x^2 - t_j^2)$$

for any $f \epsilon W^{2m}(R)$ such that $f(x) = 0(e^{t_1|x|})$, $x \to \pm\infty$. Thus

$$E = \min_{\alpha} \left\{ \int_{-\infty}^{+\infty} \left| \Omega'(-t) - \int_{-\infty}^{+\infty} \Omega(x-t) d\alpha(x) \right| dt + \epsilon \int_{-\infty}^{+\infty} |d\alpha(x)| \right\} \tag{2.28}$$

where the minimum is taken over all measures of finite total variation. We showed in [6] that for ϵ less than some ϵ_o (2.28) has a solution of the form $d\alpha_o(t) = \sum_{-\infty}^{+\infty} c_\mu \, \delta(t - \mu h + \frac{h}{2})$ where h is some function of ϵ. Thus when $\epsilon < \epsilon_o$ the best estimator for $f'(0)$ is $\int_{-\infty}^{+\infty} f(t) d\alpha_o(t)$ while for $\epsilon \geq \epsilon_o$

we showed that the best estimator for $f'(0)$ is zero.

Example 2.5 Suppose $0 \le x_1 < \ldots < x_{n+r} \le 1$ and $W_\infty^n = \{f : f^{(n-1)}$ absolutely continuous on $[0,1]$, $f^{(n)} \in L^\infty[0,1]\}$. We wish to obtain the best estimator for $f(x)$ given the function values $f(x_1), \ldots, f(x_{n+r})$ which we assume are known exactly and $f \in W^n$, $||f^{(n)}||_\infty = \text{ess sup}\, \{|f^{(n)}(x)| : 0 \le x \le 1\} \le 1$. According to corollary 2.1 the minimum error is given by

$$E = \sup_{\substack{||f^{(n)}||_\infty \le 1 \\ f(x_i)=0,\ i=1,\ldots,\ n+r}} |f(x)|$$

$$= \min_{c_1,\ldots,c_{n+r}} \ \sup_{||f^{(n)}||_\infty \le 1} \left| f(x) - \sum_{j=1}^{n+r} c_j f(x_j) \right| .$$

Using Taylor's theorem with remainder the minimum problem reduces to

$$\min_S \int_0^1 |S(t)|\,dt$$

where the minimum is taken over all functions

$$S(t) = \frac{1}{(n-1)!} \{(x-t)_+^{n-1} - \sum_{j=1}^{n+r} c_j (x_j - t)_+^{n-1}\}$$

subject to the boundary condition

$$S^{(i)}(0) = 0,\ i = 0,1,\ldots,n-1,$$

$(x_+^n = x^n,\ x \ge 0,\ \ 0$ otherwise$)$.

For a discussion of the maximum and minimum problem in this example see [5], [7].

REFERENCES

1. Aubin, J.P., Inf-Sup Inequalities, MRC Tech. Summ. Report #1425, 1974, U.S. Army, Univ. of Wisconsin

2. Bakhvalov, N.S., On the Optimality of Linear Methods for Operator Approximation in Convex Classes of Functions. USSR Computational Mathematics and Mathematical Physics, 11, #14, 244-249, 1971.

3. Holmes, R.B., A Course on Optimization and Best Approximation, Lecture Notes in Mathematics, 257, Springer, Berlin-Heidelberg, 1972.

4. Ioffe, A., and Tikhomviov, V.M., Duality of Convex Functions and Extremal Problems, Russian Math. Surveys 23 (1968), 53-124.

5. Micchelli, C.A., Best L^1-Approximation by Weak Chebyshev Systems and the Uniqueness of Interpolating Perfect Splines, IBM Research Report #5388, 1975.

6. Micchelli, C.A., On an Optimal Method for the Numerical Differentiation of Smooth Functions, to appear Journal of Approximation Theory.

7. Micchelli, C.A., Rivlin, T.J., Winograd, S., The Optimal Recovery of Smooth Functions, accepted for publication, Numerische Math.

8. Singer, I., Best Approximation in Normed Linear Spaces by Elements of Linear Subspaces. Springer, Berlin-Heidelberg, 1970.

9. Smolyak, S.A., On an Optimal Restoration of Functions and Functionals of them, Candidate Dissertation, Moscow State University, 1965.

10. Tihomirov, V.M., Best Methods of Approximation and Interpolation of Differentiable Functions in the space $C_{[-1,1]}$, Math. USSR Sbornik, 9 275-289, 1969.

11. Wiener, N., Extrapolation, Interpolation, and Smoothing of Stationary Time Series, the Technology Press of MIT and J. Wiley, New York, 1950.