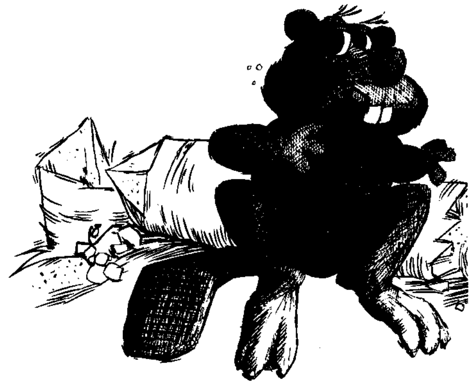
UNIVERSITY OF WATERLOO COMPUTER SCIENCE DEPARTMENT
UNIVERSITY OF WATERLOO COMPUTER SCIENCE DEPARTMENT
UNIVERSITY OF WATERLOO COMPUTER SCIENCE DEPARTMENT
UNIVERSITY OF WATERLOO COMPUTER SCIENCE DEPARTMENT



Optimal Recovery in the Asymptotic Setting

G.M. Trojan

Research Report
CS-87-16

March, 1987

OPTIMAL RECOVERY IN THE ASYMPTOTIC SETTING

G.M. Trojan

Department of Computer Science
University of Waterloo
Waterloo, Ontario N2L 3G1
Canada

ABSTRACT

We prove that the asymptotic setting for linear problems is essentially equivalent to the worst-case setting, as considered in [1]. Algorithms and information operators which are optimal in the worst-case setting lead to the best possible asymptotic convergence. We also prove, under some restrictions, that sequential choice of information does not help for linear problems.

1. Introduction

The objective of this paper is to characterize asymptotic convergence of optimal algorithms. We rely strongly on the results of the long article by Micchelli and Rivlin [1] and references therein (which are not listed here). For the convenience of the reader we recall briefly the principal definitions and results of that fundamental paper.

Let X and Z be normed spaces, U be a linear operator from X into Z (further assumptions we need will be added in Section 2). We want to recover Ux, $x \in X$, given finite information operator

$$I_n(x) = [L_1x, L_2x\ , \ .\ .\ .\ .\ ,\ L_nx], \tag{1.1}$$

where L_k, $k = 1\ , \ .\ .\ .\ .\ ,\ n$ are linear (real) functionals. So we seek an algorithm $\alpha_n : \mathbf{R}^n \to Z$ such that $\alpha_n(I_n(x))$ approximates Ux. The process is schematized in Figure 1.

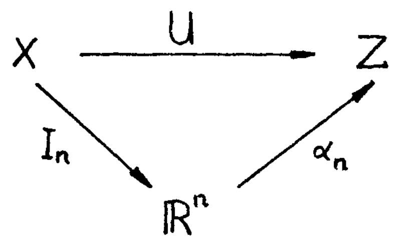


Figure 1

By considering an example of integration, where

$$Ux = \int_0^1 x(t)dt, \quad Z = \mathbf{R}, \quad I_n(x) = [x(t_1), x(t_2), \ldots, x(t_n)] \tag{1.2}$$

it is obvious that one cannot guarantee any finite bound on the error. Some additional information is required. Usually it is assumed that x is in some set B, bounded with respect to some seminorm on X. Then the error of α_n on B is defined by

$$E(\alpha_n, B) = \sup_{x \in B} ||\alpha_n(I_n(x)) - Ux||.$$

Micchelli and Rivlin prove that if B is balanced and convex then there is an intrinsic lower bound on $E(\alpha_n, B)$, namely

$$e(I_n, B) \leq \inf_{\alpha_n} E(\alpha_n, B) \leq 2e(I_n, B) \tag{1.3}$$

where

$$e(I_n, B) \triangleq \sup \{||Ux|| \mid x \in \ker I_n \cap B\}.$$

We have now two problems to solve: find an optimal information operator I_n^*, in view of (1.3) it amounts to minimizing $e(I_n, B)$ and then, given I_n^*, find an optimal algorithm α_n^*, where, by definition

$$E(\alpha_n^*, B) = \inf_{\alpha_n} E(\alpha_n, B).$$

Both problems were widely treated in literature, in particular the problem of finding optimal information operator is closely related to the theory of n-widths.

Two strategies are possible. If the functionals L_k in (1.1) are given a priori then we have nonsequential (simultaneous) information operator. However, one might expect that if we allow L_k depend on the previously computed values $L_1x, \ldots, L_{k-1}x$, $k = 2, \ldots, n$ then we may obtain better results (in our example of integration (1.2) such an approach means that the point t_k depends on $x(t_1), \ldots, x(t_k)$). We call such information operator sequential. It was proven by Gal and Micchelli, although not accepted by those who use adaptive quadrature routines, that it is not the case,

$$\inf_{\textit{sequential } I_n} e(I_n, B) = \inf_{\textit{nonsequential } I_n} e(I_n, B).$$

In our paper we address the same problems, but with a different definition of the error of an algorithm. Namely, instead of restricting the set of admissible x's to B (often the estimate on the norm of x is not available) we admit x to be an arbitrary element from X and let n tend to infinity. That is, we consider sequences of information operators $I = \{I_n\}$ and algorithms $\alpha = \{\alpha_n\}$ (from now on we call α an algorithm) and define the error of α on x as a sequence

$$\{||\alpha_n(I_n(x)) - Ux||\}. \tag{1.4}$$

We look for an algorithm α such that (1.4) converges to zero as fast as possible for all x in X. We call this approach to optimal recovery an **asymptotic setting**.

The main result of this paper, Theorem 2.1 states that (up to some technical details) the bound on the rate of convergence is essentially that of $\{e(I_n, B)\}$; and later we present an algorithm which attains this rate. That is, the sequence of intrinsic errors of the worst-case recovery problem characterizes asymptotic behavior of optimal algorithms.

Section 3 deals with optimal information. We show, in Theorem 3.3 under some restrictions, that **sequential choice of information does not help**, as in the worst-case setting. Finally, in the last section we analyze the optimal rate of convergence with respect to regularity of the class X. We show that more regular X

(smoother functions $x(t)$ in our integration example (1.2)) allows faster convergence of optimal algorithms. This is what one could expect, however such statement is not true for nonlinear operators U. It is shown in [2] that for $Ux = x^{-1}(0)$ (that is, we seek a zero of a function $x(t)$ on an interval) the n-th intrinsic error is of order 2^{-n} and does not depend on the smoothness of $x(t)$.

In the rest of this paper we will assume that U is a continuous operator. Noncontinuous case can also be tackled similarly, see [3], where the ill-posed problem of solving the equation

$$Kx = z,$$

with a compact operator K is considered (that is, $U = K^{-1}$ and this operator is not bounded).

2. Optimal Algorithm

Let X be a Banach space and Z a normed space. B_X is the unit ball in X, U is a linear continuous operator from X into Z. By X^* we denote the dual space to X, X' is a subset of X^*. The (sequential) information $I : X \to \mathbb{R}^\infty$ is defined with the help of a sequence of functions $\beta_n : \mathbb{R}^n \to X'$, $n = 0, 1, 2, \ldots,$ $\mathbb{R}^0 \triangleq \{0\}$. Given $x \in X$ we put

$$L_1 = \beta_0(0), L_2 = \beta_1(L_1x), \ldots, L_n = \beta_{n-1}(L_1x, \ldots, L_{n-1}x), \ldots \tag{2.1}$$

and

$$I(x) = [L_1x, L_2x, \ldots, L_nx, \ldots]. \tag{2.2}$$

By $I_n(x)$ we denote the first n components of $I(x)$, that is

$$I_n(x) = [L_1x, L_2x, \ldots, L_nx].$$

It is worthwhile to distinguish between the operators I_n and $I_{n,x}$. $I_n : X \to \mathbb{R}^n$ is generally nonlinear and not continuous, while $I_{n,x}$ defined by

$$I_{n,x}y = [L_1y, L_2y\,, \ldots, L_ny],$$

where L_k, $k = 1, 2, \ldots, n$ are given by (2.1), is linear and continuous.

If all β_n are constant functions then $I_n \equiv I_{n,x}$ is called nonsequential information operator. By $\Phi(X')$ we denote the family of all information operators using functionals from X'. $\Phi_N(X')$ is the family of all nonsequential information operators.

An algorithm α is a sequence of arbitrary functions $\alpha_n : I_n(X) \to Z$, $n = 1, 2, \ldots$. By $A(I)$ we denote the set of all algorithms using I. The error of an algorithm α at x is a sequence $\{E_n(\alpha, I, x)\}$, where

$$E_n(\alpha, I, x) = ||\alpha_n(I_n(x)) - Ux||.$$

Let

$$e_n(I, X, x) = \sup\,\{||Uy|| \mid y \in \ker I_{n,x} \cap B_X\}. \tag{2.3}$$

We will show that the sequence $\{e_n(I, X, x)\}$ provides an intrinsic lower bound on the error of any $\alpha \in A(I)$. Let $\{\delta_n\}$ be a sequence of positive real numbers, monotonically converging to zero (read: arbitrary slowly) and let

$$M(\alpha, \delta_n) = \{x \in X \mid E_n(\alpha, I, x) = o(\delta_n)e_n(I, X, x)\}. \tag{2.4}$$

Thus $M(\alpha, \delta_n)$ consists of all elements in X such that the error of α at x converges faster than $\{\delta_n e_n(I, X, x)\}$. If $M(\alpha, \delta_n) \neq X$ then there is an $\bar{x} \in X$ such that the error of α at $\bar{x}$ does not converge faster than $\{\delta_n e_n(I, X, \bar{x})\}$. We may then say that $\{e_n(I, X, x)\}$ provides a lower bound on the error of any α in the **asymptotic worst-case** setting. From a practical point of view we would rather like to know how massive, $M(\alpha, \delta_n)$ is in X. We will show that $M(\alpha, \delta_n)$ behaves like a set of the first category in X: a countable union of $M(\alpha, \delta_n)$ cannot be equal to X. Hence $\{e_n(I, X, x)\}$ is also a lower bound on the error in the **asymptotic average-case** setting. To be more precise, let $A_0(I)$ be a countable union of algorithms $\alpha^{(r)}$, $r = 1, 2, \ldots$, using information I. Define

$$M(A_0, \delta_n) \triangleq \bigcup_r M(\alpha^{(r)}, \delta_n) = \{x \in X \mid \exists r \; E_n(\alpha^{(r)}, I, x) = o(\delta_n) e_n(I, X, x)\}. \quad (2.5)$$

Then we have

Theorem 2.1:

For any set of continuous linear functionals X', any information $I \in \Phi(X')$, any sequence of positive real numbers $\{\delta_n\}$, monotonically convergent to zero and any countable family of algorithms $A_0(I)$ the set $M(A_0, \delta_n)$ has empty interior.

Proof: Define

$$C_n^r = \{x \in X \mid E_n(\alpha^{(r)}, I, x) < \delta_n e_n(I, X, x)\}, \quad D_n^r = \bigcap_{m \geq n} C_m^r.$$

Note that

$$M(\alpha^{(r)}, \delta_n) \subset \bigcup_n D_n^r \quad \text{and} \quad D_n^r \subset D_{n+1}^r.$$

Hence

$$M(A_0, \delta_n) = \bigcup_r M(\alpha^{(r)}, \delta_n) \subset \bigcup_r \bigcup_n D_n^r = \bigcup_n \bigcup_{r \leq n} D_n^r.$$

Let

$$G \triangleq \bigcup_m G_m = \bigcup_n \bigcup_{r \leq n} D_n^r,$$

where G_m denotes the m-th term of the sum on the right-hand side. Thus

$$G_m = \{x \in X \mid E_j(\alpha^{(l_m)}, I, x) < \delta_j e_j(I, X, x) \; \forall j \geq k_m\},$$

where

$$0 \leq k_{m+1} - k_m \leq 1, \;\; k_m \to \infty .$$

It suffices to show that G has empty interior. Let $x \in X$ and $\epsilon > 0$. We are going to show that

$$(x + \epsilon B_X) \setminus G \neq \emptyset, \quad (2.6)$$

which is equivalent to our assertion.

Let $M_0 = x + \epsilon B_X$ and m_1 be such that $4\delta_{m_1} \leq \epsilon$. There exists

$$y \in \ker I_{m_1,x}, \quad ||y|| = 3\delta_{m_1}$$

such that

$$||Uy|| \geq 2\delta_{m_1} e_{m_1}(I, X, x).$$

Since

$$I_{m_1}(x) = I_{m_1}(x + y) = I_{m_1}(x - y),$$

from the triangle inequality it follows that there is $x_1 = x + y$ (say) such that

$$E_{m_1}(\alpha^{(l_1)}, I, x_1) \geq ||Uy|| \geq 2\delta_{m_1} e_{m_1}(I, X, x_1).$$

Let $M_1 = x_1 + (\delta_{m_1} B_X \cap \ker I_{m_1,x})$. Obviously M_1 is closed. Moreover, for every $\tilde{x} \in M_1$

$$||U\tilde{x} - Ux_1|| \leq \delta_{m_1} e_{m_1}(I, X, x_1) = \delta_{m_1} e_{m_1}(I, X, \tilde{x})$$

and

$$||\tilde{x} - x|| \leq ||\tilde{x} - x_1|| + ||x_1 - x|| \leq \delta_{m_1} + 3\delta_{m_1} \leq \epsilon,$$

which yield

$$M_1 \subset M_0 \text{ and } M_1 \cap G_1 = \varnothing.$$

Now, let $m_2 \geq m_1 + 1$ (so $m_2 \geq k_2$) be such that $4\delta_{m_2} \leq \delta_{m_1}$. We repeat our construction. There exists $y \in \ker I_{m_2,x_1}$, $||y|| = 3\delta_{m_2}$ such that $||Uy|| \geq 2\delta_{m_2} e_{m_2}(I, x_1)$. As before, for $x_2 = x_1 \pm y$ we have

$$E_{m_2}(\alpha^{(l_2)}, I, x_2) \geq 2\delta_{m_2} e_{m_2}(I, X, x_2).$$

Let $M_2 = x_2 + (\delta_{m_2} B \cap \ker I_{m_2}, x)$. Then, for every $\tilde{x} \in M_2$

$$||U\tilde{x} - Ux_2|| \leq \delta_{m_2} e_{m_2}(I, X, x_2)$$

and

$$||\tilde{x} - x_1|| \leq ||\tilde{x} - x_2|| + ||x_2 - x_1|| \leq \delta_{m_2} + 3\delta_{m_2} \leq \delta_{m_1}$$

which yields

$$M_2 \subset x_1 + \delta_{m_1} B_X \text{ and } M_2 \cap G_2 = \varnothing.$$

Since $\ker I_{m_2,x_2} \subset \ker I_{m_1,x_1}$ we get $M_2 \subset M_1$.

Proceeding this way we obtain a sequence of closed sets $\{M_m\}$ with the properties

$$M_m \subset M_{m-1},\ M_m \cap G_m = \varnothing,\ \operatorname{diam} M_m \leq 2\delta_m,\ m = 1, 2, \ldots.$$

Hence

$$\bigcap_m M_m \cap \bigcup_m G_m = \varnothing.$$

To show (2.6) we apply the Cantor's theorem. □

The assumption that δ_n tend to zero is crucial. Namely, consider the following example

Example 2.1:

Let X be the space of sequences $x = \{x^k\}$ converging to zero with the max norm. We consider the approximation problem, that is $Z = X$, $U = \mathrm{id}$ with non-sequential information I, where

$$I_n(x) = [x^1, x^2, \ldots, x^n].$$

Let

$$\alpha_n(I_n(x)) = [x^1, x^2, \ldots, x^n, 0, \ldots].$$

We have

$$e_n(I, X, x) \equiv \max_{k \geq n+1} \{|x^k| \mid \max_{1 \leq l < \infty} |x^l| = 1\} = 1,$$

whereas

$$E_n(\alpha, I, x) = \max_{k \geq n+1} |x^k| = o(1)$$

for every sequence x. □

We will show now that there are algorithms with error of order $e_n(I, x)$. To this end we choose a constant $c > 1$ and denote by x_n^c any solution of

$$||x_n^c|| \leq c \inf \{||\tilde{x}|| \mid I_{n,x}\, \tilde{x} = I_n(x)\}, \quad I_{n,x}\, x_n^c = I_n(x). \tag{2.7}$$

Then we put

$$\alpha_n^c(I_n(x)) = U x_n^c. \tag{2.8}$$

Note that our definition of α^c is not unique. There are infinitely many algorithms defined by (2.7) and (2.8). However each of them produces asymptotically the same error which is given by the following theorem.

Theorem 2.2:

For any $c > 1$ and x

$$E_n(\alpha^c, I, x) \leq (1 + c)\ ||x||\ e_n(I, X, x).$$

Proof: We have

$$E_n(\alpha^c, I, x) = ||Ux_n^c - Ux|| \leq ||x_n^c - x||\ ||U \frac{x_n^c - x}{||x_n^c - x||}\ || \leq$$

$$\leq ||x - x_n^c|| \sup \{||Uy|| \mid y \in \ker I_{n,x} \cap B\} \leq$$

$$\leq (1 + c)\ ||x||\ e_n(I, X, x).$$

□

The estimate of Theorem 2 is sharp (up to a constant). Namely, let us look at another approximation problem

Example 2.2:

Let X be the space of all bounded sequences with the sup-norm, Z, U, I and α be as in Example 2.1. Note that $\alpha = \alpha^c$ for any $c > 1$.

We have here

$$e_n(I, X, x) = \sup_{k \geq n+1} \{|x^k| \; |\sup_{1 \leq l \infty} |x^l| = 1\} = 1$$

and

$$E_n(\alpha, I, x) = \sup_{k \leq n+1} |x^k|.$$

Contrary to Example 2.1, there are elements in X such that $E_n(\alpha, I, x) = \Theta(1)$. □

3. Optimal information

In this section we try to find optimal information $I \in \Phi(X')$, given X'. The question of optimality is a tricky one: we want to find the fastest convergent sequence $\{e_n(I, X, x)\}$ for every $x \in X$. Since the set of convergent sequences is only partially ordered one can not expect that this problem has a solution in general. We will show however, that under an additional assumption there exists optimal information, moreover, this information is nonsequential.

Let

$$e_n(X') = \inf_{L_1, \ldots, L_n \in X'} \sup \{||Uy|| \mid y \in \bigcap_k \ker L_k \cap B_X\}, \tag{3.1}$$

$e_n(X')$ provides a lower bound (up to a factor 2) for an optimal algorithm in the nonasymptotic setting, [1]. Let L_k^n, $k = 1, \ldots, n$ be such that

$$\sup \{||Uy|| \mid y \in \bigcap_k \ker L_k^n \cap B_X\} \leq Ce_n(X'), \tag{3.2}$$

where the constant C does not depend on n. If we put

$$I_n^* = [L_1^n, L_2^n, \ldots, L_n^n],$$

we may say that $\{I_n^*\}$ is a sequence of **nearly optimal information operators** for the nonasymptotic setting. We say that the sequence $\{I_n^*\}$ is **nested** if

$$I^*_{n+1} = [I_n^*, L_{n+1}^{n+1}] \quad n = 1, 2, \ldots. \tag{3.3}$$

The following theorem follows directly from our definitions

Theorem 3.1:

Assume that there exists a nested sequence of nearly optimal information operators using functionals from X'. Then

(i) for every $I \in \Phi(X')$ and $x \in X$

$$e_n(X') = O(1)\, e_n(I, X, x)$$

(ii) if $I(x) \triangleq [L_1^1 x, L_2^2 x, \ldots, L_n^n x, \ldots]$ then

$$e_n(I, X, x) \equiv \Theta(e_n(X')).$$

□

If we accept the definition of optimal information as given by (i) and (ii) then Theorem 3.1 states a condition for existence of optimal nonsequential information. Since $e_n(I, X, x)$ does not depend on x for $I = \Phi_N(X')$ since now on we will use an abbreviated notation $e_n(I, X)$ whenever it is clear I is nonsequential.

Example 3.1:

Let X be a Hilbert space, $Z = X$, U — a compact linear operator, $X' = X^*$. Then there are sequences of orthonormal eigenvectors $\{x^n\}$ and associated eigenvalues $\{\lambda_n^2\}$, $\lambda_n \to 0$ of the operator U^*U. The optimal information operators for the nonasymptotic case have a form

$$I_n^* = [(x, x^1), (x, x^2), \ldots, (x, x^n)].$$

Hence the sequence $\{I_n^*\}$ is nested, so there exists optimal information $I \in \Phi_N(X^*)$ in the asymptotic setting,

$$I(x) = [(x, x^1), (x, x^2), \ldots, (x, x^n), \ldots]$$

and

$$e_n(I, X) = e_n(X^*) = \lambda_n.$$

□

It may happen that the sequence $\{I_n^*\}$ is not nested, for instance in the integration problem. However, one may still find a nested sequence of nearly optimal information operators, provided that $\{e_n(X')\}$ does not tend to zero too fast.

Theorem 3.2:

Assume that

$$e_{4n}(X') = \Theta(e_n(X')). \tag{3.4}$$

Then there exists $I \in \Phi_N(X')$ such that

$$e_n(I, X) = \Theta(e_n(X')).$$

Proof: Let $\{I_n^*\}$ be any sequence such that (3.2) holds. We define

$$I(x) = [I_1^* x, I_2^* x, I_4 x, \ldots, I_{2^k} x, \ldots].$$

Assume that $2^k - 1 \leq n < 2^{k+1} - 1$, then

$$I_n(x) = [I_1^* x, \ldots, I_{2^k-1}^* x, L_1^{2^k} x, \ldots, L_s^{2^k} x], \quad s = n - 2^k + 1 \leq 2^k.$$

Hence and from (3.2)

$$e_{2^{k+1}}(X') \leq e_n(X') \leq e_n(I, X) \leq C e_{2^k-1}(X').$$

This and (3.4) proves our assertion. □

For many problems of practical interest $e_n(X') = \Theta(n^{-r})$ for some $r \geq 0$, for instance integration problem in Sobolev spaces $W^{r,p}$, see [1]. Then (3.4) is satisfied. Note that (3.4) is not satisfied if $e_n(X') = O(q^n)$, $q < 1$. This is the case of integration of analytic functions.

As a corollary of Theorem 2.1 we now have

Corollary 3.1:

Let there exist a nested sequence of optimal information operators using functionals from X' and let $I \in \Phi(X')$. Then for any sequence of positive numbers $\{\delta_n\}$, monotonically convergent to zero, and every countable family of algorithms $A_0(I) \subset A(I)$ the set $M(A_0, \delta_n)$, defined by (2.4) and (2.5) has empty interior. □

If we restrict our considerations to **continuous algorithms** (those consisting of continuous functions α_n, $n = 1, 2, \ldots$) then it is possible to show, using Baire's category method, that the sets $M(\alpha, \delta_n)$ are of the first category in X. Therefore the result in Corollary 3.1 is not the strongest possible. To get a full analogy with the continuous case we would need to show that the countable union of $M(\alpha^{(r)}, \delta_n)$ cannot be equal to X, where $\alpha^{(r)} \in A(I^{(r)})$, $r = 1, 2, \ldots$ that is we admit a countable family of (possibly different) informations $I^{(r)}$. We will show this under an additional assumption.

Let $\Psi_{\mathbf{N}}(X') \triangleq \{I^{(1)}, I^{(2)}, \ldots\} \subset \Phi(X')$ denote a countable family of informations and

$$A(\Psi_{\mathbf{N}}(X')) \triangleq \{\alpha^{(r)} \mid \alpha^{(r)} \in A(I^{(r)}), \quad r = 1, 2, \ldots\}.$$

We show that sequential choice of information does not, in general, lead to faster convergence.

Theorem 3.3:

Assume that (3.4) holds. Then, for any sets $X' \subset X^*$, $\Psi_{\mathbf{N}}(X')$, $A(\Psi_{\mathbf{N}}(X'))$ and any sequence $\{\delta_n\}$, monotonically converging to zero, the set

$$M(\Psi_{\mathbf{N}}, \delta_n) \triangleq \{x \in X \mid \exists\, r \;\; E_n(I^{(r)}, \alpha^{(r)}, x) = o(\delta_n) e_n(X')\}$$

has empty interior.

Proof: The proof resembles the proof of Theorem 2.1. First, note that (3.4) yields

$$e_{2n}(X') \geq c e_n(X') \quad n = 1, 2, \ldots \tag{3.5}$$

for some constant c, $0 < c < 1$. Define

$$D_n^r \triangleq \{x \in X \mid E_m(I^{(r)}, \alpha^{(r)}, x) < c\, \delta_m\, e_m(X'),\ \forall\ m \geq n\},$$

$$D \triangleq \bigcup_r \bigcup_n D_n^r.$$

Then $M(\Psi_{\mathbf{N}}, \delta_n) \subset D$ and it suffices to show that D does not contain a ball. Let $\{u_k\} = (1, 2, 1, 2, 3, 1, 2, 3, 4, 1, \ldots)$ and $\{m_k\}$ be any sequence that

$$m_k \geq m_1 + m_2 + \cdots + m_{k-1}. \tag{3.6}$$

Since $D_n^r \subset D_{n+1}^r$ for every n and r we get

$$D = \bigcup_k D_{m_k}^{u_k}. \tag{3.7}$$

Let $x \in X$ and ϵ be an arbitrary positive number. There exist m_1 such that $4\delta_{m_1} \leq \epsilon$ and $y \in \ker I_{m_1,x}^{(u_1)} \cap 3\delta_{m_1} B_X$ such that $||Uy|| \geq 2\delta_{m_1} e_{m_1}(I^{(u_1)}, X, x)$. As in the proof of Theorem 2.1 we may find $x_1 = x \pm y$ such that

$$E_{m_1}(I^{(u_1)}, X, x) \geq 2\delta_{m_1} e_{m_1}(I^{(u_1)}, X, x_1).$$

Let $M_1 \triangleq x + (\delta_{m_1} B_X \cap \ker I_{m_1,x_1}^{(u_1)})$. Then, since $e_{m_1}(I^{(u_1)}, X, x_1) \geq e_{m_1}(X')$,

$$M_1 \subset x + \epsilon B_X \ \text{ and } \ M_1 \cap D_{m_1}^{u_1} = \emptyset.$$

We proceed now by induction. For any $k \geq 2$ we choose m_k such that $m_k \geq 2m_{k-1}$ and $\delta_{m_k} \leq \frac{1}{4}\, \delta_{m_k}$ (hence (3.6) is satisfied). Denote by $\bar{I}^{(k)}$ the operator

$$\bar{I}^{(k)} = I^{(u_1)} \cup I^{(u_2)} \cup \cdots \cup I^{(u_k)} = [L_1^{(u_1)}, \ldots, L_{m_1}^{(u_1)}, \ldots, L_1^{(u_k)}, \ldots, L_{m_k}^{(u_k)}].$$

There exists

$$y \in \ker \tilde{I}^{(k)}_{m_1+\ldots+m_k,x} \cap 3\delta^{\,k}_{m} B_X$$

such that

$$||Uy|| \geq 2\delta_{m_k}\, e_{m_1+\ldots+m_k}\,(\tilde{I}^{(k)}, X, x_{k-1})$$

and

$$x_k = x_{k-1} \pm y$$

such that

$$E_{m_k}(I^{(u_k)}, \alpha^{(u_k)}, x_k) \geq 2\delta_{m_k}\, e_{m_1+\ldots+m_k}\,(\tilde{I}^{(k)}, X, x_k). \tag{3.8}$$

Let $M_k + x_k = \delta_{m_k} B_X \cap \ker \tilde{I}^{(k)}_{m_1+\ldots+m_k,x_k}$, then $M_k \subset M_{k-1}$. Since for any $\tilde{x} \in M_k$, $||U\tilde{x} - Ux_k|| \leq \delta_{m_k}\, e_{m_1+\ldots+m_k}\,(\tilde{I}^{(k)}, X, \tilde{x})$, from (3.8) and (3.6) we obtain

$$E_{m_k}(I^{(u_k)}, \alpha^{(u_k)}, \tilde{x}) \geq \delta_{m_k}\, e_{m_1+\ldots+m_k}\,(X') \geq \delta_{m_k}\, e_{2m_k}\,(X') \geq c\delta_{m_k}\, e_{m_k}\,(X').$$

Hence $M_k \cap D^{u_k}_{m_k} = \varnothing$.

We have constructed a sequence of closed sets $\{M_k\}$ with the properties

$$M_{k+1} \subset M_k,\ \ M_k \cap D^{u_k}_{m_k} = \varnothing \ \text{ and } \ \operatorname{diam} M_k \leq 2\delta_{m_k}.$$

Let $M_0 = \bigcap_k M_k$. Then, by Cantor's theorem, $M_0 \neq \varnothing$, moreover

$$M_0 \subset x + \epsilon B_X$$

and, by (3.7),

$$M_0 \cap D = \varnothing.$$

This completes the proof. □

4. Regularity and the rate of convergence

In this section we are concerned with the question: how regularity of the space X affects the rate of convergence of $\{e_n(I, X, x)\}$. Since in the previous section we have proven, provided $\{e_n(X')\}$ does not converge too fast, that sequential information does not help, and we suspect this is true in the general case, we restrict our considerations only to nonsequential informations.

Be begin with a trivial lemma

Lemma 4.1:

Let $I \in \Phi_N(X')$ be given by (2.2). A necessary condition for $e_n(I, X) = o(1)$ is

$$K \triangleq \bigcap_{k=1}^{\infty} \ker L_k \subset \ker U. \tag{4.1}$$

□

Let Y be a subspace of X. We say that Y is **more regular** than X if the canonical embedding $Y \to X$ is compact. A motivation for such definition is provided by the Sobolev embedding theorems. The following theorem generalizes the results of [4] and shows that more regular problems are easier to solve.

Theorem 4.1:

Let $K \subset \ker U$ and Y be more regular than X. Then for any $I \in \Phi_N(X')$

$$e_n(I, Y) \triangleq \sup\{||Uy|| \mid y \in \ker I_n \cap B_Y\} = o(1)e_n(I, X).$$

Proof: For $n = 1, 2, \ldots$ let $x_n \in \ker I_n \cap B_Y$ be such that

$$||Ux_n|| \geq e_n(I, Y)/2. \tag{4.2}$$

We prove first that

$$\mathrm{dist}_X(x_n, K) = o(1).$$

Indeed, if $\{x_n\}$ does not converge to K then, due to compactness, there is a sequence $\{n_k\}$ and an element $x^* \in X\ \ K$ such that $x_{n_k} \to x^*$. On the other hand, for $i \le n_k$ we have $L_i x_{n_k} = 0$, so continuity of L_i yields $L_i x^* = 0$, $i = 1, 2, \ldots,$ that is $x^* \in K$ and we have a contradiction.

Let $x_n' \in K$ be such that $\|x_n - x_n'\| \le 2 \operatorname{dist}_X(x_n, K)$. Then

$$\|x_n - x_n'\| = o(1). \tag{4.3}$$

We define $\bar{x}_n = (x_n - x_n')/\|x_n - x_n'\|$. We have $\|\bar{x}_n\| = 1$, and since $\bar{x}_n \in \ker I_n$,

$$\|U\bar{x}_n\| \le e_n(I, X). \tag{4.4}$$

Now, from (4.1) and (4.2) it follows that

$$\|U\bar{x}_n\| = \|Ux_n\|/\|x_n - x_n'\| \ge \frac{1}{2}\, e_n(I, Y)/\|x_n - x_n'\|.$$

This, (4.3) and (4.4) give

$$e_n(I, Y)/e_n(I, X) \le 2\|x_n - x_n'\| = o(1),$$

which completes the proof. □

On the basis of Lemma 4.1 and Theorem 4.1 we may deduce a sufficient condition for $\{e_n(I, X)\}$ to converge to zero.

Corollary 4.1:

Let $K \subset \ker U$ and let there exist a Banach space X_0 containing X such that

(i) the embedding $X \subset X_0$ is continuous

(ii) U and the functionals L_k, $k = 1, 2, \ldots$ are continuous on X_0.

Then $e_n(I, X) = o(1)$. □

References

[1] C. Micchelli, T.J. Rivlin, Lectures on Optimal Recovery, in Numerical Analysis, Lancaster 1984, Lecture Notes on Mathematics, Vol. 1129, Springer 1985.

[2] K. Sikorski, G.M. Trojan, Asymptotic Optimality of the Bisection Method, Technical Report, Department of Computer Science, Columbia University, 1984.

[3] G.M. Trojan, Asymptotically Optimal Algorithms for Ill-Posed Problems, Technical Report CS-87-15, Department of Computer Science, University of Waterloo, 1987.

[4] A. Werschulz, Does Increased Regularity Lower Complexity?, Math. Comp., 42 (1984), 69-93.