

THE ADAPTION PROBLEM FOR APPROXIMATING LINEAR OPERATORS

MARK A. KON AND ERICH NOVAK

In this note we answer two open questions on the finite-dimensional approximation of linear operators in Banach spaces. The first result establishes bounds on the ratio α of the error of adaptive approximations to the error of nonadaptive approximations of linear operators (see [PW], open problem 1); terms are defined more precisely below. This result is of interest partly because of its connection to questions regarding the error arising in parallel computational solutions of linear problems in infinite dimension (see [TWW], [PW]).

The second result concerns (possibly nonlinear) continuous finite-dimensional approximations of infinite-dimensional linear operators in Banach space (see [KW]). It is shown that such approximations can yield strictly smaller error than even optimal linear ones. This statement has been shown to be false in Hilbert space (cf. [KW]).

We defer discussion of related results to give some precise definitions. Let $S: F \rightarrow G$ be a bounded linear operator from a linear space F to a Banach space G . We wish to evaluate S at an element $f \in F$ (the "problem element"), restricted to lie in a bounded balanced convex subset B of F . The element f is uncertain to the extent that it is specified only by the value of its image $N(f)$ under a finite rank operator N (to be defined below). We induce a norm on F whose unit ball is B .

Let $N: F \rightarrow Y$ be a linear operator (*information operator*), with $Y \equiv \mathbf{R}^n$ finite dimensional. Decompose N into component linear functionals, $N = (l_1, l_2, l_3, \dots, l_n)$. The image Nf represents the (finite-dimensional) information available about the (high or infinite-dimensional) problem element f .

Received by the editors May 30, 1989 and, in revised form, December 19, 1989.
1980 *Mathematics Subject Classification* (1985 Revision). Primary 65J10, 68Q25.

The research of the first author was partially supported by the National Science Foundation.

To compute Sf from Nf we use a (possibly nonlinear) map $\phi: Y \rightarrow G$ (an *algorithm*) defined so that $\phi \circ N$ approximates S optimally.

The *radius of information* of N is the smallest possible error (over all choices of ϕ) in approximating Sf as a function of the information Nf (and not of f directly):

$$R(N) \equiv \inf_{\phi} \sup_{f \in B} \|Sf - \phi \circ Nf\| = \sup_{y \in Y} \text{rad } S(N^{-1}(y) \cap B),$$

where rad denotes radius of a set in Banach space.

The linear information operator N consists of n linear functionals l_i which are fixed *a priori* and do not depend on a particular problem element f . Consider therefore a more general information operator $N^*: F \rightarrow Y$, defined by

$$N^*(f) = (l_1^*(f), l_2^*(f), \dots, l_n^*(f)),$$

where l_i^* are linear functionals that are allowed to be chosen adaptively, so that each functional l_i^* depends on the already computed values $l_1^*(f), l_2^*(f), \dots, l_{i-1}^*(f)$. Precisely, let

$$l_2^*(f) = l_2^*(f, y_1),$$

where the dependence on f above is still linear, but *arbitrary* dependence on $y_1 \equiv l_1^*(f)$ is allowed. In general, let

$$l_i^*(f) = l_i^*(f, y_{i-1}, y_{i-2}, \dots, y_1), \\ y_i \equiv l_i^*(f);$$

the dependence of l on f is linear, and the dependence on y_{i-1}, \dots, y_1 arbitrary.

Characteristic of N^* is the fact that its elements must be calculated in sequence (i.e., $l_i(f)$ cannot be calculated until $l_1(f), \dots, l_{i-1}(f)$ have been calculated). This way of obtaining information is *adaptive*, and N^* is an adaptive information operator. Correspondingly, N is *nonadaptive*.

An important question, related to questions of implementing parallel computations to "solve the problem" S (i.e. to obtain good finite-dimensional approximation $\phi \circ N$ of S), is whether the radius of information (uncertainty) is decreased if we replace N above by a properly chosen element N^* from the more general class \mathcal{N}_n^* of adaptive information operators with range $Y = \mathbf{R}^n$. If the answer is yes, then it may in some cases be better to use N^* , even though parallel methods cannot be used to calculate it

(since each linear functional in N^* explicitly depends on previous ones.)

Fix $S: F \rightarrow G$ and the dimension n , and define \mathcal{N}_n to be the set of all nonadaptive information operators $N: F \rightarrow Y = \mathbf{R}^n$. Let \mathcal{N}_n^* denote the corresponding set of adaptive information operators. For the following definition, we set $0/0 = 1$.

We define the ratio

$$\alpha_{S,n} \equiv \inf_{N \in \mathcal{N}_n^*} R(N) / \inf_{N \in \mathcal{N}_n} R(N),$$

and take the infimum over all linear operators (problems) S between all pairs of Banach spaces F and G , and over all information cardinalities (dimensions) n :

$$\alpha \equiv \inf_{S,n} \alpha_{S,n}.$$

If $\alpha = 1$ then adaption does not help, and if $\alpha < 1$ then adaption can help to reduce the radius of information, by a factor α . Let α_2 be the corresponding infimum over all operators S whose range is a Hilbert space.

The problem of whether adaptive information helps in the evaluation of operators on Banach spaces (i.e. whether $\alpha < 1$) was solved for the case of S a linear functional in 1971 by Bakhvalov [B], where it was shown that $\alpha = 1$ for this (restricted) class. That the ratio of adaptive to nonadaptive radii of information (and hence α) is always between $\frac{1}{2}$ and 1 was later proved by Gal and Micchelli [GM] and by Traub and Woźniakowski (see [TW]). This was a consequence of the fact that the ratio of the *diameters* of the relevant sets is always 1, and that the ratio of the radius to the diameter of a set in a metric space is always between $\frac{1}{2}$ and 1. Kiefer [K] showed that the ratio used to define α is always 1 for a class of linear functionals defined by the integration problem on certain function spaces. Sukharev [S] and Zaliznyak and Ligu [ZL] showed that the ratio is unity for some nonlinear problems (mappings of functions to their maxima). Here we proved more precise bounds on the infimum of this ratio over all linear problems. We have:

Theorem 1. *The bounds $.5 \leq \alpha \leq \sqrt{.8665}$ and $\sqrt{2}/2 \leq \alpha_2 \leq \sqrt{.8665}$ hold.*

A consequence of the fact that the ratio can be less than 1 is that adaption can help for linear problems, i.e. that adaptive

information operators N^* can yield smaller radii of information than the best nonadaptive (linear) ones. This provides an answer to the adaption problem of analytic complexity theory (see, for example [PW], p. 17, Open Problem 1).

The lower bounds in the theorem follow from the results on diameters mentioned above. In particular, the $\sqrt{2}/2$ lower bound for α_2 follows from the fact that in a Hilbert space the ratio of radius to diameter of a set is always less than $\sqrt{2}/2$.

The upper bound follows from the construction of a norm on the domain F (via the definition of a unit ball B) and of a solution operator S , for which the ratio of radii of adaptive and nonadaptive information is at most $\sqrt{.8665}$. This is done explicitly for a four dimensional F . We require the following.

Lemma. *With information operators $N: F \rightarrow Y$ restricted to be of constant rank, the local radius of information $R(N, y) \equiv \text{rad } S(N^{-1}(y) \cap B)$ is a jointly continuous function of (N, y) taken in the uniform operator topology crossed with the topology of \mathbf{R}^n , in the interior of the support of R .*

Consider the problem for which $S = I$, where I is the identity on \mathbf{R}^4 . We identify the domain of S as $F = \mathbf{R}^4$, endowed with the norm whose unit ball is a convex and balanced subset $\tilde{B}_1 \subset F$, to be defined below. We identify the range of S as $G = \mathbf{R}^4$, with the ordinary Euclidean norm.

The unit ball \tilde{B}_1 of F is constructed as follows. Let B denote the convex hull of the six points consisting of

$$\{(1, 0, 0, 0), (0, 1, 0, 0), (0, 0, 1, 0)\} \subset \mathbf{R}^4,$$

and their reflections through the origin. The three-dimensional set B is enlarged to a four-dimensional set $\tilde{B} \subset \mathbf{R}^4$, whose intersection with the hyperplane spanned by the first three coordinates in \mathbf{R}^4 is exactly B . Precisely, \tilde{B} is

$$\tilde{B} = \{(x_1, x_2, x_3, x_4): (x_1, x_2, x_3) \in B \quad \text{and} \\ |x_4| \leq \frac{1}{2}L(x_1, x_2, x_3)\},$$

where $L(x_1, x_2, x_3) = \sqrt{\frac{2}{3}}|x_1 + x_2 + x_3| + \sqrt{\frac{7}{6}}|1 - x_1 - x_2 - x_3|$. Let \hat{n} be the unit vector in the $(1, 1, 1, 0)$ direction. Finally, let $R_1: F \rightarrow F$ denote the symmetric linear transformation hav-

ing one eigenvector parallel to \hat{n} with eigenvalue l , and three other orthogonal eigenvectors, all with unit eigenvalue. (Thus, R_l stretches vectors by a factor l in the \hat{n} direction.) Let $\tilde{B}_l = R_l \tilde{B}$, and F_l denote F endowed with a norm whose unit ball is \tilde{B}_l .

Let $N_l \in \mathcal{N}_2$ be the nonadaptive information operator of rank 2 on F_l which yields the smallest radius of information (error) for the problem S . The lemma and geometric arguments show that N_l has a kernel asymptotically orthogonal to \hat{n} (for l large). This provides the control necessary to explicitly construct N_l . This operator has level sets $N_l^{-1}(y)$ which must be parallel to each other, and whose maximal radius of intersection with \tilde{B}_l (in the Euclidean norm of the image space G) is as small as possible under this constraint (of being parallel).

An adaptive information operator $N_l^* \in \mathcal{N}_2^*$ can then be constructed whose radius of information is smaller than that of N_l . The operator N_l^* , being effectively nonlinear, has level sets which are not necessarily parallel, and whose radii of intersection with \tilde{B}_l can thus be adjusted with more degrees of freedom. Things can thus be arranged so that the maximal radius of intersection of the level sets $(N_l^*)^{-1}(y)$ with \tilde{B}_l is smaller than the largest of the corresponding radii of intersection for N_l . The fact that the adaptive radius (error) is smaller than the smallest nonadaptive one immediately implies that $\alpha < 1$, or that adaption can help (i.e. reduce error). The bound on α is then obtained by taking the ratio of the radii of information for the two (adaptive and nonadaptive) operators above.

We remark that the (four-dimensional) space $F = F_l$ constructed above has the lowest dimension in which adaption can help (so that our choice of dimension is optimal). This is because if adaptive and nonadaptive information are to differ, N must have at least rank 2. Thus if F is three dimensional the sets $N^{-1}(y) \cap \tilde{B}_l$ are one dimensional, so that their diameter is twice their radius. In this case the results of [GM, TW] (on the fact that adaption cannot reduce diameters if information) imply that adaption cannot help to reduce the radius of information.

In light of the above result, it is also of interest to ask whether (possibly nonlinear) continuous information operators N can yield smaller radii of information than the best linear ones [KW]. Precisely (following the above notation), if \mathcal{N}_n^{**} denotes the set of information operators $N: F \rightarrow Y = \mathbf{R}^n$ which are continuous

but not necessarily linear, one can form an analogous ratio

$$\alpha_{S,n}^* \equiv \inf_{N \in \mathcal{N}_n^{**}} R(N) / \inf_{N \in \mathcal{N}_n} R(N),$$

with $\alpha^* \equiv \inf_{S,n} \alpha_{S,n}^*$. Note that \mathcal{N}_n^* (the class of adaptive information operators) is not included in \mathcal{N}_n^{**} . since operators in \mathcal{N}_n^* need not be continuous. In this case, if $\alpha^* < 1$, then continuous information is (at least sometimes) better than (i.e. yields a smaller radius of information than) linear information. It was shown [KW] that $\alpha_2^* = 1$, where α_2^* is the corresponding constant for which the domain F and range G are restricted to be Hilbert spaces. That is, nonlinear continuous information operators do not yield any smaller radii of information than linear ones in the Hilbert case. It was asked there what the corresponding situation is for the general case of Banach spaces. The answer to this question (proved through analogous constructions) is:

Theorem 2. *The constant α^* is strictly smaller than 1. Thus there exist linear operators S for which some nonlinear continuous information operators N have smaller radii of information than any linear ones.*

It was brought to the authors' attention that P. Mathé [M] has also obtained the result of Theorem 2, using a high-dimensional construction.

REFERENCES

- [B] N. S. Bakhvalov, *On the optimality of linear methods for operator approximation in convex classes of functions*, U.S.S.R. Comput. Math. and Math. Phys. **11** (1971), 244–249.
- [GM] S. Gal and C. A. Micchelli, *Optimal sequential and non-sequential procedures for evaluating a functional*, Appl. Anal. **10** (1980), 105–120.
- [KW] B. Z. Kacwicz and G. W. Wasilkowski, *How powerful is continuous nonlinear information for linear problems?* J. Complexity **2** (1986), 306–316.
- [K] J. Kiefer, *Optimum sequential search and approximation methods under minimum regularity assumptions*, SIAM J. **5** (1957), 105–136.
- [M] P. Mathé, *s-numbers in analytic complexity*, Akademie der Wissenschaften der DDR, preprint.
- [PW] E. W. Packel and H. Woźniakowski, *Recent developments in information-based complexity*, Bull. Amer. Math. Soc. **17** (1987), 9–36.
- [S] A. G. Sukharev, *Optimal strategies for the search for an extremum*, U.S.S.R. Comput. Math. and Math. Phys. **11** (1971), 119–137.
- [TW] J. F. Traub and H. Woźniakowski, *A general theory of optimal algorithms*, Academic Press, New York, 1980.

- [TWW] J. F. Traub, G. W. Wasilkowski and H. Woźniakowski, *Information based complexity*, Academic Press, New York, 1988.
- [ZL] N. F. Zaliznyak and A. A. Ligun, *On optimum strategy in search of a global maximum of a function*, U.S.S.R. Comput. Math. and Math. Phys. **18** (1978), 314–321.

DEPARTMENT OF MATHEMATICS, BOSTON UNIVERSITY, BOSTON, MASSACHUSETTS 02215

UNIVERSITÄT ERLANGEN-NÜRNBERG, BISMARCKSTRASSE 1 $\frac{1}{2}$, D-8520 ERLANGEN, WEST GERMANY

