JOURNAL OF APPROXIMATION THEORY 80, 390–408 (1995)

Optimal Recovery and n-Widths for Convex Classes of Functions

ERICH NOVAK

Mathematisches Institut, Universität Erlangen, Bismarckstrasse 1 1/2, 91054 Erlangen, Germany

Communicated by Allan Pinkus

Received April 26, 1993; accepted October 27, 1993

We study the problem of optimal recovery in the case of a nonsymmetric convex class of functions. In particular we show that adaptive methods may be much better than nonadaptive methods. We define certain Gelfand-type widths that are useful for nonsymmetric classes and prove relations to optimal error bounds for adaptive and nonadaptive methods, respectively. © 1995 Academic Press, Inc.

1. INTRODUCTION

No approximation scheme can be good for every function f. We need some a priori information about f of the form $f \in F$. Usually one assumes that f is an element of a certain Banach space X and so might have certain smoothness properties. Then it is our task, for example, to find a good approximation of the linear operator $S: X \to G$ such that

$$\left\| S(f) - \sum_{i=1}^{n} L_i(f) \cdot g_i \right\|_G \leqslant c_n \cdot \|f\|_X$$

holds with as small a c_n as possible. Here the L_i are linear functionals, $L_i: X \to \mathbf{R}$, for example, function values or Fourier coefficients.

Often, we perform a worst case analysis on a unit ball

$$F = \{ f \in X \mid \|f\|_X \leqslant 1 \}$$

which is convex and symmetric, i.e., $-f \in F$ if $f \in F$. This approach is the usual one in numerical analysis, at least if the solution operator is linear. Also, most of the known results on optimal recovery and closely related problems on n-widths usually are studied under the assumption that the set F of problem elements is convex and symmetric. In many cases, however, we have a different type of a priori information. We give some examples.

0021-9045/95 $6.00

Sometimes we know that f is positive because, for example, f is a certain density function. In this case we should consider sets of the type

$$F = \{f \in X \mid \|f\|_X \leqslant 1, f \geqslant 0\}.$$

Observe that such a set is still convex, but not symmetric. In other cases we might know in advance that f is a monotone or convex function. This also leads us to study convex classes of functions that are nonsymmetric. The geometric information given by positivity, monotonicity, or convexity is very important in some cases. If often helps to find an effective numerical method, even if the problem is ill-posed without this information.

Therefore it is usually not a good idea to just ignore the additional information about f. However, it may seem that it is still enough to study symmetric and convex sets—at least modulo some minor details. Let us again consider the case where we want to approximate a linear operator S on F. By taking $F-F$, defined by

$$F - F = \{f_1 - f_2 \mid f_1, f_2 \in F\},$$

we clearly get a symmetric set and for each convex set F we get the error estimate

$$\inf_{S_n} \Delta^F_{\max}(S_n) \leqslant \inf_{S_n} \Delta^{F-F}_{\max}(S_n) \leqslant 4 \inf_{S_n} \Delta^F_{\max}(S_n). \tag{1.1}$$

Here the maximal error is defined in the usual way by

$$\Delta^F_{\max}(S_n) = \sup_{f \in F} \|S(f) - S_n(f)\|$$

and the infimum runs through all methods of the form

$$S_n(f) = \phi(L_1(f), ..., L_n(f)) \tag{1.2}$$

with nonadaptively chosen linear functionals L_i; see Proposition 2. In the symmetric case we know that such nonadaptive methods are almost optimal in the class of all adaptive methods that use n linear functionals.

So we know that optimal error bounds for F and for $F-F$ differ at most by a factor of 4 in the case of nonadaptive methods and adaption does not help (up to a factor of 2) for $F-F$. Therefore we can get much better error bounds on F only if we allow adaptive methods. We will see later that for some linear problems $S: X \to G$ and convex $F \subset X$ adaptive methods actually are much better than nonadaptive ones. This also proves that an inequality such as (1.1) does not hold if we allow adaptive methods.

We mention some of the literature in this subject. In the linear theory, i.e., under the assumption that F is symmetric and convex, the close connection between optimal recovery and n-widths or s-numbers is well known; see Mathé [18], Micchelli and Rivlin [19, 20], Novak [23], Pinkus [29], Traub and Woźniakowski [36], and Traub, Wasilkowski, and Woźniakowski [35]. Useful surveys on n-widths are Pietsch [27], Pinkus [28], and Tikhomirov [34].

In the nonsymmetric case not so much is known. Some of the known n-widths can also be defined in the nonsymmetric case, but there is no theory of diameters in connection with optimal recovery, in particular when adaptive methods are allowed. Some special problems, however, are studied in the literature. The problem of optimal numerical integration of monotone functions was studied by Kiefer [12] and Novak [24]. The knots t_i may be chosen adaptively, i.e., sequentially. Kiefer proved that the best method is given by the trapezoidal rule. Hence we have an affine and nonadaptive algorithm which is optimal. Observe that adaption does not help in this case. This is also known for arbitrary linear $S: F \to \mathbf{R}$ in the case where F is convex and symmetric; see Bakhalov [3].

In the present paper we study the question of whether adaption can help if F is only convex. Also, in some other papers linear problems (such as integration or optimal reconstruction in L_∞-norm) have been studied for certain nonsymmetric convex classes of monotone or convex functions. We mention the papers of Braß [6], Glinkin [8, 9], Novak [25], Petras [26], and Sonnevend [30].

Different nonsymmetric extremal problems in approximation theory were investigated by Babenko [1, 2], Gal and Micchelli [7], Ioffe and Tikhomirov [10], Korneichuk [15], Magaril-Ilyaev and Osipenko [17], Sukharev [31], and Sun [32]. We are mainly interested in the following question, where the worst case setting is studied for linear problems: Can adaption help (much) on a convex class of functions? Much is known about linear problems

$$S: X \to Y,$$

when considered on a symmetric and convex set $F \subset X$ in the worst case. A slight superiority of adaptive methods can be proven in some cases even if F is symmetric; see Kon and Novak [13, 14]. It is well known, however, that adaption cannot help much in that case. Although adaptive methods are widely used, most theoretical results show that adaption does not help under various conditions.

It is known, however, that there are examples of a convex and nonsymmetric set F, where adaption helps considerably; see Novak [25] and Section 4. In this paper we define certain new "Gelfand-type" n-widths that turn out to be important for the study of linear problems on convex domains. We study the connection between these n-widths and problems of optimal recovery.

We believe that it is important to calculate the n-widths for standard classes of nonsymmetric sets, for example, sets of the type

$$\left\{f\colon [0,1]\to \mathbf{R} \;\middle|\; \sum_{i=0}^{k} \|f^{(i)}\|_p \leq 1\right\} \cap \{f\in C^l([0,1]) \mid f^{(l)} \geq 0\}.$$

This would be useful for the construction of efficient algorithms for many practical problems.

In Section 5 we study the case where only methods of the form

$$S_n^{(\mathrm{ad})}(f) = \phi(f(t_1), ..., f(t_n)),$$

with function values instead of general linear functionals, are admissible. In this case adaption can help even more. An artificial example with exponential improvement is presented. A recent example from Korneichuk [16] which also shows that adaption can help is given.

2. Diameters for Nonsymmetric Sets

We want to know whether adaption can help for linear problems on a convex set of functions. We begin slightly more generally and first define certain diameters that are interesting in the case where F is not symmetric.

Let X be a Banach space over $\mathbf{R}$ and let $F\subset X$ be convex. We first assume that F is also symmetric, i.e., $f\in F$ implies $-f\in F$. The Kolmogorov n-width of F in X is given by

$$d_n(F) = \inf_{X_n} \sup_{f\in F} \inf_{g\in X_n} \|f-g\|, \tag{2.1}$$

where the left infimum is taken over all n-dimensional subspaces X_n of X. Similarly, the Gelfand n-width of F is given by

$$d^n(F) = \inf_{U_n} \sup_{f\in F\cap U_n} \|f\|, \tag{2.2}$$

where the infimum is taken over all closed subspaces U_n of X with codimension n. These numbers measure the "thickness" or "massivity" of F.

In the case of arbitrary (in particular: nonsymmetric) sets $F\subset X$ these definitions seem to be inadequate. The widths should be translation-invariant; therefore the Kolmogorov n-width (for arbitrary $F\subset X$) should be given by

$$d_n(F) = \inf_{X_n} \sup_{f\in F} \inf_{g\in X_n} \|f-g\|, \tag{2.3}$$

where X_n runs through all n-dimensional affine subspaces of X; see Tikhomirov [33]. For a convex and symmetric set $F \subset X$, (2.2) can be rewritten as

$$d^n(F) = \tfrac{1}{2} \cdot \inf_{U_n} \operatorname{diam}(F \cap U_n).$$

Here, diam(B) means the diameter of a set B, defined by

$$\operatorname{diam}(B) = \sup_{f, g \in B} \|f - g\|.$$

It is interesting to note that this definition (for symmetric sets) can be extended to arbitrary sets in two different ways. Both of them are interesting—at least if we are thinking of applications in the field of optimal recovery. A "global" variant of the Gelfand width (for arbitrary F) is given by

$$d^n_{\text{glob}}(F) = \tfrac{1}{2} \cdot \inf_{U_n} \sup_{f \in X} \operatorname{diam}(F \cap (U_n + f)) \tag{2.4}$$

(a slightly different notion is defined in Ioffe and Tikhomirov [10]), while a "local" variant is given by

$$d^n_{\text{loc}}(F) = \tfrac{1}{2} \cdot \sup_{f \in X} \inf_{U_n} \operatorname{diam}(F \cap (U_n + f)). \tag{2.5}$$

Both these widths are translation-invariant, and we always have

$$d^n_{\text{loc}}(F) \leqslant d^n_{\text{glob}}(F). \tag{2.6}$$

If F is convex and symmetric, then

$$d^n(F) = d^n_{\text{glob}}(F) = d^n_{\text{loc}}(F). \tag{2.7}$$

The widths defined by (2.3)–(2.5) do not increase if F is replaced by its convex hull. Therefore we can and will assume that F is convex. In this case the sup over X in (2.4) and (2.5) clearly can be replaced by a supremum over F. The global and local Gelfand widths are related to the problem of optimal recovery using nonadaptive or adaptive methods, respectively; see Section 3. Therefore for the adaption problem the following question is interesting. Can the number $d^n_{\text{loc}}(F)$ be much smaller than $d^n_{\text{glob}}(F)$?

It is useful to study the function

$$f \in F \mapsto \tfrac{1}{2} \inf_{U_n} \operatorname{diam}((U_n + f) \cap F).$$

A maximum f^* of that function is called a worst element of F. If, in addition,

$$d^n_{\text{glob}}(F) = \tfrac{1}{2} \inf_{U_n} \operatorname{diam}((U_n + f^*) \cap F),$$

then f^* is called a center of F. In this case we have

$$d^n_{\text{loc}}(F) = d^n_{\text{glob}}(F).$$

If F is convex and symmetric, then 0 is a center of F. This follows from the fact that

$$\operatorname{diam}((U_n + f) \cap F) = \operatorname{diam}((U_n - f) \cap F) \leqslant \operatorname{diam}(U_n \cap F)$$

for such an F. Not every convex set has such a center. It is interesting to know whether every convex set F can be increased slightly such that the bigger set has a center. We will see that this is not the case.

There are many papers and also books on n-widths. Nonsymmetric sets have rarely been studied so far, however. This seems to be related to the fact that for Kolmogorov widths and global Gelfand widths nonsymmetric sets do not yield very interesting results. By this we mean the following. Let $F \subset X$ be convex (and nonsymmetric). Then the symmetric set

$$F - F = \{f_1 - f_2 \mid f_i \in F\}$$

is the smallest symmetric set that "contains" F (more exactly: $F - F$ contains a translation of F). The following result says that the n-widths of F and its symmetrization $F - F$ differ at most by a constant of two.

PROPOSITION 1. *Let $F \subset X$ be convex. Then*

$$d_n(F) \leqslant d_n(F - F) \leqslant 2 d_n(F),$$

where d_n is as given in (2.3), *and*

$$d^n_{\text{glob}}(F) \leqslant d^n(F - F) \leqslant 2 d^n_{\text{glob}}(F). \tag{2.8}$$

Proof. We only give the proof of (2.8). The inequalities for the Kolmogorov widths are even easier to prove. Assume that U_n is a closed subspace of X with codimension n such that

$$\delta = \sup_{f \in X} \operatorname{diam}((F - F) \cap (U_n + f)).$$

For any fixed $f_1 \in F$ we get

$$\delta \geqslant \sup_{f \in X} \operatorname{diam}((F - f_1) \cap (U_n + f)) = \sup_{f \in X} \operatorname{diam}(F \cap (U + f)).$$

This proves $d^n_{\text{glob}}(F) \leqslant d^n(F-F)$. Let U_n be such that

$$\delta = \sup_{f \in X} \operatorname{diam}(F \cap (U_n + f)).$$

Each $f^* \in (F-F) \cap U_n$ can be written as $f^* = f_1 - f_2$ with $f_i \in F$ and also $f_i \in U_n + f$. Because of the assumption we obtain $\|f^*\| \leqslant \operatorname{diam}(F \cap (U_n + f)) \leqslant \delta$. Hence

$$\operatorname{diam}((F-F) \cap U_n) \leqslant 2\delta.$$

Because $F-F$ is symmetric we can conclude that

$$\operatorname{diam}((F-F) \cap (U_n + f)) \leqslant 2\delta$$

for each f and therefore $d^n(F-F) = d^n_{\text{glob}}(F-F) \leqslant 2d^n_{\text{glob}}(F)$. ∎

Such a result does not hold for the local widths. The following example also shows that the local widths can be much smaller than the global widths.

EXAMPLE 1. Let $X = l_\infty$ and

$$F = \left\{ x \in X \mid x_i \geqslant 0, \sum_{i=1}^{\infty} x_i \leqslant 1 \right\}.$$

Then we have

$$d^n_{\text{glob}}(F) = \tfrac{1}{2}$$

for every $n \in \mathbf{N}$ and also $d^n(F-F) = 1$. For the local widths of F, however, we obtain

$$\frac{1}{2n+2} \leqslant d^n_{\text{loc}}(F) \leqslant \frac{1}{n+1}.$$

Proof. Because $\operatorname{diam}(F) = 1$ we certainly have $d^n_{\text{glob}}(F) \leqslant \frac{1}{2}$ and $d^n(F-F) \leqslant 1$ for all $n \in \mathbf{N}$. One can also prove that $d^n(F-F) \geqslant 1$. Because I did not find a reference for this result, I sketch a proof using results and notation of Section 4. There we prove $e^n_{\text{ad}}(S|_F) = \frac{1}{2}$, the statement $e^n_{\text{ad}}(S|_{F-F}) = e^n_{\text{non}}(S|_{F-F}) = 1$ can be proved analogously or even more easily. Proposition 3 then implies $d^n(F-F) \geqslant 1$. Hence we obtain $d^n_{\text{glob}}(F) = \frac{1}{2}$ from (2.8).

Now we study the local widths. Consider the set

$$\{x \in l_\infty \mid x_i \in [0, 1/(n+1)] \text{ for } i = 1, \ldots, n+1, \text{ and } x_i = 0 \text{ for } i > n+1\} \subset F.$$

This set is (up to translation) an $(n+1)$-dimensional ball of diameter $1/(n+1)$, and it is well known that the n-width of such a set is $1/(2n+2)$. It then easily follows that

$$d^n_{\text{loc}}(F) \geqslant \frac{1}{2n+2}.$$

Now fix an $x \in F$. There is a partition $\mathbf{N} = N_1 \cup N_2 \cup \cdots \cup N_n$ with the property that for each k either N_k contains only one number or

$$L_k(x) := \sum_{i \in N_k} x_i \leqslant \frac{2}{n+1}.$$

It is clear that the N_k and the L_k depend on $x \in F$. Assume that $y \in F$ with $L_k(y) = L_k(x)$ for all k. It follows, in particular, that for any i we have $y_i = x_i$ or $y_i \in [0, 2/(n+1)]$. We obtain

$$\text{diam}(F \cap \{y \mid L_k(y) = L_k(x), k = 1, ..., n\}) \leqslant \frac{2}{n+1}$$

and this implies

$$d^n_{\text{loc}}(F) \leqslant \frac{1}{n+1}. \quad \blacksquare$$

Remark. We have seen that the local widths can be much smaller than the global widths. Is there a bound on how much smaller they can be? In other words, is there an inequality of the form

$$d^n_{\text{glob}}(F) \leqslant c_n \cdot d^n_{\text{loc}}(F) \tag{2.9}$$

with a sequence c_n that is independent of F? If this is the case then of course it would be interesting to know the best inequality of the form (2.9). Actually we conjecture that Example 1 is the most extreme example in the sense that $d^n_{\text{glob}}(F) = O(n \cdot d^n_{\text{loc}}(F))$ is always true. This is a deep problem which seems to be related to a conjecture of Mityagin and Henkin [21], see also [4], which is still open.

3. Diameters of Mappings and Optimal Recovery

Now we define the widths of $S|_F$, where $S: X \to Y$ is a continuous linear mapping into a normed space Y. We use the notation

$$d_n(S|_F) = \inf_{Y_n} \sup_{f \in F} \inf_{g \in Y_n} \|S(f) - g\|, \tag{3.1}$$

where Y_n runs over all n-dimensional affine subspaces of Y, for the Kolmogorov widths. The global Gelfand width (for arbitrary $F \subset X$) is given by

$$d^n_{\text{glob}}(S|_F) = \tfrac{1}{2} \cdot \inf_{U_n} \sup_{f \in X} \operatorname{diam}(S(F \cap (U_n + f))), \tag{3.2}$$

while a "local" variant is given by

$$d^n_{\text{loc}}(S|_F) = \tfrac{1}{2} \cdot \sup_{f \in X} \inf_{U_n} \operatorname{diam}(S(F \cap (U_n + f))). \tag{3.3}$$

Here the infimum is taken over closed subspaces U_n of X of codimension at most n. Now an f^* can be called a worst element if the supremum in (3.3) is attained for f^*. If, in addition, the local width equals the global width then f^* is a "center" of F with respect to S. This case is similar to the symmetric case insofar as then adaption can help at most by a factor of two; see Proposition 3. It is useful in what follows to define Bernstein widths. The idea of these widths was already used in the proof of Proposition 1 and in Example 1. Here the definition is

$$b_n(S|_F) = \sup\{r \mid S(F) \text{ contains an } (n+1)\text{-dimensional ball with radius } r\}.$$

Observe that in the case $S = \text{Id}: X \to X$ we obtain

$$s_n(S|_F) = s_n(F),$$

where s_n is one of the widths considered here. This means that the diameters of sets are just special cases of this more general notion.

We study the problem of optimal recovery of $S(f)$ for $f \in F \subset X$, if only (adaptive or nonadaptive) information of the form

$$N(f) = (L_1(f), L_2(f), ..., L_n(f))$$

is available. Each method is of the form $S_n^{\text{ad}} = \phi \circ N$ with some $\phi: \mathbf{R}^n \to Y$ and we want to minimize the maximal error

$$\Delta_{\max}(S_n^{\text{ad}}) = \sup_{f \in F} \|S(f) - S_n^{\text{ad}}(f)\|.$$

In this section we assume that the L_i are arbitrary linear continuous functionals $L_i: X \to \mathbf{R}$. In the adaptive case the choice of L_i may depend on $L_1(f), ..., L_{i-1}(f)$. See, for example, Traub, Wasilkowski, Woźniakowski [35] for the exact definitions and known results. If we consider only methods $S_n = \phi \circ N$ with a fixed information mapping N, then we obtain the radius of N by

$$\text{rad}(N, F) = \inf_{\phi} \Delta_{\max}(\phi \circ N).$$

In connection with Proposition 1 we have the following result.

PROPOSITION 2. *Assume that* $N: X \to \mathbf{R}^n$ *is nonadaptive information. Then we have*

$$\text{rad}(N, F) \leqslant \text{rad}(N, F-F) \leqslant 4 \,\text{rad}(N, F).$$

Proof. This just follows from

$$\begin{aligned}\text{rad}(N, F) &\leqslant \text{rad}(N, F-F) \leqslant \text{diam}(N, F-F)\\ &\leqslant 2\,\text{diam}(N, F) \leqslant 4\,\text{rad}(N, F),\end{aligned}$$

where $\text{diam}(N, F)$ is defined by

$$\text{diam}(N, F) = \sup_{y} \text{diam}\{Sf \mid Nf = y\}.$$

The constant 4 is probably not optimal here. ∎

In the following we want to compare the numbers

$$e^n_{\text{non}}(S|_F) = \inf \Delta_{\max}(S_n)$$

with the numbers

$$e^n_{\text{ad}}(S|_F) = \inf \Delta_{\max}(S_n^{\text{ad}}),$$

where S_n runs through all nonadaptive methods and S_n^{ad} runs through all adaptive methods using information N consisting of n linear functionals. We always assume that F is convex. First we note a connection between these error bounds and the Gelfand widths. We skip the simple proof.

PROPOSITION 3. *Let* $F \subset X$ *be a convex set and let* $S: X \to Y$ *be linear and continuous. Then*

$$\tfrac{1}{2} \cdot d^n_{\text{glob}}(S|_F) \leqslant e^n_{\text{non}}(S|_F) \leqslant d^n_{\text{glob}}(S|_F)$$

and

$$\tfrac{1}{2} \cdot d^n_{\text{loc}}(S|_F) \leqslant e^n_{\text{ad}}(S|_F).$$

Remarks. (a) Assume that F is convex and symmetric. Then the result is well known. Because of (2.7) we have

$$d^n_{\text{glob}}(S|_F) = d^n_{\text{loc}}(S|_F)$$

in this case and it follows that adaption can help only by a factor of two since

$$e^n_{\text{non}}(S|_F) \leqslant d^n_{\text{glob}}(S|_F) = d^n_{\text{loc}}(S|_F) \leqslant 2e^n_{\text{ad}}(S|_F).$$

This is known from [7, 36]. See Kon and Novak [13, 14] and Traub, Wasilkowski, Woźniakowski [35] for further results.

(b) In the next section we will present examples where adaptive methods are much better than nonadaptive ones.

4. On the Adaption Problem

Our next example shows that adaptive methods may be much better than nonadaptive methods. This example was constructed to demonstrate the superiority of adaptive methods. We do not know, whether it is "worst possible" for nonadaptive methods. We mention that a similar example is contained in Novak [25].

EXAMPLE 2. Let $X = l_\infty$ and

$$F = \left\{ x \in X \mid x_i \geqslant 0,\ \sum_{i=1}^{\infty} x_i \leqslant 1,\ x_k \geqslant x_{2k},\ x_k \geqslant x_{2k+1} \right\}.$$

Let e^i be the sequence defined by $e^i_k = \delta_{ik}$. For $m \in \mathbf{N}$ we obtain

$$e^i/m \in F - F, \qquad i = 1, ..., 2^{m-1}.$$

Now we use a result of Kashin [11] (see Pinkus [28]) on the Gelfand numbers of the octahedron $O_n = \{x \in \mathbf{R}^n \mid \sum |x_i| \leqslant 1\}$ in the l_∞-norm, namely

$$d^n(O_{2n}) \asymp 1/\sqrt{n}.$$

Therefore for $m \in \mathbf{N}$ we conclude that

$$d^{2^{m-2}}(F - F) \geqslant \frac{1}{m} \cdot d^{2^{m-2}}(O_{2^{m-1}}) \geqslant \frac{c'}{2^{m/2} m}.$$

From this we easily derive the lower bound

$$d^n_{\text{glob}}(F) \asymp d^n(F-F) \geqslant \frac{c}{\sqrt{n}\log n}$$

for the global Gelfand width of F. For the local widths we clearly have the estimate

$$d^n_{\text{loc}}(F) \leqslant \frac{1}{2n+2},$$

since F is contained in the set which was studied in Example 1. Assume now that we want to reconstruct $x \in F$ in l_∞-norm using (adaptively or nonadaptively) linear functionals as information. That is, $S = \text{Id}$. For the error of optimal nonadaptive methods we have the lower bound

$$e^n_{\text{non}}(S|_F) = \frac{1}{2} d^n_{\text{glob}}(F) \geqslant \frac{c}{\sqrt{n}\log n}.$$

Now we describe an adaptive method which is much better. For simplicity we assume here that $n = 2m-1$ is odd. By δ_i we mean the functional $\delta_i(x) = x_i$. First we describe the functionals L_i which are of the form $L_i = \delta_{l_i}$. Take $L_1 = \delta_1$. Suppose that $L_i(x) = x_{l_i}$ are already computed for $1 \leqslant i \leqslant 2k-1$. Define

$$J_k = \{ j \in \{l_1, ..., l_{2k-1}\} \mid 2j \notin \{l_1, ..., l_{2k-1}\}\}$$

and

$$j_k = \min\left\{ j \in J_k \mid x_j = \max_{l \in J_k} x_l \right\}.$$

Take $L_{2k} = \delta_{2j_k}$ and $L_{2k+1} = \delta_{2j_k+1}$, i.e., $l_{2k} = 2j_k$ and $l_{2k+1} = 2j_k + 1$. We obtain

$$J_{k+1} = J_k \cup \{2j_k, 2j_k+1\} \backslash \{j_k\}$$

and from $x_{j_k} \geqslant x_{2j_k}$ and $x_{j_k} \geqslant x_{2j_k+1}$ we conclude that

$$x_{j_{k+1}} \leqslant x_{j_k}. \tag{4.1}$$

We consider the adaptive information

$$N_n(x) = (L_1(x), ..., L_n(x)) = (x_{l_1}, ..., x_{l_n})$$
$$= (x_1, x_2, x_3, x_{2j_2}, x_{2j_2+1}, ..., x_{2j_{m-1}+1}).$$

The l_i are distinct because, aside from l_1, they come in pairs of the form $2j_k$ and $2j_k+1$, where $2j_k$ is always an index not chosen before.

The set J_k has the following property. Assume that l is any coordinate not contained in $\{l_1, l_2, ..., l_{2k-1}\}$. Then there is a $j \in J_k$ with $x_l \leqslant x_j$. The simple proof of this fact would use induction and a partial order on $\mathbf{N}$, defined by the dyadic expansion. Consider, for example, the case where $l = 86$ is not contained in $\{l_1, l_2, ..., l_{2k-1}\}$. Then, depending on k, exactly one of the indices 1, 2, 5, 10, 21, or 43 is in J_k and the respective coordinate is at least as large as x_l. In particular we obtain $x_l \leqslant x_{j_k}$ and $j_k \notin J_{k+1}$.

Together with (4.1) this means that $x_l \leqslant x_j$ for at least k distinct values of j and we obtain

$$x_l \leqslant \frac{1}{k+1}.$$

So for $n = 2m-1$ we get $x_l \leqslant 1/(m+1) = 2/(n+3)$ for all l different from the $x_{l_1}, ..., x_{l_n}$. Hence we can reconstruct x from the information N_n up to an error of $1/(2m+2) = 1/(n+3)$, i.e., we have found a method with

$$\Delta_{\max}(S_n^{\text{ad}}) \leqslant 1/(n+3).$$

EXAMPLE 1 (*continuation*). In Proposition 3 we obtained only a one-sided estimate of the error of optimal adaptive methods through the local widths. Can we also prove an upper bound for the error of optimal adaptive methods through the local widths? To answer this question it is enough to study $S = \text{Id}: F \to l_\infty$, where F is as in Example 1. We claim that

$$e_{\text{ad}}^n(S|_F) = \tfrac{1}{2}$$

holds for all n. This means that we have an example with $e_{\text{ad}}^n(S|_F) \asymp n \cdot d_{\text{loc}}^n(S|_F)$.

Proof. Because $\text{diam}(F) = 1$ it is enough to show that $e_{\text{ad}}^n(S|_F) \geqslant \frac{1}{2}$. Assume that $N: F \to \mathbf{R}^n$ is some adaptive information. We have to prove that

$$\sup_{z \in \mathbf{R}^n} \text{diam}\{x \in F \mid N(x) = z\} = 1.$$

For each $x \in F$, let

$$N(x) = (L_1(x), ..., L_n(x)),$$

where L_i depends on $L_1, ..., L_{i-1}$ and on the values $L_1(x), ..., L_{i-1}(x)$. The L_i can also be considered as functionals on c_0, the space of convergent sequences, and therefore are of the form

$$L_i(x) = \sum_{j=1}^{\infty} a_j^i x_j \qquad (\text{for } x \in F)$$

with $(a_j^i)_{j \in \mathbb{N}} \in l_1$. Now let j_1 be an index such that

$$|a_{j_1}^1| \geqslant |a_j^1| \qquad \text{for all } j.$$

We can assume that the functional L_1 has the form

$$L_1(x) = x_{j_1} + \sum_{j \neq j_1} a_j^1 x_j \tag{4.2}$$

with $|a_j^1| \leqslant 1$ for all j. Observe that (4.2) is true for all $x \in F := F_0$ because the first functional L_1 is independent of x. Consider the $x \in F_0$ for which $L_1(x) = c_1$ for some small positive c_1 and put

$$F_1 := \{x \in F_0 \mid L_1(x) = c_1\}.$$

For our later analysis it is convenient to assume that

$$c_1 = 3^{n-1} \cdot \delta,$$

where $\delta > 0$ is sufficiently small that

$$2n \cdot 3^{n-1}\delta < 1. \tag{4.3}$$

In this case the set F_1 is not empty. The second functional L_2 is well-defined on F_1 since it only depends on L_1 and c_1. We can assume that this linear functional is of the form $x \mapsto \sum_{j=1}^{\infty} a_j^2 x_j$ with $a_{j_1}^2 = 0$. If $a_{j_1}^2 \neq 0$ then we can consider $L_2 + \alpha L_1$ instead of the original L_2 without essentially changing the information operator. By multiplying with a constant we now even assume that

$$L_2(x) = x_{j_2} + \sum_{j \notin \{j_1, j_2\}} a_j^2 x_j$$

with $|a_j^2| \leqslant 1$ for all j. Let

$$c_2 = 3^{n-2} \cdot \delta$$

and consider the set

$$F_2 := \{x \in F_1 \mid L_2(x) = c_2\}.$$

The set F_2 is nonempty since there is a (unique) $x \in F_2$ whose only nonzero coordinates are x_{j_1} and x_{j_2}. The functional L_3 is well-defined on F_2 since it only depends on L_1, L_2, c_1, and c_2. For general k we define

$$F_k := \{x \in F_{k-1} \mid L_k(x) = c_k\},$$

where L_k is the next functional dependent on $L_1, ..., L_{k-1}$ and $c_1, ..., c_{k-1}$ and c_k is defined by

$$c_k = 3^{n-k} \cdot \delta. \tag{4.4}$$

The set F_k is not empty since there is a (unique) $x \in F_k$ whose only nonzero coordinates are $x_{j_1}, ..., x_{j_k}$.

We can and will assume that each L_k is of the form

$$L_k(x) = x_{j_k} + \sum_{j \notin \{j_1, ..., j_k\}} a_j^k x_j$$

with distinct j_k and $|a_j^k| \leqslant 1$ for all j.

Let k_n be an index for which the coordinates $a_{k_n}^i$ are small for all $i = 1, ..., n$, i.e.,

$$|a_{k_n}^i| < \delta \qquad \text{for} \quad i = 1, ..., n. \tag{4.5}$$

The sequences $(a^i)_j$ are in c_0 and therefore such an index k_n clearly exists. We consider a sequence x with nonzero coordinates $x_{j_1}, ..., x_{j_n}$ and $x_{k_n} = c \in [0, 1]$; i.e., we assume that $x_i = 0$ for all other i. We also assume that

$$L_i(x) = c_i \qquad (i = 1, ..., n).$$

It is easy to see that the sequence $x \in l_\infty$ is uniquely defined by these requirements for each given c. The nonzero coordinates of the sequence x are given by the linear system

$$\begin{aligned} x_{j_n} &= -a_{k_n}^n c + c_n \\ a_{j_n}^{n-1} x_{j_n} + x_{j_{n-1}} &= -a_{k_n}^{n-1} c + c_{n-1} \\ &\cdots \\ a_{j_n}^1 x_{j_n} + a_{j_{n-1}}^1 x_{j_{n-1}} + \cdots + a_{j_2}^1 x_{j_2} + x_{j_1} &= -a_{k_n}^1 c + c_1. \end{aligned}$$

We also have $x_{k_n} = c$ while all other coordinates are zero. By (4.5) and $0 \leqslant c \leqslant 1$ we know that all terms of the form $|a_{k_n}^i c|$ are bounded by δ. Using (4.4), for the unique solution of the system we get the estimates

$$0 < x_{j_n} < 2\delta$$
$$0 < x_{j_{n-1}} < 6\delta$$
$$\cdots$$
$$0 < x_{j_1} < 2 \cdot 3^{n-1}\delta.$$

We obtain

$$\sum_{l=1}^{n} x_{j_l} < 2n \cdot 3^{n-1}\delta$$

and conclude from (4.3) that $x \in F_n$ for all $c = x_{k_n}$ with

$$0 \leqslant c \leqslant 1 - 2n \cdot 3^{n-1}\delta.$$

Thus $\sup_{z \in \mathbf{R}^n} \operatorname{diam}\{x \in F \mid N(x) = z\} \geqslant 1 - 2n \cdot 3^{n-1}\delta$ for each $\delta > 0$ as in (4.3). Letting $\delta \to 0$, we obtain our result. ∎

5. The Case of Restricted Information

In many practical cases, X is a Banach space of functions and only certain linear functionals are available as information. Here we only consider the case where all functionals

$$\delta_x : f \in X \mapsto f(x) \in \mathbf{R}$$

are continuous and form the set of available functionals; i.e., each L_i is of the form $L_i = \delta_{x_i}$. Hence we study methods of the form

$$S_n^{(\mathrm{ad})}(f) = \phi(f(t_1), ..., f(t_n)) \tag{5.1}$$

and define the error bounds

$$\tilde{e}^n_{\mathrm{non}}(S|_F) = \inf_{S_n} \Delta_{\max}(S_n)$$

and

$$\tilde{e}^n_{\mathrm{ad}}(S|_F) = \inf_{S_n^{\mathrm{ad}}} \Delta_{\max}(S_n^{\mathrm{ad}}),$$

where the infimum runs through all nonadaptive or adaptive methods of the form (5.1), respectively. Here we only give some examples and therefore we do not define any new n-widths with a restriction to the type of admissible subspaces. The numbers $\tilde{e}^n_{\mathrm{non}}(S|_F)$ can be estimated from above by the Kolmogorov widths $d_n(F)$, even if S is nonlinear; see Novak [22, 23]. The estimate for the case of a linear functional was also found by Belinskii [5].

The following example shows that adaptive methods may be exponentially better than nonadaptive ones.

EXAMPLE 3. Again we consider $S = \mathrm{Id} : F \to l_\infty$ with

$$F = \left\{ x \in X \mid x_i \geqslant 0, \sum_{i=1}^{\infty} x_i \leqslant 1, x_k \geqslant x_{2k}, x_k \geqslant x_{2k+1} \right\},$$

but now we only allow methods of the form (5.1). We already proved that

$$\tilde{e}^n_{\mathrm{ad}}(S|_F) \leqslant \frac{1}{n+3}$$

and it is not difficult to see that the nonadaptive information

$$N_n(x) = (x_1, x_2, ..., x_n)$$

is optimal among all nonadaptive information operators. It follows, in particular, that

$$\tilde{e}^n_{\mathrm{non}}(S|_F) \asymp \frac{1}{\log(n+1)}.$$

EXAMPLE 4. We mention the following example from Korneichuk [16], which is closer to the standard classes of approximation theory. Consider the reconstruction problem $S = \mathrm{Id} : F \to L_\infty([0, 1])$ for

$$F = \{ f : [0, 1] \to [0, 1] \mid f \text{ monotone and } |f(x) - f(y)| \leqslant |x - y|^\alpha \}$$

with $0 < \alpha < 1$. This problem can be solved adaptively using the bisection method, while nonadaptive methods are worse. Korneichuk [16] proved, more exactly, that

$$\tilde{e}^n_{\mathrm{non}}(S|_F) \asymp n^{-\alpha} \qquad \text{while} \qquad \tilde{e}^n_{\mathrm{ad}}(S|_F) \asymp n^{-1} \log n.$$

ACKNOWLEDGMENTS

The author was supported by a Heisenberg scholarship of the DFG. This paper was written during a visit to the International Computer Science Institute (ICSI), Berkeley, and the University of California at Berkeley. I thank J. Feldman and S. Smale for their invitations. An earlier version of this paper contained a severe mistake. I thank A. Pinkus and an anonymous referee, who found this bug.

REFERENCES

1. V. F. BABENKO, Nonsymmetric extremal problems in approximation theory, *Soviet Math. Dokl.* **27** (1983), 346–349. Russian original, *Dokl. Akad. Nauk SSSR* **269**, No. 3 (1983), 521–524.
2. V. F. BABENKO, The best approximations of the classes of functions preset by means of the continuity modulus, *Ukrain Mat. Zh.* **44**, No. 5 (1992), 579–589. [Russian].
3. N. S. BAKHVALOV, On the optimality of linear methods for operator approximation in convex classes of functions, *USSR Comput. Math. and. Math. Phys.* **11** (1971), 244–249. Russian original, *Zh. Vychisl. Mat. i Mat. Fiz.* **11** (1971), 244–249.
4. W. BAUHARDT, Hilbert-Zahlen von Operatoren in Banachräumen, *Math. Nachr.* **79** (1977), 181–187.
5. E. S. BELINSKII, Estimates for the Kolmogorov diameters of classes of functions with conditions on the mixed difference in the uniform metric, *Math. Notes* **50** (1991), 1198–1200. Russian original, *Mat. Zametki* **50**, No. 5 (1991), 147–149.
6. H. BRAß, Zur Quadraturtheorie konvexer Funktionen, *in* "Numerical Integration" (G. Hämmerlin, Ed.), pp. 34–37, Internat. Schriftenreihe Numer. Math., Vol. 57, Birkhäuser, Basel/Boston, 1982.
7. S. GAL AND C. A. MICCHELLI, Optimal sequential and non-sequential procedures for evaluating a functional, *Appl. Anal.* **10** (1980), 105–120.
8. I. A. GLINKIN, Optimal algorithms for integrating convex functions, *Math. USSR Comput. Math and Math. Phys.* **23**, No. 2 (1983), 6–12. Russian original, *Zh. Vychisl. Mat. i Mat. Fiz.* **23**, No. 2 (1983), 267–277.
9. I. A. GLINKIN, Best quadrature formula in the class of convex functions, *Math. Notes* **35** (1984), 368–374. Russian original, *Mat. Zametki* **35** (1984), 697–707.
10. A. D IOFFE AND V. M. TIKHOMIROV, Duality of convex functions, and extremal problems, *Russian Math. Surveys* **23**, No. 6 (1968), 53–124. Russian original, *Uspekhi Mat. Nauk* **23**, No. 6 (1968), 51–116.
11. B. S. KASHIN, Diameters of some finite-dimensional sets and classes of smooth functions, *Math. USSR-Izv.* **11** (1977), 317–333. Russian original, *Izv. Akad. Nauk SSSR Ser. Math.* **41** (1977), 334–351.
12. J. KIEFER, Optimum sequential search and approximation methods under minimum regularity assumptions, *J. Soc. Indust. Appl. Math.* (1957), 105–136.
13. M. A. KON AND E. NOVAK, On the adaptive and continuous information problems, *J. Complexity* **5** (1989), 345–362.
14. M. A. KON AND E. NOVAK, The adaption problem for approximating linear operators, *Bull. Amer. Math. Soc.* **23** (1990), 159–165.
15. N. KORNEICHUK, "Exact Constants in Approximation Theory," Cambridge Univ. Press, London/New York, 1991. Russian original, Nauka, Moscow, 1987.
16. N. P. KORNEICHUK, Optimization of active algorithms for recovery of monotonic functions from Hölder's class, *J. Complexity* **10** (1994), 265–269.
17. G. G. MAGARIL-ILYAEV AND K. YU. OSIPENKO, Optimal recovery of functionals based on inaccurate data, *Math. Notes* **50**, No. 6 (1991), 1274–1279. Russian original, *Mat. Zametki* **50**, No. 6 (1991), 85–93.
18. P. MATHÉ, s-Numbers in information-based complexity, *J. Complexity* **6** (1990), 41–66.
19. 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.), Plenum, New York, 1977.
20. C. A. MICCHELLI AND T. J. RIVLIN, "Lectures in Optimal Recovery," Lecture Notes in Mathematics, Vol. 1129, Springer-Verlag, Berlin/New York, 1985.

21. B. S. MITYAGIN AND G. M. HENKIN, Inequalities between n-diameters, *in* "Proceedings, Seminar on Functional Analysis, Voronezh, 1963," Vol. 7, pp. 97–103.
22. E. NOVAK, Quadrature and widths, *J. Approx. Theory* **47** (1986), 195–202.
23. E. NOVAK, "Deterministic and Stochastic Error Bounds in Numerical Analysis," Lecture Notes in Mathematics, Vol. 1349, Springer-Verlag, Berlin/New York, 1988.
24. E. NOVAK, Quadrature formulas for monotone functions, *Proc. Amer. Math. Soc.* **115** (1992), 59–68.
25. E. NOVAK, Quadrature formulas for convex classes of functions, *in* "Numerical Integration IV" (H. Braß and G. Hämmerlin, Eds.), pp. 283–296, Internat. Schriftenreihe Numer. Math., Vol. 112, Birkhäuser, Basel/Boston, 1993.
26. K. PETRAS, Quadrature theory of convex functions, *in* "Numerical Integration IV" (H. Braß and G. Hämmerlin, Eds.), pp. 315–329, Internat. Schriftenreihe Numer. Math., Vol. 112, Birkhäuser, Basel/Boston, 1993.
27. A. PIETSCH, "Eigenvalues and s-Numbers," Cambridge Univ. Press, London/New York, 1987.
28. A. PINKUS, "n-Widths in Approximation Theory," Springer-Verlag, Berlin/New York, 1985.
29. A. PINKUS, n-Widths and optimal recovery, *in* "Approximation Theory" (C. de Boor, Ed.), pp. 51–66, Proc. Sympos. Appl. Math., Vol. 36, Amer. Math. Soc., Providence, RI, 1986.
30. G. SONNEVEND, An optimal sequential algorithm for the uniform approximation of convex functions on $[0, 1]^2$, *Appl. Math. Optim.* **10** (1983), 127–142.
31. A. G. SUKHAREV, On the existence of optimal affine methods for approximating linear functionals, *J. Complexity* **2** (1986), 317–322.
32. Y. SUN, "An optimal recovery problem for the derivative on a class of smooth functions on the real line," preprint, Beijing Normal University, 1993.
33. V. M. TIKHOMIROV, Diameters of sets in function spaces and the theory of best approximations, *Russian Math. Surveys* **15**, No. 3 (1960), 75–112. Russian original, *Uspekhi Mat. Nauk* **15**, No. 3 (1969), 81–112.
34. V. M. TIKHOMIROV, Approximation theory, *in* "Encyclopeadia of Mathematical Sciences, Vol. 14," Berlin/New York, 1990. Russian original published by Publisher VINITI, R. V. Gamkrelidze, Ed., Moscow, 1987.
35. J. F. TRAUB, G. W. WASILKOWSKI, AND H. WOŹNIAKOWSKI, "Information-Based Complexity," Academic Press, San Diego, CA, 1988.
36. J. F. TRAUB AND H. WOŹNIAKOWSKI, "A General Theory of Optimal Algorithms," Academic Press, San Diego, CA, 1980.