Optimal Recovery and n-Widths For Convex Classes of Functions

Erich Novak

TR-93-014

March 16, 1993

Abstract

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.

Optimal Recovery and n-Widths For Convex Classes of Functions

Erich Novak

Abstract. 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.

1. Introduction

No approximation scheme can be good for every function f on some domain. 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

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

holds with as small c_n as possible. Here the L_i are linear functionals, $L_i : X \to \mathbf{R}$, for example function values or Fourier coefficients. This means that we make a worst case analysis on the symmetric and convex set

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

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.
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

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

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 extra-information, given by positivity, monotonicity, or convexity, is very important in some cases. It 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

$$\inf_{S_n} \Delta_{\max}^{F}(S_n) \le \inf_{S_n} \Delta_{\max}^{F-F}(S_n) \le 4 \inf_{S_n} \Delta_{\max}^{F}(S_n). \tag{1.1}$$

Here the maximal error (over F or over $F - F$) is defined in the usual way and the infimum runs through all methods of the form

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

with nonadaptively chosen linear functionals L_i, see Proposition 3. 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 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 give some remarks to the literature. 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 (1990), Micchelli and Rivlin (1977, 1985), Novak (1988), Pinkus (1986), Traub and Woźniakowski (1980), and Traub, Wasilkowski and Woźniakowski (1988). Useful surveys on n-widths are Pietsch (1987) and Pinkus (1985).

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 also 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 (1957) and Novak (1992). 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 that F is convex and symmetric, see Bakhvalov (1971). In the present paper we study the question 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 Braß (1982), Glinkin (1983, 1984), Gorenflo and Vessella (1991), Novak (1993), Petras (1993), and Sonnevend (1983).

Different nonsymmetric extremal problems in approximation theory were investigated by Babenko (1983, 1992), Gal and Micchelli (1980), Ioffe and Tikhomirov (1968), Korneichuk (1991), Magaril-Ilyayev and Osipenko (1991), Sukharev (1986), and Sun (1993). 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 studied 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 (1989, 1990). 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 with a convex and nonsymmetric set F, where adaption helps a lot, see Novak (1993) 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

$$\{f : [0,1] \to \mathbf{R} \mid \sum_{i=0}^{k} \|f^{(i)}\|_p \le 1\} \cap \{f \in C^l([0,1]) \mid f^{(l)} \ge 0\}.$$

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

$$S_n^{(\mathrm{ad})}(f) = \phi(f(x_1), \ldots, f(t_n))$$

with function values instead of general linear functionals are admissible. In this case adaption can help even more, we present a rather extreme example. Also a recent result of Korneichuk (1993) belongs to this section and is mentioned there.

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 general 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. First we 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 not adequate. 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. For a convex and symmetric set $F \subset X$ Eq. 2.2 can be written as

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

Here, $\operatorname{diam}(B)$ 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 on applications in the field of optimal recovery. A 'global' variant of the Gelfand width (for arbitrary F) is given by

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

(a slightly different notion is defined in Joffe and Tikhomirov 1968) while a 'local' variant is given by

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

Both these widths are translation-invariant, we always have

$$d^n_{\mathrm{loc}}(F) \le d^n_{\mathrm{glob}}(F), \tag{2.6}$$

if F is convex and symmetric, then

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

The widths defined by Eqs. 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. 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 $d^n_{\mathrm{loc}}(F)$ be much smaller than $d^n_{\mathrm{glob}}(F)$? It is useful to study the function

$$f \in F \mapsto \inf_{U_n} \mathrm{diam}((U+f) \cap F).$$

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

$$d^n_{\mathrm{glob}}(F) = \inf_{U_n} \mathrm{diam}((U+f^*) \cap F),$$

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

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

If F is convex and symmetric, then 0 is a center of 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 are rarely 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

$$d_n(F) \le d_n(F-F) \le 2 d_n(F)$$

and

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

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

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

This proves $d^n_{\text{glob}}(F) \leq d^n(F - F)$. Let U_n 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 $\|f^*\| \leq \operatorname{diam}(F \cap (U_n + f)) \leq \delta$, hence

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

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

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

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

□

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

$$F = \{x \in X \mid x_i \geq 0, \ \sum_{i=1}^{\infty} x_i = 1\}.$$

Then we have

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

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

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

Proof: Because of $\operatorname{diam}(F) = 1$ we certainly have $d^n_{\text{glob}}(F) \leq 1/2$ for all $n \in \mathbf{N}$. It is also well known that $d^n_{\text{glob}}(F-F) = 1$. Therefore we obtain $d^n_{\text{loc}}(F) = 1/2$ from (2.8).

Now we study the local widths. Consider the set

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

This set is (up to translation) a $(n+1)$-dimensional ball of radius $1/(n+1)$, it is well-known that the n-width of such a set is $1/(2n+2)$. This proves that

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

Now fix a $x \in F$. We define a permutation of $\mathbf{N}$ such that

$$x_{i_1} \geq x_{i_2} \geq x_{i_3} \geq \dots .$$

We define L_1 by

$$L_1(y) = y_{i_1} + y_{i_2} + \dots + y_{i_k}$$

such that k is the first number with

$$x_{i_1} + x_{i_2} + \dots + x_{i_k} \geq \frac{1}{n+1}.$$

Similarly we define

$$L_2(y) = y_{i_{k+1}} + y_{i_2} + \dots + y_{i_l}$$

such that l is the first number with

$$x_{i_{k+1}} + x_{i_2} + \dots + x_{i_l} \geq \frac{1}{n+1},$$

and so on. Now we assume that $y, z \in F$ such that $L_i(y) = L_i(z) = L_i(x)$ for $i = 1, \dots, n$. We can conclude that

$$\|y - z\| \leq \frac{2}{n+1}$$

and this implies

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

□

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

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

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 is related to a conjecture of Mityagin and Henkin (1963) which is still open. It would follow from their conjecture that

$$d^n_{\text{glob}}(F) \le (2n+2) \cdot d^n_{\text{loc}}(F).$$

The following proposition contains a weaker result. We do not prove it here because it is a special case of the slightly more general Proposition 4.

Proposition 2. Let $F \subset X$ be a convex set. Then

$$d^n_{\text{glob}}(F) \le 2(n+1)^2 \cdot d^n_{\text{loc}}(F).$$

□

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 following notation, namely

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

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

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

while a 'local' variant is given by

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

Here the infimum is taken over closed subspaces U_n of X of codimension at most n. Now a 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 as the symmetric case insofar as adaption can help at most by a factor of two, see Proposition 5. It is useful to define Bernstein widths, here the definition is

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

Observe that in the case $S = Id : X \to X$ 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

$$N(f) = (L_1(f), L_2(f), \ldots, L_n(f))$$

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

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

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 $L_1(f), \ldots, L_{i-1}(f)$. See, for example, Traub, Wasilkowski, Woźniakowski (1988) 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

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

In connection with Proposition 1 we have the following result.

Proposition 3. Assume that $N : X \to \mathbf{R}^n$ is a nonadaptive information. Then we have

$$\mathrm{rad}(N, F) \le \mathrm{rad}(N, F - F) \le 4\,\mathrm{rad}(N, F).$$

Proof: This just follows from

$$\mathrm{rad}(N, F) \le \mathrm{rad}(N, F - F) \le \mathrm{diam}(N, F - F) \le 2\mathrm{diam}(N, F) \le 4\mathrm{rad}(N, F),$$

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

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

Probably, however, the constant 4 is not optimal here.

□

The next result in particular contains Proposition 2.

Proposition 4. Let $F \subset X$ be a convex set and let $S : X \to Y$ be linear and continuous. Then

$$d^n_{\mathrm{glob}}(S_{|F}) \le 2(n+1)^2 \cdot d^n_{\mathrm{loc}}(S_{|F}).$$

Proof: First we have

$$d^n_{\mathrm{glob}}(S_{|F}) \le d^n(S_{|F-F}).$$

Now we use the well known duality between the Gelfand numbers and the Kolmogorov numbers

$$d^n(S_{|F-F}) = d_n(S'_{|F-F}).$$

Now we use the inequality

$$d_n(S'_{|F-F}) \le (n+1)^2 h_n(S'_{|F-F})$$

of Bauhardt (1977) which is similar as the result of Mityagin and Henkin (1963). The h_n are the so called Hilbert numbers, the smallest s-numbers, see Bauhardt (1977) for details. We have

$$(n+1)^2 h_n(S'_{|F-F}) = (n+1)^2 h_n(S_{|F-F}) \le (n+1)^2 b_n(S_{|F-F})$$

and

$$(n+1)^2 b_n(S_{|F-F}) = 2(n+1)^2\, b_n(S_{|F}) \le 2(n+1)^2\, d^n_{\mathrm{loc}}(S_{|F}).$$

Taking these inequalities together we obtain our claim. Also the Bernstein widths usually are studied only for symmetric sets. It is easy to prove, however, that

$$b_n(S_{|F}) = \frac{1}{2} b_n(S_{|F-F})$$

for any convex set. We also have used the fact

$$b_n(S_{|F}) \le d^n_{\mathrm{loc}}(S_{|F})$$

which is easy to prove and well known, at least in the symmetric case.

□

In the following we want to compare the numbers

$$e^n_{\mathrm{non}}(S_{|F}) = \inf \Delta_{\max}(S_n)$$

with the numbers

$$e^n_{\mathrm{ad}}(S_{|F}) = \inf \Delta_{\max}(S_n^{\mathrm{ad}}),$$

where S_n runs through all nonadaptive methods and S_n^{ad} runs through all adaptive methods using an 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 5. Let $F \subset X$ be a convex set and let $S : X \to Y$ be linear and continuous. Then

$$\frac{1}{2} \cdot d^n_{\mathrm{glob}}(S_{|F}) \le e^n_{\mathrm{non}}(S_{|F}) \le d^n_{\mathrm{glob}}(S_{|F})$$

and

$$\frac{1}{2} \cdot d^n_{\mathrm{loc}}(S_{|F}) \le e^n_{\mathrm{ad}}(S_{|F}).$$

□

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

$$d^n_{\mathrm{glob}}(S_{|F}) = d^n_{\mathrm{loc}}(S_{|F})$$

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

$$e^n_{\mathrm{non}}(S_{|F}) \le d^n_{\mathrm{glob}}(S_{|F}) = d^n_{\mathrm{loc}}(S_{|F}) \le 2e^n_{\mathrm{ad}}(S_{|F}).$$

See Kon, Novak (1989, 1990) and Traub, Wasilkowski, Woźniakowski (1988) for further results.
b) In the next section we will present examples, where adaptive methods are much better than nonadaptive ones. Observe, however, that from Propositions 4 and 5 we easily obtain the following result which says, in a certain sense, that adaption 'does not help too much'. We do not know how the optimal inequality of this type looks like. We think that this is an interesting open question.

Proposition 6. Let $F \subset X$ be a convex set and let $S : X \to Y$ be linear and continuous. Then

$$e^n_{\mathrm{non}}(S_{|F}) \le 4(n+1)^2\, e^n_{\mathrm{ad}}(S_{|F}).$$

□

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, however, whether it is the 'worst possible' example for nonadaptive methods. We mention that a similar example is already contained in Novak (1993).

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

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

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

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

Now we use a known result (see Pinkus 1985) on the Gelfand-numbers of the octahedron $O_n = \{x \in \mathbf{R}^n \mid \sum |x_i| \leq 1\}$ in l_∞-norm, namely

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

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

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

From this we easily derive the lower bound

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

for the global Gelfand width of F. Now we prove that the local widths of F are much smaller. Let $x \in F$. We define i_1 by

$$i_1 = \min\{i \mid x_i = \max x_j\},$$

and for $k > 1$ we define

$$i_k = \min\{i \mid x_i = \max\{x_j \mid j \neq i_1, \ldots, i_{k-1}\}.$$

Now consider the space U_n, given by

$$U_n = \{x \in X \mid x_{i_k} = 0,\ k = 1, \ldots, n\}.$$

It is easy to see that $\text{diam}(F \cap (U_n + x)) \leq 1/(n+1)$ and therefore we have

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

Assume now that we want to reconstruct $x \in F$ in l_∞-norm using (adaptively or nonadaptively) linear functionals as information. That is $S = Id$. For the error of optimal nonadaptive methods we have the lower bound

$$e^n_{\text{non}}(S_{|F}) = d^n_{\text{glob}}(F) \geq \frac{c}{\sqrt{n}\log n}.$$

Now we describe an adaptive method which is much better. For simplicity we assume here that n 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 $1 \leq i \leq 2k-1$. Define

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

and

$$j^* = \min\{j \in J \mid x_j = \max_{j \in J} x_j\}.$$

Take $L_{2k} = \delta_{2j^*}$ and $L_{2k+1} = \delta_{2j^*+1}$. We consider the adaptive information

$$N_n(x) = (L_1(x), \dots, L_n(x)).$$

From the definition of N_n follows that

$$N_n(x) = (x_{l_1}, \dots, x_{l_n})$$

with pairwise different l_i. The number n is odd, say $n = 2m+1$. Then we already have picked up the largest $(m+1)$ coordinates of x. Since $x \in F$, we conclude that

$$x_k \leq 1/(m+2)$$

for all k which do not coincide with one of the l_i. Hence we can reconstruct x from the information N_n up to an error of $1/(2m+4) = 1/(n+3)$, i.e., we have found a method with

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

□

Example 1, continuation. In Proposition 5 we only got a one-sided estimate of the error of optimal adaptive methods through the local widths. Can we also proof an upper bound for the error of optimal adaptive methods through the local widths? To answer this question it is enough to study $S = Id : F \to l_\infty$, where F is as in Example 1. We claim that

$$e^n_{\text{ad}}(S_{|F}) = \frac{1}{2}$$

holds for all n. This means that we have an example with $e^n_{\text{ad}}(S_{|F}) \asymp n \cdot d^n_{\text{loc}}(S_{|F})$.

Proof: Because of $\mathrm{diam}(F) = 1$ it is enough to show that $e^n_{\mathrm{ad}}(S_{|F}) \geq 1/2$. Assume that $N : F \to \mathbf{R}^n$ is some adaptive information. We have to prove that

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

Let

$$N(x) = (L_1(x), \dots, L_n(x)),$$

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

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

with $(a^i_j)_{j \in \mathbf{N}} \in l_1$. Now let j_1 be an index such that

$$|a^1_{j_1}| \geq |a^1_j| \qquad \text{for all } j.$$

We can assume that $a^i_{j_1} = 1$ and get

$$L_1(x) = x_{j_1} + \sum_{j \neq j_1} a^1_j x_j \tag{4.1}$$

with $|a^1_j| \leq 1$ for all j. Assume that $L_1(x) = \delta$ for some small positive δ. Because of $(a^1_j)_j \in l_1$ we can find two different indices k_1 and l_1 such that $|a^1_{k_1}| < \delta$ and $|a^1_{l_1}| < \delta$. Now we define $x, y \in F$ by $L_1(x) = L_1(y) = \delta$ and x has nonzero coordinates j_1 and k_1, while y has nonzero coordinates j_1 and l_1. Observe that x and y are uniquely determined, in particular we have

$$L_1(x) = x_{j_1} + a^1_{k_1} x_{k_1} = \delta$$

and therefore $x_{j_1} < 2\delta$ and $x_{k_1} > 1 - 2\delta$, hence

$$\|x - y\| > 1 - 2\delta.$$

Under the condition that $L_1(x) = \delta$ is fixed we get a certain L_2. We can assume that this linear functional is of the form $x \mapsto \sum_{j=1}^{\infty} a^2_j x_j$ with $a_{j_1} = 0$, because if $a^2_{j_1} \neq 0$ then we can consider $L_2 + \alpha L_1$ instead of the original L_2 without changing the information operator essentially. By multiplying with a constant we even can assume that

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

with $|a_j^2| \le 1$ for all j. Again we assume that $L_2(x) = \delta$ and consider the respective L_3. So we can assume that

$$L_1(x) = L_2(x) = \ldots = L_n(x) = \delta,$$

where L_k is the information in the case $L_1(x) = \ldots = L_{k-1}(x) = \delta$, and each L_k is of the form

$$L_k(x) = x_{j_k} + \sum_{j \notin \{j_1,\ldots,j_{k-1}\}} a_j^k x_j$$

with pairwise different j_k and $|a_j^k| \le 1$ for all j and k. We prove that

$$\operatorname{diam}\{x \in F \mid L_1(x) = \ldots = L_n(x) = \delta\}$$

tends to 1 as δ tends to zero. Let k_n and l_n be two different indices such that the coordinates $a_{k_n}^i$ and $a_{l_n}^i$ are small for all $i = 1, \ldots, n$, i.e.,

$$|a_{k_n}^i| < \delta \qquad \text{and} \qquad |a_{l_n}^i| < \delta \qquad \text{for } i = 1, \ldots, n.$$

We consider $x, y \in F$ with

$$L_1(x) = \ldots = L_n(x) = L_1(y) = \ldots = L_n(y) = \delta$$

and x has nonzero coordinates $j_1, \ldots, j_n$ and k_n, while y has nonzero coordinates $j_1, \ldots, j_n$ and l_n. Because of $L_n(x) = \delta$ and $x \in F$ we obtain $x_{j_n} < 2\delta$. Using $L_{n-1}(x) = \delta$ we get a similar upper bound for $x_{j_{n-1}}$ and so on. Finally we obtain a lower bound for x_{k_n} and this also is a lower bound for $\|x - y\|$ which can be made arbitrarily close to one.

□

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 that 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

$$S_n^{(\mathrm{ad})}(f) = \phi(f(x_1), \ldots, f(t_n)) \tag{5.1}$$

and define the error bounds

$$\tilde{e}_{\mathrm{non}}^n(S_{|F}) = \inf_{S_n} \Delta_{\max}(S_n)$$

and

$$\tilde{e}^n_{\rm ad}(S_{|F}) = \inf_{S_n^{\rm ad}} \Delta_{\max}(S_n^{\rm ad}),$$

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 admisssible subspaces.

The numbers $\tilde{e}^n_{\rm non}(S_{|F})$ can be estimated from above by the Kolmogorov widths $d_n(F)$, even if S is nonlinear, see Novak (1986, 1988). The estimate for the case of a linear functional was also found by Belinskii (1991).

The following example shows that adaptive methods may be exponentially better than nonadaptive ones. In particular, the analog of Proposition 6 is wrong.

□

Example 3. Again we consider $S = Id : F \to l_\infty$ with

$$F = \{x \in X \mid x_i \geq 0, \ \sum_{i=1}^{\infty} x_i \leq 1, \ x_k \geq x_{2k}, \ x_k \geq x_{2k+1}\},$$

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

$$\tilde{e}^n_{\rm ad}(S_{|F}) \leq \frac{1}{n+3}$$

and it is not difficult to see that the nonadaptive information

$$N_n(x) = (x_1, x_2, \ldots, x_n)$$

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

$$\tilde{e}^n_{\rm non}(S_{|F}) \asymp \frac{1}{\log(n+1)}.$$

□

Example 4. We just mention the following example from Korneichuk (1993). It is not as axtreme as the last example. It is, however, closer to the standard classes of approximation theory. Consider the reconstruction problem $S = Id : F \to L_\infty([0,1])$ for

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

with $0 < \alpha < 1$. This problem can be solved adaptively using the bisection method, while nonadaptive methods are worse. Korneichuk (1993) proved, more exactly, that

$$\tilde{e}^n_{\rm non}(S_{|F}) \asymp n^{-\alpha} \qquad \text{while} \qquad \tilde{e}^n_{\rm ad}(S_{|F}) \asymp n^{-1} \log n.$$

□

Acknowledgment
The author is supported by a Heisenberg scholarship of the DFG. This paper was written during a visit of the International Computer Science Institute (ICSI), Berkeley, and the UC Berkeley. I thank J. Feldman and S. Smale for their invitations.

References

[1] V. F. Babenko (1983): Nonsymmetric extremal problems in approximation theory. Soviet Math. Doklady **27**, 346–349. [Russian original in Dokl. Akad. Nauk SSSR, Tom **269**(3), 521–524]

[2] V. F. Babenko (1992): The best approximations of the classes of functions preset by means of the continuity modulus. Ukr. Mat. Zhurn. **44**(5), 579–589. [Russian]

[3] N. S. Bakhvalov (1971): On the optimality of linear methods for operator approximation in convex classes of functions. USSR Comput. Math. Math. Phys. **11**, 244–249. [Russian original: Zh. vychisl. Mat. mat. Fiz. **11**, 244–249]

[4] W. Bauhardt (1977): Hilbert-Zahlen von Operatoren in Banachräumen. Math. Nachr. **79**, 181–187.

[5] E. S. Belinskii (1991): Estimates for the Kolmogorov diameters of classes of functions with conditions on the mixed difference in the uniform metric. Math. Notes **50**, 1198–1200. [Russian original: Mat. Zametki, Tom 50(5), 147–149]

[6] H. Braß (1982): Zur Quadraturtheorie konvexer Funktionen. In: G. Hämmerlin (ed.), Numerical integration, pp. 34–47, ISNM **57**.

[7] S. Gal, C. A. Micchelli (1980): Optimal sequential and non-sequential procedures for evaluating a functional. Appl. Anal. **10**, 105–120.

[8] I. A. Glinkin (1983): Optimal algorithms for integrating convex functions. Math. USSR Comput. Maths. Math. Phys. **23**(2), 6–12. [Russian original: Zh. vychisl. Mat. mat. Fiz. **23**(2), 267–277]

[9] I. A. Glinkin (1984): Best quadrature formula in the class of convex functions. Math. Notes **35**, 368–374. [Russian original: Mat. Zametki **35**, 697–707]

[10] R. Gorenflo, S. Vessella (1991): Abel Integral Equations. Lecture Notes in Mathematics **1461**, Springer-Verlag.

[11] A. D. Ioffe, V. M. Tikhomirov (1968): Duality of convex functions, and extremal problems. Russian Math. Surveys **23:6**, 53–124. [Russian original: Uspekhi Mat. Nauk **23:6**, 51–116]

[12] J. Kiefer (1957): Optimum sequential search and approximation methods under minimum regularity assumptions. J. Soc. Indust. Appl. Math. **5**, 105–136.

[13] M. A. Kon, E. Novak (1989): On the adaptive and continuous information problems. J. Complexity **5**, 345–362.

[14] M. A. Kon, E. Novak (1990): The adaption problem for approximating linear operators. Bull. Amer. Math. Soc. **23**, 159–165.

[15] N. Korneichuk (1991): Exact constants in approximation theory. Cambridge University Press. [Russian original: Nauka Moscow, 1987]

[16] N. P. Korneichuk (1993): Optimization of active algorithms for recovery of functions from the class Lip-α. To appear in J. Complexity

[17] G. G. Magaril-Ilyayev, K. Yu. Osipenko (1991): Optimal recovery of functionals based on inaccurate data. Math. Notes of the Academy of the USSR **50** (6), 1274–1279. [Russian original: Mat. Zametki **50**(6), 85–93]

[18] P. Mathé (1990): s-numbers in information-based complexity. J. Complexity **6**, 41–66.

[19] C. A. Micchelli, T. J. Rivlin (1977): A survey of optimal recovery. In: Optimal estimation in approximation theory, C. A. Micchelli and T. J. Rivlin (eds.), Plenum Press, New York.

[20] C. A. Micchelli, T. J. Rivlin (1985): Lectures in optimal recovery. Lecture Notes in Math. **1129**, Springer.

[21] B. S. Mityagin, G. M. Henkin (1963): Inequalities between n-diameters. In: Proc. of the Seminar on Functional Analysis **7**, 97–103, Voronezh.

[22] E. Novak (1986): Quadrature and widths. J. Approx. Th. **47**, 195–202.

[23] E. Novak (1988): Deterministic and stochastic error bounds in numerical analysis. Lecture Notes in Mathematics **1349**, Springer.

[24] E. Novak (1992): Quadrature formulas for monotone functions. Proc. Amer. Math. Soc. **115**, 59–68.

[25] E. Novak (1993): Quadrature formulas for convex classes of functions. In: H. Braß, G. Hämmerlin (eds.), Numerical Integration IV, to appear.

[26] K. Petras (1993): Quadrature theory of convex functions. In: H. Braß, G. Hämmerlin (eds.), Numerical Integration IV, to appear.

[27] A. Pietsch (1987): Eigenvalues and s-numbers. Cambridge University Press, 1987.

[28] A. Pinkus (1985): n-widths in approximation theory. Springer.

[29] A. Pinkus (1986): n-widths and optimal recovery. In: Approximation Theory, C. de Boor (ed.), Proc. of Symp. in Applied Math. **36**, 51–66, AMS.

[30] G. Sonnevend (1983): An optimal sequential algorithm for the uniform approximation of convex functions on $[0,1]^2$. Appl. Math. Optim. **10**, 127–142.

[31] A. G. Sukharev (1986): On the existence of optimal affine methods for approximating linear functionals. J. Complexity **2**, 317–322.

[32] Y. Sun (1993): An optimal recovery problem for the derivative on a class of smooth functions on the real line. Beijing Normal University, Preprint.

[33] V. M. Tikhomirov (1990): Approximation theory. In: Encyclopaedia of Mathematical Sciences, Vol. 14, Springer-Verlag. [The Russian original was published by Publisher VINITI, Moscow 1987, R. V. Gamkrelidze, ed.]

[34] J. F. Traub, G. W. Wasilkowski, H. Woźniakowski (1988): Information-based complexity. Academic Press.

[35] J. F. Traub, H. Woźniakowski (1980): A general theory of optimal algorithms. Academic Press.

Erich Novak
Mathematisches Institut
Universität Erlangen
Bismarckstr. 1 1/2
W-8520 Erlangen
Germany
novak@mi.uni-erlangen.de