Deterministic and Stochastic Error Bounds

In Numerical Analysis

Erich Novak

Erlangen, 1988

The following text is the original version, published as
Lecture Notes in Mathematics 1349, Springer Verlag, 1988.
Only the layout and the page numbers are different.

Erich Novak
Mathematisches Institut
Universität Erlangen-Nürnberg
Bismarckstr. 1 1/2
D-91054 Erlangen
Germany

novak@mi.uni-erlangen.de

http:\\www.mi.uni-erlangen.de\~novak

Contents

Acknowledgments

This is a revised version of my notes "Deterministic, stochastic, and average error bounds in optimal recovery", written in 1984–1986. The author is indebted to several people for generous advice and comments, in particular to Professor D. Kölzow who also initiated and supervised his doctoral thesis.

Erlangen, June 1988 Erich Novak

Introduction

In these notes we want to investigate different deterministic and stochastic error bounds of numerical analysis. For many computational problems (such as approximation, optimization, and quadrature) we have only partial information and consequently such problems can only be solved with uncertainty in the answer. The information-centered approach asks for optimal methods and optimal error bounds if only the type of information available is indicated. We begin with worst case error bounds for deterministic methods and consider relations between these error bounds and the n-widths of the class of problem elements (1.2). In 1.3 we give worst case error bounds for some special problems. We are mainly interested in the problems of approximation (*App*), optimization (*Opt* and *Opt**), and quadrature or integration (*Int*). We consider different function classes, for both adaptive and nonadaptive methods. First of all, I explain the information-based approach by means of an example.

Example 1. If $f : [0,1] \to \mathbf{R}$ is a continuous function (or Riemann-integrable) and

$$\widetilde{S}_n(f) = \frac{1}{n} \sum_{i=1}^{n} f\left(\frac{2i-1}{2n}\right),$$

then

$$\lim_{n\to\infty} \widetilde{S}_n(f) = \int_0^1 f(x)\,dx := S(f).$$

For the error $\Delta(\widetilde{S}_n(f), f) = |S(f) - \widetilde{S}(f)|$, the following estimates are valid:

$$\Delta(\widetilde{S}_n(f), f) \le 1/2n \quad \text{if} \quad f \in F^1 = \{f : [0,1] \to \mathbf{R} \mid \mathrm{Var}(f) \le 1\},$$

$$\Delta(\widetilde{S}_n(f), f) \le 1/4n \quad \text{if} \quad f \in F^2 = \{f : [0,1] \to \mathbf{R} \mid |f(x) - f(y)| \le |x-y|\}, \text{ and}$$

$$\Delta(\widetilde{S}_n(f), f) \le 1/192n^3 \ \text{ if } \ f \in F^3 = \{f : \mathbf{R} \to \mathbf{R} \mid f \text{ is 1-periodic and } |f'''(x)| \le 1\}.$$

It is not surprising that these error bounds strongly depend on the properties of f. At least the first two estimates are not very satisfying and we could ask whether there is a quadrature formula which yields better results for these function classes. We want to compare $\widetilde{S}_n$ with all methods $\widetilde{S} \in A_n$ which only use the values of $f \in F^i$ at n knots, i.e. methods of the form

$$\widetilde{S}(f) = \phi(f(a_1), \ldots, f(a_n))$$

with $\phi : \mathbf{R}^n \to \mathbf{R}$ and $a_k \in [0,1]$. In fact, the midpoint rule $\widetilde{S}_n$ is optimal for all three classes F^i, we have

$$\Delta_{max}(\widetilde{S}_n) = \inf_{\widetilde{S} \in A_n} \Delta_{max}(\widetilde{S})$$

for all n. Here $\Delta_{max}(\widetilde{S})$ is defined by

$$\Delta_{max}(\widetilde{S}) = \sup_{f \in F^i} \Delta(\widetilde{S}(f), f) = \sup_{f \in F^i} |S(f) - \widetilde{S}(f)|.$$

We conclude this example with two remarks:
i) The optimality of the midpoint rule $\widetilde{S}_n$ actually holds within the much bigger class A_n^{ad} of all adaptive methods. For these methods, the knots a_k may depend on $f \in F$ in the following way: The point a_1 is fixed, $a_2 = a_2(f(a_1))$, $a_3 = a_3(f(a_1), f(a_2))$ and so on. The fact that adaption does not help for this and many other problems is remarkable because the structure of nonadaptive methods is much simpler than the structure of adaptive methods.
ii) Of course, the midpoint rule is optimal only for some classes of functions. For other classes there might exist methods which are much better. It is known, however, that the midpoint rule is optimal for many classes of periodic functions, see *Motornyi* (1973), *Zhensykbaev* (1976), and A.2. The optimality of $\widetilde{S}_n$ for F^1 and F^2 is easy to prove and well known, see *Zubrzycki* (1963/64).

To be more specific, we now give some basic definitions. Let F be a set of real-valued functions on a set X and let $S : F \to M$ be an arbitrary mapping. The mapping S is to be approximated by some mapping $\widetilde{S} : F \to M$, being an element of a set A of admissible approximations. We mainly consider the classes $A = A_n^{ad}$ and $A = A_n$ ($n \in \mathbf{N}$) which are defined as follows. Let A_n^{ad} be the set of all approximations $\widetilde{S} : F \to M$ which use n values $f(a_i)$ whereby the a_i are chosen adaptively, i.e.

$$A_n^{ad} = \{\widetilde{S} = \phi \circ N \mid N(f) = (f(a_1), f(a_2(f(a_1))), \ldots, f(a_n(f(a_1), \ldots, f(a_{n-1})))),$$
$$\text{where } a_i : \mathbf{R}^{i-1} \to X \text{ and } \phi : \mathbf{R}^n \to M\}.$$

The set A_n contains all $\widetilde{S} \in A_n^{ad}$ where the a_i are nonadaptively chosen, i.e.

$$A_n = \{\widetilde{S} = \phi \circ N \mid N(f) = (f(a_1), \ldots, f(a_n)),\ a_i \in X,\ \phi : \mathbf{R}^n \to M\}.$$

We assume that an error $\Delta(m, f) \geq 0$ is defined for each $f \in F$ and $m \in M$, for example by $\Delta(m, f) = d(m, S(f))$ where d is a metric on M. The number $\Delta_{max}(\widetilde{S}) = \sup_{f \in F} \Delta(\widetilde{S}(f), f)$ is called the maximal error of $\widetilde{S}$. The numbers

$$e_n(F, S) = \inf_{\widetilde{S} \in A_n} \Delta_{max}(\widetilde{S}) \quad \text{and} \quad e_n^{ad}(F, S) = \inf_{\widetilde{S} \in A_n^{ad}} \Delta_{max}(\widetilde{S})$$

indicate how well S can be approximated on F by a $\widetilde{S} \in A_n$ and $\widetilde{S} \in A_n^{ad}$, respectively.

We now formulate the problems of approximation, optimization, and integration which fit to our basic model.

Approximation (App): Let F be a set of bounded functions on X, and let $S = App : F \to M = B(X)$ be defined by $S(f) = f$, and consider the error function $\Delta(m, f) = \|m - f\|_\infty$. Then the number $\Delta_{max}(\widetilde{S})$ of $\widetilde{S} : F \to B(X)$ is given by

$$\Delta_{max}(\widetilde{S}) = \sup_{f \in F} \|\widetilde{S}(f) - f\|_\infty.$$

Optimization (Opt and *Opt*)*: Again we assume that F is a set of bounded functions on X. We deal with two different problems: If one only wishes to know the value of $\sup f$ then $S = Opt^* : F \to M = \mathbf{R}$, $S(f) = \sup f$, and $\Delta(m, f) = |\sup f - m|$. We have $\Delta_{max}(\widetilde{S}) = \sup_{f \in F} |\sup f - \widetilde{S}(f)|$. In many cases one seeks an $x \in X$ such that $\sup f - f(x)$ is sufficiently small. In this case the approximations $\widetilde{S}$ have values in X and we consider the error function $\Delta(\widetilde{S}(f), f) = \sup f - f(\widetilde{S}(f))$. This problem is called *Opt*. We do not define a mapping $Opt : F \to X$ because there is no distinguished mapping. Observe, however, that all relevant data are given by the error function $\Delta : M \times F \to \mathbf{R}_0^+$.
Quadrature or *Integration (Int)*: We assume that $S = Int : F \to \mathbf{R}$ is linear and $\Delta(m, f) = |m - S(f)|$. In many cases $S = \lambda^s$ is the usual (s-dimensional) Lebesgue measure, but more general linear functionals are considered as well.
We observe that other problems such as differentiation, extrapolation, or different problems connected with differential and integral equations can be formulated similarly.

It should be evident that the numbers $e_n(F, S)$ and $e_n^{ad}(F, S)$ play a basic role in the worst case analysis of numerical methods. The numbers $e_n(F, S)$ describe the size of the set F. It is hardly astonishing that they are related to the n-widths $d_n(F)$ of F in $B(X)$ according to Kolmogorov. The following inequalities hold.

Proposition.

$$e_n(F, App) \leq (1 + n) \cdot d_n(F) \tag{1}$$

$$e_n(F, Opt) \leq 2 \cdot e_n(F, App) \tag{2}$$

$$e_n(F, Opt^*) \leq e_n(F, App) \tag{3}$$

$$e_n(F, Int) \leq 2 \cdot \|Int\| \cdot d_n(F) \tag{4}$$

((2) and (3) are known, see *Wasilkowski* (1984); (4) is due to *Novak* (1983); (1) is new. See 1.2.4, 1.2.5, and 1.2.8).

By means of suitable packings and coverings of X by equal balls we prove lower and upper bounds for the n-widths of certain Lipschitz classes. In some cases these bounds yield the exact values of the $d_n(F)$, see 1.3.7. In particular, we obtain a result of *Grigorian* (1973).

A problem could be called ill-posed in our setting, if the relation $\lim_{n\to\infty} e_n^{ad}(F,S) = 0$ does not hold, see 1.1.6. This means that a finite number of function values is not sufficient to get approximations $\widetilde{S}$ with $\Delta_{max}(\widetilde{S}) < \varepsilon$ for each $\varepsilon > 0$. In 1.2.7 we prove some conditions which are sufficient for $e_n(F,S) \to 0$. For the problem *App* and symmetric, convex sets $F \subset B(X)$ we show that $e_n(F,App) \to 0$ if and only if $F \subset K + V$, where $K \subset B(X)$ is compact and $V \subset B(X)$ is a finite dimensional subspace.

In Section 1.3 we deal with the numbers $e_n^{(ad)}(F,S)$ for some special problems. We begin with a linear operator $S : F \to M$, where $F \subset B(X)$ is a symmetric and convex set and M is a normed space. We assume that the error function is given by $\Delta(m,f) = \|S(f) - m\|$. Such a problem is called a linear problem. It is known that $e_n^{ad}(F,S) \geq \frac{1}{2} e_n(F,S)$ in this case, but it was an unsolved problem for some years, whether $e_n^{ad}(F,S) < e_n(F,S)$ is indeed possible, see *Packel, Woźniakowski* (1986). The first example, where adaption helps for a linear problem in the worst case, was independently found by Mark A. Kon and the author, see 1.3.1.

Then we discuss several results for the problems *App*, *Opt*, *Opt**, and *Int*. We mention one of them which is from *Novak* (1983), see 1.3.12.

Proposition.
Let $k, s \in \mathbf{N}$, $1 \leq p \leq \infty$, and $pk > s$. Then

$$\sup_{\|Int\| \leq 1} e_n(W_p^{k,s}, Int) \asymp n^{-k/s}.$$

There are problems in numerical analysis which are intractable in the sense that the numbers $e_n(F,S)$ and $e_n^{ad}(F,S)$ tend to zero very slowly (with $n \to \infty$). We give an example which is due to Sukharev (see *Sukharev* (1971, 1978, 1979, and 1984)).

Example 2.
Let $s \in \mathbf{N}$, $F_s = \{f : [0,1]^s \to \mathbf{R} \mid |f(x) - f(y)| \leq \max |x_i - y_i|\}$, and $n = m^s$ (with $m \in \mathbf{N}$). Then

$$e_n^{ad}(F_s, App) = e_n(F_s, App) = 1/2 \cdot n^{-1/s} \tag{1}$$

$$e_n^{ad}(F_s, Opt) = e_n(F_s, Opt) = 1/2 \cdot n^{-1/s} \tag{2}$$

$$e_n^{ad}(F_s, Opt^*) = e_n(F_s, Opt^*) = 1/4 \cdot n^{-1/s} \tag{3}$$

$$e_n^{ad}(F_s, \lambda^s) = e_n(F_s, \lambda^s) = \frac{s}{2(s+1)} \cdot n^{-1/s} \tag{4}$$

The following features of this example should be emphasized:
a) Adaption does not help for any of the four problems within the worst case model.
b) Although the problems *Opt*, *Opt**, and *Int* seem to be easier than the problem *App*, this is not the case within the worst case model: The numbers $e_n^{ad}(F_s,S) = e_n(F_s,S)$ are very similar for the four different problems.

c) The numbers $e_n(F_s,S)$ tend to zero very slowly. Let $s=10$, for example. Then $e_1(F_s,App) = e_1(F_s,Opt) = 1/2$ and $e_1(F_s,\lambda^s) = 5/11$. For $n = 2^{20} > 10^6$ we get $e_n(F_s,App) = e_n(F_s,Opt) = 1/8$ and $e_n(F_s,\lambda^s) = 5/44$.

If one wants to approximate $S(f_0)$ by $\widetilde{S}(f_0)$, only knowing that $f_0 \in F$, one has the estimate $\Delta(\widetilde{S}(f_0), f_0) \leq \Delta_{max}(\widetilde{S})$, which is best possible, of course, if one insists on an exact deterministic error bound. On the other hand, it may be desirable to distinguish between two possiblilities which are imaginable in practical situations:
a) For all or for most $f \in F$ the error $\Delta(\widetilde{S}(f), f)$ of $\widetilde{S}$ is near $\Delta_{max}(\widetilde{S})$.
b) The error of $\widetilde{S}$ is much smaller than $\Delta_{max}(\widetilde{S})$ for most $f \in F$. Only for some extreme $f \in F$, the error is near $\Delta_{max}(\widetilde{S})$.
We could also ask whether adaptive methods are better than nonadaptive ones in a stochastic sense.

Chapter 2 begins with general remarks concerning stochastic error bounds. Then we discuss the meaning of Monte Carlo methods. These methods seem to be reliable if the population of approximations from which the sample is chosen, is finite. More exactly, we give the following definition. Let F,S,A, and Δ be given. A random variable $Q = Q(\omega)_{\omega\in\Omega}$ with values in A is called a Monte Carlo method in A, if $\{Q(\omega) \mid \omega \in \Omega\}$ is a finite subset of A and the probability $m(\widetilde{S})$ of each $\widetilde{S}$ is a rational number. The set of all those Monte Carlo methods is denoted by $C(A)$. The number

$$\sigma(Q) = \sup_{f\in F} E(\Delta(Q(\omega)(f), f))$$

(where E is the expectation of a random variable) is called the error of Q. Finally we define

$$\sigma_n(F,S) = \inf_{Q\in C(A_n)} \sigma(Q) \quad \text{and} \quad \sigma_n^{ad}(F,S) = \inf_{Q\in C(A_n^{ad})} \sigma(Q).$$

It is easy to see that $\sigma_n(F,S) \leq e_n(F,S)$ and $\sigma_n^{ad}(F,S) \leq e_n^{ad}(F,S)$.

Concerning *Example 2*, we prove the following statements (see 2.2.6; the upper bound (4) for the problem *Int* is due to *Haber* (1966)).

Example 2, continued.

$$\sigma_n^{ad}(F_s,App) \geq e_{n+1}(F_s,App) \tag{1}$$

$$\sigma_n^{ad}(F_s,Opt) \geq (1/2-\varepsilon)\cdot 2^{-1/s}\cdot e_n(F_s,Opt) \tag{2}$$

$$\sigma_n^{ad}(F_s,Opt^*) \geq (1/2-\varepsilon)\cdot 2^{-1/s}\cdot e_n(F_s,Opt^*) \tag{3}$$

$$\sigma_n(F_s,\lambda^s) \leq (12^{-1/s}+\varepsilon)\cdot n^{-1/2-1/s} \tag{4}$$

$$\sigma_n^{ad}(F_s,\lambda^s) \geq \frac{1-\varepsilon}{24(s+1)}\cdot n^{-1/2-1/s} \tag{5}$$

((1) holds for all $n \in \mathbf{N}$, (2)–(5) are valid for all $\varepsilon > 0$ and $n > n(\varepsilon)$.)

The lower bounds are proved by means of general lower bounds for the problems *App*, *Opt*, *Opt**, and *Int* and do not depend on our relatively restricted notion of a Monte Carlo method. These lower bounds are established in 2.2.2–2.2.4. Observe that the numbers $\sigma_n(F_s,S)$ and $\sigma_n^{ad}(F_s,S)$ are only slightly smaller than the respective deterministic error bounds for the problems *App*, *Opt*, and *Opt**. This is true not only for the classes F_s but also for many other classes (see 2.2.8). We conclude that Monte Carlo methods do not seem to be very useful for those problems and can be replaced by deterministic methods. This may be interesting particularly for the problem *Opt* because stochastic search methods for finding the global maximum of a function are widely used. The numbers $\sigma_n(F_s,\lambda^s)$ (which differ only slightly from the $\sigma_n^{ad}(F_s,\lambda^s)$) tend to zero much faster than the $e_n(F_s,\lambda^s)$. This shows that in the case $S = Int = \lambda^s$ there are approximations $\widetilde{S} \in A_n$ with the property that $\Delta(\widetilde{S}(f),f)$ is much smaller than $\Delta_{max}(\widetilde{S})$ for most $f \in F_s$. Similar results hold for other classes of functions, see 2.2.9.

Monte Carlo methods together with their stochastic error bounds seem to be useful if the numbers $\sigma_n(F,S)$ are much smaller than the respective $e_n(F,S)$. This is often the case for the problem *Int* but depends on the function class F as the following example shows. (See 2.2.7. The worst case error bound is due to *Zubrzycki* (1963/64); similar stochastic error bounds are proved in *Novak* (1983)).

Example 3.
Let $F_{\mathrm{Var}} = \{f : [0,1] \to \mathbf{R} \mid \mathrm{Var}(f) \le 1\}$ and $S(f) = \int_0^1 f(x)dx$. Then

$$e_n(F_{\mathrm{Var}},S) = \frac{1}{2n} \quad \text{and} \quad \frac{1}{8n-4} \le \sigma_n^{ad}(F_{\mathrm{Var}},S) \le \frac{1}{2n}.$$

In 2.1.8 we treat the question whether $\sigma_n(F,S) \to 0$ implies that also $e_n(F,S) \to 0$. We prove that this conjecture holds in the normed space setting (i.e. $\Delta(m,f) = \|m - S(f)\|$), but we also show that it does not hold in general.

In 2.1.10 and 2.2.10 we deal with Monte Carlo methods $Q = (Q_\omega)_{\omega\in\Omega}$ with varying cardinality. Here the Q_ω are adaptive deterministic methods where the number $n_\omega(f)$ of function values may depend on $f \in F$. We consider the class

$$C(A_n^{var}) = \{(Q_\omega)_{\omega\in\Omega} \mid \sup_{f\in F} \int_\Omega n_\omega(f)\, d\omega \le n\} \supset C(A_n^{ad})$$

and show that methods with varying cardinality are not significantly better than methods from $C(A_n^{ad})$ or even $C(A_n)$ for many problems. In particular, we prove a conjecture of *Wasilkowski* (1987) concerning the error of stochastic quadrature formulas $Q \in C(A_n^{var})$.

In Section 3.1 we consider average errors defined by

$$\Delta_\mu(\widetilde{S}) = \int_F \Delta(\widetilde{S}(f), f)\ d\mu(f),$$

where μ is a probability measure on F. Analogously to the $e_n^{(ad)}(F,S)$ and $\sigma_n^{(ad)}(F,S)$ we define

$$e_n^{(ad)}(F,S,\mu) = \inf_{\widetilde{S}\in A_n^{(ad)}} \Delta_\mu(\widetilde{S}).$$

A crucial point of this approach is the choice of a reasonable measure μ. We introduce two concepts of a "fair" probability measure on F and discuss some of their properties (see 3.1.5–3.1.6). *Graf, Novak* (1987) consider a "natural" probability measure on the set

$$F = \{f : [0,1] \to \mathbf{R} \mid f \text{ continuous},\ f(0) = 0,\ \mathrm{Var}(f) \le 1\}$$

and prove estimates for the average error of quadrature formulas, see 3.1.7.

Two questions concerning the $e_n(F,S,\mu)$ and $e_n^{ad}(F,S,\mu)$ are of main interest:
i) How much do the numbers $e_n^{(ad)}(F,S,\mu)$ differ from the $e_n^{(ad)}(F,S)$?
ii) How much do the $e_n^{ad}(F,S,\mu)$ differ from the $e_n(F,S,\mu)$, i.e. how much does adaption help on the average?
The answer to both questions of course depends on the measure μ. There are problems, however, where the numbers $e_n^{(ad)}(F,S,\mu)$ are much smaller than the $e_n^{(ad)}(F,S)$ for every probability measure. Concerning the function class F_s of *Example 2* we prove:
a) For each $n \in \mathbf{N}$ there is a probablity measure μ on F_s such that $e_n^{ad}(F_s,S,\mu)$ is only slightly smaller than $e_n(F_s,S)$ if $S = App$, Opt, or Opt^*. This is interesting particularly for the problems Opt and Opt^* because it seems to be intuitively clear that adaptive methods are much better on the average than nonadaptive methods.
b) For the problem $Int = \lambda^s$ we prove that the numbers $e_n(F_s,S,\mu)$ are much smaller than the $e_n(F_s,S)$. To be more exact, we have

$$e_n(F_s,\lambda^s,\mu) \le 12^{-1/s} \cdot n^{-1/2-1/s} \qquad (n = m^s,\ m \in \mathbf{N})$$

for every probability measure μ on F_s, while adaptive methods are again only slightly better. We also show that this bound is best possible in the following sense. If $\beta > 1/2 + 1/s$, then there is a measure μ on F_s such that $e_n^{ad}(F_s,\lambda^s,\mu) \ge c \cdot n^{-\beta}$ for some $c > 0$.
We also prove similar results for different function classes and for methods with varying cardinality $n(f)$ and $\int_F n(f)\,d\mu(f) \le n$, see 3.1.10.

In Section 3.2 we consider the average a posteriori error of a method $\widetilde{S}$, defined in *Novak* (1986b, c). The estimate $\Delta(\widetilde{S}(f), f) \le \Delta_{max}(\widetilde{S})$ is the best one, valid for all

$f \in F$. Knowing $N(f) = x \in \mathbf{R}^n$, of course

$$\Delta(\widetilde{S}(f), f) \leq \sup\{\Delta(\widetilde{S}(g), g) \mid g \in F,\ N(g) = x\} =: \alpha(x)$$

is valid. This a posteriori error bound may be much smaller than $\Delta_{max}(\widetilde{S})$ for many $f \in F$. If μ is a probability measure on $N(F)$, we can define an average a posteriori error of $\widetilde{S} = \phi \circ N$ by $\int_{N(F)} \alpha(x)\ d\mu(x)$. Using the (suitably normed) Lebesgue measure on $N(F)$ we obtain the number $\Delta_{aver}^{(ad)}(\phi, N)$. Finally, we define

$$a_n^{(ad)}(F,S) = \inf_{\widetilde{S} = \phi \circ N \in A_n^{(ad)}} \Delta_{aver}^{(ad)}(\phi, N).$$

By means of suitable examples we show that there are linear problems with

$$a_n^{ad}(F,S) = a_n(F,S) = e_n(F,S)$$

for all $n \in \mathbf{N}$, i.e. the a posteriori error bound $\Delta(\widetilde{S}(f), f) \leq \alpha(N(f))$ is not better than the worst case error bound $e_n(F,S)$ for any $f \in F$, even if $\widetilde{S} \in A_n^{ad}$ is allowed to be adaptive. On the other side there are linear problems with

$$a_n(F,S) \to 0 \quad \text{but} \quad e_n(F,S) = 1$$

for all n.

Concerning the class F_1 of *Example 2* we determine the optimal nonadaptive information $N_{aver}^* \in I_n$ for the problems *App* and *Int* according to our average case model. It is interesting that this optimal average case information differs from the optimal worst case information

$$N_{max}^*(f) = \left(f\left(\frac{1}{2n}\right), f\left(\frac{3}{2n}\right), \ldots, f\left(\frac{2n-1}{2n}\right) \right)$$

for all $n > 1$. We obtain

$$\frac{1}{2\,(n+1)} < a_n(F_1, App) < \frac{1}{2n}$$

and

$$a_n(F_1, \lambda) = \frac{1}{6n-2},$$

hence the a posteriori error bound is only slightly better than the a priori error bound for most $f \in F_1$. It is easy to see that $a_n^{ad}(F_1,S) < a_n(F_1,S)$ for all $n > 2$ and S being one of the problems *App*, *Opt*, *Opt** or $Int = \lambda$ but we conjecture that the numbers $a_n^{ad}(F_1,S)$ are considerably smaller than $a_n(F_1,S)$ only for the nonlinear problems *Opt* and *Opt**.

In an appendix we deal with different problems concerning the existence and uniqueness of optimal algorihms. We always assume that $S : F \to M$, where M is a metric space and $\Delta(\widetilde{S}(f),f) = d(S(f),\widetilde{S}(f))$. First of all we could ask whether there is a $S_n^* \in A_n$ with

$$\Delta_{max}(S_n^*) = e_n(F,S).$$

This problem can only be solved in some cases using special methods. Therefore we mainly treat a more special problem: We assume that an information operator $N : F \to \mathbf{R}^n$ is given and ask whether there is an optimal algorithm $\phi^* \circ N$ in the class $A(N) = \{\phi \circ N \mid \phi : \mathbf{R}^n \to M\}$. First we deal with the worst case and indicate some results using well known facts on Chebyshev centers. Then we introduce the notion of a center of a probability measure in a metric space and prove statements on its existence and uniqueness. Similarly as in the worst case these results can be used to prove the existence and uniqueness of optimal average case algorithms. These results are due to *Novak, Ritter* (1987) and use the definition of the average error of *Wasilkowski* (1983). We also show that nonmeasurable algorithms are not better than measurable ones.

1. Deterministic error bounds

1.1 Basic assumptions and definitions

1.1.1 Definition
Many problems of numerical analysis may be abstractly formulated as follows: Let sets F and M be given and let $S : F \to M$ be a mapping which is to be approximated by a mapping $\widetilde{S} : F \to M$. For every $f \in F$ and $m \in M$ an error $\Delta(m, f) \in \mathbf{R}_0^+$ of m with respect to f is defined. Then the number

$$\Delta(\widetilde{S}(f), f) \in \mathbf{R}_0^+$$

is the error of $\widetilde{S}$ for some $f \in F$, and

$$\Delta_{max}(\widetilde{S}) = \sup_{f \in F} \Delta(\widetilde{S}(f), f)$$

is called the *maximal error* of $\widetilde{S}$.

Remarks
a) Of course these definitions only make sense if the error function $\Delta : M \times F \to \mathbf{R}_0^+$ is reasonably connected with S. If, for example, M is a metric space, one can define

$$\Delta(\widetilde{S}(f), f) = d(\widetilde{S}(f), S(f)).$$

There are important problems, however, where the error function Δ can not be defined this way. We observe that only the error function $\Delta : M \times F \to \mathbf{R}_0^+$ and not the function S is important for our further definitions and discussions. Indeed, for some problems it is not possible to define an appropriate mapping S.
We give the following example. Let F be a set of bounded functions on a set X. The problem *Opt* of global optimization of a function $f \in F$ consists in finding an $x \in X$ such that $\sup f - f(x)$ is sufficiently small. Hence for $\widetilde{S} : F \to X = M$, the error is defined by

$$\Delta(\widetilde{S}(f), f) = \sup f - f(\widetilde{S}(f)).$$

Here the error function Δ can not be given by a metric on $M = X$ and no solution mapping S is clearly distinguished (actually we do not assume that there is an $x \in X$ with $\sup f = f(x)$).
On the other hand it seems to be more vivid to indicate the solution mapping S where possible; in many cases the mapping S yields an error function Δ in a natural way.
b) A very similar concept of errors is discussed in *Traub, Wasilkowski, Woźniakowski* (1983) (see p.10). Our notion of an error function Δ is slightly more general since we do not assume that there is an $m \in M$ with $\Delta(m, f) = 0$.

1.1.2 Definition

A mapping $\widetilde{S} : F \to M$ is called an ε-approximation if $\Delta_{max}(\widetilde{S}) \leq \varepsilon$. The notion of an ε-approximation is fundamental for numerical analysis. In many situations one seeks an ε-approximation with certain additional properties:
– The computation of $\widetilde{S}$ should be easy and fast;
– The approximation $\widetilde{S}$ should be an element of a set A of admissible approximations, for example, $\widetilde{S}$ should use only a certain amount of information of $f \in F$.
We assume that, in general, the element $f \in F$ is not known. Instead we have some *information* on f given by $N(f)$, where $N : F \to H$. We say that the approximation $\widetilde{S} : F \to M$ only uses the information N, if $\widetilde{S}$ may be written as $\widetilde{S} = \phi \circ N$, where $\phi : H \to M$ is an arbitrary mapping which is called an (*idealized*) *algorithm*. We write

$$A(N) = \{\widetilde{S} = \phi \circ N \mid \phi : H \to M\}$$

for the set of all approximations which use the information N.
Now we ask for the maximal error of approximations $\widetilde{S} \in A(N)$. Particularly we are interested in the quantity $\inf_{\widetilde{S} \in A(N)} \Delta_{max}(\widetilde{S})$. For this, we introduce the *(maximal) radius* $r(N) = r_{max}(N)$ *of the information* N by

$$r(N) = \sup_{f \in F} \inf_{m \in M} \sup_{g \in F} \{\Delta(m, g) \mid N(f) = N(g)\}.$$

It is not difficult to prove that

$$r(N) = \inf_{\widetilde{S} \in A(N)} \Delta_{max}(\widetilde{S}),$$

see Theorem 3.1 of *Traub,Wasilkowski,Woźniakowski* (1983).

1.1.3 Definition

In these notes we consider the following more special case. We assume that F is a set of real-valued functions which are defined on a set X. We consider the following classes of information operators. Let $n \in \mathbf{N}$,

$$I_n^{ad} = \{N : F \to \mathbf{R}^n \mid N(f) = (f(a_1), f(a_2(f(a_1))), \ldots, f(a_n(f(a_1), \ldots, f(a_{n-1})))),$$
$$\text{where } a_i : \mathbf{R}^{i-1} \to X\,\}$$

and

$$I_n = \{N : F \to \mathbf{R}^n \mid N(f) = (f(a_1), \ldots, f(a_n)), \text{ where } a_i \in X\}.$$

We say that $N \in I_n^{ad}$ is the information of n function values $f(a_i)$, where the a_i are chosen *adaptively*, whereas $N \in I_n$ gives the information of n function values, where

the a_i are chosen *nonadaptively*. It is evident that I_n^{ad} contains many more information operators than I_n.
The respective classes of approximations are denoted by A_n^{ad} and A_n, i.e.

$$A_n^{ad} = \{\widetilde{S} : F \to M \mid \widetilde{S} = \phi \circ N \text{ with } \phi : \mathbf{R}^n \to M \text{ and } N \in I_n^{ad}\}$$

and

$$A_n = \{\widetilde{S} : F \to M \mid \widetilde{S} = \phi \circ N \text{ with } \phi : \mathbf{R}^n \to M \text{ and } N \in I_n\}.$$

Now we define the numbers $e_n^{(ad)}(F,S)$ which show how well the mapping $S : F \to M$ can be approximated by $\widetilde{S} \in A_n^{ad}$ and $\widetilde{S} \in A_n$, respectively. For $n \in \mathbf{N}$ let

$$e_n^{ad}(F,S) = \inf_{\widetilde{S} \in A_n^{ad}} \Delta_{max}(\widetilde{S}) = \inf_{N \in I_n^{ad}} r(N)$$

and

$$e_n(F,S) = \inf_{\widetilde{S} \in A_n} \Delta_{max}(\widetilde{S}) = \inf_{N \in I_n} r(N).$$

Remarks
a) The numbers $\Delta_{max}(\widetilde{S})$, $r(N)$, $e_n^{ad}(F,S)$, and $e_n(F,S)$ are related to a worst case analysis of numerical methods: We are interested in exact or deterministic error bounds, i.e. the error bounds are valid for all $f \in F$.
b) We assume that the information $N(f)$ is given exactly and that $\widetilde{S}(f) = \phi \circ N(f)$ may be computed without errors. So we do not consider rounding errors or errors in measurement. Both limitations may be removed in the basic model, see *Traub,Wasilkowski,Woźniakowski* (1983) and *Nemirovsky,Yudin* (1983).

1.1.4 Main problems treated in this paper

a) Approximation (*App*)
We consider the solution operator $S = App = Id : F \to B(X)$, where $F \subset B(X)$. The error function $\Delta : M \times F \to \mathbf{R}_0^+$ is given by $\Delta(m,f) = \|m - f\|_\infty$, i.e. $\Delta_{max}(\widetilde{S}) = \sup_{f \in F} \|\widetilde{S}(f) - f\|_\infty$.
b) Optimization (*Opt*) and (*Opt**)
We deal with two different problems: If only $\sup f$ is requested then $S = Opt^* : F \to \mathbf{R}$, $S(f) = \sup f$, and $\Delta(m,f) = |\sup f - m|$, i.e. $\Delta_{max}(\widetilde{S}) = \sup_{f \in F} |\sup f - \widetilde{S}(f)|$.
More often one seeks an $x \in X$ such that $\sup f - f(x)$ is small. In this case the approximations $\widetilde{S}$ have values in X and $\Delta(\widetilde{S}(f), f) = |\sup f - f(\widetilde{S}(f))|$. Again we stress that there is no clearly distinguished solution operator S for this problem (which we call *Opt*).
c) Quadrature or Integration (*Int*)
We assume that $S = Int : F \to \mathbf{R}$ is a linear functional and $\Delta(m,f) = |m - S(f)|$. In many of our examples $S = \lambda^s$ is the usual Lebesgue integral on a set $X \subset \mathbf{R}^s$, but more general linear functionals are considered as well.

1.1.5 Remark: Differential and integral equations

Different problems connected with the approximate solution of differential and integral equations can be formulated within the setting of this paper. We mention one result concerning a Fredholm equation of the second kind, see *Werschulz* (1985a, b and 1986), *Pereverzev* (1986), and *Kacewicz* (1987) for further results. We assume that an equation

$$u(x) - \int_0^1 k(x,y)u(y)\,dy = f(x)$$

is given, where k is r-times continuously differentiable and 1 is not an eigenvalue of the integral operator. Then the equation has a unique solution $S(f)$ for each $f \in F = W_2^r = \{f : [0,1] \to \mathbf{R} \mid f^{(r-1)}$ is absolutely continuous and $\|f^{(r)}\|_2 \leq 1\}$. If the error is measured in the L_2-norm, then

$$e_n(F,S) \asymp n^{-r}.$$

It turns out that the finite element method (together with a Gauss quadrature) is asymptotically optimal if the spline space is chosen suitably.

1.1.6 Remark: Ill-posed problems

A problem can be called ill-posed within our setting, if the relation

$$\lim_{n\to\infty} e_n^{ad}(F,S) = 0$$

does not hold. This means that a finite number of function values is not sufficient to get an approximation $\widetilde{S}$ with $\Delta_{max}(\widetilde{S}) < \varepsilon$ for each $\varepsilon > 0$. Otherwise a problem is called well-posed. In 1.2.7 we prove some conditions which are sufficient for well-posedness.

Assume that $F = \{f \in V \subset B(X) \mid \|f\| \leq 1\}$ and $S : V \to M$ is an unbounded linear operator, where M is a normed space and V is a subspace of $B(X)$. Let Δ be given by $\Delta(\widetilde{S}(f), f) = \|S(f) - \widetilde{S}(f)\|$. A source of such problems is the solution of a Fredholm equation of the first kind. *Werschulz* (1987a) proved that $e_n^{ad}(F,S) = \infty$ for any $n \in \mathbf{N}$ and hence such a problem (which is ill-posed in the sense of Hadamard) is also ill-posed in a very severe sense within our setting. See *Werschulz* (1987b) for related results.

Exploiting additional a priori information seems to be necessary for any efficient numerical treatment of ill-posed problems. Technically spoken, $S : F \to M$ must be replaced by a suitable ("regularized") well-posed problem $S^* : F \to M$. For example, the regularization method of Tikhonov is widely used. We are mainly interested in error bounds for well-posed problems, see *Baumeister* (1987), *Groetsch* (1984), *Hofmann* (1987), *Morozov* (1984), *Talenti* (1986), and *Tikhonov, Arsenin* (1979) for surveys on ill-posed problems and the method of regularization.

1.1.7 Notes and references

Our basic model of "optimal recovery" or "information-based complexity" traces back to investigations which were done independently in the east and in the west, beginning in the fifties with papers by Kiefer, Nikolskii, Sard, Bakhvalov, and others. See *Traub, Woźniakowski* (1980) for historical remarks and for a detailed annoted bibliography. Monographs and survey papers on a worst case analysis of numerical methods are *Micchelli, Rivlin* (1977), *Traub, Woźniakowski* (1980), *Traub, Wasilkowski, Woźniakowski* (1983), *Nemirovsky, Yudin* (1983), *Traub, Woźniakowski* (1984), *Micchelli, Rivlin* (1985), *Woźniakowski* (1986), and *Packel, Woźniakowski* (1987).

Different investigations concerning the efficiency of algorithms of analysis are contained in *Smale* (1985). The problems *App*, *Opt*, *Opt**, and *Int* are also discussed in *Sukharev* (1984). In this paper the class $F = \{f : X \to \mathbf{R} \mid |f(x) - f(y)| \leq d(x,y)\}$ is considered, where (X,d) is a metric space. *Sukharev* (1984) also explains the underlying ideas and studies sequentially-optimal algorithms which are not discussed in our paper.

Let us remark that we do not consider the notion of complexity in these notes. Technically spoken, we only study the information cost of algorithms and not the computational cost. This assumption is suitable for many problems but there are counterexamples as well, see *Woźniakowski* (1986) for a discussion of this point.

Added in the proof: A main reference on information-based complexity is the recent book *Traub, Wasilkowski, Woźniakowski* (1988) which contains a thourough introduction to the whole subject with many references and new results.

1.2 Deterministic error bounds and n-widths

1.2.1 Definition

We assume that F is a set of bounded real-valued functions on a set X. The n-width $d_n(F)$ of $F \subset B(X)$ according to Kolmogorov is defined by

$$d_n(F) = \inf_V \sup_{f \in F} \inf_{v \in V} \|f - v\|_\infty,$$

where V runs through all n-dimensional subspaces of $B(X)$.

In a certain way, both $e_n(F,S)$ and $d_n(F)$ are a measure for the size of F and it is hardly astonishing that these quantities are related.
The notions of n-widths of various kinds play an important role in the theory of optimal worst case error bounds, see 1.2.9 for references.

1.2.2

We begin with the following lemma.

Lemma
Let $n, m \in \mathbf{N}$ with $n \leq m$. Further let X be a set with *card* $(X) = m$ and $V \subset B(X)$ with *dim* $V = n$. Then there are $a_1, \dots, a_n \in X$ and $f_1, \dots, f_n \in V$ with $f_i(a_j) = \delta_{ij}$ and $\|f_i\| = 1$.

Proof:
The statement of the lemma is trivial for $n = m$. Hence we assume $k = m - n > 0$. The subspace V can be described by k linear independent equations

$$\sum_{j=1}^{m} c_{ij} f(a_j) = 0 \qquad (i = 1, \dots, k).$$

We consider all $k \times k$ submatrices of $C = (c_{ij})_{i,j}$. Because of *rank* $A = k$, there is such a submatrix whose determinant has maximum absolute value which is not zero. Without restriction we assume that $\widetilde{C} = (c_{ij})$ with $i = 1, \dots, k$ and $j = n+1, \dots, m$ is such a submatrix. Then V has a unique basis of the form $\{f_i \mid i = 1, \dots, n\}$ with $f_i(a_j) = \delta_{ij} \quad (i, j = 1, \dots, n)$. Using Cramer‘s rule we obtain $|f_i(a_j)| \leq 1$ for $j = n+1, \dots, m$.

If X is an arbitrary set, $\varepsilon > 0$, and $V \subset B(X)$ with *dim* $V = n$, then there is a finite subset $K \subset X$ such that

$$\|f_{|K}\|_\infty \geq (1 - \varepsilon) \cdot \|f\|_\infty$$

for all $f \in V$ (see the proof of Lemma 1.2.8). As a consequence, we obtain the next lemma.

Lemma
Let X be an arbitrary set and $V \subset B(X)$ with $dim\, V = n$. Further let $\varepsilon > 0$. Then there are $a_1, \ldots, a_n \in X$ and $f_1, \ldots, f_n \in V$ such that $f_i(a_j) = \delta_{ij}$ and $\|f_i\|_\infty \leq 1 + \varepsilon$.

Remark
This statement is not true for $\varepsilon = 0$, a simple counterexample is the following:
$n = 1,\ X = \mathbf{N},\ f^* \in V$ where $f^*(m) = 1 - 1/m$ for all $m \in \mathbf{N}$.
If we further assume that X is a compact space and $V \subset C(X)$, then the lemma is valid even for $\varepsilon = 0$.

1.2.3 Proposition
Let $V \subset B(X)$ with $dim\, V = n$ and let $\varepsilon > 0$. Then there are $a_1, \ldots, a_n \in X$ such that

$$\|f\|_\infty \leq (n + \varepsilon) \cdot \sup_{i=1,\ldots,n} |f(a_i)| \qquad \text{for all } f \in V.$$

Proof:
Let $a_1, \ldots, a_n \in X$ and $f_1, \ldots, f_n \in V$ such that $f_i(a_j) = \delta_{ij}$ and $\|f_i\|_\infty \leq 1 + \varepsilon/n$ (see 1.2.2). Each $f \in V$ may be written as $f = \sum_{i=1}^{n} c_i f_i$ (with $c_i \in \mathbf{R}$) and it is easy to see that

$$\|f\|_\infty \leq \sup_{i=1,\ldots,n} |f(a_i)| \cdot n \cdot (1 + \varepsilon/n)$$

and the statement follows.

Remark
We show that the constant n in this proposition is optimal. It is enough to consider the following example: $n \in \mathbf{N},\ X = \{1, 2, \ldots, n+1\}$, and $V = \{f \mid \sum_{i=1}^{n+1} f(i) = 0\}$. Then for $\{a_1, \ldots, a_n\} \subset X$ the function $f \in V$ with $f(a_i) = 1$ ($i = 1, \ldots, n$) has norm $\|f\| = n$.

1.2.4 Definition
Many problems S on $B(X)$ – formally given by an error function $\Delta : M \times B(X) \to \mathbf{R}_0^+$ – are *solvable* and *lipschitz* according to the following definition. A problem $S : B(X) \to M$ is said to be
a) solvable, if $\inf_{m \in M} \Delta(m, f) = 0$ for each $f \in B(X)$;
b) lipschitz with lipschitz constant $L \geq 0$, if

$$|\Delta(m, f_1) - \Delta(m, f_2)| \leq L \cdot \|f_1 - f_2\|$$

for all $m \in M$ and $f_1, f_2 \in B(X)$.

Remarks
It is easy to see that the problems *App*, *Opt*, and *Opt** are both solvable and lipschitz (with $L = 1$ for *App* and *Opt** and $L = 2$ for *Opt*), although there is in general no $m \in M = X$ with $\Delta(m, f) = 0$ for the problem *Opt*.
If $S = Int : B(X) \to \mathbf{R}$ is a bounded linear functional, then the problem *Int* is also solvable and lipschitz with $L = \|S\|$. The following proposition shows that *App* is "worst" of all problems S on $F \subset B(X)$ which are solvable and lipschitz with $L = 1$. We also give an upper bound for $e_n(S,F)$ in terms of $d_n(F)$.

1.2.5 Proposition
Let $F \subset B(X)$ and let S be a problem on $B(X)$ which is both solvable and lipschitz with $L = 1$. Then the following inequalities hold:

$$e_n(F,S) \leq e_n(F,App) \leq (1+n) \cdot d_n(F)$$

and

$$e_n^{ad}(F,S) \leq e_n^{ad}(F,App).$$

The constant $(1+n)$ is optimal in the following sense: For each n there is a set F with $e_n(F,App) = (1+n)\ d_n(F)$.

Proof:
First we prove that $e_n(F,S) \leq e_n(F,App)$, the same consideration is valid for e_n^{ad}.
Let $\varepsilon > 0$ and let $\widetilde{S} : F \to B(X)$ be an approximation for the problem *App* with $\widetilde{S} \in A_n$ and

$$\Delta_{max}(\widetilde{S}) \leq e_n(F,App) + \varepsilon.$$

The mapping $\widetilde{S}$ can be written as $\widetilde{S}(f) = \phi(N(f)) = \phi(f(a_1), \ldots, f(a_n))$ where $a_i \in X$ and $\phi : \mathbf{R}^n \to B(X)$. Now we define $S^* : F \to M$ by

$$S^*(f) = \phi^*(N(f))$$

where $\phi^*(N(f))$ is an element of M satisfying

$$\Delta(\phi^*(N(f)), \widetilde{S}(f)) \leq \varepsilon.$$

Then we have

$$\Delta(S^*(f), f) \leq \varepsilon + \|\widetilde{S}(f) - f\| \leq \varepsilon + \Delta_{max}(\widetilde{S}) \leq e_n(F,App) + 2\varepsilon$$

and the statement follows.
Now we prove $e_n(F,App) \leq (1+n)\ d_n(F)$.
Let $\varepsilon > 0$ be given. There is a subspace $V \subset B(X)$ with $dim\ V = n$ and

$$\inf_{g \in V} \|f - g\| \leq d_n(F) + \varepsilon$$

for every $f \in F$. Because of 1.2.3 there are $a_1, \ldots, a_n \in X$ with

$$\|g\| \leq (n+\varepsilon) \sup_{i=1,\ldots,n} |g(a_i)|$$

for all $g \in V$. For each $f \in B(X)$ there is a unique $g(f) \in V$ with

$$g(f)(a_i) = f(a_i) \quad (i = 1, \ldots, n).$$

We define $\widetilde{S} \in A_n$ by $\widetilde{S}(f) = g(f)$. For $f \in F$ we have the estimate $\|f - h\| \leq d_n(F) + \varepsilon$ for some $h \in V$. Hence

$$\sup_i |\widetilde{S}(f)(a_i) - h(a_i)| \leq d_n(F) + \varepsilon$$

and

$$\|\widetilde{S}(f) - h\| \leq (n+\varepsilon)\ (d_n(F) + \varepsilon)$$

are valid. Therefore we get

$$\|\widetilde{S}(f) - f\| \leq (n+\varepsilon+1)\ (d_n(F) + \varepsilon)$$

and the statement follows.
We show that the constant $(n+1)$ is optimal. Let $X = \{1, \ldots, n+1\}$ and

$$F = \{f \in B(X) \mid \|f - g\| \leq c \quad \text{for some } g \text{ with } \sum_{i=1}^{n+1} g(i) = 0\},$$

where $c > 0$. It is easy to see that $d_n(F) = c$ and that each $\widetilde{S} \in A_n$ is constant on a subset

$$\{f \mid f(a_i) = 0 \text{ for } i = 1, \ldots, n \text{ and } f(a_{n+1}) \in [-(n+1)c,\ (n+1)c]\ \} \subset F.$$

Hence $\Delta_{max}(\widetilde{S}) \geq (n+1)\ c$ and the proposition is proved.

1.2.6
It often occurs that $F \subset B(X)$ is not an arbitrary set, but rather convex and symmetric (i.e. $f \in F \Rightarrow -f \in F$). In this case adaption does not help for the linear problems *App* and *Int* (see 1.3.1 for other linear problems and 1.3.2 for the nonlinear problems *Opt* and *Opt**). It is also known by Smolyak‘s lemma that there is a linear optimal error algorithm (again see 1.3.1 for other linear problems). So for the problems *App* and *Int* the following statements are valid.

Proposition
Let $F \subset B(X)$ be convex and symmetric and let $n \in \mathbf{N}$. Then

$$e_n^{ad}(F,App) = e_n(F,App) = \inf_{a_i \in X,\ i=1,\ldots,n}\ \inf_{f_i \in B(X)}\ \sup_{f \in F} \|f - \sum_{i=1}^{n} f(a_i) f_i\|_\infty =$$

$$= \inf_{a_i \in X,\ i=1,\dots,n} \sup_{f\in F,\ f(a_i)=0} \|f\|_\infty \tag{1}$$

and

$$e_n^{ad}(F,Int) = e_n(F,Int) = \inf_{a_i \in X,\ i=1,\dots,n} \inf_{c_i\in \mathbf{R}} \sup_{f\in F} |S(f) - \sum_{i=1}^{n} c_i f(a_i)| =$$

$$= \inf_{a_i \in X,\ i=1,\dots,n} \sup_{f\in F,\ f(a_i)=0} S(f), \tag{2}$$

where $S = Int : B(X) \to \mathbf{R}$ is linear.

Remarks
Under the conditions of Proposition 1.2.6 arbitrary approximations $\widetilde{S} \in A_n^{ad}$ are not better than the most simple nonadaptive and linear ones, if one considers the maximum error.
By 1.2.5 and 1.2.6 we get the following result:
Let $F \subset B(X)$ be convex and symmetric and let $d_n^L(F)$ be the n-th linear width of F, defined by

$$d_n^L(F) = \inf_{P_n} \sup_{f\in F} \|f - P_n(f)\|_\infty,$$

where P_n runs through all linear $P_n : B(X) \to B(X)$ with *rank* $P_n \le n$. Then the inequality

$$d_n^L(F) \le (1+n)\cdot d_n(F)$$

is valid. This inequality is not optimal. By a theorem of *Kadec, Snobar* (1971) one can prove that

$$d_n^L(F) \le (1+\sqrt{n})\cdot d_n(F),$$

see *Pinkus* (1985). Observe, however, that we only consider those P_n, which are of the form $P_n(f) = \sum_{i=1}^n f(a_i) f_i$ for some $a_i \in X$ and $f_i \in B(X)$.

1.2.7
It is well known (although it is difficult to give a suitable reference) that the following statements are equivalent:
i) $F \subset K + V$, where $K \subset B(X)$ is compact and $V \subset B(X)$ is a finite dimensional subspace;
ii) $\lim_{n\to\infty} d_n(F) = 0$.
Now we want to study the analogous problem for the numbers $e_n(F,S)$.

Proposition
Let $F \subset B(X)$ with

$$\lim_{n\to\infty} d_n(F) = 0.$$

Then

$$\lim_{n\to\infty} e_n(F,App) = 0$$

is valid.

Proof:
We assume that $\lim_{n\to\infty} d_n(F) = 0$. Then $F \subset V + K$, where $V \subset B(X)$ has dimension $m \in \mathbf{N}$ and $K \subset B(X)$ is symmetric, convex, compact, and contains $\{f \in V \mid \|f\| \leq 1\}$. Let $\varepsilon > 0$ be given. There is a finite number of $f_1,\ldots,f_r \in K$ such that

$$\min_i \|f - f_i\| < \varepsilon$$

for all $f \in K$.
There is also a finite number of $a_1,\ldots,a_s \in X$ such that

$$\max_k |f_i(a_k) - f_j(a_k)| \geq \|f_i - f_j\| \cdot (1-\varepsilon)$$

for $i,j = 1,\ldots,r$. Assume now that $f \in K$ and $N(f) = (f(a_1),\ldots,f(a_s)) \in \mathbf{R}^s$ is given. We define

$$F_{N(f)} = \{f_i \mid i = 1,\ldots,r \text{ and } \max_k |f(a_k) - f_i(a_k)| < \varepsilon\}.$$

The set $F_{N(f)}$ is nonempty and hence there is a $\widetilde{S}_1 = \phi \circ N : K \to \{f_i \mid i = 1,\ldots,r\}$ with $\widetilde{S}_1(f) \in F_{N(f)}$. If $f_i, f_j \in F_{N(f)}$ then $\max_k |f_i(a_k) - f_j(a_k)| < 2\varepsilon$ and hence

$$\|f_i - f_j\| < \frac{2\varepsilon}{(1-\varepsilon)} .$$

Therefore we get

$$\|\widetilde{S}_1(f) - f\| < \varepsilon + \frac{2\varepsilon}{(1-\varepsilon)}$$

for all $f \in K$ and have proved that $e_n(K,App)$ tends to zero.
Because of Proposition 1.2.6 for each $\varepsilon > 0$ there is a linear $\widetilde{S}_2 \in A_{n(\varepsilon)}$ with $\|\widetilde{S}_2(f) - f\| < \varepsilon$ for all $f \in K$. Now the mapping $S^+ = S - \widetilde{S}_2 = Id - \widetilde{S}_2$ is linear and defines a problem which is solvable and lipschitz with a lipschitz constant less than ε. By Proposition 1.2.5 we get

$$e_m(V+K,S^+) \leq (1+m) \cdot d_m(V+K) \cdot \varepsilon,$$

where $d_m(V+K)$ is some finite number. Hence there is a $\widetilde{S}_3 \in A_m$ with

$$\|\widetilde{S}_3(f) - S^+(f)\| < (1+m) \cdot d_m(V+K) \cdot \varepsilon + \varepsilon$$

for all $f \in V+K$. For $\widetilde{S} = \widetilde{S}_2 + \widetilde{S}_3 \in A_{n(\varepsilon)+m}$ we get

$$\|\widetilde{S}(f) - f\| = \|\widetilde{S}_3(f) - S^+(f)\| < (1+m) \cdot d_m(V+K) \cdot \varepsilon + \varepsilon$$

and hence $\lim_{n\to\infty} e_n(V+K,App) = 0$ which proves the statement.

Remarks

a) Let $F \subset B(X)$ be of the form $F \subset K+V$ where $K \subset B(X)$ is compact and $V \subset B(X)$ is finite dimensional. If $S : F \to M$ is both solvable and lipschitz, then

$$\lim_{n\to\infty} e_n(F,S) = 0.$$

This follows from 1.2.5 and 1.2.7.

b) It is easy to construct an example with $e_n(F,App) \to 0$ but $d_n(F)$ does not tend to zero. This is not possible if we assume that F is convex and symmetric. By 1.2.6 we get $e_n(F,App) \geq d_n^L(F) \geq d_n(F)$ and hence

$$\lim_{n\to\infty} d_n(F) = 0 \quad \Longleftrightarrow \quad \lim_{n\to\infty} e_n(F,App) = 0.$$

c) Let $S : F \to M$ be a linear problem (see 1.3.1 for the definition) with $\lim e_n^{ad}(F,S) = 0$. It can be proved that in this case $S(F) \subset M$ is of the form $S(F) \subset K+V$ where K is compact and V is finite dimensional. This condition, however, is not sufficient for $\lim e_n^{ad}(F,S) = 0$. Consider, for example, the problem $S = Int = \lambda$ with $F = \{f \in C([0,1]) \mid \|f\|_\infty \leq 1\}$.

1.2.8 The problem *Int*

Let $S = Int : B(X) \to \mathbf{R}$ be a bounded and linear mapping. By means of Proposition 1.2.5 we have

$$e_n(F,S) \leq (1+n)\cdot \|S\| \cdot d_n(F)$$

for each $F \subset B(X)$. This estimate is not optimal, instead we can prove the following stronger result (see *Novak* (1983 and 1986a)).

Proposition

Let $S = Int : B(X) \to \mathbf{R}$ be bounded and linear. Then

$$e_n(F,S) \leq 2\cdot \|S\| \cdot d_n(F)$$

for each $F \subset B(X)$ and the constant 2 is the best possible.

By means of this proposition one can prove known error estimates as well as some new error estimates for quadrature formulas. In *Novak* (1983 and 1986a) one can find applications to Hölder, Sobolev, and Besov spaces (see also Section 1.3).

To prove this inequality, we need the following lemma which is a version of a well known interpolation theorem, see *Shapiro* (1971).

Lemma. Let $V \subset B(X)$ be a vector space with $dim\ V = n$ and let $L : V \to \mathbf{R}$ be a linear mapping. Then for each $\varepsilon > 0$ there exist $a_1, \ldots, a_n \in X$ and $c_1, \ldots, c_n \in \mathbf{R}$ with

$$L(f) = \sum_{i=1}^{n} c_i f(a_i)$$

for all $f \in V$ and

$$\|L\| \leq \sum_{i=1}^{n} |c_i| \leq \|L\| + \varepsilon.$$

Proof:

If X is compact and V consists of continuous functions, this statement even holds for $\varepsilon = 0$ (this is the mentioned interpolation theorem, see *Shapiro* (1971)). We reduce the general case to this special case: The set $M = \{f \in V \mid \|f\| = 1\}$ is compact, for $\delta_1 > 0$ exist $f_1, \ldots, f_m \in M$ with

$$M = \bigcup_{i=1}^{m} \{f \in M \mid \|f - f_i\| \leq \delta_1\}.$$

For $\delta_2 > 0$ let b_i be given so that $|f_i(b_i)| \geq 1 - \delta_2$ ($i = 1, \ldots, m$).
By eventually making $K' = \{b_1, \ldots, b_m\}$ larger we can assume that for $V' = \{f_{|K} \mid f \in V\}$ with K finite and $K' \subset K$ the statement $dim\ V' = n$ holds. Now we apply the special case to K, V', and L' defined by $L'(f_{|K}) = L(f)$.
Because of

$$\|f_{|K}\| \geq \|f\| \cdot (1 - \delta_1 - \delta_2)$$

for all $f \in V$ we get

$$L(f) = \sum_{i=1}^{n} c_i f(a_i)$$

(for all $f \in V$ and suitable $c_i \in \mathbf{R}$, $a_i \in K$) with

$$\|L\| \leq \sum_{i=1}^{n} |c_i| = \|L'\| \leq \|L\| \cdot (1 - \delta_1 - \delta_2)^{-1}$$

and the statement follows.

Proof of the Proposition:

Let F, S, n and $\varepsilon > 0$ be given. There is a vector space $V \subset B(X)$ with $dim\ V = n$ and

$$\sup_{f \in F} \inf_{g \in V} \|f - g\| \leq d_n(F) + \varepsilon.$$

Because of the last lemma there is a $L : B(X) \to \mathbf{R}$ of the form $L(f) = \sum_{i=1}^{n} c_i f(a_i)$ with $L(f) = S(f)$ for all $f \in V$ and $\|L\| \leq \|S\| + \varepsilon$. Therefore

$$|L(f) - S(f)| \leq (d_n(F) + \varepsilon) \cdot (\|L\| + \|S\|) \leq (d_n(F) + \varepsilon) \cdot (2\|S\| + \varepsilon)$$

for all $f \in F$ and the estimate

$$e_n(F, S) \leq 2\|S\| d_n(F)$$

follows.

We have to show that the constant 2 is best possible. Let $n, m \in \mathbf{N}$ and let $X = \{a_i \mid i = 1, \ldots, n \cdot m\}$ be a set with $n \cdot m$ elements. We consider $F \subset B(X)$ given by

$$F = \{f \in B(X) \mid f(a_1) = \ldots = f(a_m),\ f(a_{m+1}) = \ldots = f(a_{2m}), \ldots,$$
$$f(a_{(n-1)m+1}) = \ldots = f(a_{nm})\} + \ \{f \in B(X) \mid |f(a_i)| \leq 1 \ \text{ for all } i\}.$$

We have $d_n(F) = 1$ and for

$$S(f) = \sum_{i=1}^{nm} f(a_i)$$

we get $e_n(F, S) = 2n(m-1)$ and hence

$$e_n(F, S) = (2 - 2/m) \cdot \|S\| \cdot d_n(F).$$

1.2.9 Notes and references

Classical surveys on n-widths are *Tikhomirov* (1960) and Chapter 9 of *Lorentz* (1966), a more recent one is contained in *Tikhomirov* (1976). A monograph on n-widths is *Pinkus* (1985). Many interesting relations between n-widths of various kinds and optimal (worst case) error bounds for numerical methods are discussed in *Micchelli, Rivlin* (1977) and in *Traub, Woźniakowski* (1980). See also *Korneicuk* (1981), *Micchelli, Rivlin* (1985) and *Pinkus* (1986). This section contains results of *Novak* (1983 and 1986a), where the problem *Int* is treated. Both papers also deal with stochastic error bounds for Monte Carlo methods and relations to the (Kolmogorov) n-widths in the space L_2 (see 2.2.9). Other widths and their connections to problems of numerical analysis are considered in *Babenko* (1985).

1.3 Results for special problems

We begin with arbitrary linear problems (1.3.1). Then we study the worst case error bounds $e_n(F,S)$ and $e_n^{ad}(F,S)$ for the problems *App*, *Opt*, *Opt**, and *Int* and different classes F of functions. These results are of practical interest. In addition, they show certain limits of a pure worst case analysis. In later chapters we will compare these error bounds with different stochastic error bounds.

1.3.1 Linear problems

A problem $S : F \to M$ is called linear, if the following conditions are satisfied:

1) F is a convex and symmetric (i.e. $f \in F$ implies $-f \in F$) set;

2) M is a normed space and $S : F \to M$ is linear;

3) the error function Δ is given by $\Delta(\widetilde{S}(f), f) = \|S(f) - \widetilde{S}(f)\|$.

It is well known that $e_n^{ad}(F,S) \geq \frac{1}{2} e_n(F,S)$ for any linear problem, see *Gal, Micchelli* (1980) and *Traub, Woźniakowski* (1980). It is easy to sharpen this inequality using the notion of Jung's constant.

Definition

Let M be a normed space. The Jung constant $c(M)$ is defined by

$$c(M) = \sup\{rad(D)/diam(D) \mid \emptyset \neq D \subset M \text{ is bounded}\}.$$

Here $rad(D)$ and $diam(D)$ are defined by

$$rad(D) = \inf_{x \in M} \sup_{y \in D} d(x,y) \qquad \text{and} \qquad diam(D) = \sup_{x,y \in D} d(x,y).$$

Proposition 1

Let $S : F \to M$ be a linear problem. Then

$$\frac{1}{2} \leq \frac{1}{2c(M)} \leq \frac{e_n^{ad}(F,S)}{e_n(F,S)} \leq 1.$$

Proof:

We use the following notation. Let $N \in I_n^{ad}$ be given by

$$N(f) = (f(a_1), f(a_2(f(a_1))), \ldots, f(a_n(f(a_1), \ldots, f(a_{n-1})))).$$

Then we define the nonadaptive information $N_g \in I_n$ by

$$N_g(f) = (f(a_1), f(a_2(g(a_1))), \ldots, f(a_n(g(a_1), \ldots, g(a_{n-1})))).$$

The mapping N_0 is linear and for $x = N_0(f)$, $f \in F$, we have $S(N_0^{-1}x) = \{S(\tilde{f}) \in M \mid \tilde{f} \in F,\ N_0(\tilde{f}) = N_0(f)\}$. We get

$$diam(S(N_0^{-1}(x))) \leq 2\sup\{\|S(f)\| \mid f \in F,\ N_0(f) = 0\} = diam(S(N_0^{-1}(0))).$$

Thus

$$diam(S(N_0^{-1}(0))) \geq diam(S(N_0^{-1}(x))) \geq \frac{1}{c(M)} rad(S(N_0^{-1}(x))).$$

Since this holds for all x we obtain $diam(S(N_0^{-1}(0))) \geq \frac{1}{c(M)} r_{max}(N_0)$. Thus

$$r_{max}(N) \geq rad(S(N^{-1}(0))) = rad(S(N_0^{-1}(0))) = \frac{1}{2} diam(S(N_0^{-1}(0))) \geq$$
$$\geq \frac{1}{2c(M)} r_{max}(N_0).$$

Together with $c(M) \leq 1$ this completes the proof.

The following results concerning $c(M)$ are well known, see *Holmes* (1972) and *Amir* (1985). Together with Proposition 1 these results yield lower bounds for the adaption constant $\alpha(M)$, which we define by

$$\alpha(M) = \inf\{e_n^{ad}(F,S)/e_n(F,S) \mid S : F \to M \text{ is a linear problem and } n \in \mathbf{N}\}.$$

In particular, we obtain $\alpha(L_\infty(\mu)) = 1$, i.e. adaption can not help in the case $M = L_\infty(\mu)$. This result is essentially due to *Bakhvalov* (1971).

Proposition 2

1) $1/2 \leq c(M) \leq 1$ for every normed space M;
2) $c(l_2^m) = \left(\frac{m}{2m+2}\right)^{1/2}$, where l_2^m is the euclidean space $\mathbf{R}^m$;
3) $c(H) = 2^{-1/2}$, where H is an infinite dimensional Hilbert space;
4) $c(C[0,1]) = 1$ and more general $c(C(K)) = 1$ if K is a compact space which is not extremally disconnected;
5) $c(L_\infty(\mu)) = 1/2$ and more general: $c(M) = 1/2$ if and only if M is a $\mathcal{P}_1$-space, i.e. M is isometric with some $C(K)$, where K is a extremally disconnected compact space.

Example

It was an unsolved problem for some years, whether there is a linear problem where adaptive methods are really better than nonadaptive ones, see Problem 1 of *Packel, Woźniakowski* (1986). The following example was independently found by Mark A. Kon and the author in 1987, see *Kon, Novak* (1988) for more details.

Let $K \subset \mathbf{R}^3$. We define

$$R_x(K) = \sup_{d \in \mathbf{R}} rad\{k \in K \mid k \cdot x = d\} \qquad (x \in \mathbf{R}^3,\ x \neq 0)$$

and

$$R(K) = \inf_{x \in \mathbf{R}^3, x \neq 0} R_x(K).$$

We assume that $T \subset \mathbf{R}^3$ is the convex hull of the following six points

$$(0, \sqrt{3}/3, \pm 1/4) \qquad (\pm 1/2, -\sqrt{3}/6, \pm 1/4).$$

It can be proved that $R_{(0,0,1)}(T) = \sqrt{3}/3 \approx 0.577$, $R_{(1,0,0)}(T) = 1/2$, and $R(T) = R_{(1,0,0)} = 1/2$. For $M_0 = \{(t_1 + t_2)/2 \mid t_1 \in T, t_2 \in -T\}$ we get $R(M_0) = R_{(0,0,1)}(M_0) = 1/2$. The set M_0 is the convex hull of the points

$$(\pm 1/4, \pm\sqrt{3}/4, \pm h_0) \quad \text{and} \quad (\pm 1/2, 0, \pm h_0),$$

where $h_0 = 1/4$. If we replace h_0 by $h = \sqrt{21}/12 \approx 0.382$ we obtain a convex and symmetric set $M \subset \mathbf{R}^3$ with $R(M) = R_{(0,0,1)}(M) = 1/2$ and we also have $R_{(1,0,0)}(M) = \sqrt{3}/3 \approx 0.577$. Now we consider the information

$$x_\lambda = \begin{cases} (0,0,1), & \text{if } \lambda \leq \frac{7}{12} \\ (1,0,0), & \text{if } \lambda > \frac{7}{12} \end{cases}$$

for the convex body $M_\lambda = \lambda T + (1-\lambda)M$. We obtain

$$R(M_\lambda) \leq R_{x_\lambda}(M_\lambda) < 0.530 \qquad (0 \leq \lambda \leq 1). \tag{1}$$

It can be shown, however, that there is no single $x \in \mathbf{R}^3$ with

$$R_x(M_\lambda) < \sqrt{3}/3 \approx 0.577 \tag{2}$$

for all λ. Now we can give an example for a linear problem such that adaption helps in the worst case. Let $F \subset \mathbf{R}^4$ be the convex hull of the points

$$\{\pm(x_1,x_2,x_3,x_4) \mid (x_1,x_2,x_3) \in T \text{ and } x_4 = k\},\ \ \{(x_1,x_2,x_3,0) \mid (x_1,x_2,x_3) \in M\},$$

where $k > 0$. We consider the linear problem

$$S : F \to l_2^4, \qquad S(x) = x$$

(observe that F is a symmetric and convex set). It follows from (1) that

$$e_2^{ad}(F,S) < 0.530$$

(we use the information $I_1(x) = x_4$ and $I_2(x) = x_3$ if $|x_4| \leq \frac{7k}{12}$ and $I_2(x) = x_1$ if $|x_4| > \frac{7k}{12}$). Each nonadaptive method $\widetilde{S} \in A_2$, however, satisfies

$$\Delta_{max}(\widetilde{S}) > \sqrt{3}/3 - \varepsilon \approx 0.577 - \varepsilon,$$

where $\varepsilon > 0$ and $k > k(\varepsilon)$. This follows from (2).

Remark
As a corollary, we obtain $0.790 \leq \alpha(l_2^4) \leq 0.918$ (and it is easy to prove that $\alpha(M) = 1$ if $dim(M) < 4$). Our example uses a construction in a four dimensional space. Therefore it is interesting to get estimates for higher dimensions. Here we prove the following.

Proposition 3
Let M be a subspace of the normed space $\widetilde{M}$ such that there is a norm-1-projection $P : \widetilde{M} \to M$. Then we have

$$c(\widetilde{M}) \geq c(M) \qquad \text{and} \qquad \alpha(\widetilde{M}) \leq \alpha(M).$$

Proof:
Let $\varepsilon > 0$. Then there is a bounded set $D \subset M$ with

$$\frac{rad(D)}{diam(D)} \geq c(M) - \varepsilon.$$

Here $rad(D) = rad_M(D)$ is the radius of D in M. Because of the existence of a norm-1-projection $P : \widetilde{M} \to M$, we have

$$rad_M(D) = rad_{\widetilde{M}}(D)$$

(see Garkavi (1964)) and hence

$$\frac{rad_{\widetilde{M}}(D)}{diam(D)} \geq c(M) - \varepsilon.$$

We conclude that $c(\widetilde{M}) \geq c(M)$. Now let $S : F \to M$ be a linear problem and $n \in \mathbf{N}$. Then we also have $S_1 : F \to \widetilde{M}$ with the same operator $S_1 = S$. We show that

$$e_n^{(ad)}(F, S_1) = e_n^{(ad)}(F, S)$$

which proves our statement. Let $N : F \to \mathbf{R}^n$ be some (adaptive or nonadaptive) information operator. Then the radius $r_M(N)$ of N with respect to the problem $S : F \to M$ is given by

$$r_M(N) = \sup_{x \in \mathbf{R}^n} rad_M S(N^{-1}(x))$$

and an analogous formula holds for $r_{\widetilde{M}}(N)$. It follows that $r_M(N) = r_{\widetilde{M}}(N)$.

The following Proposition 4 deals with L_p-spaces. Our result implies that $\alpha(M) < 1$ (and adaption helps for certain linear problems) if M is a L_p-space with $dim M > 3$ and $|p - 2| < \delta$ for some $\delta > 0$. A more careful analysis of our example in four dimensions would probably show that $\alpha(l_p^4) < 1$ for all $1 < p < \infty$. It would be interesting to have more information on the numbers $\alpha(l_p^m)$.

Proposition 4
i) $c(l_p^m)$ and $\alpha(l_p^m)$ depend (for fixed m) continuously on p, $1 \leq p \leq \infty$.
ii) $\alpha(l_p^m) \geq \alpha(l_p^{m+1}) \geq \alpha(l_p^{\mathbf{N}}) \geq \alpha(L_p(\mathbf{R}))$.

Proof:
First we prove i). Let $D \subset \mathbf{R}^m$ be a bounded set and let $rad_p(D)$ and $diam_p(D)$ denote its radius and diameter in l_p^m. It is easy to prove the following estimates for $1 \leq p < \tilde{p} \leq \infty$.

$$rad_p(D) \geq rad_{\tilde{p}}(D) \geq m^{1/\tilde{p}-1/p} \cdot rad_p(D)$$

and

$$diam_p(D) \geq diam_{\tilde{p}}(D) \geq m^{1/\tilde{p}-1/p} \cdot diam_p(D).$$

It follows that $p \mapsto c(l_p^m)$ is a continuous function, $1 \leq p \leq \infty$. Let $S : F \to \mathbf{R}^m$ be a given problem, let $N : F \to \mathbf{R}^n$ be a given information and let $r_p(N)$ be the radius of N with respect to the norm of l_p^m in $\mathbf{R}^m$. It follows that

$$r_p(N) \geq r_{\tilde{p}}(N) \geq m^{1/\tilde{p}-1/p} \cdot r_p(N)$$

and the continuity of $p \mapsto \alpha(l_p^m)$ can easily be concluded. The inequalities in ii) follow from Proposition 3 because the projections

$$P : l_p^{m+1} \to l_p^m \qquad P((x_1,\ldots,x_{m+1})) = (x_1,\ldots,x_m),$$

$$P : l_p^{\mathbf{N}} \to l_p^m \qquad P((x_1,x_2,\ldots)) = (x_1,\ldots,x_m),$$

$$P : L_p(\mathbf{R}) \to l_p^{\mathbf{N}} \qquad P(f) = (\int_0^1 f dx, \int_1^2 f dx, \int_2^3 f dx, \ldots)$$

all have norm 1.

Remark on linear algorithms
We have seen that adaptive methods are only slightly better than nonadaptive ones for linear problems. Let $N \in A_n$ be a given nonadaptive information,

$$N(f) = (f(a_1), \ldots f(a_n)).$$

We could ask whether there is a linear method

$$\widetilde{S}(f) = \phi \circ N(f) = \sum_{i=1}^{n} f(a_i) \cdot m_i$$

(with $m_i \in M$) such that $r(N) = \Delta_{max}(\widetilde{S})$. This is true for the problems *App* and *Int* (see 1.2.6) and more general in the case $M = \mathbf{R}$ or $M = B(X)$ (Smolyak‘s lemma, see *Bakhvalov* (1971), *Sukharev* (1986), and *Osipenko* (1976) for the complex case). However, there are linear problems for which no linear optimal algorithm exists, see *Packel* (1986) and *Werschulz, Woźniakowski* (1986).

1.3.2 Proposition
Let $F \subset B(X)$ be convex and symmetric, and let $n \in \mathbf{N}$. Then

$$e_{n+1}(F,App) \le e_n^{ad}(F,Opt) \le e_n(F,Opt) \le 2 \cdot e_n(F,App).$$

Proof: The inequalities $e_n^{ad}(F,Opt) \le e_n(F,Opt) \le 2e_n(F,App)$ follow from 1.2.5. Also we know from 1.2.6 that

$$e_{n+1}(F,App) = \inf_{a_i \in X i=1,\dots,n+1} \sup_{f \in F f(a_i)=0} \|f\|_\infty.$$

Let $\widetilde{S} \in A_n^{ad}$ be an approximation for the problem *Opt*, $\widetilde{S} = \phi \circ N$ with $N_0(f) = (f(a_1),\dots,f(a_n))$ and $\phi(N_0(0)) = \phi(0) = a_{n+1} \in X$. We have

$$\begin{aligned}
\Delta_{max}(\widetilde{S}) \quad &\ge \ \sup\,\{\sup f - f(a_{n+1}) \mid f \in F,\ f(a_i) = 0 \ \text{ for } i = 1,\dots,n\} \\
&\ge \ \sup\,\{\sup f \mid f \in F,\ f(a_i) = 0, \ \text{ for } i = 1,\dots,n+1\} \\
&= \ \sup\,\{\|f\| \mid f \in F,\ f(a_i) = 0 \ \text{ for } i = 1,\dots,n+1\} \\
&\ge \ e_{n+1}(F,App).
\end{aligned}$$

Remarks
a) The paper *Wasilkowski* (1984) contains a slightly different result, similar problems have been treated in *Sukharev* (1971), *Zaliznyak, Ligun* (1978), and in *Sukharev* (1985).
b) We observe two properties of the problems *Opt* and *App* which follow from 1.3.2 (see also 1.2.6 for the problems *App* and *Int* and 1.3.3 for the problem *Opt**).
i) Although the problem *Opt* seems to be easier than *App*, this is not the case when considering convex symmetric sets $F \subset B(X)$ and worst case errors.
ii) Under the same conditions adaption does not help for both problems. The frequent use of adaptive methods for the problem *Opt* can not be justified by a worst case analysis of deterministic methods if the underlying class of functions is convex and symmetric.

The next proposition which is due to *Wasilkowski* (1984) treats the problem Opt^* for which similar statements hold.

1.3.3 Proposition
Let $F \subset B(X)$ be convex and symmetric. Then

$$1/2 \cdot e_n(F,App) \le e_n^{ad}(F,Opt^*) \le e_n(F,Opt^*) \le e_n(F,App).$$

Proof:
Let $\widetilde{S} \in A_n^{ad}$, $\widetilde{S} = \phi \circ N$ and $\phi(N(0)) = m \in \mathbf{R}$. For all $f \in F$ with $N(f) = 0$ we have $\widetilde{S}(f) = m$ and hence

$$\Delta_{max}(\widetilde{S}) \geq 1/2 \cdot \sup_{f \in F, N(f)=0} \sup f \geq 1/2 \cdot e_n(F, App),$$

and therefore $e_n^{ad}(F, Opt^*) \geq 1/2 \cdot e_n(F, App)$. The inequality $e_n(F, Opt^*) \leq e_n(F, App)$ follows from 1.2.5.

1.3.4 Remark
The statements 1.2.6, 1.3.2, and 1.3.3 show the important role of the numbers

$$\inf_{a_1,\dots,a_n} \sup_{f \in F f(a_i)=0} \|f\|_\infty$$

for the problems *App*, *Opt*, and Opt^*, if $F \subset B(X)$ is a convex and symmetric set. These numbers $e_n(F, App)$ have been calculated or estimated for many classes F of functions.
In the following, we specify those results for some classes of Lipschitz, Hölder, and Sobolev functions. We also treat the problem *Int* for these classes.
We begin with some general error bounds which are easy to prove but useful for certain applications.

1.3.5 Proposition
Let $F \subset B(X)$ and $n \in \mathbf{N}$.
a) Assume that $f_1, \dots, f_{n+1} \in F$ have mutually disjoint supports, $-f_i \in F$, and $\|f_i\| \geq \varepsilon$ for all $i = 1, \dots, n+1$. Then

$$e_n^{ad}(F, App) \geq \varepsilon, \qquad e_{n-1}^{ad}(F, Opt) \geq \varepsilon,$$

and

$$e_n^{ad}(F, Opt^*) \geq 1/2 \cdot \varepsilon.$$

b) Assume that $f_1, \dots, f_m$ (where $m > n$) have mutually disjoint supports and

$$\sum_{i=1}^m \delta_i f_i \in F$$

for all $\delta_i \in \{-1, 1\}$. Let $Int(f_i) \geq \varepsilon$ for $i = 1, \dots, m$. Then

$$e_n^{ad}(F, Int) \geq (m-n) \cdot \varepsilon.$$

c) Assume that $f_1, \dots, f_{n+1}$ have mutually disjoint supports, $Int(f_i) \geq \varepsilon$, and $\pm f_i \in F$ for $i = 1, \dots, n+1$. Then

$$e_n^{ad}(F, Int) \geq \varepsilon.$$

Proof:
The proof of these lower bounds can easily be given by means of the radius of information, see 1.1.2. To give an example, we show b). Let $\widetilde{S} \in A_n^{ad}$ and $f = \sum_{i=1}^m f_i$. Then $\widetilde{S}$ uses some information $N_f(f) = (f(a_1), \ldots, f(a_n))$ to compute $\widetilde{S}(f) = \phi \circ N_f(f)$. Renumbering the f_i we can assume that $N_f(f^-) = N_f(f)$, where

$$f^- = \sum_{i=1}^n f_i - \sum_{i=n+1}^m f_i.$$

Because of $|S(f) - S(f^-)| \geq 2 \cdot (m-n) \cdot \varepsilon$ and $\widetilde{S}(f) = \widetilde{S}(f^-)$ the maximal error of $\widetilde{S}$ must be at least $(m-n) \cdot \varepsilon$.

1.3.6 The problems *App*, *Opt*, and *Opt** for classes of Lipschitz functions and related classes of functions

Let (X,d) be a bounded metric space and let $f : X \to \mathbf{R}$ be a continuous function. We put

$$\omega(f,h) = \sup\{|f(x) - f(y)| \mid d(x,y) \leq h\}.$$

The function $\omega(f,.) : \mathbf{R}^+ \to \mathbf{R}_0^+ \cup \{\infty\}$ is called the modulus of continuity of f. If the function f is uniformly continuous then $\omega(f,.)$ has the following properties:

(1) $\omega(f,.)$ is nondecreasing, finite, and continuous;
(2) $\lim_{h \to 0} \omega(f,h) = 0$;
(3) $\omega(f, h_1 + h_2) \leq \omega(f,h_1) + \omega(f,h_2)$.

In the following we always assume that $\omega : \mathbf{R}^+ \to \mathbf{R}^+$ is a function which satisfies (1)–(3). We consider the set

$$F_\omega = \{f : X \to \mathbf{R} \mid \omega(f,h) \leq \omega(h) \ \text{ for all } h > 0\}.$$

First we note that F_ω is a convex and symmetric subset of $B(X)$. Let $n \in \mathbf{N}$. We define the n-th covering constant c_n of X by

$$c_n = \inf_{a_1,\ldots,a_n \in X} \sup_{a \in X} \inf_i d(a,a_i).$$

Proposition

$$e_n^{ad}(F_\omega, App) = e_n(F_\omega, App) = \omega(c_n) \tag{1}$$

$$e_n^{ad}(F_\omega, Opt) = e_n(F_\omega, Opt) = \omega(c_n) \tag{2}$$

$$e_n^{ad}(F_\omega, Opt^*) = e_n(F_\omega, Opt^*) = 1/2 \cdot \omega(c_n). \tag{3}$$

Proof:
(1) Using 1.2.6 we get

$$e_n^{ad}(F_\omega,App) = e_n(F_\omega,App) = \inf_{a_1,\ldots,a_n\in X} \sup_{f\in F_\omega,\ f(a_i)=0} \|f\|_\infty$$
$$= \inf_{a_1,\ldots,a_n\in X} \sup \{\omega(\inf_i d(a,a_i)) \mid a\in X\} = \omega(c_n).$$

(2) Let $\widetilde{S} = \phi\circ N \in A_n^{ad}$ for the problem Opt, let $N_0(f) = (f(a_1),\ldots,f(a_n))$ be the information which is taken for $g=0$ and let $\widetilde{S}(0) = a_{n+1}$. We assume that $\Delta_{max}(\widetilde{S}) = \omega(c_n)-\varepsilon$ with $\varepsilon>0$. We define

$$f_i(a) = \omega(d(a,a_i)) \qquad (i=1,\ldots,n)$$
$$f_{n+1}(a) = \omega(d(a,a_{n+1})) - \omega(\min_{i=1,\ldots,n} d(a_i,a_{n+1})) \quad \text{and}$$
$$f(a) = \min_{i=1,\ldots,n+1} f_i(a).$$

We have $f_i\in F_\omega$, hence $f\in F_\omega$, and $f(a_i)=0$ for $i=1,\ldots,n$ and therefore $\widetilde{S}(f) = a_{n+1}$ and

$$\sup f(a) - f(a_{n+1}) = \sup f(a) + \omega(\min_{i=1,\ldots,n} d(a_i,a_{n+1})) \le$$
$$\le \Delta_{max}(\widetilde{S}) = \omega(c_n)-\varepsilon.$$

Without restriction we assume that

$$d(a_1,a_{n+1}) = \min_{i=1,\ldots,n} d(a_i,a_{n+1}).$$

It follows that for each $a\in X$ there is some j with

$$f_j(a) + \omega(d(a_1,a_{n+1})) \le \omega(c_n)-\varepsilon.$$

If $j\in\{1,\ldots,n\}$ then

$$\omega(d(a,a_j)) + \omega(d(a_1,a_{n+1})) \le \omega(c_n)-\varepsilon$$

and

$$\omega(\min_{i=2,\ldots,n+1} d(a,a_i)) \le \omega(c_n)-\varepsilon.$$

This last inequality is also valid if $j=n+1$ and therefore it holds for all $a\in X$.
This contradicts the definition of the c_n and thus our assumption $\Delta_{max}(\widetilde{S}) < \omega(c_n)$ must be wrong and hence $e_n^{ad}(F_\omega,Opt)\ge\omega(c_n)$.
To prove $e_n(F_\omega,Opt)\le\omega(c_n)$ it is sufficient to observe that for given $a_i\in X$ the method

$$\widetilde{S}(f) = a_i,$$

where i is the smallest number for which

$$f(a_i) = \max_{k=1,\ldots,n} f(a_k),$$

satisfies $\widetilde{S} \in A_n$ and

$$\Delta_{max}(\widetilde{S}) = \sup_{a \in X} \inf_i \omega(d(a, a_i)).$$

(3) In view of 1.3.3 it is enough to show that $e_n(F_\omega, Opt^*) \leq 1/2 \cdot \omega(c_n)$.
Let $a_1, \ldots, a_n \in X$ and define

$$\widetilde{S}(f) = \max_i f(a_i) + 1/2 \cdot \omega(c_n).$$

Then $\widetilde{S} \in A_n$ and

$$\Delta_{max}(\widetilde{S}) = \sup_{a \in X} \inf_i \omega(d(a, a_i)) - 1/2 \cdot \omega(c_n)$$

which proves the statement.

Remarks
Statement (1) is due to *Sukharev* (1978), where the problem *App* is treated in more detail.
The problem *Opt* is studied in *Sukharev* (1984). Observe, however, that only methods of the form

$$\widetilde{S}(f) = a_{i_0}, \qquad \text{where} \quad N(f) = (f(a_1), \ldots, f(a_n)) \quad \text{and} \quad f(a_{i_0}) = \max f(a_i),$$

are considered. These methods are in general not optimal, we give a simple example:
Let

$$F_1 = \{f : [0,1] \to \mathbf{R} \mid |f(x) - f(y)| \leq |x - y|\}$$

and $N(f) = f(a_1)$. Then the method $\widetilde{S}(f) = 1/2$ is optimal — independent of a_1 (see *Novak* (1986c), where this example is discussed in more detail).
Methods of the form $\widetilde{S}(f) = \max_{i=1,\ldots,n} f(a_i)$ for the problem Opt^* are treated in *Sukharev* (1971), *Sukharev* (1972), *Sukharev* (1981b), *Niederreiter* (1983), and *Niederreiter* (1985a). The optimal methods are indicated in *Sukharev* (1984).
Some of these papers also deal with estimates for the numbers c_n. A classical reference for covering problems in $\mathbf{R}^n$ is *Rogers* (1964). See also the surveys *Kolmogorov, Tikhomirov* (1959) and Chapter 10 of *Lorentz* (1966).
For later reference we give a very simple example for Proposition 1.3.6.

Example
Let $s \in \mathbf{N}$ and $X = [0,1]^s$ with $d(x,y) = \max_i |x_i - y_i|$.
Then $c_n = 1/2 \cdot m^{-1}$ for $n = m^s$ up to $n = (m+1)^s - 1$, where $m \in \mathbf{N}$. In special, we get for

$$F_s = \{f : [0,1]^s \to \mathbf{R} \mid |f(x) - f(y)| \leq \max_i |x_i - y_i|\} :$$

$e_n(F_s, App) = 1/2 \cdot m^{-1}$ for $n = m^s$ up to $n = (m+1)^s - 1$ ($m \in \mathbf{N}$) and hence

$$e_n(F_s, App) \approx 1/2 \cdot n^{-1/s},$$

where $a_n \approx b_n$ denotes the strong equivalence of sequences, i.e.

$$\lim_{n\to\infty} a_n/b_n = 1.$$

We see that the maximal error of optimal methods $S^* \in A_n$ for solving the problem *App*, *Opt*, or *Opt** may tend to zero (with $n \to \infty$) rather slowly. The use of adaptive methods does not improve these error bounds.

1.3.7 n-widths for classes of Lipschitz functions and related classes of functions

We have seen that the problems *App*, *Opt*, and *Opt** for Lipschitz classes are intimately connected with a covering problem for the set X.
Similarly, the $d_n(F_\omega)$ can be estimated (and can be exactly computed in some cases) by means of packing and covering X with equal spheres.
First we define the n-th packing constant p_n of X by

$$p_n = \sup_{a_i \in X,\ i=1,\dots,n} \inf_{i \neq j} 1/2 \cdot d(a_i, a_j).$$

Now we can prove the following proposition.

Proposition

$$1/2 \cdot \omega(2p_{n+1}) \leq d_n(F_\omega) \leq 1/2 \cdot \omega(2c_n).$$

Proof:
Let $\varepsilon > 0$ and $a_1, \dots, a_{n+1} \in X$ with

$$\inf_{i\neq j} 1/2 \cdot d(a_i, a_j) \geq p_{n+1} - \varepsilon.$$

For each choice of $\delta_i \in \{-1, 1\}$ $(i = 1, \dots, n+1)$ we define a function f by

$$f(x) = 1/2 \cdot \delta_i \cdot \omega(2(p_{n+1} - \varepsilon) - 2d), \text{ if } d = d(x, a_i) < p_{n+1} - \varepsilon \text{ for some } i$$

and 0 otherwise.
It is easy to see that $f \in F_\omega$ and $f(a_i) = 1/2 \cdot \omega(2(p_{n+1} - \varepsilon)) \cdot \delta_i$.
By a well known theorem of *Tikhomirov* (1960) and *Lorentz* (1960) it follows that $d_n(F_\omega) \geq 1/2 \cdot \omega(2p_{n+1})$.
To prove the other inequality we assume that $\varepsilon > 0$ and $a_1, \dots, a_n \in X$ with

$$\sup_{x\in X} \inf_i d(x, a_i) < c_n + \varepsilon.$$

We also assume that $\{X_1,\ldots,X_n\}$ forms a partition of X such that

$$X_i \subset \{x \mid d(x,a_i) < c_n + \varepsilon\}.$$

Clearly the diameter of X_i is less than $2(c_n+\varepsilon)$ and therefore each $f \in F_\omega$ can be approximated by a g such that

$$\|f-g\| < 1/2 \cdot \omega(2(c_n+\varepsilon))$$

and $g_{|X_i}$ is a constant for each i (define g by $g_{|X_i} = 1/2 \cdot (\sup_{x\in X_i} f(x) + \inf_{x\in X_i} f(x))$).

Example
Let $X = [0,1]^s$ with $d(x,y) = \max_i |x_i - y_i|$.
It is easy to prove that $p_n = 1/2 \cdot m^{-1}$ for $n = m^s+1$ up to $n = (m+1)^s$, where $m \in \mathbf{N}$.
Hence we have $c_n = p_{n+1}$ for all $n \in \mathbf{N}$ and Proposition 1.3.7 yields the exact values of $d_n(F_\omega)$ for all $n \in \mathbf{N}$:

$$1/2 \cdot \omega(2c_n) = d_n(F_\omega) = 1/2 \cdot \omega(2p_{n+1}).$$

Only the special case $s = 1$ of this statement seems to be known, see *Grigorian* (1973).

Remark
It is interesting to compare this Proposition with 1.3.6:
We have $e_n(F_\omega, App) = \omega(c_n)$ and $d_n(F_\omega) \le 1/2 \cdot \omega(2c_n)$ and hence

$$d_n(F_\omega) \le e_n(F_\omega, App) \le 2 \cdot d_n(F_\omega).$$

The case $F_s = \{f : [0,1]^s \to \mathbf{R} \mid |f(x) - f(y)| \le \max_i |x_i - y_i|\}$ is of special interest because of

$$d_n(F_s) = e_n(F_s, App) = c_n$$

for all $n \in \mathbf{N}$.
Hence an optimal approximation method can be obtained by a simple sampling of function values $f(a_1),\ldots,f(a_n)$.
Similar results concerning function classes "where sampling is optimal" are surveyed in *Micchelli, Rivlin* (1985).

Example
Let X be the unit circle (with the spherical distance). Then F_ω may be considered as the set

$$\{f : \mathbf{R} \to \mathbf{R} \mid f \text{ is } 2\pi\text{-periodic with } \omega(f,h) \le \omega(h)\}.$$

It is easy to see that $c_n = p_n = \pi/n$ for all $n \in \mathbf{N}$ (and $n \neq 1$ for p_n). Proposition 1.3.7 yields

$$1/2 \cdot \omega(2\pi/n) \geq d_n(F_\omega) \geq 1/2 \cdot \omega(2\pi/(n+1))$$

for all $n \in \mathbf{N}$. In this case we do not obtain the exact values of $d_n(F_\omega)$ but instead a rather tight estimate.
The exact values of $d_n(F_\omega)$ are known in the case where ω is a concave function (see *Lorentz* (1966) for odd n and *Ruban* (1974) in the general case), the result is the following:

$$\begin{aligned} d_n(F_\omega) &= 1/2 \cdot \omega\left(\tfrac{2\pi}{n+1}\right) && \text{for odd } n \text{ and} \\ d_n(F_\omega) &= 1/2 \cdot \omega\left(\tfrac{2\pi}{n}\right) = d_{n-1}(F_\omega) && \text{for even } n. \end{aligned}$$

1.3.8 The problem *Int* for classes of Lipschitz functions and related classes of functions

Let $S = Int : B(X) \to \mathbf{R}$ be linear and bounded. It is easy to prove (and follows from 1.2.5 and 1.3.6) that

$$e_n(F_\omega, S) \leq \|S\| \cdot \omega(c_n),$$

but of course $e_n(F_\omega, S)$ can be much smaller than $\|S\| \cdot \omega(c_n)$. We give a lower bound for the numbers

$$\sup_{\|Int\| \leq 1} e_n^{ad}(F_\omega, Int)$$

in terms of the packing numbers p_n of X.

Proposition

$$\sup_{\|Int\| \leq 1} e_n^{ad}(F_\omega, Int) \geq 1/2 \cdot \omega(p_{2n}).$$

Proof:
Let $\varepsilon > 0$ and $a_1, \ldots, a_{2n} \in X$ with

$$\inf_{i \neq j} 1/2 \cdot d(a_i, a_j) \geq p_{2n} - \varepsilon.$$

We consider the functional *Int*, defined by

$$Int(f) = \frac{1}{2n} \sum_{i=1}^{2n} f(a_i).$$

Define f_i by

$$f_i(x) = \begin{cases} \omega(p_{2n} - \varepsilon - d), & \text{if } d = d(x, a_i) < p_{2n} - \varepsilon,; \\ 0, & \text{otherwise.} \end{cases}$$

It is easy to see that $f_i \in F_\omega$, it follows by means of 1.3.5 that

$$e_n^{ad}(F_\omega, Int) \geq (2n-n) \cdot \frac{1}{2n} \cdot \omega(p_{2n}) = 1/2 \cdot \omega(p_{2n}).$$

Example

The case $X \subset \mathbf{R}^s$ and $Int = \lambda^s : F_\omega \to \mathbf{R}$ is studied in *Sukharev* (1979). We give only one result.

Let $F_s = \{f : [0,1]^s \to \mathbf{R} \mid |f(x) - f(y)| \leq \max_i |x_i - y_i|\}$ and $n = m^s$ $(m \in \mathbf{N})$. Then

$$e_n(F_s, \lambda^s) = \frac{s}{2s+2} \cdot n^{-1/s}.$$

Comparing this result with 1.3.6 it can be seen that $e_n(F_s, \lambda^s)$ is only slightly smaller than $e_n(F_s, App)$. This is typical for many classes F. On the other hand, there are also examples which behave very differently:

For $F_{\mathrm{Var}} = \{f : [0,1] \to \mathbf{R} \mid \mathrm{Var}(f) \leq 1\}$ we get $d_n(F_{\mathrm{Var}}) = 1/2$ for all $n \in \mathbf{N}$, $e_1(F_{\mathrm{Var}}, App) = 1$ and $e_n(F_{\mathrm{Var}}, App) = 1/2$ for $n > 1$, but $e_n(F_{\mathrm{Var}}, \lambda) = 1/(2n)$ (the last result is due to *Zubrzycki* (1963/64), see *Novak* (1983)).

1.3.9 Hölder classes of functions

We define the Hölder classes $C_s^{k,\alpha}$ by

$$C_s^{k,\alpha} = \{f : [0,1]^s \to \mathbf{R} \mid \ |D^{(l)} f(x) - D^{(l)} f(y)| \leq \max_i |x_i - y_i|^\alpha$$
$$\text{for all partial derivatives of order } k\}.$$

The results of this section are valid for all $s \in \mathbf{N}$, $k \in \mathbf{N}_0$ and $0 < \alpha \leq 1$, but of course for $s = 1$ or $k = 0$ much better estimates are known.

The n-widths of $C_s^{k,\alpha}$ have been estimated by *Tikhomirov* (1960), the numbers $e_n(C_s^{k,\alpha}, \lambda^s)$ by *Bakhvalov* (1959), and the numbers $e_n(C_s^{k,\alpha}, Opt)$ (for the case $\alpha = 1$) by *Ivanov* (1971).

We summarize these results using 1.3.3.

Proposition

Let $s \in \mathbf{N}$, $0 < \alpha \leq 1$, and $k \in \mathbf{N}_0$. Then

$$d_n(C_s^{k,\alpha}) \asymp e_n(C_s^{k,\alpha}, \lambda^s) \asymp e_n(C_s^{k,\alpha}, App) \asymp e_n(C_s^{k,\alpha}, Opt) \asymp n^{-\frac{k+\alpha}{s}},$$

where $a_n \asymp b_n$ denotes the weak equivalence of sequences, i.e. $c < a_n/b_n < C$ for certain $0 < c < C$ and all n. The same estimates are valid for the e_n^{ad} and also for the problem Opt^*.

Proof:
The lower bounds can be obtained by 1.3.5 in the following way. Let

$$\Phi(x) = a \cdot \prod_{i=1}^{s} (1 - x_i^2)^{k+1}$$

for $|x_i| \leq 1$ $(i = 1, \ldots, s)$ and $\Phi(x) = 0$ elsewhere. We can choose $a > 0$ such that $\Phi \in C_s^{k,\alpha}$ holds.
Consider the functions

$$f_i(x) = \varepsilon \cdot \Phi\left((x - x_i) \cdot \varepsilon^{-1/(k+\alpha)}\right),$$

where $i = 1, \ldots, n$ and pairs of x_i are separated by at least $2\varepsilon^{1/(k+\alpha)}$. We can choose n of order

$$n \asymp (1/\varepsilon)^{\frac{s}{k+\alpha}}.$$

Because of $f_i \in C_s^{k,\alpha}$, the lower bounds now follow from 1.3.5. To prove the upper bounds we can use known methods using piecewise polynomials, see *Ivanov* (1971).

Remarks
a) *Bückner* (1950a,b) seems to be the first to obtain similar lower bounds for the problem *Int*.
b) For $\alpha = 1$ and $k = 0$ or $s = 1$ we know that

$$d_n(C_s^{k,\alpha}) = e_n(App, C_s^{k,\alpha}),$$

see 1.3.7 and 1.3.10. It would be interesting to know whether this holds for arbitrary k and s. (What are perfect splines in several variables?)
c) By means of 1.2.8 it is easy to prove that

$$\sup_{\|Int\| \leq 1} e_n(C_s^{k,\alpha}, Int) \asymp n^{-\frac{k+\alpha}{s}}$$

(where *Int* runs through all Radon measures $Int : C([0,1]^s) \to \mathbf{R}$ with $\|Int\| \leq 1$).
It is much more difficult to compute the analogous quantities for Sobolev classes (see 1.3.12).

1.3.10 n-widths and the problem *App* for the class W_∞^k

We now give a famous result of *Tikhomirov* (1969) concerning the classes

$$W_\infty^k = C_1^{k-1,1} = \{f : [0,1] \to \mathbf{R} \mid f^{k-1} \text{ is absolutely cont. and } |f^{(k)}(x)| \leq 1 \ a.e.\},$$

where $k \in \mathbf{N}$. A thorough discussion of this result is given in *Traub, Woźniakowski* (1980), see also *Micchelli, Rivlin, Winograd* (1976).

The problem *Opt** for the same classes is treated in *Pevnyi* (1982).

Proposition
Let $k \in \mathbf{N}$. Then

$$d_n(W_\infty^k) = e_n(W_\infty^k, App) = K_k(\pi n)^{-k} \cdot (1 + o(1)),$$

where $K_k \in [1, \pi/2]$ is the Farvard constant defined by

$$K_k = \frac{4}{\pi} \sum_{i=0}^{\infty} \frac{(-1)^{i(k+1)}}{(2i+1)^{k+1}} \; .$$

Remark: The problem *Int* for W_∞^k and more general Sobolev classes of functions will be treated in 1.3.12.

1.3.11 n-widths and the problem *App* for Sobolev classes of functions
We define the Sobolev classes $W_p^{k,s}$ by

$$W_p^{k,s} = \Big\{ f : [0,1]^s \to \mathbf{R} \mid \sum_{|\alpha|=k} \|f^{(\alpha)}\|_p \le 1 \Big\},$$

where $1 \le p \le \infty$, $s, k \in \mathbf{N}$, α is a multiindex, and $f^{(\alpha)}$ is the respective distributional derivative. If $pk > s$ then $W_p^{k,s}$ may be regarded as a subset of $C([0,1]^s)$ ("imbedding condition").
If, in addition, $2k > s$ then the following is known:

$$d_n(W_p^{k,s}) \asymp \begin{cases} n^{-k/s}, & \text{for } p \ge 2 \\ n^{-k/s-1/2+1/p}, & \text{for } 1 \le p < 2 \end{cases}$$

and

$$d_n^L(W_p^{k,s}) \asymp \begin{cases} n^{-k/s+1/p}, & \text{for } p \ge 2 \\ n^{-k/s+1/2}, & \text{for } 1 \le p < 2. \end{cases}$$

(See *Kashin* (1977, 1980), *Höllig* (1979, 1980), and *Pinkus* (1985)). It is much easier to prove the following statement.

Proposition
Let $pk > s$. Then

$$e_n(W_p^{k,s}, App) \asymp n^{-k/s+1/p}.$$

Proof:
The lower bound can be proved by 1.3.5a. Let

$$\Phi(x) = \begin{cases} a \cdot \prod_{i=1}^{s} (1-x_i^2)^k, & \text{if } |x_i| \le 1 \text{ for all } i = 1, \ldots, n \\ 0, & \text{elsewhere.} \end{cases}$$

We choose $a > 0$ such that $f \in W_p^{k,s}$. For $m \in \mathbf{N}$ we divide $[0,1]^s$ in m^s cubes with edge length $1/m$ and midpoints y_j ($j = 1, \ldots, m^s$). Now we define

$$f_j(x) = \left(\frac{1}{2m}\right)^k \cdot \Phi(2mx - 2my_j) \cdot m^{s/p}.$$

An easy computation yields $f_j \in W_p^{k,s}$ and the lower bound follows from 1.3.5a. The upper bound can again be obtained by known methods using piecewise polynomials.

1.3.12 The problem *Int* for Sobolev classes
First we treat the case of the Lebesgue measure $Int = \lambda^s$.

Proposition 1
Let $kp > s$ or $p = 1$ and $k = s$. Then

$$e_n(W_p^{k,s}, \lambda^s) \asymp n^{-k/s}.$$

To prove this statement we need the following lemma which is well known (see *Dahmen, DeVore, Scherer* (1980) and *Maz'ja* (1985) for this and similar inequalities).

Lemma 1
Let $1 \le p \le \infty$, $s \in \mathbf{N}$, $k \in \mathbf{N}$ and $kp > s$ or $p = 1$ and $k = s$. Then there is a $c > 0$ such that for each $f \in W_p^{k,s}([a,b]^s)$ there is a polynomial $g \in P(k,s)$ (i.e. the degree of g is less than k) with

$$\|f - g\|_\infty \le c \cdot (b-a)^{k-s/p} \cdot \sum_{|\alpha|=k} \|D^\alpha f\|_p.$$

Proof of Proposition 1
Again the lower bound follows from 1.3.5. To prove the upper bound we divide the cube $[0,1]^s$ into m^s little cubes with edge length $1/m$ ($m \in \mathbf{N}$). Consider a quadrature formula on each of the little cubes with $dim\, P(k,s)$ knots and norm m^{-s} which is exact for all $p \in P(k,s)$ (such a formula exists, see 1.2.8). Let $\widetilde{S}$ be the sum of these quadrature formulas, let

$$n = \binom{s+k-1}{k-1} \cdot m^s,$$

and let $\|D^\alpha f\|_p^{(i)}$ denote the p-norm on the ith subcube ($i = 1, \ldots, m^s$).
Then $\widetilde{S} \in A_n$ and for $f \in W_p^{k,s}$ we get by Lemma 1:

$$\begin{aligned}
|S(f) - \widetilde{S}(f)| &\leq c \cdot m^{-s-k+s/p} \sum_{i=1}^{m^s} \sum_{|\alpha|=k} \|D^\alpha f\|_p^{(i)} \\
&\leq c \cdot m^{-s-k+s/p} \cdot m^{s-s/p} \cdot \sum_{|\alpha|=k} \|D^\alpha f\|_p \\
&\leq c \cdot m^{-k},
\end{aligned}$$

which proves the statement.

Remark
The special case $p = 2$ of Proposition 1 is treated in *Sobolev* (1965) with Hilbert space methods. The case $1 < p < \infty$ is considered in *Polovinkin* (1974). See *Besov* (1980,1981) for further results.
It is not difficult to prove the following stronger result.

Proposition 2
Let $kp > s$ or $p = 1$ and $k = s$. Then

$$\lim_{n \to \infty} n^{k/s} \cdot e_n(W_p^{k,s}, \lambda^s) \in \mathbf{R}^+$$

exists.

Proof:
Let

$$h(n) = n^{k/s} \cdot e_n(W_p^{k,s}, \lambda^s).$$

First we have $\inf_{n \in \mathbf{N}} h(n) = c > 0$. Let $\varepsilon > 0$ and $n_0 \in \mathbf{N}$ with $c \leq h(n_0) \leq c + \varepsilon$. By means of composed quadrature formulas we can show that $c \leq h(n_0 \cdot m^s) \leq c + \varepsilon$ for all $m \in \mathbf{N}$ and the statement $\lim_{n \to \infty} h(n) = c$ follows (see *Novak* (1983) for more details).

Remark
This result is well known for $s = 1$. In this case, the respective limits and quadrature formulas are also known. See *Levin,Girshovich* (1979), *Strauss* (1979), and *Zhensykbaev* (1981).
Until now, we have only considered the Lebesgue measure $S = \lambda^s$. The following theorem, which is due to *Novak* (1983), gives an error bound for all (bounded) linear functionals on $C([0,1]^s)$.

Proposition 3
Let $k,s \in \mathbf{N}$, $1 \leq p \leq \infty$, and $pk > s$. Then

$$\sup_{\|Int\| \leq 1} e_n(W_p^{k,s}, Int) \asymp n^{-k/s}.$$

Remark
The proof of Proposition 3 is rather complicated because it is not sufficient to divide (as in Proposition 1) the cube $[0,1]^s$ into subcubes of the same size—this method would imply only the weaker estimate

$$\sup_{\|Int\| \leq 1} e_n(W_p^{k,s}, Int) \leq c \cdot n^{-k/s+1/p}.$$

In the case $p \geq 2$ we can prove Proposition 3 by means of 1.2.8 and 1.3.11. For $1 \leq p < 2$ the relation

$$\sup_{\|Int\| \leq 1} e_n(W_p^{k,s}, Int) \asymp d_n(W_p^{k,s})$$

does not hold and we have to partition measures μ on $[0,1]^s$ according to the following lemma.

Lemma 2
Let μ be a (signed) Borel measure on $[0,1]^s$ with $\|\mu\| \leq 1$ and let $\alpha > 0$, $0 < \delta < 1$. Then there are measures μ_i $(i = 1, \ldots, N)$ with the following properties:

i) $$\sum_{i=1}^{N} \mu_i = \mu \quad \text{and} \quad \sum_{i=1}^{N} \|\mu_i\| \leq 1;$$

ii) the support of μ_i is contained in a cube with edge length a_i and

$$\|\mu_i\| \cdot a_i^\alpha \leq \delta \quad \text{for all } i, \quad \text{and} \quad \sum_{i=1}^{N} a_i^s = 1;$$

iii) $$N < c \cdot \delta^{-\frac{s}{\alpha+s}},$$

where $c > 0$ does not depend on $0 < \delta < 1$ or μ.

Proof:
Let be given a Borel measure μ on $[0,1]^s$ with $\|\mu\| \leq 1$ and $\alpha > 0$, $0 < \delta < 1$. We consider the following partition Z_1 of $[0,1]^s$ in subcubes:
Beginning with $[0,1]^s$ each (sub–) cube is divided into 2^s cubes with half the edge length provided that for the respective cube the inequality

$$\|\mu_i\| \cdot a_i^\alpha \leq \delta \tag{1}$$

is not yet satisfied. Here μ_i is the restriction of μ on the ith subcube and a_i is its edge length. We continue this process until (1) is valid for each subcube W_i ($i = 1,\ldots,N(\delta)$). We have

$$\sum_{i=1}^{N(\delta)} \mu_i = \mu, \qquad \sum_{i=1}^{N(\delta)} \|\mu_i\| \leq 1 \tag{2}$$

and

$$\sum_{i=1}^{N(\delta)} a_i^s = 1. \tag{3}$$

It is sufficient to show that there is a $c > 0$ (independent of μ and δ) such that

$$N(\delta) \leq c \cdot \delta^{-\frac{s}{s+\alpha}}. \tag{4}$$

To prove (4) we consider the partition Z_2 which is defined slightly different:
A (sub–) cube is divided into 2^s smaller ones only if the inequality $\|\mu_i\| \cdot a_i^\alpha > \delta$ still holds for at least one of the smaller cubes. In this way we obtain the partition $Z_2 = \{M_i \mid i = 1,\ldots,\widetilde{N}(\delta)\}$ with

$$\widetilde{N}(\delta) \geq 2^{-s} \cdot N(\delta). \tag{5}$$

Clearly it is sufficient to prove (4) for $\widetilde{N}(\delta)$ instead of $N(\delta)$. Let M be the number of the M_i ($i = 1,\ldots,\widetilde{N}(\delta)$) for which

$$\|\mu_i\| \cdot a_i^\alpha > \delta. \tag{6}$$

Let $l_1 = 2^{-a}$ be the edge length of the biggest cube in Z_2, and let $l_2 = 2^{-b}$ be the edge length of the smallest cube. Then

$$\widetilde{N}(\delta) \leq (b-a) \cdot 2^s \cdot M. \tag{7}$$

Let $M_a \in Z_2$ have edge length l_1 and let $M_b \in Z_2$ have edge length l_2. We further assume that (6) is valid for M_b.
We define a measure μ' by the following properties:

$$\mu' = \mu \ \text{ outside of } M_a \text{ and } M_b,\ \ \mu' = 0 \ \text{ on } M_b \ \text{ and } \mu' = \|\mu\|_{M_b} \cdot 2^{as} \cdot \lambda^s \ \text{ on } M_a. \tag{8}$$

An easy computation yields

$$\|\mu'\| \leq 1. \tag{9}$$

We apply the technique which yields to the partition Z_2 to the measure μ' and obtain a partition

$$Z_2' = \{M_i' \mid i = 1,\ldots,\widetilde{N}'(\delta)\}.$$

Now we compare $\widetilde{N}'(\delta)$ with $\widetilde{N}(\delta)$:

First it can be seen that by the change of the measure on M_b the number of the cubes becomes smaller at most by $(b-a)\cdot 2^s$. On the other hand new cubes may be added by the partition of M_a. First we have

$$\mu' = k\lambda^s \quad \text{with} \ \ k > \delta \cdot 2^{b\alpha + as} \quad \text{on } M_a. \tag{10}$$

Now we define $m \in \mathbf{N}$ by

$$(2^s)^{m-2} < b-a \leq (2^s)^{m-1} \tag{11}$$

and want to give conditions under which M_a is divided m-times into $(2^s)^m$ little cubes. This is certainly the case if

$$(2^{-a}\cdot 2^{-m})^{\alpha}\cdot \delta \cdot 2^{b\alpha - sm} \geq \delta, \tag{12}$$

i.e. if $(b-a)\cdot\alpha \geq m(s+\alpha)$. It follows from (11) that there is a $c_{s,\alpha} > 0$ such that (12) is valid if

$$b-a \geq c_{s,\alpha}. \tag{13}$$

Hence we have shown that

$$\widetilde{N}'(\delta) \geq \widetilde{N}(\delta) \tag{14}$$

if (13) is valid. Applying this construction repeatedly, we obtain the following result:

$$\text{There is a } \mu' \ \text{ with } \|\mu'\| \leq 1,\ b'-a' < c_{s,\alpha}, \text{ and } \ \widetilde{N}'(\delta) \geq \widetilde{N}(\delta). \tag{15}$$

By means of (7) we get

$$\widetilde{N}'(\delta) \leq c_{s,\alpha}\cdot 2^s \cdot M'. \tag{16}$$

To give an estimate of M' it is enough to consider the following inequalities:

$$\|\mu_i\|\cdot a_i^{\alpha} \geq \delta \qquad \sum_{i=1}^{M'} \|\mu_i\| \leq 1 \qquad \sum_{i=1}^{M'} a_i^s \leq 1. \tag{17}$$

Using the well known inequality

$$\frac{1}{u+v}(ux+vy) \geq (x^u\cdot y^v)^{\frac{1}{u+v}}, \tag{18}$$

which is valid for $u,v,x,y > 0$ (see *Mitrinović* (1970, p.30)), we obtain

$$M' \leq \delta^{-\frac{s}{\alpha+s}}. \tag{19}$$

By (5), (15), (16) and (19) we get

$$N(\delta) \leq 2^{2s}\cdot c_{s,\alpha}\cdot \delta^{-\frac{s}{\alpha+s}}$$

and hence the lemma is proved.

Proof of Proposition 3:
Let be given a Borel measure μ on $[0,1]^s$ with $\|\mu\| \leq 1$, $0 < \delta < 1$ and k, p such that $pk > s$.
We put $\alpha = k - s/p > 0$ and apply Lemma 2. As in the proof of Proposition 1 we approximate each μ_i $(i = 1, \ldots, N)$ by a quadrature formula $\tilde{\mu}_i$ with $dim\, P(k,s)$ knots and norm $\|\tilde{\mu}_i\| \leq \|\mu_i\|$ which is exact for all $p \in P(k,s)$. Using Lemma 1 we obtain the following error estimate for the resulting quadrature formula

$$\widetilde{S}(f) = \sum_{i=1}^{N} \tilde{\mu}_i \qquad (f \in W_p^{k,s}) :$$

$$\begin{aligned}
|\widetilde{S}(f) - \mu(f)| &\leq \textstyle\sum_{i=1}^N |\tilde{\mu}_i(f) - \mu_i(f)| \\
&\leq \textstyle\sum_{i=1}^N 2c' \|\mu_i\| a_i^{k-s/p} \cdot (\sum_{|\alpha|=k} \|D^\alpha f\|_p^{(i)}) \\
&\leq 2c' \cdot \delta \cdot \textstyle\sum_{i=1}^N \sum_{|\alpha|=k} \|D^\alpha f\|_p^{(i)} \\
&\leq 2c' \cdot \delta \cdot N^{1-1/p} \cdot \textstyle\sum_{|\alpha|=k} \|D^\alpha f\|_p \\
&\leq 2\tilde{c}\delta \cdot \delta^{\frac{s(p-1)}{s-sp-pk}} \cdot \textstyle\sum_{|\alpha|=k} \|D^\alpha f\|_p \\
&= 2\tilde{c}\delta^{-\frac{pk}{s-sp-kp}} \cdot \textstyle\sum_{|\alpha|=k} \|D^\alpha f\|_p.
\end{aligned}$$

We have proved that

$$\sup_{\|Int\| \leq 1} e_n(W_p^{k,s}, Int) \leq 2\tilde{c} \cdot \delta^{-\frac{pk}{s-sp-kp}}$$

for arbitrary $0 < \delta < 1$ and

$$n = \binom{s+k-1}{k-1} \cdot \delta^{-\frac{sp}{kp-s+sp}}$$

and the statement follows.

1.3.13 Notes and references

There are many papers dealing with a worst case analysis for the problems *App*, *Opt*, *Opt**, and especially *Int*. The aim of this section was not to survey these papers but rather to indicate some basic results which we want to compare with different stochastic error bounds in the following chapters.
Many references can be found in *Traub, Woźniakowski* (1980). Optimization problems are surveyed in *Archetti, Szegö* (1980), *Sukharev* (1981a) and in *Nemirovsky, Yudin* (1983). Monographs and surveys on the problem *Int* are *Haber* (1970), *Stroud*

(1971), *Braß* (1977), *Engels* (1980), and *Davis, Rabinowitz* (1984). Additional results concerning the problem *App* can be found in *Osipenko* (1972, 1976), *Temlyakov* (1985), and in *Micchelli, Rivlin* (1985).
The results of 1.3.1 are from *Kon, Novak* (1988), those of 1.3.12 are from *Novak* (1983), see also *Novak* (1987a) for the decomposition technique for measures in $\mathbf{R}^n$. I observed only recently that *Birman, Solomjak* (1967) obtained similar results by the same technique much earlier. See also *Birman, Solomjak* (1980).

2. Error bounds for Monte Carlo methods

2.1 Basic properties of Monte Carlo methods

2.1.1 Remark
First we repeat two results of the first chapter:
i) There are important problems for which each ε-approximation is very difficult, i.e. the sequence $e_n^{ad}(F,S)$ tends to zero very slowly.
ii) For many problems adaptive methods are not better than nonadaptive ones within a worst case analysis.
These observations lead to the following questions:
i) Are there problems for which the $e_n^{ad}(F,S)$ converge to zero very slowly (or do not converge to zero at all) though there are approximations $\widetilde{S} \in A_n^{ad}$ such that $\Delta(\widetilde{S}(f),f)$ is reasonable small for most $f \in F$? A positive answer means that it makes sense to weaken the notion of an ε-approximation: If every ε-approximation is very complicated one could be content with an approximation $\widetilde{S} \in A_n^{ad}$ which is relatively simple (n is small) and satisfies $\Delta(\widetilde{S}(f),f) \leq \varepsilon$ with a high probability.
ii) If $e_n^{ad}(F,S) = e_n(F,S)$ holds for a certain problem then this could still mean that there is a method $\widetilde{S} \in A_n^{ad}$ such that $\Delta(\widetilde{S}(f),f)$ is much smaller than the error of an arbitrary nonadaptive method for most $f \in F$ and hence we can ask, for which problems adaption helps on the average.
There are two basically different possibilities to establish an average case analysis for numerical methods:
i) We can consider means over a class of approximations (so called stochastic, Monte Carlo, or mixed methods).
ii) By means of expectations over the class F of problem elements and similar notions one defines the average error of deterministic (or pure) methods.
In this second chapter we deal with Monte Carlo methods. In 3.1 we treat average case errors using a probability measure on F. In 3.2 we introduce the notion of the average a posteriori error, using an averaging process over the set of information $N(F) \subset \mathbf{R}^n$. (Of course the average error of Monte Carlo methods can also be defined. We will consider this notion in 3.1.3).
One point of our discussion is the conceptual difficulties of the different averaging methods. These notional problems lead us to introduce new concepts such as Monte Carlo methods in our restricted sense, equidistributions on compact metric spaces, and average a posteriori errors.
We want to apply methods from probability theory to the investigation of numerical methods. Even though the notion of probability was well established by Kolmogorov in 1933 in a formal sense, it still contains difficulties as to the contents. Neither the substance of statements on probabilities is sufficiently clear nor the method of

obtaining such assertions. Not even the range of matters to which those statements can be applied to is undisputed. There are mathematicians who consider Monte Carlo methods to be a conceptual misunderstanding. Though they have serious reasons for this conviction, we want to show that it is not necessary to give up that notion totally. Probabilities describe a knowledge of facts which are uncertain. We need not assume that a fact is principally indeterministic (whatever this may be) in order that the application of stochastic methods makes sense. This is illustrated by successful applications of stochastic reasoning in classical physics or medicine.
A discussion of the foundations of probability theory can be found in *Harper, Hocker* (1977), *Dinges, Rost* (1982), *Stegmüller* (1973), and in *Borovcnik* (1984).

2.1.2 Definition
Let F, S, A, and Δ be given as in Chapter 1.
A random variable $Q = (Q(\omega))_{\omega\in\Omega}$ is called a *generalized Monte Carlo method* in A if (Ω, B, m) is a probability space and $Q(\omega) \in A$ for all $\omega \in \Omega$. The set of all generalized Monte Carlo methods in A is denoted by ${}^*C(A)$. The number

$$\Delta(Q,f) = \int_{\Omega}^{*} \Delta(Q(\omega)(f), f)\ dm(\omega)$$

is called the error of Q with respect to f.
(In order that $\Delta(Q,f)$ is always defined we take the upper integral $\int^* \ldots\ dm(\omega)$ which always exists.)
The number

$$\Delta_{max}(Q) = \sup_{f\in F} \Delta(Q,f)$$

is called the *error of* Q. Analogously to the $e_n(F,S)$ and $e_n^{ad}(F,S)$ we define

$${}^*\sigma_n(F,S) = \inf_{Q\in{}^*C(A_n)} \Delta_{max}(Q)$$

and

$${}^*\sigma_n^{ad}(F,S) = \inf_{Q\in{}^*C(A_n^{ad})} \Delta_{max}(Q).$$

Remarks
i) It is common to define Monte Carlo methods in this or a similar way. Many authors demand some additional properties so that $\Delta(Q(\omega)(f), f)$ is m-integrable for each $f \in F$. (See *Nemirovsky, Yudin* (1983).)
ii) Later we will define Monte Carlo methods in a more restricted sense, the notations $C(A)$, $\sigma_n(F,S)$, and $\sigma_n^{ad}(F,S)$ are reserved for this definite definition of a Monte Carlo method.

iii) Instead of the error of Q often the *dispersion*

$$\widetilde{\Delta}(Q,f) = \left(\int_{\Omega}^{*} \Delta(Q(\omega)(f),f)^2 \; dm(\omega) \right)^{1/2}$$

of Q with respect to f is considered. Of course we then can define

$$\widetilde{\Delta}_{max}(Q) = \sup_{f \in F} \widetilde{\Delta}(Q,f)$$

("dispersion of Q") and

$${}^{*}\tilde{\sigma}_n^{(ad)}(F,S) = \inf_{Q \in {}^{*}C(A_n^{(ad)})} \widetilde{\Delta}_{max}(Q).$$

2.1.3 A first discussion of Monte Carlo methods

We begin our discussion of Monte Carlo methods with a most simple example. Let $D \subset \mathbf{R}^s$ with $\lambda^s(D) = 1$ and let

$$F = \left\{ f : D \to \mathbf{R} \mid f \in L_2(D), \; \int_D f^2(x) \; d\lambda^s(x) \leq 1 \right\}.$$

We consider the linear problem $S = Int = \lambda^s : F \to M = \mathbf{R}$, defined by

$$S(f) = \int_D f(x) \; d\lambda^s(x).$$

The classical Monte Carlo method $Q_n \in {}^{*}C(A_n)$ is given by

$$Q_n(\omega)(f) = \frac{1}{n} \sum_{i=1}^{n} f(x_i(\omega)),$$

where the x_i are uniformly distributed over the domain D (i.e. $x_i : \Omega \to D$ is a random variable with distribution ${\lambda^s}_{|D}$) and independent. It is easy to see that

$$\begin{aligned} \widetilde{\Delta}_{max}(Q) \quad &= \sup\nolimits_{f \in F} \left(\int \left(\tfrac{1}{n} \sum\nolimits_{i=1}^{n} f(x_i(\omega)) - \int_D f(x) \; dx \right)^2 \; dm(\omega) \right)^{1/2} = \\ &= n^{-1/2} \cdot \sup\nolimits_{f \in F} \left(\int_D (f(x) - \int_D f(y) \; dy)^2 \; dx \right)^{1/2} = \\ &= n^{-1/2}. \end{aligned}$$

By the supporters of the Monte Carlo method, this result is interpreted as follows: Choose $x_1, \ldots, x_n \in D$ independently according to the distribution ${\lambda^s}_{|D}$ and compute the sum

$$Q(f) = \frac{1}{n} \sum_{i=1}^{n} f(x_i).$$

Then, for example, $3 \cdot \widetilde{\Delta}_{max}(Q) = 3 \cdot n^{-1/2}$ is an estimate for the error $|Q(f) - S(f)|$. Though the true error may be larger, this will happen relatively seldom. The random choice of the x_i occurs by means of a suitably chosen algorithm ("random number generator"). The following idea seems to be the basis of this point of view:
Arbitrary probability measures can be simulated by means of (deterministic) algorithms, i.e. there is an algorithm producing elements $\omega_i \in \Omega$ "randomly" according to a given distribution. The word 'randomly' at least means that the laws of statistics should be applicable to this sample. This assessment is expressed very decidedly in *Coveyou* (1969).
Some authors (see *Zaremba* (1968)) have a point of view which is totally different: A number (or a quadrature formula) which is given by some fixed algorithm is deterministic and fixed and hence it is in no reasonable sense random and accessible to stochastic considerations, because probability theory only deals with events which are random. All Monte Carlo methods which use deterministic random number generators are absurd: Laws of stochastic are wrongly applied to phenomena that are not governed by those rules.
(There is another criticism of Monte Carlo methods: Many authors believe that well chosen deterministic methods should converge faster than Monte Carlo methods. We deal with this objection in 2.1.7).
Before we discuss the view of *Zaremba* (1968), we want to show that the idea of Monte Carlo methods is not tenable in such a general form as in 2.1.2.
Let F and S as in our example and let

$$Q(\omega)(f) = \frac{1}{n} \sum_{i=1}^{n} f(x_i(\omega))$$

with some "random" $x_i(\omega)$ which are generated by some algorithm. Of course the $x_i(\omega)$ are recursive and hence are elements of the countable set $R \subset D$ of all recursive points of D. Now consider the function $f = 1_{|R}$. We have $Q(\omega)(f) = 1$ for all $\omega \in \Omega$ and hence $\widetilde{\Delta}(Q, f) = 1$ in contrast to $\widetilde{\Delta}_{max}(Q) = n^{-1/2}$.
This simple example shows that the classical notion of a Monte Carlo method is too general, it does not confirm to the possibilities of computation. Before we introduce our concept of *(restricted) Monte Carlo methods* we have to deal with the criticism of Zaremba.
The basis of Zaremba's criticism seems to be the following:
There are random events and deterministic ones – and nothing between. All results of computations are deterministic. Probability theory can only be applied to random events. Hence it is nonsense to apply the laws of statistics to "random numbers" which are generated by a deterministic algorithm. A random variable can not be simulated by a computer because of these reasons.
I think that this conception too strongly limits the applications of statistics and leaves the following critical points open:

i) How can random events be distinguished from those which are deterministic? From this point of view randomness should be independent of our knowledge and a property of the events as they are – whatever this may be.
ii) How does this conception fit to the fact that statistics is extensively applied to phenomena which are not considered to be random in such a strong sense?

2.1.4

Now we introduce the notion of *(restricted) Monte Carlo methods*. With this concept we can
i) avoid the difficulties of the conception of a random sampling from an arbitrary set and
ii) prove just the practically relevant results on classical (i.e. generalized) Monte Carlo methods.
The generalized Monte Carlo methods seem to involve an idealization which is too strong, related to the applications. On the other side it seems to be unnecessary to totally abandon the idea of random sampling by means of algorithms. The difficulties result from the fact that each practical sampling – for example by a computer – is a sampling out of a finite set. It is not possible to choose a random $x \in [0,1]$ according to the Lebesgue measure but the idea of a random sampling of an $x \in \{0,1,\ldots,n\}$ according to the uniform distribution seems to be sound.

Definition
A generalized Monte Carlo method $Q = (Q(\omega))_{\omega\in\Omega}$ in A is called a *(restricted) Monte Carlo method* in A if $\{Q(\omega) \mid \omega \in \Omega\} \subset A$ is a finite set and the probability $m(\widetilde{S}) := m\{\omega \mid Q(\omega) = \widetilde{S}\}$ is a rational number for each $\widetilde{S} \in A$. We denote the set of all Monte Carlo methods in A by $C(A)$. Finally we define

$$\sigma_n(F,S) = \inf_{Q\in C(A_n)} \Delta_{max}(Q)$$

and

$$\sigma_n^{ad}(F,S) = \inf_{Q\in C(A_n^{ad})} \Delta_{max}(Q).$$

The numbers $\tilde{\sigma}_n(F,S)$ and $\tilde{\sigma}_n^{ad}(F,S)$ (see 2.1.2) are defined analogously.

2.1.5 Remark: Mathematical theories of chance and random number generators

Our definition of Monte Carlo methods is based on the idea that a random sampling of $x \in \{0,1,\ldots,n\}$ according to the uniform distribution is possible by means of a deterministic algorithm. A discussion of this idea must include the answer of the question of how such a random sampling should be done. Can the choice of a special algorithm be justified by means of mathematical arguments, i.e. is there a mathematical theory of chance? We do not discuss these questions thoroughly but only give some remarks and references.

First of all, randomness is not a property of a result $x \in \{0, 1, \ldots, 2^m - 1\}$ (or equivalently $x_i \in \{0, 1\}$, $i = 1, \ldots, m$), it is not even a unique quality of the procedure of choice but also depends on our knowledge and our understanding of the whole situation. Therefore a mathematical theory of chance hardly seems to be possible because every such theory can only deal with the results of certain sampling methods: In such a theory certain 0-1-sequences (for example) are defined to be random and the situation, in which such a sequence occurs, is not considered at all. On the other hand the results of random procedures most likely have certain properties which can be investigated by means of mathematics. Sequences of random numbers are expected to have a high complexity (they are difficult to describe) and have strong properties with respect to uniform distribution of subsequences (principle of the exclusion of gambling systems). Hence a mathematical theory which describes these features of randomness is quite possible, see *Schnorr* (1971), (1977), *Knuth* (1981), *Zvonkin, Levin* (1970) and *Kranakis* (1986).

The practical construction of random numbers is faced with the following contradiction: While sequences of random numbers (as theoretically defined) are very difficult to describe, a good algorithm for constructing pseudo random numbers should be simple and fast. The usual method confronting this dilemma is to construct sequences by means of simple algorithms (for example, the linear congruential method $x_{n+1} = ax_n + c \pmod{m}$ with well chosen natural numbers a, c, m, and x_0 is widely used) and to test the resulting numbers by means of certain statistical tests. These pseudo random sequences are suitable for many applications although it seems to be not really understood how to adapt the tests to special problems.

Surveys on the construction of random numbers are *Knuth* (1981) and *Niederreiter* (1978). See also *Zieliński* (1978), *Schmitz, Lehmann* (1976), *Fishman, Moore* (1986), *Niederreiter* (1984, 1985b), *Devroye* (1986), and several of the papers in *Afflerbach, Lehn* (1986).

2.1.6 Remark on Quasi Monte Carlo methods

We want briefly discuss the relation between Monte Carlo methods and the so called Quasi Monte Carlo methods. The latter methods are deterministic ones, they use deterministic copies of stochastic properties of sequences of random numbers. These number theoretical methods of numerical analysis were developed in the late fifties by Korobov, Hlawka, and Bakhvalov. Surveys are *Keng, Yuan* (1981), *Hlawka et al.* (1981), and *Niederreiter* (1978). Also see *Niedereiter* (1986) for the problem *Int*. The factual Quasi Monte Carlo methods hardly differ from Monte Carlo methods but there is a crucial difference in the theoretical analysis which leads to deterministic error bounds for Quasi Monte Carlo methods instead of stochastic error bounds for Monte Carlo methods.

2.1.7 Remark

After having clarified certain theoretical problems concerning Monte Carlo methods we have to deal with the question of for which problems these methods are favourable, more carefully: Which conditions have to be satisfied in order that Monte Carlo methods with their stochastic errors can be taken into account?

There is a widespread belief that well chosen deterministic methods are always better than stochastic methods. This is of course not true, the following inequalities are always valid:

$$^*\sigma_n(F,S) \leq \sigma_n(F,S) \leq e_n(F,S)$$

and

$$^*\sigma_n^{ad}(F,S) \leq \sigma_n^{ad}(F,S) \leq e_n^{ad}(F,S).$$

These inequalities follow from the simple fact that each deterministic method $\widetilde{S} \in A$ can be considered as a Monte Carlo method $\widetilde{S} \in C(A)$.

(Nevertheless it is true that the simplest Monte Carlo methods converge like $n^{-1/2}$ which is rather slow.)

On the other side, Monte Carlo methods and stochastic error bounds have the disadvantage that only uncertain statements on the error can be made in the individual case. Therefore it is only reasonable to take Monte Carlo methods into account if the respective numbers $\sigma_n^{(ad)}$ are much smaller than the $e_n^{(ad)}$. In this latter case, the stochastic errors of Monte Carlo methods may give more realistic information on the error than the pessimistic worst case errors.

We will give examples for problems where the numbers $\sigma_n(F,S)$ tend to zero much faster than the $e_n(F,S)$. For many problems, however, the $\sigma_n(F,S)$ are only slightly smaller than the $e_n(F,S)$ and hence Monte Carlo methods seem to have little use.

2.1.8

Our example in 2.1.3 satisfies $^*\sigma_n(F,S) \leq n^{-1/2}$ and $e_n(F,S) = 1$ for all n. It is also easy to see that $\sigma_n(F,S) = 1$ for all n. Hence this problem is tractable neither by Monte Carlo nor by deterministic methods but only by the unrealistic generalized Monte Carlo methods. Now we treat the problem whether $\sigma_n(F,S) \to 0$ implies $e_n(F,S) \to 0$.

Example

Let $F = \{f : \mathbf{N} \to \{0,1\}\}$, $M = F$, and $S = Id$. The error function Δ is defined by

$$\Delta(f,g) = \begin{cases} 1, & \text{if } f(i) = g(i) \text{ only for a finite number of } i \\ 0, & \text{else.} \end{cases}$$

Then $e_n(F,S) = 1$ for all n but $\sigma_n(F,S) = 0$ for all n.

Proof:
Let $N \in \mathbf{N}$ and $f_1, \dots, f_N \in F$ with $f_i(k) \neq f_j(k)$ for an infinite number of k ($i \neq j$). Now define $Q(i)(f) = f_i$ where $i \in \{1, \dots, N\} = \Omega$ is uniformly distributed. We clearly have $Q \in C(A_n)$ for all n and $\Delta_{max}(Q) = N^{-1}$, hence $\sigma_n(F,S) \leq N^{-1}$.

This example shows that $\sigma_n(F,S) \to 0$ does not imply that $e_n(F,S) \to 0$ in general. In a normed space setting, however, this is indeed the case and not more problems are tractable by Monte Carlo methods than by deterministic ones.

Proposition
Let M be a normed space and $\Delta(m, f) = \|m - S(f)\|$. Then

$$\lim_{n\to\infty} \sigma_n(F,S) = \lim_{n\to\infty} e_n(F,S)$$

and

$$\lim_{n\to\infty} \sigma_n^{ad}(F,S) = \lim_{n\to\infty} e_n^{ad}(F,S).$$

Proof:
Let $r \in \mathbf{N}$, $m_i \geq 0$ and $\sum_{i=1}^r m_i = 1$. Then for the mean value $\tilde{x} = \sum_{i=1}^r m_i x_i$ the inequality

$$(*) \qquad \sum_{i=1}^{r} \|x_i - x\| \cdot m_i \geq \|x - \tilde{x}\|$$

is valid for all $x \in M$. Let $\lim_{n\to\infty} \sigma_n(F,S) = \delta \in \mathbf{R}$ and let $\varepsilon > 0$. Then there is a $n \in \mathbf{N}$ and a $Q \in C(A_n)$ with $\Delta_{max}(Q) \leq \delta + \varepsilon$. Let $\Omega = \{\omega_1, \dots, \omega_r\}$ with $m(\{\omega_i\}) = m_i$ be the probability space that is taken as a basis. For each $f \in F$ we have

$$\Delta(Q, f) = \sum_{i=1}^{r} \|Q(\omega_i)(f) - S(f)\| \cdot m_i \leq \delta + \varepsilon.$$

By means of $(*)$ and with

$$\widetilde{S}(f) = \sum_{i=1}^{r} m_i Q(\omega_i)(f)$$

we get

$$\|\widetilde{S}(f) - S(f)\| \leq \delta + \varepsilon \qquad (f \in F).$$

It is easy to see that $\widetilde{S} \in A_{nr}$ and therefore we have $e_{nr}(F,S) \leq \delta + \varepsilon$ which proves the first inequality. The proof of the second one is analogous.

Remark
The problems *Int*, *App*, and *Opt** satisfy the conditions of our theorem. We show that both inequalities also hold for the problem *Opt*.

Let $\lim_{n\to\infty}\sigma_n(F,Opt)=\delta\in\mathbf{R}$ and let $\varepsilon>0$. Then there is a $n\in\mathbf{N}$ and a $Q\in C(A_n)$ with $\Delta_{max}(Q)\le\delta+\varepsilon$. Let $\Omega=\{\omega_1,\dots,\omega_r\}$ with $m(\{\omega_i\})=m_i$ be the probability space that is taken as a basis. For each $f\in F$ we have

$$\Delta(Q,f)=\sum_{i=1}^{r}|f(Q(\omega_i)(f))-\sup f|\cdot m_i\le\delta+\varepsilon.$$

We define $\widetilde{S}\in A_{nr}$ by

$$\widetilde{S}(f)=Q(\omega_i)(f)$$

where $i\in\{1,\dots,r\}$ is the least number with

$$f(Q(\omega_i)(f))=\max_j f(Q(\omega_j)(f)).$$

Now we have $|f(\widetilde{S}(f))-\sup f|\le\delta+\varepsilon$ for all $f\in F$ which proves the first inequality. The proof of the second one is analogous.

2.1.9 Lower bounds of the $\sigma_n^{(ad)}(F,S)$

Estimates from above of the numbers $\sigma_n^{(ad)}(F,S)$ can be given by the more or less explicit indication of suitable Monte Carlo methods. Because only generalized Monte Carlo methods have been investigated up to now, one often has to discretize known methods appropriately.
In this section we deal with lower bounds of the $\sigma_n^{(ad)}(F,S)$. These estimates only use properties of deterministic methods. First we define the numbers $e_n^{(ad)}(F,S,\mu)$ which will be discussed in the third chapter.

Definition
Let μ be a (Borel) probability measure on F and let $\int^* h\,d\mu$ denote the upper integral of a function $h:F\to\mathbf{R}$. The number

$$\Delta_\mu(\widetilde{S})=\int_F^*\Delta(\widetilde{S}(f),f)\,d\mu(f)$$

is called the *average error* of the (deterministic) method $\widetilde{S}:F\to M$ with respect to the measure μ. Analogously to the $e_n^{(ad)}$ we define

$$e_n^{(ad)}(F,S,\mu)=\inf_{\widetilde{S}\in A_n^{(ad)}}\Delta_\mu(\widetilde{S}).$$

Because of $\Delta_\mu(\widetilde{S})\le\Delta_{max}(\widetilde{S})$ we have

$$e_n^{(ad)}(F,S,\mu)\le e_n^{(ad)}(F,S).$$

By the following proposition it follows that the $e_n^{(ad)}(F,S,\mu)$ are also lower bounds for the $\sigma_n^{(ad)}(F,S)$ in many cases.

Proposition
Let μ be of the form $\mu = \sum_{i=1}^{m} c_i \varepsilon_{f_i}$ with $c_i \geq 0$, $\sum c_i = 1$, and $f_i \in F$ (where ε_f denotes the respective Dirac measure). Then the inequalities

$$e_n^{(ad)}(F,S,\mu) \leq {}^*\sigma_n^{(ad)}(F,S) \leq \sigma_n^{(ad)}(F,S)$$

are valid.

Proof:
Let $Q \in {}^*C(A_n^{(ad)})$. Then the following inequalities hold:

$$\begin{aligned}
\Delta_{max}(Q) \quad &= \sup_{f\in F} \textstyle\int_\Omega^* \Delta(Q(\omega)(f), f)\, dm(\omega) \\
&\geq \textstyle\int_F \int_\Omega^* \Delta(Q(\omega)(f), f)\, dm(\omega) d\mu(f) \\
&\geq \textstyle\int_\Omega^* \int_F \Delta(Q(\omega)(f), f)\, d\mu(f) dm(\omega) \\
&\geq \textstyle\int_\Omega^* e_n^{(ad)}(F,S,\mu)\, dm(\omega) \\
&= e_n^{(ad)}(F,S,\mu).
\end{aligned}$$

(We have used the well known formula $\int^*(g+h)\, dm \leq \int^* g\, dm + \int^* h\, dm$).

Remarks
i) Applications of the last proposition to the problems *App*, *Opt*, *Opt**, and *Int* can be found in Section 2.2.
ii) It is easy to prove analogous estimates for the numbers $\tilde{\sigma}_n^{(ad)}(F,S)$. Under the same conditions, for $Q \in {}^*C(A_n^{(ad)})$ the inequalities

$$\begin{aligned}
(\widetilde{\Delta}_{max}(Q))^2 \quad &= \sup_{f\in F} \textstyle\int_\Omega^* \Delta(Q(\omega)(f), f)^2\, dm(\omega) \\
&\geq \textstyle\int_F \int_\Omega^* \Delta(Q(\omega)(f), f)^2\, dm(\omega) d\mu(f) \\
&\geq \textstyle\int_\Omega^* \int_F \Delta Q(\omega)(f), f)^2\, d\mu(f) dm(\omega) \\
&\geq \inf_{\widetilde{S}\in A_n^{(ad)}} \textstyle\int_F \Delta(\widetilde{S}(f), f)^2\, d\mu(f)
\end{aligned}$$

hold and hence

$$\inf_{\widetilde{S}\in A_n^{(ad)}} \mu(\Delta(\widetilde{S}(f), f)^2)^{1/2} \leq {}^*\tilde{\sigma}_n^{(ad)}(F,S) \leq \tilde{\sigma}_n^{(ad)}(F,S).$$

2.1.10 Monte Carlo methods with varying cardinality

There are some papers which consider Monte Carlo methods with varying cardinality. I explain the basic notions intuitively, see *Wasilkowski* (1986b and 1987) and *Nemirovsky, Yudin* (1983) for exact definitions. We want to approximate a mapping

$S : F \to M$ by a method $\widetilde{S}(f) = \phi(f(a_1), \dots, f(a_n))$, where the a_i are chosen adaptively and moreover $n = n(f)$ may depend on $f \in F$. Now we consider the class $C(A_n^{var})$ of all Monte Carlo methods $Q = (Q(\omega))_{\omega\in\Omega}$ where $Q(\omega)$ is a method with varying cardinality and

$$\sup_{f\in F} \int_\Omega n_\omega(f)\, dm(\omega) \le n.$$

We also define

$$\sigma_n^{var}(F,S) = \inf_{Q\in C(A_n^{var})} \Delta_{max}(Q) = \inf_{Q\in C(A_n^{var})} \sup_{f\in F} \int_\Omega \Delta(Q(\omega)(f), f)\, dm(\omega).$$

Because of $C(A_n^{ad}) \subset C(A_n^{var})$ we have $\sigma_n^{var}(F,S) \le \sigma_n^{ad}(F,S)$. *Wasilkowski* (1987) conjectured that varying cardinality does not help significantly for the problem *Int* and certain function classes. We shall prove this conjecture in 2.2.10.

2.1.11 Notes and references

The aim of this section was not to give an introduction to or a survey of Monte Carlo methods (see *Ermakow* (1975), *Granovskii, Ermakov* (1977), *Hammersley, Handscomb* (1964), *Hengartner,Theodorescu* (1978), *Niederreiter* (1978), *Sobol* (1985), and *Kalos, Whitlock* (1986)) but to discuss some methodological difficulties which are related to these methods.
We introduced the notion of restricted Monte Carlo methods earlier in *Novak* (1985 and 1986b). The results of 2.1.8 are due to *Novak* (1986b), statements similar to those of 2.1.9 are known, see *Nemirovsky, Yudin* (1983).

2.2 Results for special problems

2.2.1 Remark

We have seen that Monte Carlo methods are interesting only for such problems $S : F \to M$, for which
i) the numbers $e_n(F,S)$ and $e_n^{ad}(F,S)$, respectively, are rather large for reasonable small $n \in \mathbf{N}$ and
ii) the stochastic error bounds $\sigma_n(F,S)$ or $\sigma_n^{ad}(F,S)$ are essentially smaller than the corresponding deterministic error bounds.
In this section we will show that both conditions are satisfied in some cases. On the other hand it will become apparent that Monte Carlo methods are used in practice even for those problems where the stochastic error bounds $\sigma_n^{(ad)}$ are only slightly smaller than the $e_n^{(ad)}$. In these cases stochastic methods seem to be of no special use and should be replaced by deterministic ones (such as Quasi Monte Carlo methods).
We begin with general lower bounds for the problems *App*, *Opt*, *Opt**, and *Int*. These estimates are analogous to those of 1.3.5 and follow from 2.1.9.

2.2.2 Lower bounds for the problem *App*

Proposition 1

Let $F \subset B(X)$ and f_i $(i = 1, \ldots, n+1)$ with the following properties:
i) the f_i have disjoint supports and satisfy $\sup f_i \geq \varepsilon$ for $i = 1, \ldots, n+1$;
ii) $\sum_{i=1}^{n+1} \delta_i f_i \in F$ for all $\delta_i \in \{-1, 1\}$.
Then

$$\sigma_n(F,App) \geq \sigma_n^{ad}(F,App) \geq {}^*\sigma_n^{ad}(F,App) \geq \varepsilon, \qquad (i)$$

$$\tilde{\sigma}_n(F,App) \geq \tilde{\sigma}_n^{ad}(F,App) \geq {}^*\tilde{\sigma}_n^{ad}(F,App) \geq \varepsilon, \qquad (ii)$$

and

$$e_n^{ad}(F,App) \geq \varepsilon. \qquad (iii)$$

(It is known (see *Lorentz* (1960) or *Tikhomirov* (1960)) that under the same conditions $d_n(F) \geq \varepsilon$ holds).

In some cases, the following estimates yield better results.

Proposition 2

Let $F \subset B(X)$ and f_i $(i = 1, \ldots, 2n)$ with the following properties:
i) the f_i have disjoint supports and satisfy $\sup f_i \geq \varepsilon$;
ii) $\pm f_i \in F$ for $i = 1, \ldots, 2n$.
Then

$${}^*\sigma_n^{ad}(F,App) \geq \frac{1}{2}\varepsilon, \qquad (i)$$

$$^*\tilde{\sigma}_n^{ad}(F,App) \geq \frac{1}{2}\sqrt{2}\varepsilon, \tag{ii}$$

and

$$e_{2n-1}(F,App) \geq \varepsilon. \tag{iii}$$

Proof:
Let $\widetilde{S} \in A_n^{ad}$ and $\widetilde{F} = \{\sum_{i=1}^{n+1} \delta_i f_i \mid \delta_i \in \{-1,1\}\}$. For $f \in \widetilde{F}$ let N_f be the information $N_f(g) = (g(x_1),\ldots,g(x_n))$ which is taken for f. If $\tilde{f}(x_i) = f(x_i)$, $i = 1,\ldots,n$, then $N_{\tilde{f}} = N_f$ and hence $\widetilde{S}(\tilde{f}) = \widetilde{S}(f)$.
For $\tilde{f} \in \widetilde{F}$, $\tilde{f} \neq f$, we get $\|f - \tilde{f}\| \geq 2\varepsilon$ and hence

$$\|\widetilde{S}(f) - f\| + \|\widetilde{S}(\tilde{f}) - \tilde{f}\| \geq 2\varepsilon.$$

Using 2.1.9 and the equidistribution on $\widetilde{F}$ we get

$$^*\sigma_n^{ad}(F,App) \geq \varepsilon$$

and the other statements of Proposition 1. The proof of Proposition 2 is quite similar.

2.2.3 Lower bounds for the problems *Opt* and *Opt**

Proposition 1
Let $F \subset B(X)$ and f_i $(i = 1,\ldots,2n)$ with the following properties:
i) the f_i have disjoint supports and satisfy $\sup f_i \geq \varepsilon$ for $i = 1,\ldots,2n$.
ii) $f_i \in F$ for $i = 1,\ldots,2n$.
Then

$$^*\sigma_n^{ad}(F,Opt) \geq \frac{n-1}{2n}\cdot\varepsilon, \tag{i}$$

$$^*\tilde{\sigma}_n^{ad}(F,Opt) \geq \sqrt{\frac{n-1}{2n}}\cdot\varepsilon, \tag{ii}$$

and

$$e_{2n-2}^{ad}(F,Opt) \geq \varepsilon. \tag{iii}$$

Proposition 2
If we additionally assume that $-f_i \in F$ for $i = 1,\ldots,2n$, then the following bounds for Opt^* hold:

$$^*\sigma_n^{ad}(F,Opt^*) \geq \frac{1}{4}\cdot\varepsilon, \tag{i}$$

$$^*\tilde{\sigma}_n^{ad}(F,Opt^*) \geq \frac{1}{4}\sqrt{2}\cdot\varepsilon, \tag{ii}$$

and

$$e_{2n-1}^{ad}(F,Opt^*) \geq \frac{1}{2}\cdot\varepsilon. \tag{iii}$$

Proof:
We show ${}^*\sigma_n^{ad}(F,Opt) \geq \frac{n-1}{2n} \cdot \varepsilon$ under the conditions of Proposition 1. Consider

$$\sum_{f \in \widetilde{F}} \sup f - f(\widetilde{S}(f)),$$

where $\widetilde{S} \in A_n^{ad}$ and $\widetilde{F} = \{f_i \mid i = 1, \dots, 2n\}$. Let $N_0(f) = (f(x_1), \dots, f(x_n))$ be the information which is taken for $g = 0$. Then

$$\widetilde{S}(f_i) = \phi(N_0(0)) = \tilde{x}$$

for at least n different $i = 1, \dots, 2n$ and hence

$$\sum_{f \in \widetilde{F}} \sup f - f(\widetilde{S}(f)) \geq (n-1) \cdot \varepsilon.$$

The statement follows by means of 2.1.9 using the equidistribution on $\widetilde{F}$. The proof of the other inequalities is analogous.

2.2.4 Lower bounds for the problem *Int*

Proposition 1
Let $F \subset B(X)$ and f_i $(i = 1, \dots, 2n)$ with the following conditions:
i) the f_i have disjoint supports and satisfy $Int(f_i) = S(f_i) \geq \varepsilon$ for $i = 1, \dots, 2n$.
ii) for all $\delta_i \in \{1, -1\}$ the function $\sum_{i=1}^{2n} \delta_i f_i$ is an element of F.
Then

$$ {}^*\sigma_n^{ad}(F,Int) \geq \frac{1}{4}\sqrt{2} \cdot \varepsilon \cdot n^{1/2}, \qquad (i)$$

$$ {}^*\tilde{\sigma}_n^{ad}(F,Int) \geq \frac{1}{2} \cdot \varepsilon \cdot n^{1/2}, \qquad (ii)$$

and

$$ e_n^{ad}(F,Int) \geq n \cdot \varepsilon. \qquad (iii)$$

In some cases, another estimate yields better results:

Proposition 2
If instead of ii) (in Proposition 1) the property
ii') for all $i = 1, \dots, 2n$ the functions $\pm f_i$ are an element of F
holds, then

$$ {}^*\sigma_n^{ad}(F,Int) \geq \frac{1}{2} \cdot \varepsilon, \qquad (i)$$

$$ {}^*\tilde{\sigma}_n^{ad}(F,Int) \geq \frac{1}{2}\sqrt{2} \cdot \varepsilon, \qquad (ii)$$

and

$$e^{ad}_{2n-1}(F,Int) \geq \varepsilon. \qquad (iii)$$

Proof:
To prove the inequality ${}^*\tilde{\sigma}^{ad}_n(F,Int) \geq \frac{1}{2}\varepsilon n^{1/2}$ under the conditions of Proposition 1, we assume that $\widetilde{S} \in A^{ad}_n$ and consider

$$\sum_{f\in\widetilde{F}} |S(f)-\widetilde{S}(f)|^2,$$

where

$$\widetilde{F} = \{\sum_{i=1}^{2n} \delta_i f_i \mid \delta_i \in \{1,-1\}\}.$$

Collecting those $f \in \widetilde{F}$ with the same information we see that

$$\sum_{f\in\widetilde{F}} |S(f)-\widetilde{S}(f)|^2 \geq 2^n \cdot \sum_{i=0}^{n} \binom{n}{i} (i-\frac{n}{2})^2 \varepsilon^2 = 2^{2n}\varepsilon^2 \cdot \frac{n}{4}.$$

If μ denotes the equidistribution on $\widetilde{F}$ we get

$$\inf_{\widetilde{S}\in A^{ad}_n} \mu(\Delta(\widetilde{S}(f),f)^2)^{1/2} \geq \frac{1}{2}\varepsilon \cdot n^{1/2}$$

and the statement follows by means of 2.1.9.
Similarly we see that

$$\begin{aligned}\sum_{f\in\widetilde{F}} |S(f)-\widetilde{S}(f)| \geq \quad & 2^n \cdot \textstyle\sum_{i=0}^{n} \binom{n}{i} |i-\frac{n}{2}| \cdot \varepsilon \geq \\ \geq \quad & 2^{2n} \cdot \varepsilon \cdot \frac{1}{4}\sqrt{2} \cdot n^{1/2}\end{aligned}$$

which proves the inequality

$${}^*\sigma^{ad}_n(F,Int) \geq \frac{1}{4}\sqrt{2} \cdot \varepsilon \cdot n^{1/2}.$$

To prove Proposition 2 one has to consider the quantities

$$\sum_{f\in\widetilde{F}} |S(f)-\widetilde{S}(f)|^2$$

and

$$\sum_{f\in\widetilde{F}} |S(f)-\widetilde{S}(f)|$$

with $\widetilde{S} \in A^{ad}_n$ and $\widetilde{F} = \{\pm f_i \mid i = 1,\ldots,2n\}$ in the same way.

2.2.5 Remark
We will give applications of 2.2.2–2.2.4 to special function classes in the following sections.
Bakhvalov (1959) seems to be the first to obtain similar lower bounds for the error of Monte Carlo methods. His result can be regarded as a (weaker) version of 2.2.4.1.

2.2.6 The problems *App*, *Opt*, *Opt, and *Int* for classes of Lipschitz functions**
In this section we treat Monte Carlo methods for the class F_s ($s \in \mathbf{N}$), defined by

$$F_s = \{f : [0,1]^s \to \mathbf{R} \mid |f(x) - f(y)| \leq \max_i |x_i - y_i|\}.$$

We compare the results with the respective worst case error bounds for deterministic methods, see 1.3.6 and 1.3.8.
First we consider the problems *App*, *Opt*, and *Opt** for the more general classes F_ω, see 1.3.6 for the respective worst case results. We shall see that for these problems the $e_n(F_\omega, S)$ are tight upper bounds for the $\sigma_n(F_\omega, S)$ (and also for the ${}^*\sigma_n^{ad}(F_\omega, S)$). The lower bounds are easily obtained by 2.2.2.1 and 2.2.3.

Proposition 1

$$\omega(c_{n+1}) \leq {}^*\sigma_n^{ad}(F_\omega, App); \tag{i}$$

$$\frac{n-1}{2n} \cdot \omega(c_{2n}) \leq {}^*\sigma_n^{ad}(F_\omega, Opt); \tag{ii}$$

$$1/4 \cdot \omega(c_{2n}) \leq {}^*\sigma_n^{ad}(F_\omega, Opt^*). \tag{iii}$$

Example

$$e_{n+1}(F_s, App) \leq {}^*\sigma_n^{ad}(F_s, App) \leq e_n(F_s, App); \tag{i}$$

$$(1/2 - \varepsilon) \cdot 2^{-1/s} \cdot e_n(F_s, Opt) \leq {}^*\sigma_n^{ad}(F_s, Opt); \tag{ii}$$

$$(1/2 - \varepsilon) \cdot 2^{-1/s} \cdot e_n(F_s, Opt^*) \leq {}^*\sigma_n^{ad}(F_s, Opt^*). \tag{iii}$$

((1) holds for all $n \in \mathbf{N}$, (2) and (3) are valid for all $\varepsilon > 0$ and $n > n(\varepsilon)$.)

Remark: Statements (2) and (3) or similar ones are known, see *Nemirovsky, Yudin* (1983).

Now we treat the problem *Int* for F_s and the usual Lebesgue measure $S = Int = \lambda^s$. We see that the error of suitable Monte Carlo methods is much smaller than the error of respective deterministic methods. The upper bound for ${}^*\tilde{\sigma}_n^{ad}(F_s, Int)$ is due to *Haber* (1966), *Bakhvalov* (1959) obtained slightly weaker results.

Proposition 2

$$\tilde{\sigma}_n(F_s,\lambda^s) \leq 12^{-1/2} \cdot n^{-1/2-1/s} \qquad (i)$$

for $n = m^s$, $m \in \mathbf{N}$.

$$^*\tilde{\sigma}_n^{ad}(F_s,\lambda^s) \geq (8 \cdot 2^{1/s} \cdot (s+1))^{-1} \cdot n^{-1/2-1/s} \qquad (ii)$$

for $n = 2^{s-1}m^s$, $m \in \mathbf{N}$.

$$^*\sigma_n^{ad}(F_s,\lambda^s) \geq (8 \cdot 2^{1/2+1/s} \cdot (s+1))^{-1} \cdot n^{-1/2-1/s} \qquad (iii)$$

for $n = 2^{s-1}m^s$, $m \in \mathbf{N}$.

Proof:
We consider the following generalized Monte Carlo method $Q \in {}^*C(A_n)$, where $n = m^s$, $m \in \mathbf{N}$. Let

$$Q(\omega)(f) = \frac{1}{n}\sum_{i=1}^{n} f(x_i(\omega)),$$

where the x_i are independent and each x_i is equidistributed on a cube C_i with edge length $1/m$ and $[0,1]^s$ is the union of the C_i. The random variable $Q(f)$ is a random Riemann sum. Because of

$$\widetilde{\Delta}(Q,f)^2 = \frac{1}{n}\sum_{i=1}^{n}\int_{C_i}\left(f(x) - n\int_{C_i} f(y)\,dy\right)^2 dx$$

we can conclude (see *Haber* (1966)) that

$$\widetilde{\Delta}(Q,f)^2 \leq \frac{1}{12} \cdot n^{-1-2/s} \qquad (f \in F_s)$$

and hence

$$^*\tilde{\sigma}_n^{ad}(F_s,\lambda^s) \leq 12^{-1/2} \cdot n^{-1/2-1/s}.$$

Since the set F_s is compact we can replace Q by disrete approximations $Q^l \in C(A_n)$. This proves the upper bound. The lower bounds follow from 2.2.4.1.

Remark
We give a special example. Let $s = 10$. Then $e_1(F_{10},\lambda^{10}) = 5/11$ and

$$e_n(F_{10},\lambda^{10}) = \frac{5}{44} \quad \text{for} \quad n = 2^{20} > 10^6.$$

We see that $e_n(F,S)$ tends to zero very slowly. Because of $\tilde{\sigma}_1(F_{10},\lambda^{10}) \leq 0.2887$ and

$$\tilde{\sigma}_n(F_{10},\lambda^{10}) \leq 7.1667 \cdot 10^{-5} \quad \text{for} \quad n = 2^{20}$$

the stochastic errors of suitable Monte Carlo methods are much smaller in this case.

2.2.7 Integration of functions of bounded variation

We consider the problem $Int = \lambda$ for the class

$$F_{\mathrm{Var}} = \{f : [0,1] \to \mathbf{R} \mid \mathrm{Var}(f) \le 1\}.$$

Proposition

$$e_n(F_{\mathrm{Var}},\lambda) = \frac{1}{2n}, \tag{i}$$

$$\frac{1}{8n-4}\cdot\sqrt{2} \le {}^*\tilde{\sigma}_n^{ad}(F_{\mathrm{Var}},\lambda) \le \frac{1}{2n}, \tag{ii}$$

$$\frac{1}{8n-4} \le {}^*\sigma_n^{ad}(F_{\mathrm{Var}},\lambda) \le \frac{1}{2n}. \tag{iii}$$

Remark

The first statement is due to *Zubrzycki* (1963/64), see also *Braß* (1977) (the midpoint rule is optimal).

The other estimates follow from 2.2.4.2 taking

$$f_1 = 1_{[0,1/(4n-2)]},\ f_2 = \tfrac{1}{2}\cdot 1_{[1/(4n-2),3/(4n-2)]},\ldots,\ f_{2n} = 1_{[(4n-3)/(4n-2),1]}.$$

It is not difficult to obtain similar results for more general measures on $[0,1]$ instead of λ, see *Novak* (1983).

The function class F_{Var} is a simple example where $\sigma_n(F,Int)$ is only slightly smaller than $e_n(F,Int)$.

2.2.8 The problems *App*, *Opt*, and *Opt** for Hölder and Sobolev classes of functions

Proposition

Let F be one of the classes $C_s^{k,\alpha}$ ($s\in\mathbf{N}$, $k\in\mathbf{N}_0$, $0<\alpha\le 1$) or $W_p^{k,s}$ with $kp>s$. Then

$$\sigma_n^{ad}(F,App) \asymp \sigma_n(F,App) \asymp \sigma_n^{ad}(F,Opt) \asymp \sigma_n(F,Opt) \asymp e_n(F,App).$$

Proof:

The lower bounds follow from 2.2.2.2 and 2.2.3 in the same way as the lower bounds for the $e_n(F,App)$ from 1.3.5a (see 1.3.9 and 1.3.11). Of course these bounds are also valid for the problem Opt^* and for the ${}^*\sigma_n^{(ad)}$ instead of the $\sigma_n^{(ad)}$. The upper bound is trivial.

Remarks
i) For similar results see *Nemirovsky,Yudin* (1983).
ii) Random search methods for solving the problem of global optimization are used by many authors, see *Brooks* (1958, 1959), *Rubinstein* (1981), *Schwefel* (1977), *Zieliński, Neumann* (1983), and *Müller,Nollau,Polovinkin* (1986) for a survey.

2.2.9 The problem *Int* for Hölder and Sobolev classes

We begin with lower bounds.

Proposition 1

$$^*\sigma_n^{ad}(C_s^{k,\alpha},\lambda^s) \geq c \cdot n^{-\frac{k+\alpha}{s} - \frac{1}{2}}, \qquad (i)$$

$$^*\sigma_n^{ad}(W_p^{k,s},\lambda^s) \geq c \cdot n^{-\frac{k}{s} - \frac{1}{2}} \qquad (p \geq 2), \qquad (i)$$

$$^*\sigma_n^{ad}(W_p^{k,s},\lambda^s) \geq c \cdot n^{-\frac{k}{s} - 1 + \frac{1}{p}} \qquad (1 \leq p < 2), \qquad (iii)$$

where $c > 0$ depends on the class F.

Proof:
These lower bounds follow from 2.2.4. The f_i are taken similar as in the proofs of 1.3.9 and 1.3.11. To prove the last statement, we take 2.2.4.2, whereas the other estimates follow from 2.2.4.1.

Remarks
i) Proposition 1 is known, see *Bakhvalov* (1959) for the first statement. The other estimates are stated without proof in *Bakhvalov* (1962).
ii) To prove the respective upper bounds (not only for $Int = \lambda^s$, but for arbitrary measures on $[0,1]^s$), we need
— a stochastic analog of the interpolation theorem 1.2.8 (Lemma 1)
— inequalities similar to Lemma 1 of 1.3.12 (Lemma 2)
— the partitioning lemma for measures on $[0,1]^s$ (see Lemma 2 of 1.3.12).

Lemma 1
Let (X,A,μ) be a measure space with a positive finite measure μ and let $V \subset L_2(X,\mu)$ be a vector space with $dim\ V = n$ and $1 \in V$. Then there is a stochastic quadrature formula $Q \in {}^*C(A_n)$ with the following properties:

$$m(Q(f)) = \mu(f) \qquad \text{for all } f \in L_1(X,\mu); \qquad (i)$$

$$Q(f) = \mu(f) \qquad \text{for all } f \in V; \qquad (ii)$$

$$m((Q(f) - \mu(f))^2) \leq \|\mu\| \cdot \inf_{g \in V} \mu((f-g)^2) \qquad \text{for all } f \in L_2(X,\mu). \qquad (iii)$$

Remarks

i) Lemma 1 is due to *Ermakov, Zolotukhin* (1960), see also *Ermakow* (1975) and *Hammersley, Handscomb* (1964). Examples can be found in *Handscomb* (1964). Stochastic quadrature formulas which are exact for certain low order polynomials are also investigated in *Bogues, Morrow, Patterson* (1981) and *Siegel, O'Brian* (1985).

ii) Let μ as in Lemma 1 and $F \subset L_2(X,\mu)$. Then

$$\sigma_{n+1}(F,\mu) \leq d_{n,2}(F,\mu) \cdot \|\mu\|^{1/2},$$

where $d_{n,2}(F,\mu)$ denotes the n-width in the space $L_2(X,\mu)$. This inequality follows from Lemma 1, see *Novak* (1983, 1986a), and may help to give estimates of the $\sigma_n(F,\mu)$.

Lemma 2

Let $a,b \in \mathbf{R}$ with $a < b$ and let $s \in \mathbf{N}$.

a) Let $k \in \mathbf{N}_0$ and $0 < \alpha \leq 1$. Then there is a $c > 0$ such that for every $f \in C_s^{k,\alpha}([a,b]^s)$ there is a $p \in P(k+1,s)$ with

$$\|f-p\|_\infty \leq c \cdot (b-a)^{k+\alpha}.$$

b) Let $k \in \mathbf{N}$. Then there is a $c > 0$ such that for every $f \in W_2^{k,s}([a,b]^s)$ there is a $p \in P(k,s)$ with

$$\|f-p\|_2 \leq c \cdot (b-a)^k \sum_{|\alpha|=k} \|D^\alpha f\|_2.$$

(See Lemma 1 of 1.3.12. Statement b) is known as Poincaré's inequality).

Proposition 2

$$^*\tilde{\sigma}_n(C_s^{k,\alpha},\lambda^s) \asymp n^{-\frac{k+\alpha}{s}-\frac{1}{2}}; \qquad (i)$$

$$^*\tilde{\sigma}_n(W_p^{k,s},\lambda^s) \asymp n^{-\frac{k}{s}-\frac{1}{2}} \qquad \text{if } p \geq 2; \qquad (ii)$$

$$^*\tilde{\sigma}_n(W_p^{k,s},\lambda^s) \asymp n^{-\frac{k}{s}-1+\frac{1}{p}} \qquad (iii)$$

if $1 \leq p < 2$ and $k - s/p + s/2 \geq 0$.

Proof:

The lower bounds follow from Proposition 1. Because of known imbedding theorems (see, for example, *Kufner et al.* (1977)) iii) follows from ii) and ii) only has to be proved for $p = 2$. Similarly as in the case of deterministic methods we use composed stochastic quadrature formulas. Because the proofs of i) and ii) are very similar, we only consider the case of Hölder classes.

Let $k \in \mathbf{N}_0$, $0 < \alpha \leq 1$, $s \in \mathbf{N}$ and $m \in \mathbf{N}$. By Lemma 1 there is a stochastic quadrature formula $Q \in {}^*C(A_l)$, where $l = dim\, P(k+1,s)$ such that

$$E(Q(f)) = \int_{[0,1/m]^s} f(x)\, d\lambda^s \qquad (f \in L_1), \tag{i}$$

$$Q(f) = \int_{[0,1/m]^s} f(x)\, d\lambda^s \qquad (f \in P(k+1,s)), \tag{ii}$$

and

$$E\left((Q(f) - \int_{[0,1/m]^s} f(x)\, d\lambda^s)^2\right) \leq m^{-s} \cdot \inf_{p \in P(k+1,s)} \|f-p\|_2^2 \tag{iii}$$

(for $f \in L_2$). Applying this method independently to each of the m^s subcubes yields a method $Q_n \in {}^*C(A_n)$, $n = l \cdot m^s$. The error of Q_n can be estimated as follows.

$$\begin{aligned} E\left((Q_n(f) - \int_{[0,1]^s} f(x)\, d\lambda^s)^2\right) &\leq m^{-s} \cdot \textstyle\sum_{j=1}^{m^s} \inf_{p \in P(k+1,s)} \|f-p\|_2^{2(j)} \\ &\leq m^{-s} \cdot \textstyle\sum_{j=1}^{m^s} m^{-s} c^2 m^{-2k-2\alpha} \\ &= c^2 \cdot m^{-2k-2\alpha-s} \end{aligned}$$

for $f \in C_s^{k,\alpha}$ and hence

$${}^*\tilde{\sigma}_n(C_s^{k,\alpha}, \lambda^s) \leq \tilde{c} \cdot n^{-\frac{k+\alpha}{s} - \frac{1}{2}}.$$

Remarks

a) Proposition 2 is known, see *Bakhvalov* (1959) for i) and *Bakhvalov* (1962), where ii) and iii) are stated without proof. Composed stochastic quadrature formulas are also considered in *Haber* (1969).

b) Until now we have only considered the numbers ${}^*\sigma_n$ and ${}^*\tilde{\sigma}_n$ which are unrealistic. It is not difficult to show that the respective generalized Monte Carlo methods can be suitably approximated by our restricted Monte Carlo methods if $F \subset C([0,1]^s)$ is a compact set. As in the case of deterministic methods (see Proposition 3 of 1.3.12) we can consider arbitrary (finite Borel) measures and obtain the following result.

Proposition 3

$$\sup_{\|Int\| \leq 1} \tilde{\sigma}_n(C_s^{k,\alpha}, Int) \asymp n^{-\frac{k+\alpha}{s} - \frac{1}{2}}; \tag{i}$$

$$\sup_{\|Int\| \leq 1} \tilde{\sigma}_n(W_p^{k,s}, Int) \asymp n^{-\frac{k}{s} - \frac{1}{2}} \qquad (p \geq 2,\ kp > s); \tag{ii}$$

$$\sup_{\|Int\| \leq 1} \tilde{\sigma}(W_p^{k,s}, Int) \asymp n^{-\frac{k}{s} - 1 + \frac{1}{p}} \qquad (1 \leq p < 2,\ kp > s). \tag{iii}$$

Sketch of proof:
The lower bound follows from Proposition 1. To show the upper bound we use composed stochastic quadrature formulas (as in the proof of Proposition 2) but with a more complicated partition of $[0,1]^s$ using Lemma 2 of 1.3.12. Now the proof is very similar to that of Proposition 3 of 1.3.12.

2.2.10 Monte Carlo methods with varying cardinality
Now we want to prove some error bounds for Monte Carlo methods with varying cardinality. These lower bounds show that varying cardinality does not help significantly for certain problems. For simplicity, we only deal with the problem

$$S(f) = \int_{[0,1]^s} f(x)\, d\lambda^s.$$

(The reader will observe, however, that the problems *App*, *Opt*, and *Opt** can be treated in quite the same way.) In particular, we want to prove a conjecture of *Wasilkowski* (1987). We begin with some lower bounds which are similar to Lemma 2.2.4.

Lemma 1
a) Let F be a set of integrable functions and f_i $(i = 1,\ldots,4n)$ with the following conditions:
i) the f_i have disjoint supports and satisfy $S(f) = \int f_i \geq \varepsilon$ for $i = 1,\ldots,4n$.
ii) for all $\delta_i \in \{1,-1\}$ the function $\sum_{i=1}^{4n} \delta_i f_i$ is an element of F.
We define $F_0 = \{\sum_{i=1}^{4n} \delta_i f_i \mid \delta_i \in \{1,-1\}\}$ and consider the equidistribution μ on $F_0 \subset F$. Let be given a deterministic method $\widetilde{S} = \phi \circ N$ with $\int_F n(f)\, d\mu(f) \leq n$. Then

$$\Delta_\mu(\widetilde{S}) \geq \frac{1}{4} \cdot \varepsilon \cdot n^{1/2}.$$

b) We assume that (in statement a)) instead of ii) the property
ii') for all $i = 1,\ldots,4n$ the functions $\pm f_i$ are an element of F
holds. We define $F_0 = \{\pm f_i \mid i = 1,\ldots,4n\}$ and consider the equidistribution μ on F_0. Let be given a method $\widetilde{S} = \phi \circ N$ with $\int_F n(f)\, d\mu(f) \leq n$. Then

$$\Delta_\mu(\widetilde{S}) > \frac{2-\sqrt{2}}{2} \cdot \varepsilon.$$

Proof:
First we prove the statement a). From $\int_F n(f)\, d\mu(f) \leq n$ we get

$$\mu\{f \mid n(f) \leq 2n\} \geq 1/2.$$

To give an estimation for $\Delta_\mu(\widetilde{S})$, we collect those $f \in F_1 = \{f \in F_0 \mid n(f) \leq 2n\}$ which yield the same information $N(f)$. We may assume that F_1 contains 2^{4n-1} functions of the form

$$f = \sum_{i=1}^{2n} \delta_i f_i + \sum_{i=2n+1}^{4n} \delta_i f_i,$$

whereby the signs in the first sum are known and the signs in the second sum are unknown and arbitrary. We get

$$\sum_{f \in F_0} |S(f) - \phi(N(f))| \geq \frac{1}{2} \cdot 2^{2n} \cdot \sum_{i=0}^{2n} \binom{2n}{i} \cdot |i - n| \cdot \varepsilon \geq \frac{1}{4} \cdot 2^{4n} \cdot \varepsilon \cdot n^{1/2}$$

and this proves the inequality

$$\Delta_\mu(\widetilde{S}) \geq \frac{1}{4} \cdot \varepsilon \cdot n^{1/2}.$$

Now we prove b). The optimal information satisfies $n(f) = 1$ for two functions, $n(f) = 2$ for two functions and so on till $n(f) = k$ for two functions. Because of $\int_F n(f)\, d\mu(f) \leq n$ we get $k < \sqrt{8} \cdot n$ and therefore $N(f) = 0$ with probability greater than $\frac{4-\sqrt{8}}{4}$. It follows that

$$\Delta_\mu(\widetilde{S}) > \frac{2-\sqrt{2}}{2} \cdot \varepsilon.$$

To prove lower bounds for different function classes, one can use Lemma 1 together with the known estimates of Lemma 2.

Lemma 2
a) The assumptions of Lemma 1a can be satisfied with

$$\varepsilon(n) \asymp n^{-1-k/s}, \qquad \varepsilon(n) \asymp n^{-1-(k+\alpha)/s}, \qquad \varepsilon(n) \asymp n^{-1}$$

if $F = W_p^{k,s}$, $F = C_s^{k,\alpha}$ and $F = F_{\mathrm{Var}}$, respectively.
b) The assumptions of Lemma 1b can be satisfied with

$$\varepsilon(n) \asymp n^{-k/s-1/p}, \qquad \varepsilon(n) \asymp n^{-(k+\alpha)/s}, \qquad \varepsilon(n) \asymp n^{-1}$$

if $F = W_p^{k,s}$, $F = C_s^{k,\alpha}$ and $F = F_{\mathrm{Var}}$, respectively.

Proposition

$$\sigma_n^{var}(W_p^{k,s}, \lambda^s) \asymp n^{-k/s-r}, \tag{i}$$

where $kp > s$ and $r = 1/2$ if $p \geq 2$ and $r = 1 - 1/p$ if $1 \leq p < 2$;

$$\sigma_n^{var}(C_s^{k,\alpha}, \lambda^s) \asymp n^{-(k+\alpha)/s-1/2}; \tag{ii}$$

and

$$\sigma_n^{var}(F_{\mathrm{Var}}, \lambda) \asymp n^{-1}. \tag{iii}$$

Proof:
The upper bounds follow from 2.2.7 and 2.2.9. We prove the lower bounds. Let $Q \in C(A^{var}_{n/2})$, i.e.

$$\sup_{f \in F} \int_\Omega n_\omega(f)\, dm(\omega) \le n/2.$$

It follows that $\int_F \int_\Omega n_\omega(f)\, dm(\omega) d\mu(f) \le n/2$ and hence $m\{\omega \mid \int_F n_\omega(f)\, d\mu(f) \le n\} \ge 1/2$. By Lemma 1a [Lemma 1b] we conclude that

$$m\{\omega \mid \Delta_\mu(Q(\omega)) \ge \tfrac{1}{4}\varepsilon(n) n^{1/2}\} \ge 1/2 \qquad [m\{\omega \mid \Delta_\mu(Q(\omega)) \ge \tfrac{2-\sqrt{2}}{2}\varepsilon(n)\} \ge 1/2]$$

and therefore

$$\int_\Omega \int_F |S(f) - Q(\omega)(f)|\, d\mu(f) dm(\omega) \ge \tfrac{1}{8}\varepsilon(n) n^{1/2} \qquad [\ldots \ge \tfrac{2-\sqrt{2}}{2}\varepsilon(n)]$$

which yields

$$\Delta_{max}(Q) \ge \frac{1}{8}\varepsilon(n) n^{1/2} \qquad [\Delta_{max}(Q) \ge \frac{2-\sqrt{2}}{4}\varepsilon(n)].$$

Using Lemma 2a for $C_s^{k,\alpha}$ and $W_p^{k,s}$ for $p \ge 2$ and Lemma 2b for F_{Var} and $W_p^{k,s}$ in the case $1 \le p < 2$ we obtain our statements.

2.2.11 Notes and references

The lower bounds 2.2.2–2.2.4 and some of its consequences seem to be new although similar error estimates have been used beginning with *Bakhvalov* (1959). Proposition 3 of 2.2.9 gives new upper bounds for the error of stochastic quadrature formulas, see *Novak* (1983) for similar results. The results of 2.2.10 are from *Novak* (1987b).

3. Average error bounds

3.1 Averages over the class of problem elements

3.1.1 Remark
Let a problem $S : F \to M$ and an error function $\Delta : M \times F \to \mathbf{R}_0^+$ be given. One may argue that the maximal error

$$\Delta_{max}(\widetilde{S}) = \sup_{f \in F} \Delta(\widetilde{S}(f), f)$$

of a deterministic method $\widetilde{S}$ is a very pessimistic estimate for the true error $\Delta(\widetilde{S}(f_0), f_0)$ for some given $f_0 \in F$. Hence it seems to be reasonable, additionally to consider the average error

$$\int_F \Delta(\widetilde{S}(f), f)\ d\mu(f),$$

where μ is a suitable probability measure on F.
In Chapters 1 and 2 we have seen that adaption does not help for many problems if we consider the maximal error (over F) of deterministic or stochastic methods. This might change if we consider average case errors.

Example
Let $F_1 = \{f : [0,1] \to \mathbf{R} \mid |f(x) - f(y)| \le |x - y|\}$ and $n \in \mathbf{N}$. In Chapter 1 and 2 we have proved the following facts:

$$e_n^{ad}(F_1, App) = e_n(F_1, App) = e_n^{ad}(F_1, Opt) = e_n(F_1, Opt) = \frac{1}{2n}, \tag{i}$$

$$e_n^{ad}(F_1, Opt^*) = e_n(F_1, Opt^*) = e_n^{ad}(F_1, \lambda) = e_n(F_1, \lambda) = \frac{1}{4n}, \tag{ii}$$

$$\frac{1}{2n+2} \le \sigma_n^{ad}(F_1, App) \le \sigma_n(F_1, App) \le \frac{1}{2n}, \tag{iii}$$

$$\frac{n-1}{8n^2} \le \sigma_n^{ad}(F_1, Opt) \le \sigma_n(F_1, Opt) \le \frac{1}{2n}, \tag{iv}$$

$$\frac{1}{16n} \le \sigma_n^{ad}(F_1, Opt^*) \le \sigma_n(F_1, Opt^*) \le \frac{1}{4n}, \tag{v}$$

and

$$\frac{\sqrt{2}}{64} \cdot n^{-3/2} \le \sigma_n^{ad}(F_1, \lambda) \le \sigma_n(F_1, \lambda) \le 12^{-1/2} \cdot n^{-3/2}. \tag{vi}$$

We mention one further result concerning the problem $Int = \lambda$ and the class F_1. Let $r_n(f) = |S(f) - \widetilde{S}(f)|$ be the error of the trapezoidal rule with n knots. *Braß* (1979) proved that

$$\sum_{n=1}^{\infty} r_n(f) \cdot n^{\beta} < \infty \qquad (f \in F_1)$$

holds for each $\beta < 1/2$ and concluded that the error of the trapezoidal rule bahaves "on the average" as $n^{-3/2}$.

The conjecture seems to be sound that adaption helps in an average case setting for the problems *Opt* and *Opt**. A crucial point is finding a suitable measure on the set F. We will discuss this problem in 3.1.4–3.1.7 and begin with basic definitions.

3.1.2 Definitions

Let μ be a (Borel) probability measure on F. The number

$$\Delta_\mu(\widetilde{S}) = \int_F^* \Delta(\widetilde{S}(f), f)\, d\mu(f)$$

is called the *average error* of the deterministic method $\widetilde{S} : F \to M$ with respect to μ. Analogously to the $e_n^{(ad)}(F,S)$ we define

$$e_n^{(ad)}(F,S,\mu) = \inf_{\widetilde{S} \in A_n^{(ad)}} \Delta_\mu(\widetilde{S}).$$

Remarks

a) Instead of $\int^* \Delta(\widetilde{S}(f), f)\ d\mu(f)$ often $\widetilde{\Delta}_\mu(\widetilde{S}) := (\int^* \Delta(\widetilde{S}(f), f)^2\, d\mu(f))^{1/2}$ is taken as the average error of $\widetilde{S}$. We define $\tilde{e}_n^{(ad)}(F,S,\mu) = \inf_{\widetilde{S} \in A_n^{(ad)}} \widetilde{\Delta}_\mu(\widetilde{S})$.

b) Using the upper integral, $\Delta_\mu(\widetilde{S})$ is defined for all $\widetilde{S}$. *Wasilkowski* (1983) gives a slightly different definition introducing the notion of the local average error, see A.5.

3.1.3 The average error of Monte Carlo methods

Similar definitions could be given for Monte Carlo methods:
Let $Q = Q(\omega)_{\omega \in \Omega}$ be a Monte Carlo method and let

$$\Delta_\mu(Q) = \int_F^* \left(\int_\Omega \Delta(Q(\omega)(f), f)\ dm(\omega) \right) d\mu(f)$$

be the average error of Q. Then we define

$$\sigma_n^{(ad)}(F,S,\mu) = \inf_{Q \in C(A_n^{(ad)})} \Delta_\mu(Q).$$

If $\int \int \Delta(Q(\omega)(f), f) dm(\omega) d\mu(f)$ exists for a $Q \in C(A_n^{(ad)}))$ and all $\omega \in \Omega$ then there is a $\widetilde{S} \in A_n^{(ad)}$ with $\Delta_\mu(\widetilde{S}) \le \Delta_\mu(Q)$. Therefore the numbers $\sigma_n^{(ad)}(F,S,\mu)$ should be the same as the corresponding $e_n^{(ad)}(F,S,\mu)$ in most practical cases and hence will not be further discussed.

Remark

In *Voronin, Skalyga* (1984), the authors compare deterministic and stochastic methods for the problem *Int* and compute average errors relating to the Wiener measure. Unfortunately the authors only consider the simplest Monte Carlo method. Our considerations show that such a comparison cannot prove a superiority of either classical or stochastic methods.

3.1.4 Remark

We want to discuss the problem of how a probability measure μ on F should be chosen. First of all, each special choice of μ yields some further information on the problem given by $S : F \to M$ and Δ. In some practical situations a suitable μ may be at hand. In many cases, however, fixing μ may be difficult. It was pointed out by *Nemirovsky,Yudin* (1983, page 26–27) that even in simple cases an empirical approach to the construction of the a priori distribution μ on F is hopeless. As a consequence, one should look for measures which are distinct in a theoretical way. One natural class of probability measures is that of the Gaussian measures in Hilbert spaces. Also Wiener measures on Banach spaces have been considered by several authors, see 3.1.11 for references.

We take an alternative approach. When speaking of an average error on F, we should at least postulate that μ has full support. If $F = \{f_1,\dots,f_n\}$ is a finite set then the equidistribution $\mu(\{f_i\}) = 1/n$ is clearly distinguished and naturally is useful for averaging. This measure is the unique one with maximal entropy and it seems to be adequate to relate the average error to this measure. We try to proceed similarly in more general cases and look for "fair" measures on F: Each special averaging process yields some extra information about the problem. We want to make this information as small as possible (principle of maximal entropy, see *Guiasu, Shenitzer* (1985) and *Jaynes* (1979)). Many classes F of functions used in numerical analysis are of the form

$$F = K + V,$$

where $K \subset B(X)$ is compact and $V \subset B(X)$ is a finite dimensional vector space. In 3.1.5 and 3.1.6 we study approaches to how to define a "fair" measure μ on F in the case where F is a compact metric space. Because of geometric properties of infinite dimensional spaces which are related to the concentration of measure phenomenon, see *Milman, Schechtman* (1986), our results are mainly negative because we do not find a reasonable measure on sets such as

$$F_1' = \{f : [0,1] \to [-1,1] \mid |f(x) - f(y)| \le |x-y|\}.$$

3.1.5 Homogeneous measures on compact metric spaces

We suppose that F is a compact metric space. One could postulate that for a fair measure μ on F two balls with the same radius should have the same probability. This claim, however, is so strong that such a measure only exists on very special spaces. Therefore we weaken this claim in the following way.

Definition
Let F be a compact metric space and let μ be a (Borel) probability measure on F. The measure μ is called homogeneous if there is a $c > 0$ such that for all $r > 0$ and $f_1, f_2 \in F$ the following inequality holds:

$$\mu(\{g \mid d(g, f_1) \leq r\} \geq c \cdot \mu(\{g \mid d(g, f_2) \leq r\}).$$

We show that on F_1' no such measure exists. More general, the following proposition holds.

Proposition
Let $F \subset B(X)$ be compact and convex and assume that F contains a maximal element f_0 and an element g_0 with $\inf(f_0 - g_0) > 0$. Then there is a homogeneous measure μ on F if and only if F is finite dimensional.

Proof
a) If F is compact, convex, and finite dimensional, the normed Lebesgue measure on F is a homogenous measure. We do not give the formal arguments, because we are mainly interested in the case where F is infinite dimensional.
b) Now let $F \subset B(X)$ with the given properties and let μ be a homogenous measure on F. Let $f_0 \in F$ be maximal and $g_0 \in F$ with $a = \inf(f_0 - g_0) > 0$, $b = \sup(f_0 - g_0)$. Then for $0 < r \leq a + b$

$$\{h \mid d(f_0, h) \leq r\} \subset \left\{h \mid d\left(f_0 + \frac{r}{a+b} \cdot (g_0 - f_0), h\right) \leq \frac{br}{a+b}\right\}$$

holds. Therefore we have

$$\begin{aligned} \mu(\{h \mid d(g,h) \leq r\} &\leq \tfrac{1}{c} \cdot \mu(\{h \mid d(f_0, h) \leq r\}) \\ &\leq \tfrac{1}{c} \cdot \mu\left(\left\{h \mid d\left(f_0 + \tfrac{r}{a+b} \cdot (g_0 - f_0), h\right) \leq \tfrac{br}{a+b}\right\}\right) \\ &\leq \tfrac{1}{c^2} \cdot \mu\left(\left\{h \mid d(f, h) \leq \tfrac{br}{a+b}\right\}\right) \end{aligned}$$

for arbitrary $f, g \in F$ and $0 < r \leq a + b$.
It follows that the numbers p_n (see 1.3.7) can be estimated by

$$p_n \leq k \cdot r^{-\alpha}$$

where $k > 0$ and $\alpha > 0$. Because of this bound, the set F must be finite dimensional, see *Lorentz* (1966, page 167, problem 5).

Remarks

a) I do not know if there is any infinite dimensional compact convex set in a Banach space on which there is a homogeneous measure.

b) There is another notion of a homogeneous measure on a metric space in the literature. Let F be a compact metric space and let μ be a Borel measure on F. We say that μ satisfies the doubling condition if

$$\mu(\{g \mid d(g,f) \leq 2r\}) \leq C \cdot \mu(\{g \mid d(g,f) \leq r\}) \qquad (f \in F,\ r > 0)$$

for some $C > 0$. It is known, see *Volberg, Konyagin* (1984), that there is a nontrivial measure with the doubling condition on any compact subset of $\mathbf{R}^n$ and more general: There is a nontrivial measure with the doubling condition on F if and only if the metric dimension of F is finite, i.e. $p_n \leq k \cdot n^{-\alpha}$ for some $k > 0$ and $\alpha > 0$.

3.1.6 Equidistributions on compact metric spaces

After this failed attempt, we consider another approach to define "fair" measures on a compact metric space. Continuing with the principle of maximal entropy, one has to define the entropy of a probability measure on compact metric spaces. This is done with the notion of ε-entropy which was defined and information theoretically founded in *Posner, Rodemich* (1971, 1972) and *McEliece, Posner* (1971, 1973). Now we could consider probability measures with maximal ε-entropy, where $\varepsilon > 0$. Although this idea seems to be sound, the choice of ε is difficult, however, and therefore we only consider the limit $\varepsilon \to 0$. This leads to the following definition.

Definition

Let F be a compact metric space and let μ be a (Borel) probability measure. The measure μ is called an *equidistribution* on F, if there is a positive null sequence $(r_i)_{i\in\mathbf{N}}$ with the following properties:

For r_i there is a minimal covering of F by balls $\{f \mid d(f,g^i_j) \leq r_i\}$ $(j = 1,\ldots,n(i))$ so that for the measures

$$\mu_i = \frac{1}{n(i)} \sum_{j=1}^{n(i)} \varepsilon_{g^i_j}$$

the sequence $(\mu_i)_{i\in\mathbf{N}}$ converges (weakly) to μ.

Remark

The existence of an equidistribution on F follows from the fact that each sequence of probability measures has a convergent subsequence. For many special F there exists exactly one equidistribution but in general uniqueness does not hold. (I conjecture that uniqueness holds in the case where F is a compact convex subset of a Banach space).

The following proposition gives a characterization of the Lebesgue measure on Jordan measurable sets $F \subset \mathbf{R}^n$. (Here we assume that F is a metric space, where the metric is given by some norm in $\mathbf{R}^n$).

Proposition 1
Let $F \subset \mathbf{R}^n$ be compact and Jordan measurable with $\lambda^n(F) > 0$. Then there is a unique equidistribution μ on F. Moreover, μ is the normed Lebesgue measure on F:

$$\mu = \frac{1}{\lambda^n(F)} \cdot \lambda^n{}_{|F}.$$

Proof:
The following covering property of Jordan measurable sets was proved by *Hlawka* (1949) and independently by *Kolmogorov,Tikhomirov* (1959):
If $X \subset \mathbf{R}^n$ is Jordan measurable and $C(X,\varepsilon)$ is the number of ε-balls which are necessary to cover X, then

$$\lim_{\varepsilon \to 0} C(X,\varepsilon)/\varepsilon^n = k \cdot \lambda^n(X),$$

where k is a constant, only depending on n and the norm in $\mathbf{R}^n$. Our proposition follows by means of this theorem.

Remark
Because of this result one could hope that equidistributions have nice properties on other spaces, too. However, we have the following negative result concerning the class

$$F_1' = \{f : [0,1] \to [-1,1] \mid |f(x) - f(y)| \leq |x-y|\}.$$

Proposition 2
The equidistribution μ on F_1' is unique. The measure μ is identical with the one dimensional normed Lebesgue measure on the constants $\{f = c \mid c \in [-1,1]\}$.

Proof:
For $\varepsilon > 0$ let M_ε be the set of all functions $f : [0,1] \to \mathbf{R}$ with the following properties:
a) $f(0)$ is a multiple of 2ε and $|f(0)| < 1+\varepsilon$;
b) between $\varepsilon \cdot k$ and $\varepsilon \cdot (k+1)$, where $k \in \mathbf{N}_0$, the function f is linear with gradient 1 or -1;
c) f is continuous.

The number of elements of M_ε may easily be computed, one also can show (see *Kolmogorov,Tikhomirov* (1959)) that M_ε is a relatively economical ε-covering of F_1', more exactly, the inequality

$$card\ M_\varepsilon \leq 18 \cdot C(F_1', \varepsilon)$$

holds for all $0 < \varepsilon < 1/16$. (Again $C(A, \varepsilon)$ denotes the least number of ε-balls that are necessary to cover A). By means of easy combinatoric considerations one can show: M_ε consists in the case $\varepsilon = 1/n$ with even n of $(n+1) \cdot 2^n$ functions and of $n \cdot 2^n$ if n is odd. If $f_0 \in M_\varepsilon$, in both cases the number of the $f \in M_\varepsilon$ for which $\sup(f - f_0)$ has a fixed value $l \cdot \varepsilon$ (with l even) is at most 2^n. The same is valid for $\inf(f - f_0)$ instead of $\sup(f - f_0)$. Therefore, we have

$$\limsup_{\varepsilon \to 0} C(\{f \mid r - \delta \leq d(f, f_0) \leq r + \delta\}, \varepsilon) \,/\, card\ M_\varepsilon \leq 4\delta$$

for all $f_0 \in F_1'$ and $\delta > 0$. It follows that

$$\limsup_{\varepsilon \to 0} C(\{f \mid r - \delta \leq d(f, f_0) \leq r + \delta\}, \varepsilon) \,/\, C(F_1') \leq 72\delta$$

and we get the following statement.
The equidistribution μ on F_1' is unique and for each ball $A = \{f \mid d(f, f_0) \leq r\}$ (with $f_0 \in F_1'$ and $r > 0$)

$$\mu(A) = \lim_{\varepsilon \to 0} C(A, \varepsilon) / C(F_1', \varepsilon)$$

is valid. The special form of μ now follows from well known limit theorems of probability theory: One can show that for small $\varepsilon > 0$ most $f \in M_\varepsilon$ are very close to a constant function. I skip the details.

Remark
For our purpases, Proposition 2 is a purely negative one: It is evident that μ is by no means suitable to define an average error for approximations that are defined on F_1'.

3.1.7 A "natural" measure for functions of bounded variation

Graf, Mauldin, Williams (1986) constructed a "natural" measure μ on the set

$$H = \{h : [0,1] \to [0,1] \mid h \text{ nondecreasing and continuous, } h(0) = 0,\ h(1) = 1\}.$$

This measure μ can be described by the following method to generate at random an element $h \in H$. We first choose the value of $h(\frac{1}{2})$ according to the uniform distribution over $[0,1]$. Then we choose $h(\frac{1}{4})$ according to the uniform distribution over $[0, h(\frac{1}{2})]$. Independently we choose $h(\frac{3}{4})$ according to the uniform distribution over $[h(\frac{1}{2}), 1]$. Continue this point process. It turns out that with probability one, a function so constructed from the set of dyadic rationals into $[0,1]$ extends to a $h \in H$. Thus, this process defines a probability measure μ on H. The resulting measure μ can be described by the formula

$$(*) \qquad \int_H G(h)\, d\mu(h) = \int_H \int_H \int_0^1 G([f,g]_y)\, dy\, d\mu(f)\, d\mu(g)$$

which holds for each integrable $G : H \to \mathbf{R}$, where

$$[f,g]_y(t) = \begin{cases} yf(2t) & t \le 1/2 \\ y + (1-y)g(2t-1) & t > 1/2. \end{cases}$$

We are interested in the set

$$F = \{f : [0,1] \to \mathbf{R} \mid f \text{ continuous}, \ f(0) = 0, \ \mathrm{Var}(f) \le 1\}.$$

We observe that F can be written as

$$F = \{z_1 h_1 - z_2 h_2 \mid z_i \ge 0, h_i \in H, z_1 + z_2 \le 1\}.$$

A "natural" probability measure $\tilde{\mu}$ on F is given by

$$\int_F G(f)\, d\tilde{\mu}(f) = 2 \int_Z \int_H \int_H G(z_1 h_1 - z_2 h_2)\, d\mu(h_1)\, d\mu(h_2)\, dz_1\, dz_2,$$

where $Z = \{(z_1, z_2) \mid z_i \ge 0, z_1 + z_2 \le 1\}$. Using formula $(*)$ we can compute the average error

$$\widetilde{\Delta}_{\tilde{\mu}}(S_n) = \left(\int_F \left(S_n(f) - \int_0^1 f(x)\, dx \right)^2 d\tilde{\mu}(f) \right)^{1/2}$$

of the trapezoidal rule

$$S_n(f) = \frac{1}{2n} f(1) + \sum_{i=1}^{n-1} \frac{1}{n} f\left(\frac{i}{n}\right)$$

for $n = 2^m$ $(m \in \mathbf{N}_0)$.
We obtain $\widetilde{\Delta}_{\tilde{\mu}}(S_1) = 120^{-1/2}$ and

$$\widetilde{\Delta}_{\tilde{\mu}}(S_{2n}) = \frac{1}{\sqrt{6}} \cdot \widetilde{\Delta}_{\tilde{\mu}}(S_n) \qquad (n \in \mathbf{N}).$$

Furthermore we can prove that S_n $(n = 2^m)$ is not far from being optimal in the sense that

$$\tilde{e}_n(F, Int, \tilde{\mu}) \asymp n^{-\log 6/\log 4},$$

(where $\log 6/\log 4 = 1.29248\ldots$), see *Graf, Novak* (1987). It is interesting to compare this result with the worst case error bound

$$e_n(F, Int) = \frac{1}{2n+1}.$$

See also Proposition 2 of 3.1.9 for a related result.

3.1.8 "Can adaption help on the average?"

We have seen that adaption does not help for many problems if we consider the maximal error of deterministic or stochastic methods. *Wasilkowski, Woźniakowski* (1984) proved that even in an average case setting adaption does not help for certain linear problems.

Let F and M be real separable Hilbert spaces and let $S : F \to M$ be a continuous linear operator. Let the average error of $\widetilde{S} : F \to M$ be defined by

$$\widetilde{\Delta}_\mu(\widetilde{S}) = \left(\int \|S(f) - \widetilde{S}(f)\|^2 \, d\mu(f) \right)^{1/2},$$

where μ is a (Borel) probability measure on F. *Wasilkowski, Woźniakowski* (1984) show that adaption does not help for a wide class of measures including Gaussian measures. See *Wasilkowski* (1985) and *Packel, Woźniakowski* (1987) for a survey of related results.

There are also linear problems, however, where adaption helps on the average for many probability measures.

Example

We consider a simple discrete analog of the problem *Int* for the class F_1^l and show that adaption helps on the average for every probability measure with full support. (The reader will observe, however, that adaptive methods are only slightly better than nonadaptive ones.)

Let $m \in \mathbf{N}$, $m \geq 5$, and

$$F = \{f : \{0, 1/m, \ldots, 1\} \to \{0, \pm 1/m, \ldots, \pm 1\} \mid |f(x) - f(y)| \leq |x - y|\}.$$

We consider the problem

$$S(f) = \frac{1}{m+1} \sum_{i=0}^{m} f(i/m).$$

Let μ be a probability measure on F with full support and let $5 \leq n \leq m$. Each $\widetilde{S} \in A_n$ is of the form $\widetilde{S}(f) = \phi(f(a_1), \ldots, f(a_n))$, where $\{a_1, \ldots, a_n\}$ is a subset of $\{0, 1/m, \ldots, 1\}$. We may assume that $a_1 < a_2 < \cdots < a_n$ and can choose an $i \in \{1, 2, \ldots, n-2\}$ such that there is a point $x \in \{0, 1/m, \ldots, 1\}$ with $x \neq a_k$ for all k and $x < a_i$ or $x > a_{i+2}$. This is possible because of $m \geq 5$ and $5 \leq n \leq m$. Consider the following adaptive information $N_a \in I_n^{ad}$:

$$N_a(f) = \begin{cases} (f(a_1), \ldots, f(a_n)) & \text{if} \quad |f(a_{i+2}) - f(a_i)| < |a_{i+2} - a_i| \\ (f(a_1), \ldots, f(a_i), f(a_{i+2}), \ldots, f(a_n), f(x)) & \text{else.} \end{cases}$$

It is easy to see that the information N_a admits an approximation $\widetilde{S}' = \phi' \circ N_a$ such that

$$\Delta(\widetilde{S}'(f), f) \leq \Delta(\widetilde{S}(f), f)$$

for all $f \in F$ and $\Delta(\widetilde{S}'(f), f) < \Delta(\widetilde{S}(f), f)$ for some $f \in F$. As a consequence, each $\widetilde{S}^* \in A_n^{ad}$ with least average error is adaptive.

3.1.9 Bounds for the $e_n^{(ad)}(F, S, \mu)$

Two questions concerning the $e_n(F,S,\mu)$ and $e_n^{ad}(F,S,\mu)$ are of main interest:
i) How much do the numbers $e_n^{(ad)}(F,S,\mu)$ differ from the $e_n^{(ad)}(F,S)$?
ii) How much do the $e_n^{ad}(F,S,\mu)$ differ from the $e_n(F,S,\mu)$, i.e. how much does adaption help on the average?
The answer to both questions of course depends on the measure μ. There are problems, however, where the numbers $e_n^{(ad)}(F,S,\mu)$ are much smaller than the $e_n^{(ad)}(F,S)$ for every probability measure.
For the class F_s, as well as for many other classes, the following statements hold:
a) For each $n \in \mathbf{N}$ there is a probability measure μ on F such that $e_n^{ad}(F,S,\mu)$ is only slightly smaller than $e_n(F,S)$ if $S = App$, Opt, or Opt^*. For such an unfavourable measure the average error of optimal methods is not much smaller than the maximal error and adaption does not help on the average. This is interesting particularly for the problems *Opt* and *Opt** because one could guess that adaptive methods are better than nonadaptive ones on the average.
b) Concerning the problem $Int = \lambda^s$, the numbers $e_n(F,S,\mu)$ are much smaller than $e_n(F,S)$ for arbitrary measures μ, while adaptive methods are again only slightly better: For all four problems the numbers

$$\sup_{\mu} e_n^{(ad)}(F,S,\mu)$$

hardly differ from the numbers

$$\sup_{\mu} e_n(F,S,\mu),$$

where μ runs through all (Borel) probability measures on F.
Using results from Chapter 2, we can give lower and upper bounds of these numbers. First we consider the classes

$$F_s = \{f : [0,1]^s \to \mathbf{R} \mid |f(x) - f(y)| \le \max_i |x_i - y_i|\}.$$

Proposition 1

$$e_{n+1}(F_s, App) \le \sup_{\mu} e_n^{ad}(F_s, App, \mu) \le e_n(F_s, App); \qquad (i)$$

$$(1/2 - \varepsilon) \cdot 2^{-1/s} \cdot e_n(F_s, Opt) \le \sup_{\mu} e_n^{ad}(F_s, Opt, \mu) \le e_n(F_s, Opt); \qquad (ii)$$

$$(1/2-\varepsilon)\cdot 2^{-1/s}\cdot e_n(F_s,Opt^*)\le \sup_\mu e_n^{ad}(F_s,Opt^*,\mu)\le e_n(F_s,Opt^*);\qquad (iii)$$

$$\sup_\mu e_n(F_s,\lambda^s,\mu)\le 12^{-1/2}\cdot n^{-1/2-1/s},\qquad (iv)$$

for $n=m^s$, $m\in\mathbf{N}$;

$$\sup_\mu e_n^{ad}(F_s,\lambda^s,\mu)\ge (8\cdot 2^{1/2+1/s}\cdot(s+1))^{-1}\cdot n^{-1/2-1/s},\qquad (v)$$

for $n=2^{s-1}\cdot m^s$, $m\in\mathbf{N}$.

Proof:
Observe that we have proved the same estimates for the numbers $\sigma_n^{(ad)}(F_s,S)$ in 2.2.6. The lower bounds follow from the considerations in 2.2.2–2.2.4, where we actually proved lower bounds for the numbers $\sup_\mu e_n^{ad}(F,S,\mu)$. The upper bounds are trivial for the problems *App*, *Opt*, and *Opt**. To prove statement iv) we proceed as in the proof of Proposition 2 of 2.2.6.
Let

$$Q(\omega)(f)=\frac{1}{n}\sum_{i=1}^{n} f(x_i(\omega))$$

be a Monte Carlo method with

$$\Delta(Q,f)\le 12^{-1/2}\cdot n^{-1/2-1/s}+\varepsilon$$

(for all $f\in F_s$, where $n=m^s$, $m\in\mathbf{N}$) and let μ be a probability measure on F_s. The integral

$$\Delta_\mu(Q(\omega))=\int_{F_s}\Delta(Q(\omega(f),f)\,d\mu(f)$$

clearly exists for each $\omega\in\Omega$ and because of

$$\begin{aligned}\int_\Omega \Delta_\mu(Q(\omega))\,dm(\omega) &= \int\int \Delta(Q(\omega)(f),f)\,d\mu(f)\,dm(\omega)\\ &= \int_F \Delta(Q,f)\,d\mu(f)\\ &\le \sup_{f\in F}\Delta(Q,f)=\Delta_{max}(Q)\\ &\le 12^{-1/2}\cdot n^{-1/2-1/s}+\varepsilon\end{aligned}$$

there is a $\omega\in\Omega$ with

$$\Delta_\mu(Q(\omega))\le 12^{-1/2}\cdot n^{-1/2-1/s}+\varepsilon.$$

Remark
Using results from Chapter 2, we can prove similar results for other function classes. We do not repeat the proofs which are quite similar to that of Proposition 1.

Proposition 2
Let $F_{\mathrm{Var}} = \{f : [0,1] \to \mathbf{R} \mid \mathrm{Var}(f) \le 1\}$. Then

$$\frac{1}{8n-4} \le \sup_{\mu} e_n^{ad}(F_{\mathrm{Var}}, \lambda, \mu) \le \frac{1}{2n}$$

for all $n \in \mathbf{N}$.

Proposition 3
Let F be one of the classes $C_s^{k,\alpha}$ (with $s \in \mathbf{N}$, $k \in \mathbf{N}_0$, and $0 < \alpha \le 1$) or $W_p^{k,s}$ (with $kp > s$). Then

$$\sup_{\mu} e_n^{ad}(F, App, \mu) \asymp \sup_{\mu} e_n^{ad}(F, Opt, \mu) \asymp e_n(F, App).$$

Proposition 4

$$\sup_{\mu} e_n^{(ad)}(W_p^{k,s}, \lambda^s, \mu) \asymp n^{-k/s-1/2}, \qquad (i)$$

where $kp > s$ and $p \ge 2$;

$$\sup_{\mu} e_n^{(ad)}(W_p^{k,s}, \lambda^s, \mu) \asymp n^{-k/s-1+1/p}, \qquad (ii)$$

where $kp > s$ and $1 \le p < 2$;

$$\sup_{\mu} e_n^{(ad)}(C_s^{k,\alpha}, \lambda^s, \mu) \asymp n^{-(k+\alpha)/s-1/2}, \qquad (iii)$$

and the same estimates are valid for

$$\sup_{\mu} \sup_{\|Int\| \le 1} e_n^{(ad)}(F, Int, \mu)$$

instead of $\sup_{\mu} e_n^{(ad)}(F, \lambda^s, \mu)$.

3.1.10 Methods with varying cardinality and “unfavourable” measures
In 3.1.9 we gave estimates of the numbers $\sup_{\mu} e_n^{(ad)}(F, S, \mu)$ for different problems. It can be proved that for all those problems there is an “unfavourable” measure μ^* such that $e_n^{(ad)}(F, S, \mu^*)$ is only slightly smaller than $\sup_{\mu} e_n^{(ad)}(F, S, \mu)$ for all $n \in \mathbf{N}$. Actually this is true even if we allow methods $\widetilde{S}$ with varying cardinality (see 2.1.10). For simplicity we only consider the problem $Int = \lambda^s$, but similar results are valid for the problems App, Opt, and Opt^*, too. Given a probability measure μ we consider the class of all methods with varying cardinality and

$$cost_{\mu}(\widetilde{S}) = \int_F n(f)\, d\mu(f) \le n.$$

Proposition
i) Let $\varepsilon > 0$, $kp > s$, and $p \geq 2$. Then there is a measure μ on $W_p^{k,s}$ such that

$$\inf_{cost_\mu(\widetilde{S}) \leq n} \Delta_\mu(\widetilde{S}) \geq c \cdot n^{-k/s-1/2-\varepsilon}$$

for some $c > 0$.
ii) Let $\varepsilon > 0$, $kp > s$, and $1 \leq p < 2$. Then there is a measure μ on $W_p^{k,s}$ such that

$$\inf_{cost_\mu(\widetilde{S}) \leq n} \Delta_\mu(\widetilde{S}) \geq c \cdot n^{-k/s-1+1/p-\varepsilon}$$

for some $c > 0$.
iii) Let $\varepsilon > 0$. Then there is a measure μ on $C_s^{k,\alpha}$ such that

$$\inf_{cost_\mu(\widetilde{S}) \leq n} \Delta_\mu(\widetilde{S}) \geq c \cdot n^{-(k+\alpha)/s-1/2-\varepsilon}$$

for some $c > 0$.

Proof:
We use Lemma 1 and Lemma 2 of 2.2.10. We assume that F is a class satisfying the assumptions of Lemma 1a [Lemma 1b, respectively] with

$$\delta(n) \asymp n^{-\beta-1/2} \qquad [\delta(n) \asymp n^{-\beta}]$$

and $\beta > 0$. Let $(c_m)_{m \in \mathbf{N}}$ be a sequence of positive numbers with

$$\sum_{m=1}^{\infty} c_m = 1. \tag{1}$$

For $m \in \mathbf{N}$ let μ_m be an equidistribution

$$\mu_m = \frac{1}{2^{4m}} \cdot \sum_{i=1}^{2^{4m}} \varepsilon_{f_i} \qquad [\mu_m = \frac{1}{8m} \cdot \sum_{i=1}^{8m} \varepsilon_{f_i}]$$

as considered in Lemma 1a [Lemma 1b], where ε_f denotes the respective Dirac measure. Now we define the measure μ by

$$\mu = \sum_{m=1}^{\infty} c_m \mu_m.$$

Clearly μ is a probability measure on F. Let $\widetilde{S} = A \circ N$ with $\int_F n(f)\, d\mu(f) \leq n$. We conclude that

$$\sum_{m=1}^{\infty} c_m \cdot \int_F n(f)\, d\mu_m(f) \leq n. \tag{2}$$

It follows from Lemma 1 that

$$\Delta_{\mu_m}(\widetilde{S}) \geq c \cdot m^{-\beta} \qquad \text{if} \quad \int_F n(f)\, d\mu_m(f) \leq m. \tag{3}$$

Also it is clear that

$$\Delta_\mu(\widetilde{S}) = \sum_{m=1}^{\infty} c_m \cdot \Delta_{\mu_m}(\widetilde{S}). \tag{4}$$

Let $k(n)$ be the largest number such that

$$\sum_{m=1}^{k(n)} (m+1) \cdot c_m \le n. \tag{5}$$

By means of (2) – (5) we get

$$\Delta_\mu(\widetilde{S}) \ge c \cdot \sum_{m=k(n)+1}^{\infty} c_m \cdot m^{-\beta}. \tag{6}$$

Now we take

$$c_m = k \cdot m^{-1-\alpha} \tag{7}$$

with $0 < \alpha < 1$ and $k > 0$ such that $\sum_{m=1}^{\infty} c_m = 1$. It follows from (5) that

$$k(n) \asymp n^{1/(1-\alpha)} \tag{8}$$

in this case and by (6) we get

$$\Delta_\mu(\widetilde{S}) \ge \tilde{c} \cdot n^{-(\alpha+\beta)/(1-\alpha)}. \tag{9}$$

To conclude the proof take Lemma 1a and Lemma 2a for the classes $C_s^{k,\alpha}$ and $W_p^{k,s}$ for $p \ge 2$ and take Lemma 1b and Lemma 2b for $W_p^{k,s}$ in the case $1 \le p < 2$.

3.1.11 Notes and references

The average case analysis of algorithms for problems which are defined in infinite dimensional spaces began with the papers *Suldin* (1959, 1960) and *Larkin* (1972). These papers only consider linear algorithms, nonadaptive information, and linear problems such as *App* and *Int*. Arbitrary algorithms and adaptive information for linear problems in a finite dimensional space are studied by *Traub, Wasilkowski, Woźniakowski* (1984). The analysis of the infinite dimensional setting requires the techniques of measure theory in abstract spaces, see *Kuo* (1975), *Parthasarathy* (1967), and *Skorohod* (1974). Linear problems in Hilbert and Banach spaces are studied in *Wasilkowski, Woźniakowski* (1984), *Micchelli* (1984), *Farber* (1984), *Wasilkowski* (1986), *Voronin, Temirgaliev* (1986), *Lee* (1986), and in *Lee, Wasilkowski* (1986), see *Wasilkowski* (1985) and *Packel, Woźniakowski* (1987) for a survey.
The results of 3.1.10 are from *Novak* (1987b).

3.2 The average over the set of information

3.2.1 Remark
Let there be given a mapping $S : F \to M$ and an error function $\Delta : M \times F \to \mathbf{R}_0^+$. We want to define the average a posteriori error of an approximation $\widetilde{S} = \phi \circ N$ ($\widetilde{S} \in A_n^{ad}$, $N \in I_n^{ad}$). The estimate

$$\Delta(\widetilde{S}(f), f) \leq \Delta_{max}(\widetilde{S})$$

is the best one, valid for all $f \in F$. This estimate is an a priori error bound, it does not depend on the information $N(f)$. Knowing $N(f) = x \in \mathbf{R}^n$, of course

$$\Delta(\widetilde{S}(f), f) \leq \sup\{\Delta(\phi(x), g) \mid g \in F,\ N(g) = x\} =: \alpha(x)$$

is valid. This a posteriori error bound may be much smaller than $\Delta_{max}(\widetilde{S})$ for many $f \in F$.
If μ is a probability measure on $N(F)$ we can define an average a posteriori error by

$$\int_{N(F)} \alpha(x)\, d\mu(x).$$

In this case the choice of μ should strongly depend on N. If, for instance, we change N by a linear mapping, then μ should be appropriately changed. If an a priori measure μ_F on F is known then μ should be taken as $\mu = \mu_F N^{-1}$. In the following, we take a suitably normed Lebesgue measure, because the Lebesgue measure is distinguished in many ways (see, for example, 3.1.6). We begin with nonadaptive methods $\widetilde{S} \in A_n$.

3.2.2 Definition
First case: Let $N \in I_n$ and let $N(F) \subset \mathbf{R}^n$ be measurable with positive finite Lebesgue measure. Then let

$$\Delta_{aver}(\phi, N) = \frac{1}{\lambda^n(N(F))} \int_{N(F)} \alpha(x)\, d\lambda^n$$

if this integral exists, else take the upper integral which is defined always.
We regard this definition of the *average a posteriori error* of $\widetilde{S}$ as satisfactory if $N(F)$ has a positive finite measure. Often a somewhat more general case occurs which we want to cover in the second case (which formally includes the first case).
Second case: Let $N(F)$ be of the form $N(F) = P + V$ with a vector space V and a set P that has for an $m \leq n$ a positive finite λ^m-measure. Furthermore, let α have the property $\alpha(p + v_1) = \alpha(p + v_2)$ for all $p \in P$ and $v_1, v_2 \in V$. Then we define

$$\Delta_{aver}(\phi, N) = \frac{1}{\lambda^m(P)} \int_P \alpha(x)\, d\lambda^m$$

(and take again the upper integral, if this integral does not exist.)

In imitation of the numbers $e_n(F,S)$, $\sigma_n(F,S)$, and $e_n(F,S,\mu)$ we define

$$a_n(F,S) = \inf_{\phi\circ N\in A_n} \Delta_{aver}(\phi,N).$$

The infimum has to be taken over all $\widetilde{S} = \phi\circ N \in A_n$, for which $\Delta_{aver}(\phi,N)$ exists.

Remarks
a) This definition is due to *Novak* (1986b, c).
b) We give a simple example for the second case of this definition:
Consider the problem $S = Int = \lambda$ on $F = F_1$. For $N(f) = (f(0), f(1))$ and $\widetilde{S}(f) = 1/2(f(0)+f(1))$ we have

$$N(F_1) = \{(x,y) \in \mathbf{R}^2 \mid |x-y| \le 1\}$$

and $\alpha((x,y)) = 1/4 - 1/4|x-y|^2$. Since $\lambda^2(N(F)) = \infty$ we can not define $\Delta_{aver}(\phi, N)$ as in the first case but because of

$$\alpha((x,y)) = ((0,y-x))$$

it seems to be reasonable to define

$$\Delta_{aver}(\phi,N) = \frac{1}{2}\int_{-1}^{1} \alpha((0,t))\,dt = 1/6$$

as in the second case.
c) It should be stressed that in the case $\widetilde{S} = \phi_1\circ N_1 = \phi_2\circ N_2$ the equality $\Delta_{aver}(\phi_1,N_1) = \Delta_{aver}(\phi_2, N_2)$ does not hold in general. An approximation $\widetilde{S}$ is related with different average a posteriori errors, depending on the information that is taken as a basis. This is a necessary consequence from the fact that we form the averaging process regarding to the known information in each case.

With the help of the given definition of $\Delta_{aver}(\phi, N)$ we are now ready to define some more important notions. This is mainly done in imitation of the ideas of *Traub et al.* (1980, 1983).

Definition
Let $N : F \to \mathbf{R}^n$ be some information and let

$$A = \{\widetilde{S} = \phi\circ N \mid \phi : \mathbf{R}^n \to M\}.$$

a) A method $S^*_{aver} = \phi^*\circ N \in A$ is called a method with least average error (related to N), if $\Delta_{aver}(\phi^*,N)$ exists and $\Delta_{aver}(\phi^*,N) > \Delta_{aver}(\phi,N)$ for no $\phi\circ N$.
b) A method $S_c \in A$ is called central in the average case, if for almost all $x \in N(F)$ and all $\widetilde{S} = \phi\circ N$ the inequality $\alpha_{S_c}(x) \le \alpha_{\widetilde{S}}(x)$ holds.

c) The number $r_{aver}(N) = \inf \Delta(\phi, N)$ is called the average case radius of the information N. The infimum is to be taken over all $\phi \circ N \in A$ for which the average a posteriori error is defined.

It is easy to prove the following proposition.

Proposition
Let $S_c = \phi_c \circ N$ be central in the average and let $\Delta(\phi_c, N) \in \mathbf{R}$ exist. Then a method $\widetilde{S} \in A$ has least average error if and only if $\widetilde{S}$ is central in the average.

3.2.3 Examples

The inequality $a_n(F,S) \le e_n(F,S)$ seems to be rather obvious and it holds if we postulate that $\Delta_{aver}(\phi, N)$ is defined for suitable $\widetilde{S} \in A_n$.
Our first example shows that there are linear problems such that

$$a_n(F,S) = e_n(F,S)$$

for all $n \in \mathbf{N}$, whereas the second example shows that there are linear problems with

$$a_n(F,S) \to 0 \qquad \text{but} \qquad e_n(F,S) = 1$$

for all n.

Example 1
Let $a_1 > a_2 > \dots$ be a sequence of positive numbers with $\sum_{i=1}^{\infty} a_i < \infty$ and let

$$F = \{f : \mathbf{N} \to \mathbf{R} \mid |f(i)| \le a_i\}.$$

We consider $S_1(f) = f$ with $\Delta_1(m, f) = \|m - f\|_1$ and $S_2(f) = \sum_{i=1}^{\infty} f(i)$ with

$$\Delta_2(m, f) = |m - S_2(f)|.$$

It is easy to see that in both cases

$$e_n(F, S_i) = \sum_{i=n+1}^{\infty} a_i \qquad \text{and} \qquad a_n(F, S_i) = e_n(F, S_i)$$

hold for all n.
An algorithm $S^* \in A_n$ which is optimal both in the worst case and in the average case is

$$S^*(f) = (f(1), \dots, f(n), 0, \dots) \qquad \text{for } S_1$$

and

$$S^*(f) = \sum_{i=1}^{n} f(i) \qquad \text{for } S_2.$$

We also see that the optimal algorithm (and hence the optimal information) is unique in each case.

This is an example where the knowledge of $N(f)$ does not yield a better a posteriori error bound for any $f \in F$.

Example 2
Let F be defined by

$$F = \{f : [0,1[\to \mathbf{R} \mid f \text{ is continuous}, f(0) = 0, \text{ and } \mathrm{Var}(f) \leq 1\}.$$

We consider the problem

$$S(f) = \lim_{x \to 1} f(x)$$

with $\Delta(m, f) = |m - S(f)|$. It is easy to see that $e_n(F,S) = 1$ for all $n \in \mathbf{N}$. We compute $a_n(F,S)$. Let $S^* = \phi^* \circ N \in A_n$ be defined by $S^*(f) = f(a_n)$, where $N(f) = (f(a_1), \ldots, f(a_n))$ with $0 < a_1 < \ldots < a_n < 1$. We get $\alpha(x) = (1 - |x_1| - |x_1 - x_2| - \ldots - |x_{n-1} - x_n|)$ for

$$x \in N(F) = \{x \in \mathbf{R}^n \mid |x_1| + |x_1 - x_2| + \ldots + |x_{n-1} - x_n| \leq 1\}.$$

An easy computation of the integral yields

$$\Delta_{aver}(\phi^*, N) = \frac{1}{n+1}.$$

For every other $\widetilde{S} = \phi \circ N$ we have $\alpha_{\widetilde{S}}(x) \geq \alpha_{S^*}(x)$ and hence S^* is optimal.

3.2.4 Nonadaptive methods for the problem *App* for Lipschitz functions

We consider the problem *App* for the class F_1. Let $n \in \mathbf{N}$ and

$$N(f) = (f(a_1), \ldots, f(a_n)),$$

where $0 \leq a_1 < \ldots < a_n \leq 1$. The set of information is given by

$$N(F_1) = \{(x_1, \ldots, x_n) \mid |x_{i+1} - x_i| \leq |a_{i+1} - a_i|\}$$

and we have

$$\alpha(x) = \max_i \left(a_1, 1 - a_n, \frac{1}{2}(a_{i+1} - a_i) - \frac{1}{2}|x_{i+1} - x_i| \right).$$

Comparing the given information with

$$\widetilde{N}(f) = \left(\ldots, f(a_i), f\left(\frac{a_i + a_{i+2}}{2}\right), f(a_{i+2}), \ldots \right)$$

we can see that $r_{aver}(\widetilde{N}) \le r_{aver}(N)$ and hence the optimal information N^*_{aver} satisfies

$$a_1 = a, \quad a_2 = a+b, \quad \dots, \quad a_n = a+(n-1)b = 1-a.$$

Writing $y_i = x_{i+1} - x_i$ and $P = \{(y_1,\dots,y_{n-1}) \mid |y_i| \le b\}$ we get $\alpha(y) = \max_i (a, b/2 - |y_i|/2)$ and hence

$$\Delta_{aver}(\phi^*, N) = \frac{n-1}{b^{n-1}} \int_0^b h(z) z^{n-2}\, dz$$

for such an information, where

$$h(z) = \begin{cases} a & \text{for } z \le 2a \\ z/2 & \text{for } z > 2a. \end{cases}$$

Calculating this integral we get

$$\Delta_{aver}(\phi^*, N) = \frac{2^{n-1}a^n}{n \cdot b^{n-1}} + \frac{(n-1)\cdot b}{2n},$$

where $2a = 1-(n-1)b$ and $2a < b$. We denote by c_n the optimal value of $2a/b$ and obtain the equation

$$1 - c_n^n - n c_n^{n-1} = 0.$$

It is easy to obtain the estimates

$$(n+1)^{-1/n} < c_n < (n+1)^{-1/(n-1)}$$

and

$$a_n(F_1, App) > \frac{1}{2(n+1)}.$$

We see that $a_n(F,S)$ is only slightly smaller than $e_n(F,S)$ for this example. The optimal information for $n=2$ is given by $N^*_{aver}(f) = (f(1/2 - 1/4\cdot\sqrt{2}), f(1/2 + 1/4\cdot\sqrt{2}))$ and we have $a_2(F_1, App) = 1/2\cdot\sqrt{2} - 1/2$.

3.2.5 Nonadaptive methods for the problem *Opt* for Lipschitz functions

We treat the problem *Opt* for the class F_1 and consider the fixed nonadaptive information $N(f) = (f(a_1),\dots,f(a_n))$ with $0 \le a_1 < \dots < a_n \le 1$. It is easy to see that in the case $n=1$ the method $S^*(f) = 1/2$ has (independent from N) both least maximal and least average a posteriori error and that

$$e_1(F_1, Opt) = a_1(F_1, Opt) = \frac{1}{2}.$$

Now we assume $n > 1$ and define an approximation $S^* = \phi^* \circ N$ as follows. Let $x \in N(F_1)$, $m_1 = x_1 + a_1$, $m_2 = \max(1/2\cdot(x_1+x_2) + 1/2\cdot(a_2-a_1), \dots, 1/2\cdot(x_{n-1} +$

$x_n) + 1/2 \cdot (a_n - a_{n-1}))$, and $m_3 = x_n + (1 - a_n)$. Let i_0 be the least number such that $x_{i_0} = \max x_k$. If we put $\phi(x) = a_{i_0}$ then the local error $\alpha(x)$ amounts to

$$\alpha_1 = \max(m_1, m_2, m_3) - x_{i_0}.$$

The following facts are easy to prove:
i) For no other choice $\phi(x) \in [a_1, a_n]$ the respective local error is less than α_1.
ii) If $m_2 \geq m_1$ and $m_2 \geq m_3$ then the error $\alpha(x)$ is less than α_1 for no choice $\phi(x) \in [0,1]$ and hence we define $\phi^*(x) = a_{i_0}$.
Now we consider the case $m_1 > m_2$ and $m_1 \geq m_3$ (the third case, $m_3 > m_1$ and $m_3 > m_2$, is treated analogously). Here the choice $\phi(x) = a_{i_0}$ is not always best possible because there could be a better $\tilde{a} < a_1$. For $\tilde{a} < a_1$ the error $\alpha(x)$ of $\phi(x) = \tilde{a}$ is given by

$$\alpha(x) = \max(x_1 - a_1 + 2\tilde{a}, m_2, m_3) - (x_1 - a_1 + \tilde{a})$$

and hence the optimal choice $a^* < a_1$ satisfies

$$x_1 - a_1 + 2a^* = \max(m_2, m_3)$$

with $\alpha(x) := \alpha_2 = a^*$. It follows that the choice $\phi(x) = a^*$ is better than $\phi(x) = a_{i_0}$ iff $a_2 < a_1$ or

$$\max(m_2, m_3) < 3x_1 + a_1 - 2x_{i_0}.$$

We define $\phi^*(x)$ accordingly and have proved the following proposition.

Proposition 1
In this situation the approximation $S^* = \phi^* \circ N$ as defined above has both least maximal and least average error.

Remark
Usually only the approximation $\widetilde{S}(f) = a_{i_0}$ is considered, see for example *Sukharev* (1984). We have seen that this method is not best possible, even in the worst case setting. The unique optimal information $N^*_{max} \in I_n$ in the worst case setting is given by $N^*_{max}(f) = (f(\frac{1}{2n}), \ldots, f(\frac{2n-1}{2n}))$. The optimal information N^*_{aver} with respect to the average a posteriori error is different for $n > 1$. To give an example, we treat the case $n = 2$.

Proposition 2
In the case $n = 2$, the optimal information N^*_{aver} is unique and given by $N^*_{aver}(f) = (f(1/2 - 1/6 \cdot \sqrt{3}), f(1/2 + 1/6 \cdot \sqrt{3}))$. Furthermore, $a_2(F_1, Opt) = 1/4 \cdot (\sqrt{3} - 1)$ holds.

Proof:
The optimal information N^*_{aver} has to be of the form $N(f) = (f(a), f(1-a))$ with $a \leq 1/4$. First we indicate the optimal approximation and the local error $\alpha(x)$ for

such an information according to Proposition 1. Because of the symmetry we assume $x_2 = 0$ and $x_1 \geq 0$. Then we have

$$S^*(f) = \begin{cases} 1/4 - 1/4 \cdot x_1 & \text{if } x_1 > 1 - 4a \\ a & \text{if } x_1 \leq 1 - 4a \end{cases}$$

and

$$\alpha(x) = \begin{cases} 1/4 - 1/4 \cdot x_1 & \text{if } x_1 > 1 - 4a \\ 1/2 - 1/2 \cdot x_1 - a & \text{if } x_1 \leq 1 - 4a. \end{cases}$$

The radius $r_{aver}(N)$ is given by

$$r_{aver}(N) = \frac{1}{1-2a} \int_0^{1-2a} \alpha((x_1, 0))\, dx_1$$

and an easy calculation yields

$$r_{aver}(N) = \frac{1 - 4a + 6a^2}{4 - 8a}$$

from which the statement follows.

Remark
It seems to be very difficult to compute the optimal information N^*_{aver} for large n. Numerical calculations show that — similar as for the problem *App* — the optimal worst case information is nearly optimal. Furthermore we conjecture that the asymptotic estimate

$$a_n(F_1, Opt) \approx \frac{1}{3n}$$

is true.

3.2.6 Nonadaptive methods for the problem *Int* for Lipschitz functions

We treat nonadaptive methods for the problem $Int = \lambda$ and the Lipschitz class F_1. As information, we consider $N(f) = (f(a_1), \ldots, f(a_n))$ where $0 \leq a_1 < a_2 < \ldots < a_n \leq 1$ are fixed. In this situation the following statements hold.

Proposition 1
There is exactly one method S^*_{aver} using N with least average error and S^*_{aver} is given by $S^*_{aver} = \sum_{i=1}^n c_i \cdot f(a_i)$, where

$$c_1 = a_1 + 1/2(a_2 - a_1),$$

$$c_i = 1/2(a_i - a_{i-1}) + 1/2(a_{i+1} - a_i) \qquad \text{for } i = 2, \ldots, n-1,$$

and

$$c_n = 1/2(a_n - a_{n-1}) + (1 - a_n).$$

Moreover $r_{aver} = 1/2a_1^2 + 1/2(1-a_n)^2 + 1/6 \cdot \sum_{i=1}^{n-1}(a_{i+1} - a_i)^2$. The method S^*_{aver} also has least maximal error and

$$r_{max}(N) = 1/2a_1^2 + 1/2(1-a_n)^2 + 1/4 \cdot \sum_{i=1}^{n-1}(a_{i+1} - a_i)^2.$$

Proof:
For given $x \in N(F)$ the set $\{S(f) \mid f \in F, N(f) = x\}$ is an interval in $\mathbf{R}$. Therefore, there is a unique method $S^*_{aver} = \phi^* \circ N$ which is central in the average and ϕ^* is the midpoint of this interval. One easily shows that $\phi^*(x) = \sum_{i=1}^{n} c_i x_i$ with the given c_i. It is well known that S^*_{aver} is a method with least maximal error and that $r_{max}(N)$ is as given. Now we have to show only the given formula for $\Delta_{aver}(\phi^*, N) = r_{aver}(N)$.
First we have $\alpha(f + \text{ const}) = \alpha(f)$ for each $f \in F$ and hence

$$\Delta_{aver}(\phi^*, N) = \frac{1}{\lambda^{n-1}(P)} \int_P \alpha(x_1, \ldots, x_n)\, dx_2 \ldots dx_n$$

with $P = \{(x_1, \ldots, x_n) \mid x_1 = 0 \text{ and } |x_{i+1} - x_i| \le a_{i+1} - a_i \text{ for } i = 1, \ldots, n-1\}$. Now one shows that

$$\alpha(x) = 1/2 \cdot a_1^2 + 1/2 \cdot (1-a_n)^2 + 1/4 \cdot \sum_{i=1}^{n-1} \left[(a_{i+1} - a_i)^2 - (x_{i+1} - x_i)^2\right]$$

and integration over P leads to the given value of $r_{aver}(N)$.

Remark
For fixed N the optimal S^*_{aver} is unique (up to a change of ϕ^* on some null set of course) and at the same time a method with least maximal error (S^*_{max} is not unique, however). Now we ask for the optimal knots, i.e. we want to find the optimal information N^*_{max} and N^*_{aver} in the class I_n. It is interesting to observe that

$$N^*_{max} \neq N^*_{aver}$$

is valid for all $n > 1$.

Proposition 2
In the class I_n the optimal N^*_{max} and N^*_{aver} are unique and given by

$$N^*_{max}(f) = \left(f\left(\frac{1}{2n}\right), f\left(\frac{3}{2n}\right), \ldots, f\left(\frac{2n-1}{2n}\right)\right)$$

and

$$N^*_{aver}(f) = \left(f\left(\frac{1}{3n-1}\right), f\left(\frac{4}{3n-1}\right), \ldots, f\left(\frac{3n-2}{3n-1}\right)\right).$$

Proof:
Proposition 2 follows from Proposition 1 by means of easy calculations. The following values are also easy to compute:

$$e_n(F_1, Int) = r_{max}(N^*_{max}) = \frac{1}{4n}, \qquad r_{aver}(N^*_{max}) = \frac{1}{6n} + \frac{1}{12n^2}$$

and

$$r_{max}(N^*_{aver}) = \frac{9n-5}{4(3n-1)^2}, \qquad a_n(F_1, Int) = r_{aver}(N^*_{aver}) = \frac{1}{6n-2}.$$

The optimal information N^*_{max} for the worst case is of course well known.

3.2.7 The definition of the average a posteriori error of adaptive methods

Up to now, the average a posteriori error $\Delta_{aver}(\phi, N)$ is defined only for nonadaptive methods $\widetilde{S} = \phi \circ N \in A_n$. The following example shows that the same definition is inadequate in the case of adaptive methods.

Example
Let $F = \{f : [0,1] \to \mathbf{R} \mid f(0) = 0 \text{ and } |f(x) - f(y)| \leq |x - y|\}$ and $S(f) = \int_0^1 f(x)\,dx$. Consider the adaptive information $N(f) = (f(1), f(a))$ with

$$a = \begin{cases} 1/2 & \text{if } |f(1)| \geq 1 - \delta \\ \varepsilon & \text{if } |f(1)| < 1 - \delta, \end{cases}$$

where $\delta, \varepsilon < 1/2$. The local error of the optimal method $S^* = \phi \circ N$ is given by

$$\alpha((x_1, x_2)) = 1/4 \cdot (a^2 - x_2^2) + 1/4 \cdot ((1-a)^2 - (x_1 - x_2)^2),$$

where $a = 1/2$ if $|x_1| \geq 1 - \delta$ and $a = \varepsilon$ if $|x_1| < 1 - \delta$. For suitably chosen small $\delta, \varepsilon > 0$ we can obtain methods $\widetilde{S} \in A_2^{ad}$ such that the error $\Delta_{aver}(\phi, N)$ according to the definition in 3.2.2 is arbitrary small.

We conclude that the averaging method of 3.2.2 is not appropriate in this case. According to our idea of a fair averaging process it seems to be more sensible to assume that the distribution of $f(a_1) = f(1) \in [-1, 1]$ is the normed Lebesgue measure $1/2 \cdot \lambda_{[-1,1]}$. Similar, for fixed $f(a_1)$ we assume that the distribution of $f(a_2)$ is the normed Lebesgue measure in the respective interval

$$V = \{\tilde{f}(a_2) \mid \tilde{f} \in F, \tilde{f}(a_1) = f(a_1)\}.$$

Proceeding in this way, we obtain the following definition.

Definition
Let $\widetilde{S} = \phi \circ N \in A_n^{ad}$ and let $\alpha : N(F) \to \mathbf{R}_0^+$ be the respective local error.
i) For $x = (x_1, \ldots, x_n) \in N(F)$ we write

$$\alpha(x_1, \ldots, x_n) := \alpha(x).$$

ii) For $1 \leq m \leq n$ we define $\alpha(x_1, \ldots, x_{m-1}) = \alpha(x_1, \ldots, x_m)$ if $\alpha(x_1, \ldots, x_{m-1}, .)$ is constant and

$$\alpha(x_1, \ldots, x_{m-1}) = \frac{1}{\lambda(Z)} \int_Z \alpha(x_1, \ldots, x_m)\, dx_m$$

if the domain Z of the function $\alpha(x_1, \ldots, x_{m-1}, .)$ has a finite positive measure.
In this way we define $\alpha(x_1, \ldots, x_m)$ for all $(x_1, \ldots, x_m)$ such that $(x_1, \ldots, x_n) \in N(F)$ for certain $x_{m+1}, \ldots, x_n$. In the last step we define

$$\alpha = \Delta_{aver}^{ad}(\phi, N)$$

which we call the average a posteriori error of the (adaptive) method $\widetilde{S} = \phi \circ N$. Again we define

$$a_n^{ad}(F, S) = \inf_{\phi \circ N \in A_n^{ad}} \Delta_{aver}^{ad}(\phi, N),$$

where the infimum is taken over all $\widetilde{S} = \phi \circ N \in A_n^{ad}$ for which $\Delta_{aver}^{ad}(\phi, N)$ exists.

3.2.8 Examples
There are linear problems where the a posteriori error bound

$$\Delta(\widetilde{S}(f), f) \leq \alpha(N(f))$$

is not better than the worst case error bound $e_n(F, S)$ for any $f \in F$, even if $\widetilde{S} \in A_n^{ad}$ is allowed to be adaptive. As a consequence, we have

$$a_n^{ad}(F, S) = a_n(F, S) = e_n(F, S)$$

for such a problem. We give an example.

Example 1
We continue with Example 1 of 3.2.3. Let $N(f) = (f(i_1), \ldots, f(i_n))$, where we can assume that the i_j are different. It is easy to see that

$$\alpha(N(f)) \geq \sum_{i=1}^{\infty} a_i - \sum_{j=1}^{n} a_{i_j} \geq \sum_{i=n+1}^{\infty} a_i$$

holds for both S_1 and S_2 and arbitrary $\widetilde{S}(f)$. Therefore the a posteriori error bound $\Delta(\widetilde{S}(f), f)$ can not be better than $e_n(F, S_i)$ for any $f \in F$, even if $\widetilde{S} \in A_n^{ad}$ is adaptive.

Example 2
To see the difference of the definition in 3.2.7 to the definition of $\Delta_{aver}(\widetilde{S})$ in the nonadaptive case, we consider again the example in 3.2.7. Let $N(f) = (f(a_1), f(a_2))$ with $a_2 = a_2(f(a_1))$. The local error of the optimal method $S^* = \phi^* \circ N$ is given by

$$\alpha((x_1,x_2)) = \frac{1}{4} \cdot (a_1^2 - x_1^2 + (a_2 - a_1)^2 - (x_2 - x_1)^2) + \frac{1}{2} \cdot (1 - a_2)^2$$

if $0 < a_1 < a_2 \leq 1$.
First of all, we treat the nonadaptive case and can assume that $0 < a_1 < a_2 \leq 1$. As in the proof of Proposition 1 in 3.2.6 we compute

$$r_{aver}(N) = \Delta_{aver}(\phi^*, N) = \frac{1}{6} \cdot (a_1^2 + (a_2 - a_1)^2) + \frac{1}{2} \cdot (1 - a_2)^2.$$

The optimal nonadaptive information is given by $a_1 = 3/7$ and $a_2 = 6/7$ and we have

$$a_2(F, Int) = \frac{4}{49} = 0.0816\ldots.$$

Now we consider adaptive methods. If $a_2 < a_1$, then the optimal choice of a_2 is $a_2 = 1/2 \cdot a_1$ and

$$\alpha(x_1) = \frac{7}{12}a_1^2 - \frac{1}{12}x_1^2 + \frac{1}{2} - a_1.$$

If $a_2 > a_1$, then the optimal choice of a_2 is given by $4a_2 = 3 + a_1$ and

$$\alpha(x_1) = \frac{3}{8}a_1^2 - \frac{1}{4}x_1^2 + \frac{1}{8} - \frac{1}{4}a_1,$$

where $\alpha(x_1)$ is computed according to the definition in 3.2.7. It follows that the optimal $a_2 = a_2(f(a_1))$ is given by

$$a_2 = \begin{cases} \frac{1}{2}a_1 & \text{if } x_1^2 \leq -5/4 \cdot a_1^2 + 9/2 \cdot a_1 - 9/4 \\ \frac{1}{4}(3 + a_1) & \text{if } x_1^2 > -5/4 \cdot a_1^2 + 9/2 \cdot a_1 - 9/4. \end{cases}$$

Now we can compute the a posteriori error $\Delta_{aver}^{ad}(\phi^*, N)$ of the optimal method. We have

$$\begin{aligned} \Delta_{aver}^{ad}(\phi^*, N) = \quad & \tfrac{1}{a_1} \int_0^h \tfrac{7}{12}a_1^2 - \tfrac{1}{12}x_1^2 + 1/2 - a_1 \, dx_1 + \\ & + \tfrac{1}{a_1} \int_h^{a_1} \tfrac{3}{8}a_1^2 - \tfrac{1}{4}x_1^2 - \tfrac{1}{4}a_1 + 1/8 \, dx_1, \end{aligned}$$

where $h^2 = -5/4 \cdot a_1^2 + 9/2 \cdot a_1 - 9/4$. By means of a tedious calculation we see that the optimal a_1 is given by $a_1 = 0.9200\ldots$ and we get

$$a_2^{ad}(F, Int) = 0.052188\ldots$$

which is considerably smaller than $a_2(F, Int)$.

3.2.9 Remarks and conjectures

Concerning the class F_1, the following facts are easy to prove:

$$a_n^{ad}(F_1,S) = a_n(F_1,S) \quad \text{for } n = 1,2$$

and

$$a_n^{ad}(F_1,S) < a_n(F_1,S) \quad \text{for all } n > 2$$

and each of the problems *App*, *Opt*, *Opt**, and *Int*. Nevertheless, the linear problems *App* and *Int* behave very different from the nonlinear optimization problems. Indeed we conjecture that adaption "does not help much" for *App* and *Int* in the sense that

$$a_n^{ad}(F_1,S) > c_\varepsilon \cdot n^{-1-\varepsilon}$$

for these problems (and every $\varepsilon > 0$), whereas adaption "helps much" for the problems *Opt* and *Opt** in the sense that

$$a_n^{ad}(F_1,S) < c \cdot \alpha^n$$

for some suitable $c > 0$ and $0 < \alpha < 1$.

3.2.10 Notes and references

Some of the results of 3.2.1–3.2.6 are contained in *Novak* (1986b, 1986c). The definition of the average a posteriori error for adaptive methods is new.

Appendix: Existence and uniqueness of optimal algorithms

A.1 Remark

In this appendix we deal with different problems concerning the existence and uniqueness of optimal algorihms. We always assume that $S : F \to M$, where M is a metric space and $\Delta(\widetilde{S}(f), f) = d(S(f), \widetilde{S}(f))$. First of all we could ask whether there is a $S_n^* \in A_n$ with

$$\Delta_{max}(S_n^*) = e_n(F, S).$$

We consider this problem for $S(f) = \int_0^1 f(x)\,dx$ and $F \subset C([0,1])$ in A.2. Even in this relatively simple case, the problem of existence of an optimal S_n^* can only be solved in some cases using special methods (i.e. theory of monosplines). Therefore we treat a more special problem in the following sections: We assume that an information operator $N : F \to \mathbf{R}^n$ is given and ask whether there is an optimal algorithm $\phi^* \circ N$ in the class $A(N) = \{\phi \circ N \mid \phi : \mathbf{R}^n \to M\}$. In A.3 we deal with the worst case and indicate some results using well known facts on Chebyshev centers. Then we introduce the notion of a center of a probability measure in a metric space and prove statements on its existence and uniqueness. Similarly as in the worst case these results can be used to prove the existence and uniqueness of optimal average case algorithms. These results are due to *Novak, Ritter* (1987) and use the definition of the average error of *Wasilkowski* (1983). We also show that nonmeasurable algorithms are not better than measurable ones.

A.2 Existence and uniqueness of optimal quadrature formulas

There are many papers dealing with the existence and uniqueness of optimal quadrature formulas. The following proposition is a summary of the results of different authors, see *Bojanov* (1986) and *Zhensykbaev* (1981).

Proposition

Let F be one of the classes $W_p^k = \{f : [0,1] \to \mathbf{R} \mid \|f^{(k)}\|_p \le 1\}$ or $\widetilde{W}_p^k = \{f \in W_p^k \mid f^{(r)}(0) = f^{(r)}(1) \text{ for } r = 0, \dots, k-1\}$, where $1 \le p \le \infty$ and $k \in \mathbf{N}$.

a) There is a $S_n^* \in A_n$ with $\Delta_{max}(S_n^*) = e_n(F, \lambda)$.

b) S_n^* can be chosen linear, i.e. $S_n^*(f) = \sum_{i=1}^n c_i f(a_i)$.

c) In the case $F = W_p^k$ and $n \ge k$ the optimal linear method is unique.

d) In the case $F = \widetilde{W}_p^k$ the only optimal linear method (up to a rigid shift) is the rectangle formula.

We would like to stress that the proof of these results is difficult and relies heavily on the theory of monosplines. The following example shows that the existence of an optimal quadrature formula can not be shown by means of a simple compactness argument.

Example
Let F be the closure of the set $\{f : [0,1] \to \mathbf{R} \mid f(0) = 0,\ \|f'\|_\infty \le 1,\ \|f''\|_\infty \le 1,\ f'(0) = \int_0^1 f(x)\,dx\}$. Then F is a compact convex and symmetric set with $e_1(F,\lambda) = 0$ but there is no $\widetilde{S} \in A_1$ with $\Delta_{max}(\widetilde{S}) = 0$.

Proof:
Let $f \in C^2([0,1])$ and $f \in F$. We also assume that $0 < x \le 1$. We have $f'(x) = f'(0) + \int_0^x f''(t)\,dt$, hence $|f'(x) - \int_0^1 f(t)\,dt| \le x$ and also

$$|f'(\xi) - \int_0^1 f(t)\,dt| \le x \qquad \text{for } 0 \le \xi \le x.$$

Because of $f(x) - f(0) = x \cdot f'(\xi)$ with some $\xi \in [0,x]$ we get

$$|\tfrac{1}{x} f(x) - \int_0^1 f(t)\,dt| \le x \qquad (0 < x \le 1).$$

This estimate is valid for all $f \in F$ and we have $\Delta_{max}(\widetilde{S}_x) \le x$ for $0 < x \le 1$ and $\widetilde{S}_x(f) = \frac{1}{x} f(x)$. Hence $e_1(F,\lambda) = 0$. It is easy to see that there is no $\widetilde{S} \in A_1$ with $\Delta_{max}(\widetilde{S}) = 0$. Actually it could be proved that $\Delta_{max}(\widetilde{S}) > 0$ for all $\widetilde{S} \in A_n$ and $n \in \mathbf{N}$.

Remark
It is widely believed that the order of exactness is a good measure of the quality of an quadrature formula, at least if F is a class of analytic functions. There are situations, however, where optimal methods are much better than Gauss quadrature. An example is the class

$$F = \{f : [-1, 1] \to \mathbf{R} \mid \text{ there is an analytic extension } \tilde{f} : \{z \in \mathbf{C} \mid |z| < 1\} \to \mathbf{C} \text{ with } \|\tilde{f}\|_\infty \le 1\}.$$

For the problem $S(f) = \int_{-1}^1 f(x)\,dx$ it is known (see *Kowalski, Werschulz, Woźniakowski* (1985)) that

$$e_n(F,\lambda) \le \exp\left(-\pi\sqrt{n/2}\right)$$

while the error of the Gauss method $\widetilde{S}_n$ satisfies

$$\Delta_{max}(\widetilde{S}_n) \ge c \cdot n^{-2}.$$

A.3 Chebyshev centers and optimal worst case algorithms

Let (M,d) be a Banach space and let $D \subset M$ be a bounded set. We define

$$rad(D,x) = \sup_{y \in D} d(x,y)$$

and

$$rad(D) = rad_M(D) = \inf_{x \in M} rad(D,x).$$

The number $rad(D)$ is called the Chebyshev radius of D and a point $x \in M$ is a Chebyshev center of D if

$$rad(D) = rad(D,x).$$

The notion of a Chebyshev center is well established, see *Garkavi* (1964) and *Amir* (1984).

We assume that $S : F \to M$ and $N : F \to \mathbf{R}^n$ are given. An algorithm $S^* \in A(N)$ is called optimal (in the worst case) if

$$\Delta_{max}(S^*) = \inf_{\widetilde{S} \in A(N)} \Delta_{max}(\widetilde{S}).$$

A method $\phi \circ N \in A(N)$ is called central if $\phi(x)$ is a Chebyshev center of $S(N^{-1}(x))$ for each x. It is easy to see that every central algorithm is optimal and

$$r_{max}(N) = \sup_{x \in \mathbf{R}^n} rad\, S(N^{-1}(x)).$$

If M is a subspace of a Banach space $\widetilde{M}$ (for example $M \subset M^{**} = \widetilde{M}$ by the natural imbedding) then $rad(D,x)$ with $x \in \widetilde{M}$ and $rad_{\widetilde{M}}(D) = \inf_{x \in \widetilde{M}} rad(D,x)$ are defined as well. The following proposition is a summary of well known facts. We shall see in A.4 that some of our results for the average case are similar, but there are interesting differences as well.

Proposition

1) In the case $M = C(K)$ (where K is compact) a Chebyshev center exists for every $D \subset M$.

2) In the case $M = B(T)$ (where T is arbitrary) a Chebyshev center exists for every $D \subset M$.

3) Let $M = C(K)$ and $\widetilde{M} = B(K)$. Then $rad_M(D) \leq rad_{\widetilde{M}}(D)$ for each $D \subset M$ and this inequality is strict for some D.

4) Let M be the dual space of a normed linear space. Then a Chebyshev center exists for every $D \subset M$.

5) Let $P : \widetilde{M} \to M$ be a projection with $\|P\| = 1$. Then we have

$$rad_M(D) = rad_{\widetilde{M}}(D)$$

for every $D \subset M$ and the projection of a Chebyshev center in $\widetilde{M}$ is a Chebyshev center in M.

6) The Chebyshev center of any bounded nonempty set $D \subset M$ is unique if and only if M is uniformly convex in every direction.

Remarks

The first statement is due to *Kadets, Zamyatin* (1968), the second one is trivial. For the third statement consider $D = \{f \mid \|f\| \leq 1,\ x \cdot f(x) \geq 0\} \subset C([-1,1])$. Then $rad_{C([-1,1])}(D) = 1$ but $rad_{B([-1,1])}(D) = \frac{1}{2}$. The statements 4)–6) are due to *Garkavi* (1964). We mention that every uniformly convex space is uniformly convex in every direction and therefore every bounded nonempty set in a uniformly convex space has a unique Chebyshev center by 4), 6), and a well known theorem of Milman.

A.4 Centers of probability measures

We are mainly interested in optimal average case algorithms. To establish results on the existence and uniqueness of optimal algorithms in this setting we introduce the notion of a center of a probability measure in a metric space.
Let m be a (finite Borel) measure on the metric space M and let $1 \leq p < \infty$. We define

$$rad^p(m,x) = \left(\int_M d(x,y)^p \, dm(y) \right)^{1/p}$$

and assume that $rad^p(m,.)$ is a finite function, i.e. $rad^p(m,x) < \infty$ for some $x \in M$ (and hence for any $x \in M$). The number

$$rad^p(m) = rad^p_M(m) = \inf_{x \in M} rad^p(m,x)$$

is called the p-radius of m and a point $x \in M$ is called a p-center of m if

$$rad^p(m) = rad^p(m,x).$$

We use the following notations. A measure on a metric space is understood to be a finite Borel measure. Measurability of a mapping means measurability with respect to the Borel sets. We assume that M is a Banach space and that m is a measure on M with $rad^p(m,.) < \infty$ for some given p, $1 \leq p < \infty$. If M is a closed subspace of the Banach space $\widetilde{M}$ then $rad^p(m,x)$ with $x \in \widetilde{M}$ and $rad^p_{\widetilde{M}}(m) = \inf_{x \in \widetilde{M}} rad^p(m,x)$ are defined as well. Now we state our results concerning the existence of p-centers.

Proposition

1) In the case $M = C([0,1])$ a p-center does not always exist.

2) Let $M = B(T)$ be the space of bounded functions on a set T, and let m be a tight measure on M. Then a p-center always exists.

3) Let $M = C(K)$ and $\widetilde{M} = B(K)$, where K is a compact metric space. Then

$$rad_M^p(m) = rad_{\widetilde{M}}^p(m).$$

4) Let M be the dual space of a normed linear space and m be a tight measure on M. Then a p-center always exists.

5) Let $P : \widetilde{M} \to M$ be a projection with $\|P\| = 1$. Then we have

$$rad_M^p(m) = rad_{\widetilde{M}}^p(m)$$

and the projection of a p-center in $\widetilde{M}$ is a p-center in M.

6) Let $1 < p < \infty$. Then the p-center of every measure with $rad_M^p(m, \ ) < \infty$ is unique if and only if M is strictly convex.

Remarks and Proofs:
A measure m on a metric space M is said to be tight if for each $\varepsilon > 0$ there exists a compact set $D_\varepsilon \subset M$ such that $m(M \setminus D_\varepsilon) < \varepsilon$. If M is a complete separable metric space, then every measure on M is tight, see *Parthasarathy* (1967, p. 29).
The statements 2), 4), and 5), are analogous to the respective results on Chebyshev centers. Also the proofs are similar for 4) (it can be shown that the function $x \mapsto rad^p(m,x)$ is weak* lower semicontinuous and hence takes its infimum on M) and 5). For the proof of 2) we use the fact that $x \mapsto rad^p(m,x)$ is lower semicontinuous with respect to the product topology on $[-c,c]^T$. If M is strictly convex then $rad^p(m, .)$ is a strictly convex mapping for $1 < p < \infty$ and 6) follows. (The case $p = 1$ is slightly more complicated, see *Novak, Ritter* (1987)).
We should remark that 4) (for the case where the support of m is compact and X is the dual of a separable space) is due to *Beauzamy, Maurey* (1977). The expectation in metric spaces (i.e. the case $p = 2$) is also studied in *Pick* (1987).
Now we prove the statements 1) and 3) which are different as in the worst case.

Proof of 1). We construct a probability measure m on $M = C([0,1])$ with the following properties: The support of m is bounded and hence $rad^p(m,x) < \infty$ for each $x \in M$ and $1 \leq p < \infty$. There is no $x \in M$ with $rad^p(m,x) = rad^p(m)$. First we define the functions $x_n, y_n \in M$ for $n \in \mathbf{N}$. Let

$$x_n(t) = \begin{cases} 0, & \text{if } t \leq \frac{1}{2} \\ 1, & \text{if } t \geq \frac{1}{2} + \frac{1}{2n} \\ 2n(t - \frac{1}{2}), & \text{if } \frac{1}{2} < t < \frac{1}{2} + \frac{1}{2n} \end{cases}$$

and

$$y_n(t) = \begin{cases} -1, & \text{if } t \leq \frac{1}{2} - \frac{1}{2n} \\ 0, & \text{if } t \geq \frac{1}{2} \\ 2n(t - \frac{1}{2}), & \text{if } \frac{1}{2} - \frac{1}{2n} < t < \frac{1}{2}. \end{cases}$$

Now we consider the probability measure

$$m = \frac{1}{2} \cdot \sum_{n=1}^{\infty} \frac{1}{2^n} \cdot (\varepsilon_{x_n} + \varepsilon_{y_n}),$$

where ε_x is the respective Dirac measure. Let $z \in M = C([0,1])$. Then

$$\begin{aligned} rad^p(m,z) \quad &= \left(\sum_{n=1}^{\infty} \tfrac{1}{2^n} \cdot \left(\tfrac{1}{2}\|z - x_n\|^p + \tfrac{1}{2}\|z - y_n\|^p\right)\right)^{1/p} \\ &\geq \left(\sum_{n=1}^{\infty} \tfrac{1}{2^n} \cdot \left(\tfrac{1}{2} \cdot (\|z - x_n\| + \|z - y_n\|)\right)^p\right)^{1/p} \\ &\geq \left(\sum_{n=1}^{\infty} \tfrac{1}{2^n} \cdot \left(\tfrac{1}{2}\,(|z(0)| + |1 + z(0)|)\right)^p\right)^{1/p} \geq \tfrac{1}{2}. \end{aligned}$$

Considering the function $z_n = \frac{1}{2}(x_n + y_n)$ (for $n \to \infty$), we see that $rad^p(m) = \frac{1}{2}$. Now we assume that $r^p(m,z^*) = \frac{1}{2}$ for some $z^* \in M$. The above estimate shows that

$$\|z^* - x_n\| + \|z^* - y_n\| = 1$$

for all n. If $z^*(\frac{1}{2}) < 0$ then $\|z^* - x_n\| > 1$ for large n. If $z^*(\frac{1}{2}) > 0$ then $\|z^* - y_n\| > 1$ for large n. If $z^*(\frac{1}{2}) = 0$ then $\|z^* - x_n\| \geq \frac{2}{3}$ and $\|z^* - y_n\| \geq \frac{2}{3}$ for large n. Hence we have proved that such a z^* does not exist.

Proof of 3). Let $z^* \in \widetilde{M} = B(K)$ be a p-center of m in $\widetilde{M}$, i.e.

$$rad^p(m,z^*) = rad^p_{\widetilde{M}}(m),$$

and let $\varepsilon > 0$. There is a compact set $D \subset C(K)$ with

$$\left(\int_D \|z - x\|^p \, dm(x)\right)^{1/p} \geq rad^p(m,z) - \varepsilon$$

for all $z \in B(K)$, $\|z\| \leq \|z^*\|$. We may assume that D is of the form

$$D = \{x \in C(K) \mid \|x\| \leq k,\ \omega(x) \leq \omega\}.$$

The function $\omega(x)$ is the modulus of continuity of x, defined by

$$\omega(x)(h) = \sup\{|x(t_1) - x(t_2)| \mid d(t_1,t_2) \le h\},$$

and $\omega : \mathbf{R}^+ \to \mathbf{R}^+$ is a continuous function with the following properties: (a) $\omega(h) \to 0$ for $h \to 0$; (b) ω is increasing; (c) $\omega(h_1 + h_2) \le \omega(h_1) + \omega(h_2)$. It is enough to show that there is a $\tilde{z} \in C(K)$ with $\|\tilde{z}\| \le \|z^*\|$ such that

$$\left(\int_D \|\tilde{z} - x\|^p \, dm(x)\right)^{1/p} \le \left(\int_D \|z^* - x\|^p \, dm(x)\right)^{1/p}.$$

Now we construct such a function $\tilde{z}$. Let (t_n) be a dense sequence in K and let $z_0 = z^*$. Let the function z_n $(n \in \mathbf{N})$ be given by

$$z_n(t) = \begin{cases} z_{n-1}(t_n), & \text{if } t = t_n \\ \min(z_{n-1}(t), z_{n-1}(t_n) + \omega(d(t,t_n))), & \text{if } z_{n-1}(t) \ge z_{n-1}(t_n) \\ \max(z_{n-1}(t), z_{n-1}(t_n) - \omega(d(t,t_n))), & \text{if } z_{n-1}(t) < z_{n-1}(t_n). \end{cases}$$

Then the following assertions hold:

i) $\|z_n - x\| \le \|z_{n-1} - x\|$ for each $n \in \mathbf{N}$ and $x \in D$;

ii) $|z_n(t_i) - z_n(t)| \le \omega(d(t_i,t))$ for $i = 1, \dots, n$;

iii) $z_n(t_i) = z_i(t_i)$ if $n \ge i$.

We prove these statements:
i) Let $\delta > 0$, $x \in D$, and $|z_n(t) - x(t)| \ge \|z_n - x\| - \delta$ for some $t \in K$. We assume that $|z_n(t) - x(t)| > |z_{n-1}(t) - x(t)|$ and discuss the case $z_{n-1}(t_n) \le z_{n-1}(t)$. Then $z_{n-1}(t) - z_{n-1}(t_n) > \omega(d(t,t_n))$, $z_n(t) = z_{n-1}(t_n) + \omega(d(t,t_n))$, and $x(t) > z_n(t)$. Because of $x(t_n) \ge x(t) - \omega(d(t,t_n))$ we get $x(t_n) - z_{n-1}(t_n) \ge x(t) - z_n(t) \ge \|x - z_n\| - \delta$.
ii) Let $s,t \in K$ and $m \in \mathbf{N}$ such that

$$|z_m(s) - z_m(t)| \le \omega(d(s,t)).$$

We want to prove that $|z_{m+1}(s) - z_{m+1}(t)| \le \omega(d(s,t))$. Because this is trivial if $z_{m+1}(s) = z_m(s)$ and $z_{m+1}(t) = z_m(t)$, we assume that $z_m(s) \ne z_{m+1}(s)$ and discuss the case $z_m(s) > z_m(s_{m+1})$. We have

$$z_{m+1}(s) = z_m(t_{m+1}) + \omega(d(s,t_{m+1})).$$

If $z_{m+1}(s) \le z_{m+1}(t)$ then we get

$$\begin{aligned} |z_{m+1}(s) - z_{m+1}(t)| &\le z_m(t_{m+1}) + \omega(d(t,t_{m+1})) - z_m(t_{m+1}) - \omega(d(s,t_{m+1})) \\ &\le \omega(d(s,t)). \end{aligned}$$

If $z_{m+1}(s) > z_{m+1}(t)$ and $z_m(t) \le z_{m+1}(t)$ then we get

$$|z_{m+1}(s) - z_{m+1}(t)| < z_m(s) - z_m(t) \le \omega(d(s,t)).$$

If $z_{m+1}(s) > z_{m+1}(t)$ and $z_m(t) > z_{m+1}(t)$ then we get

$$\begin{aligned} |z_{m+1}(s) - z_{m+1}(t)| &= z_m(t_{m+1}) + \omega(d(s,t_{m+1})) - z_m(t_{m+1}) - \omega(d(t,t_{m+1})) \\ &\le \omega(d(s,t)). \end{aligned}$$

Now the statement follows from $|z_i(t_i) - z_i(t)| \le \omega(d(t_i,t))$ by induction.
iii) For arbitrary $n > i$ we have $|z_{n-1}(t_i) - z_{n-1}(t_n)| \le \omega(d(t_i,t_n))$ by ii), hence $z_n(t_i) = z_{n-1}(t_i)$, and the statement follows by induction.
Because the sequence (t_n) is dense in K, we conclude that the z_n converge uniformly to a function $\tilde{z} \in C(K)$ with $\|\tilde{z}\| \le \|z^*\|$. Because of $\|\tilde{z} - x\| \le \|z^* - x\|$ for all $x \in D$ we have

$$\left(\int_D \|\tilde{z} - x\|^p \, dm(x)\right)^{1/p} \le \left(\int_D \|z^* - x\|^p \, dm(x)\right)^{1/p}.$$

Remark
In the last proof we showed that for

$$D = \{x \in C(K) \mid \|x\| \le k,\ \omega(x) \le \omega\} \subset C(K) \subset B(K)$$

and arbitrary $z^* \in B(K)$ there is a $\tilde{z} \in D$ with $\|\tilde{z} - x\| \le \|z^* - x\|$ for all $x \in D$. Thus we have solved the problem of best coapproximation for the set $D \subset B(K)$, see *Papini, Singer* (1979) and *Beauzamy, Maurey* (1977).

A.5 Optimal average case algorithms

Henceforth we assume that F and M are complete separable metric spaces. Suppose we want to approximate $S(f)$, where $f \in F$ and $S : F \to M$ is a measurable operator. We assume that $N(f)$ is known, where $N : F \to \mathbf{R}^n$ is a measurable information operator and $N(F) = Y \subset \mathbf{R}^n$ is closed. Then we approximate S by a $\widetilde{S} = \phi \circ N$. Let μ be a probability measure on F and let $1 \le p < \infty$. If $f \mapsto d(S(f), \widetilde{S}(f))$ is measurable then we define the p-average error $\Delta_\mu^p(\widetilde{S})$ by

$$\Delta_\mu^p(\widetilde{S}) = \left(\int_F d(S(f), \widetilde{S}(f))^p \, d\mu(f)\right)^{1/p}.$$

(In 3.1 we consider $\Delta_\mu(\widetilde{S}) = \Delta_\mu^1(\widetilde{S})$ and $\tilde{\Delta}_\mu(\widetilde{S}) = \Delta_\mu^2(\widetilde{S})$, see 3.1.2). Because of technical reasons (we want to apply our results on p-centers), we extend this definition to arbitrary algorithms in a slightly different way as in 3.1.2. This method is due to *Wasilkowski* (1983). Let $\nu = N\mu$ be the probability measure on Y induced by N, i.e.

$\nu(B) = \mu(N^{-1}(B))$ for any Borel set $B \subset Y$, and let $\mu(\,.\,\|y)$, $y \in Y$, be a family of probability measures on F such that
1) $\mu(N^{-1}\{y\}\|y) = 1$ for ν-almost all $y \in Y$;
2) $\mu(B|\,.\,)$ is measurable for any Borel set $B \subset F$;
3) $\mu(B) = \int_Y \mu(B\|y)\,d\nu(y)$ for any Borel set $B \subset F$.

The family $\mu(\,.\,\|y)$ is called a regular conditional probability distribution. The existence and uniqueness ν-a.e. of $\mu(\,.\,\|y)$ follows from *Parthasarathy* (1967, p. 147). It is well known that $y \mapsto \int_F G(f)\,d\mu(f\,\|y)$ is measurable and

$$\int_F G(f)\,d\mu(f) = \int_Y \int_F G(f)\,d\mu(f\|y)\,d\nu(y)$$

holds for each measurable $G : F \to \mathbf{R}_0^+$.

Definition
1) If $\phi : Y \to M$ is an arbitrary algorithm and $\int^*$ denotes the upper integral then

$$\begin{aligned}\Delta_\mu^p(\widetilde{S}) &= \left(\textstyle\int_Y^* \int_F^* d(S(f),\widetilde{S}(f))^p\,d\mu(f\|y)\,d\nu(y)\right)^{1/p}\\ &= \left(\textstyle\int_Y^* \int_F d(S(f),\phi(y))^p\,d\mu(f\|y)\,d\nu(y)\right)^{1/p}\end{aligned}$$

is called p-average error of $\widetilde{S} = \phi \circ N$.
2) The number

$$r_\mu^p(N) = \inf_\phi \Delta_\mu^p(\phi \circ N)$$

is called p-average radius of N and the p-radius

$$rad^p(S\mu(\,.\,\|y)) = \inf_{x \in M}\left(\int_F d(S(f),x)^p\,d\mu(f\|y)\right)^{1/p}$$

of $S\mu(\,.\,\|y)$ is called local p-average radius of N in $y \in Y$.
3) An algorithm is called p-optimal if $\Delta_\mu^p(\phi \circ N) = r_\mu^p(N)$.

Next we cite a theorem of *Wasilkowski* (1983), which is stated here in terms of p-centers.

Proposition
The local p-average radius $y \mapsto rad^p(S\mu(\,.\,\|y))$ is measurable and for the p-average radius of N the following equation holds

$$r_\mu^p(N) = \left(\int_Y rad^p(S\mu(\,.\,\|y))^p\,d\nu(y)\right)^{1/p}.$$

If $r_\mu^p(N)$ is finite, then an algorithm ϕ is p-optimal iff $\phi(y)$ is a p-center of $S\mu(\,.\,\|y)$ for ν-almost all $y \in Y$.

Using the results of A.4 we can prove statements concerning the existence and uniqueness (i.e. uniqueness ν-a.e.) of p-optimal algorithms. We give an example.

Proposition
Assume that $S : F \to M$ and $N : F \to \mathbf{R}^n$ are measurable, where F and M are complete separable metric spaces and $N(F)$ is closed. We further assume that μ is an arbitrary probability measure on F. If M is a strictly convex dual space and $r_\mu^p(N) < \infty$ for some p, $1 < p < \infty$, then a p-optimal algorithm exists and is unique.

Remarks
a) Optimal algorithms can be given explicitly for many linear problems, see *Lee, Wasilkowski* (1986) and *Packel, Woźniakowski* (1987). It turns out that suitable spline algorithms are optimal in the case of an orthogonally invariant measure, see also *Novak, Ritter* (1987).
Gaussian measures μ are orthogonally invariant, and if N is an adaptive linear information operator then the measures $\mu(\,.\,\|y)$ are Gaussian too. Observing that continuous linear images of Gaussian measures are Gaussian measures we conclude that spline algorithms are the unique p-optimal algorithms for linear problems with Gaussian measures.
b) The radius $r(N)$ and $r_\mu^p(N)$ of an information can get smaller if we consider $S : F \to \widetilde{M}$ instead of $S : F \to M$, i.e. if we allow all algorithms of the form $\widetilde{S} = \phi \circ N$ with $\phi : \mathbf{R}^n \to \widetilde{M}$. It is interesting that this can happen for $M = C(K)$ and $\widetilde{M} = B(K)$ in the worst case, but not in the average case. This follows from A.3.3 and A.4.3.

A.6 Measurability of optimal algorithms.
By A.5 an algorithm ϕ is p-optimal iff

$$\Delta_\mu^p(\phi \circ N) = \left(\int_Y \int_F d(S(f), \phi(y))^p \, d\mu(f|y) d\nu(y) \right)^{1/p} = r_\mu^p(N),$$

i.e. the upper integral can be replaced by an ordinary integral in this case. (We cannot conclude, however, that every p-optimal algorithm is measurable. Actually it is easy to give counterexamples.) We prove that nonmeasurable algorithms can not be better than measurable ones. For the proof we need the selection theorem of Yankov and von Neumann, see *Wagner* (1980).

Proposition: Suppose Y and M are Polish, $\Phi \subset Y \times M$ is analytic such that $\{x \in M \mid (y,x) \in \Phi\} \neq \emptyset$ for all $y \in Y$. Then there is a mapping $\phi : Y \to M$ with $(y, \phi(y)) \in \Phi$ for all $y \in Y$ which is μ-measurable for all probability measures μ on Y.

Proposition
Assume that $S : F \to M$ and $N : F \to \mathbf{R}^n$ are measurable, where F and M are complete separable metric spaces and $N(F) = Y \subset \mathbf{R}^n$ is closed. We further assume that μ is an arbitrary probability measure on F. Then

$$r_\mu^p(N) = \inf\{\Delta_\mu^p(\phi \circ N) \mid \phi : Y \to M \text{ measurable}\}.$$

Assume that a p-optimal algorithm exists. Then there exists a p-optimal measurable algorithm.

Proof. We may assume that $r_\mu^p(N) < \infty$. Let $y \in Y$ and $x \in M$. If

$$rad^p(S\mu(\,.\,\|y),x) = \left(\int_F d(S(f),x)^p\, d\mu(f\|y)\right)^{1/p} < \infty$$

for some $x \in M$ then this quantity is finite for all $x \in M$ and $x \mapsto rad^p(S\mu(\,.\,\|y),x)$ is continuous. Let x_i, $i \in \mathbf{N}$, be a dense sequence in M and let $a \in \mathbf{R}$. We have

$$A = \{(y,x) \in Y \times M \mid rad^p(S\mu(\,.\,\|y),x) \le a\}$$

$$= \bigcap_{k\in\mathbf{N}} \bigcup_{i\in\mathbf{N}} \{(y,x) \in Y \times M \mid rad^p(S\mu(\,.\,\|y),x) \le a + \tfrac{1}{k},\ d(x,x_i) \le \tfrac{1}{k}\}$$
$$= \bigcap_{k\in\mathbf{N}} \bigcup_{i\in\mathbf{N}} \{y \in Y \mid rad^p(S\mu(\,.\,\|y),x_i) \le a + \tfrac{1}{k}\} \times \{x \in M \mid d(x,x_i) \le \tfrac{1}{k}\}.$$

The set A is a Borel set in $Y \times M$. This proves the measurability of $R : Y \times X \to \mathbf{R}_0^+ \cup \{\infty\}$, defined by $R(y,x) = rad^p(S\mu(\,.\,\|y),x)$. Let $B = \{y \in Y \mid R(y,\,.\,) < \infty\}$. We have $\nu(B) = 1$. Define

$$h(y,x) = \begin{cases} R(y,x) - rad^p(S\mu(\cdot\|y)), & \text{if } y \in B \\ 0, & \text{if } y \in Y \setminus B \end{cases}$$

and

$$H_k = \{(y,x) \in Y \times X \mid h(y,x) \le \tfrac{1}{k}\}$$

for $k \in \mathbf{N}$. Then $H_k \subset Y \times M$ is a Borel set. Since Y and M are complete separable metric spaces it follows that H_k is analytic, see *Parthasarathy* (1967, p. 16). For each $y \in Y$ there is a point $x \in M$ such that $(y,x) \in H_k$. Now the selection theorem shows the existence of a measurable algorithm ϕ with $(y,\phi(y)) \in H_k$ for ν-almost all y. We obtain

$$\begin{aligned} \Delta_\mu^p(\phi \circ N) &= \left(\textstyle\int_F d(S(f),\phi \circ N(f))^p\, d\mu(f)\right)^{1/p} \\ &\le \left(\textstyle\int_B \left(rad^p(S\mu(\,.\,\|y)) + \tfrac{1}{k}\right)^p d\nu(y)\right)^{1/p} \\ &\le r_\mu^p(N) + \tfrac{1}{k}, \end{aligned}$$

which shows $r^p_\mu(N) = \inf\{\Delta^p_\mu(\phi \circ N) \mid \phi : Y \to M \text{ measurable}\}$.
Now let ϕ be a p-optimal algorithm. Then there is a Borel set $C \subset Y$ with $\nu(C) = 0$ and $R(y, \phi(y)) = rad^p(S\mu(\,.\,\|y)) < \infty$ for all $y \in Y \setminus C$. Define

$$H = \bigcap_{k \in \mathbf{N}} H_k \cup (C \times M).$$

Then H is analytic and $(y, \phi(y)) \in H$ for all $y \in Y$. Using the same argument as above the existence of a measurable p-optimal algorithm follows.

Remark
A different definition of the average error is

$${}^*\Delta^p_\mu(\widetilde{S}) = \left(\int_F^* d(S(f), \widetilde{S}(f))^p \, d\mu(f) \right)^{1/p},$$

see 3.1.2. We get ${}^*\Delta^p_\mu(\widetilde{S}) \geq \Delta^p_\mu(\widetilde{S})$ but under the conditions of the last proposition we have $\inf_\phi {}^*\Delta^p_\mu(\phi \circ N) = r^p_\mu(N)$ and the notion of the p-average radius of N is independent of this difference.

Bibliography

L. Afflerbach, J. Lehn, eds. (1986): Kolloquium über Zufallszahlen und Simulationen (1986, Darmstadt). Teubner, Stuttgart.

D. Amir (1984): Best simultaneous approximation (Chebyshev centers). In: Parametric optimization and approximation, Conf. Oberwolfach 1983, ISNM **72**, 19–35.

D. Amir (1985): On Jung's constant and related constants in normed linear spaces. Pacific J. Math. **118**, 1–15.

F. Archetti, G. P. Szegö (1980): Global optimization algorithms. In: Nonlinear optimization (L. C. W. Dixon, E. Spedicato, G. P. Szegö, eds.), Birkhäuser Verlag, Boston, pp. 429–469.

K. I. Babenko (1985): Some problems in approximation theory and numerical analysis. Russian Math. Surv. **40**(1), 1–30.

N. S. Bakhvalov (1959): On approximate computation of integrals. *(Russian)*. Vestnik MGV, Ser. Math. Mech. Astron. Phys. Chem. **4**, 3–18.

N. S. Bakhvalov (1962): On a rate of convergence of indeterministic integration processes within the functional classes $W_p^{(l)}$. Theor. Probability Appl. **7**, 227.

N. S. Bakhvalov (1971): On the optimality of linear methods for operator approximation in convex classes of functions. U.S.S.R. Comp. Math. Math. Phys. **11**, 244–249.

P. Basso (1985): Optimal search for the global maximum of functions with bounded seminorm. SIAM J. Numer. Anal. **22**, 888–903.

J. Baumeister (1987): Stable solution of inverse problems. Vieweg Advanced Lectures, Braunschweig.

B. Beauzamy, B. Maurey (1977): Points minimaux et ensembles optimaux dans les espaces de Banach. J. Functional Anal. **24**, 107–139.

O. V. Besov (1980): Intercellular averages and an error estimate for cubature formulas in Sobolev spaces and their generalizations. Proc. of the Steklov Inst. of Math. 1980, 1, 45–60.

O. V. Besov (1981): Error estimates for cubature formulas in terms of smoothness of functions. Proc. of the Steklov Inst. of Math. 1981/4, 11–23.

M. Š. Birman, M. Z. Solomjak (1967): Piecewise-polynomial approximations of functions of the class W_p^α. Math. USSR – Sbornik **2**, 295–317.

M. Š. Birman, M. Z. Solomjak (1980): Quantitative analysis in Sobolev imbedding theorems and applications to spectral theory. AMS Translations **2, 114**.

K. Bogues, C. R. Morrow, T. N. L. Patterson (1981): An implementation of the method of Ermakov and Zolotukhin for multidimensional integration and interpolation. Numer. Math. **37**, 49–60.

B. Bojanov (1986): Comparison theorems in optimal recovery. In: Optimal Algorithms (Bl. Sendov, Ed.), Sofia, 1986.

M. Borovcnik (1984): Was bedeuten statistische Aussagen? Teubner Verlag, Wien und Stuttgart.

H. Braß (1977): Quadraturverfahren. Vandenhoeck und Ruprecht, Göttingen.

H. Braß (1979): Der Wertebereich des Trapezverfahrens. In: Numerische Integration (G. Hämmerlin, ed.), Birkhäuser, 98–108.

S. H. Brooks (1958): A discussion of random methods for seeking maxima. Operations Research **6**, 244–251.

S. H. Brooks (1959): A comparison of maximum-seeking methods. Operations Research **7**, 430–457.

H. Bückner (1950a,b): Bemerkungen zur numerischen Quadratur I,II. Math. Nachr. **3**, 142–145, 146–151.

R. R. Coveyou (1969): Random number generation is too important to be left to chance. Studies in Applied Math. **3**, 70–111.

W. Dahmen, R. DeVore, K. Scherer (1980): Multi-dimensional spline approximation. SIAM J. Numer. Anal. **17**, 380–402.

P. J. Davis, P. Rabinowitz (1984): Methods of numerical integration. (Second edition). Academic Press, London.

L. Devroye (1986): Non-uniform random variate generation. Springer, New York.

H. Dinges, H. Rost (1982): Prinzipien der Stochastik. Teubner Verlag, Stuttgart.

H. Engels (1980): Numerical quadrature and cubature. Academic Press, London.

S. M. Ermakov, V. G. Zolotukhin (1960): Polynomial approximations and the Monte Carlo method. Theor. Prob. Appl. **5**, 428–431.

S. M. Ermakov (1975): Die Monte-Carlo-Methode und verwandte Fragen. Oldenbourg Verlag, München.

M. Sh. Farber (1984): Optimal quadrature formulas and the Wiener measure. *(Russian).* Izv. Akad. Nauk Azerbaidzhan. SSR 1984, No. 6, 123–128.

G. S. Fishman, L. R. Moore (1986): An exhaustive analysis of multiplicative congruential random number generators with modulus $2^{31} - 1$. SIAM J. Sci. Stat. Comput. **7**, 24–45.

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

A. L. Garkavi (1964): The best possible net and the best possible cross-section of a set in a normed space. AMS Translations **2, 39**, 111–132.

S. Graf, R. D. Mauldin, S. C. Williams (1986): Random homeomorphisms. Advances in Math. **60**, 239–359.

S. Graf, E. Novak (1987): The average error of quadrature formulas for functions of bounded variation. Rocky Mountain J. Math., to appear.

B. L. Granovskii, S. M. Ermakov (1977): The Monte Carlo method. J. Sov. Math. **7**, 161–192.

Yu. I. Grigorian (1973): Diameters of certain sets in function spaces. Math. Notes **13**, 383–388.

C. W. Groetsch (1984): The theory of Tikhonov regularization for Fredholm equations of the first kind. Research Notes in Mathematics, Pitman.

S. Guiasu, A. Shenitzer (1985): The principle of maximum entropy. Math. Intell. **7**, 42–48.

S. Haber (1966): A modified Monte-Carlo quadrature. Math. Comp. **20**, 361–368.

S. Haber (1969): Stochastic quadrature formulas. Math. Comp. **23**, 751–764.

S. Haber (1970): Numerical evaluation of multiple integrals. SIAM Reviews **12**, 481–526.

J. M. Hammersley, D. C. Handscomb (1964): Monte-Carlo-methods. Methuen, London.

D. C. Handscomb (1964): Remarks on a Monte Carlo integration method. Numer. Math. **6**, 261–268.

W. L. Harper, C. A. Hocker (eds.)(1977): Foundations of probability theory, statistical inference, and statistical theories of science. Vol 1–3. Reidel, Dordrecht-Boston.

W. Hengartner, R. Theodorescu (1978): Einführung in die Monte-Carlo-Methode. Hanser Verlag, München.

E. Hlawka (1949): Ausfüllung und Überdeckung konvexer Körper durch konvexe Körper. Monatsh. Math. **53**, 81–131.

E. Hlawka, F. Firneis, P. Zinterhof (1981): Zahlentheoretische Methoden in der Numerischen Mathematik. Oldenbourg Verlag, Wien.

B. Hofmann (1986): Regularization for applied inverse and ill-posed problems. Teubner Texte zur Mathematik, Leipzig.

R. B. Holmes (1972): A course on optimization and best approximation. Lecture Notes in Math. **257**.

K. Höllig (1979): Approximationszahlen von Sobolev-Einbettungen. Math. Ann. **242**, 273–281.

K. Höllig (1980): Diameters of classes of smooth functions. In: Quantitative approximation. Proc. Symp. 1979 (R. A. DeVore, K. Scherer, eds.), Academic Press, New York.

V. V. Ivanov (1971): On optimum minimization algorithms in classes of differentiable functions. Sov. Math. Dokl. **12**, 1669–1673.

E. T. Jaynes (1979): Where do we stand on maximum entropy? In: The maximum entropy formalism. MIT press, Cambridge, 15–118.

B. Z. Kacewicz (1987): Optimal solution of ordinary differential equations. J. Complexity **3**, 451–465.

M. Kadec, S. Snobar (1971): Certain functionals on the Minkowski compactum. Math. Notes **10**, 694–696.

M. I. Kadets, V. N. Zamyatin (1968): Chebyshev centers in the space $C[a,b]$. Teor. Funk., Funkc. Anal. Prilŏz. **7**, 20–26.

M. H. Kalos, P. A. Whitlock (1986): Monte Carlo methods. Vol. 1. Wiley, New York.

B. S. Kashin (1977): Diameters of some finite-dimensional sets and classes of smooth functions. Math. USSR Izvestija **11**(2), 317–333.

B. S. Kashin (1980): On estimates of diameters. In: Quantitative approximation. Proc. Symp. 1979 (R. A. DeVore, K. Scherer, eds.), Academic Press, New York.

H. L. Keng, W. Yuan (1981): Applications of number theory to numerical analysis. Springer Verlag, Berlin.

D. E. Knuth (1981): The art of computer programming, Vol. 2.: Seminumerical algorithms. 2nd ed. Addison-Wesley, Reading.

A. N. Kolmogorov, V. M. Tikhomirov (1959): ε-entropy and ε-capacity of sets in functional spaces. Amer. Math. Soc. Transl. **17**, 277–364.

M. A. Kon, E. Novak (1988): On the adaption problem. Preprint.

N. P. Korneicuk (1981): Widths in L_p of classes of continuous functions and optimal methods of coding and recovering functions and their derivatives. Math. USSR Izvestija **18**(2), 227–247.

M. A. Kowalski, A. G. Werschulz, H. Woźniakowski (1985): Is Gauss quadrature optimal for analytic functions? Numer. Math. **47**, 89–98.

E. Kranakis (1986): Primality and Cryptography. Teubner and Wiley.

A. Kufner, O. John, S. Fučik (1977): Function spaces. Noordhoff, Leyden.

H. Kuo (1975): Gaussian measure in Banach spaces. Lecture Notes in Math. **463**, Springer.

F. M. Larkin (1972): Gaussian measure in Hilbert space and applications in numerical analysis. Rocky Mount. J. Math. **2**, 379–421.

D. Lee (1986): Approximation of linear operators on a Wiener space. Rocky Mount. J. Math. **16**, 641–659.

D. Lee, G. W. Wasilkowski (1986): Approximation of linear functionals on a Banach space with a Gaussian measure. J. Complexity **2**, 12–43.

M. Levin, J. Girshovich (1979): Optimal quadrature formulas. Teubner Verlag, Leipzig.

G. G. Lorentz (1960): Lower bounds for the degree of approximation. Trans. AMS **97**, 25–34.

G. G. Lorentz (1966): Approximation of functions. Holt, Rinehart and Winston, New York. (Second edition: Chelsea Publ. Comp., 1986).

V. G. Maz'ja (1985): Sobolev spaces. Springer–Verlag, Berlin.

R. J. McEliece, E. C. Posner (1971): Hide and seek, data storage, and entropy. Ann. Math. Statist. **42**, 1706–1716.

R. J. McEliece, E. C. Posner (1973): Hiding and covering in a compact metric space. Ann. of Statist. **1**, 729–739.

C. A. Micchelli (1985): Orthogonal projections are optimal algorithms. J. Approx. Theory **40**, 101–110.

C. A. Micchelli, T. J. Rivlin, S. Winograd (1976): The optimal recovery of smooth functions. Numer. Math. **26**, 191–200.

C. A. Micchelli, T. J. Rivlin (1977): A survey of optimal recovery. In: Optimal Estimation in Approximation Theory (C. A. Micchelli, T. J. Rivlin, eds.). Plenum Press, New York.

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

V. D. Milman, G. Schechtman (1986): Asymptotic theory of finite dimensional normed spaces. Lecture Notes in Math. **1200**.

D. S. Mitrinović (1970): Analytic inequalities. Springer Verlag, Berlin.

V. A. Morozov (1984): Methods for solving incorrectly posed problems. Springer-Verlag.

V. P. Motornyi (1973): On the best quadrature formula of the form $\sum_{k=1}^{n} p_k f(x_k)$ for some classes of periodic differentiable functions. Sov. Math. Dokl. **14**, 1180–1183.

P. H. Müller, V. Nollau, A. I. Polovinkin (1986): Stochastische Suchverfahren. VEB Fachbuchverlag Leipzig.

A. S. Nemirovsky, D. B. Yudin (1983): Problem complexity and method efficiency on optimization. John Wiley and Sons, Chicester.

H. Niederreiter (1978): Quasi-monte carlo methods and pseudo-random numbers. Bull. AMS **84**, 957–1041.

H. Niederreiter (1983): A quasi-Monte Carlo method for the approximate computation of the extreme values of a function. Studies in Pure Math. to the memory of Paul Turan. Akadémiai Kiadó, Budapest.

H. Niederreiter (1984): Number-theoretic problems in pseudo-random number generation. Proc. Symp. Appl. of Number Theory to Numerical Analysis (Kyoto, 1984). Preprint.

H. Niederreiter (1985a): Quasi-Monte Carlo methods for global optimization. In: Math. Statistics and Appl. , Proc. 4th Pannonian Symp. on Math. Statist., 1983. (W. Grossmann, G. Pflug, I. Vincze, W. Wertz, eds.) Reidel, 251–267.

H. Niederreiter (1985b): The serial test for pseudo-random numbers generated by the linear congruential method. Numer. Math. **46**, 51–68.

H. Niederreiter (1986): Multidimensional numerical integration using pseudorandom numbers. Math. Progr. Study **27**, 17–38.

E. Novak (1983): Zur unteren Fehlergrenze von Quadraturverfahren. Doctoral thesis, Erlangen.

E. Novak (1985): Eingeschränkte Monte Carlo-Verfahren zur numerischen Integration. In: Math. Statistics and Appl. , Proc. 4th Pannonian Symp. on Math. Statist., 1983. (W. Grossmann, G. Pflug, I. Vincze, W. Wertz, eds.) Reidel, 269–282.

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

E. Novak (1986b): On average case errors in numerical analysis. J. Complexity **2**, 229–238.

E. Novak (1986c): The average a posteriori error of numerical methods. Numer. Math. **50**, 245–252.

E. Novak (1987a): A decomposition of measures in euclidean space yielding error bounds for quadrature formulas. Math. Z. **196**, 407–413.

E. Novak (1987b): Stochastic properties of quadrature formulas. Numer. Math., to appear.

E. Novak, K. Ritter (1987): Centers of probability measures and optimal average case algorithms. Preprint, Erlangen.

K. Yu. Osipenko (1972): Optimal interpolation of analytic functions. Math. Notes **12**, 712–719.

K. Yu. Osipenko (1976): Best approximation of analytic functions from information about their values at a finite number of points. Math. Notes **19**, 17–23.

E. W. Packel (1986): Linear problems (with extended range) have linear optimal algorithms. Aequationes Math. **31**, 18–25.

E. W. Packel, H. Woźniakowski (1987): Recent developments in information-based complexity. Bulletin AMS **17**, 9–36.

P. L. Papini, I. Singer (1979): Best coapproximation in normed linear spaces. Mh. Math. **88**, 27–44.

K. R. Parthasarathy (1967): Probability measures on metric spaces. Academic Press, New York.

S. V. Pereverzev (1986): On optimization of approximate methods of solving integral equations. Soviet Math. Dokl. **33**, 347–351.

A. B. Pevnyi (1982): On optimal search strategies for the maximum of a function with bounded highest derivative. USSR Comput. Math. Math. Phys. **22**(5), 38–44.

R. Pick (1987): Expectation in metric spaces. Studia Sci. Math. Hungarica **22**, 347–350.

A. Pinkus (1985): n-widths in approximation theory. Springer, Berlin.

A. Pinkus (1986): n-widths and optimal recovery. In: Approximation Theory (C. de Boor, ed.), Proc. Symp. Appl. Math. **36**, 51–66.

V. I. Polovinkin (1974): Sequences of functionals with a boundary layer. Siberian Math. J. **15**, 296–308.

E. C. Posner, E. R. Rodemich (1971): Epsilon entropy and data compression. Ann. Math. Statist. **42**, 2079–2125.

E. C. Posner, E. R. Rodemich (1972): Epsilon entropy of probability distributions. Proc. 6th Berkeley Symp. Math. Statist. Prob. (Univ. Calif., 1970) **3**, 699–707.

C. A. Rogers (1964): Packing and covering. Cambridge University Press, London.

V. I. Ruban (1974): Even diameters of the class $W^r H_\omega$ in the space $C_{2\pi}$. Math. Notes **15**, 222–225.

R. Y. Rubinstein (1981): Simulation and the Monte Carlo method. Wiley and Sons, New York.

C. P. Schnorr (1971): Zufälligkeit und Wahrscheinlichkeit. Eine algorithmische Begründung der Wahrscheinlichkeitstheorie. Lecture Notes in Math. **218**. Springer Verlag, Berlin.

C. P. Schnorr (1977): A survey of the theory of random sequences. In: Basic problems in methodology and linguistics. (R. Butts, J. Hintikka, eds.), Reidel, Dordrecht, 193–211.

H.-P. Schwefel (1977): Numerische Optimierung von Computer Modellen mittels der Evolutionsstrategie. Birkhäuser Verlag, Stuttgart.

H. S. Shapiro (1971): Topics in approximation theory. Lecture Notes in Math. **187**, Springer Verlag, Berlin.

A. F. Siegel, F. O'Brian (1985): Unbiased Monte Carlo integration methods with exactness for low order polynomials. SIAM J. Sci. Stat. Comput. **6**, 169–181.

A. V. Skorohod (1974): Integration in Hilbert space. Springer, New York.

S. Smale (1985): On the efficiency of algorithms of analysis. Bull. AMS **13**(2), 87–121.

I. M. Sobol (1985): Die Monte-Carlo-Methode. Deutsch Verlag, Frankfurt.

S. L. Sobolev (1965): On the order of convergence of cubature formulas. Soviet Math. Dokl. **6**, 808–811.

W. Stegmüller (1973): Probleme und Resultate der Wissenschaftstheorie und Analytischen Philosophie, Band 4: Personelle und statistische Wahrscheinlichkeit. Springer Verlag, Berlin.

H. Strauss (1979): Optimal quadrature formulas. In: Approximation in Theorie und Praxis. (G. Meinardus, ed.), Bibliographisches Institut, Mannheim, 239–250.

A. H. Stroud (1971): Approximate calculation of multiple integrals. Prentice-Hall, New Jersey.

A. G. Sukharev (1971): Optimal stategies of the search for an extremum. USSR Comput. Math. Math. Phys. **11**, 119–137.

A. G. Sukharev (1972): Best sequential search strategies for finding an extremum. USSR Comput. Math. Math. Phys. **12**, 39–59.

A. G. Sukharev (1978): Optimal method of constructing best uniform approximations for functions of a certain class. USSR Comput. Math. Math. Phys. **18**(2), 21–31.

A. G. Sukharev (1979): Optimal numerical integration formulas for some classes of functions of several variables. Sov. Math. Dokl. **20**, 472–475.

A. G. Sukharev (1981a): The global extremum and methods of finding it. *(Russian)*. In: Mathematical methods in operations research, Izdvo MGU, Moscow.

A.G.Sukharev (1981b): A stochastic algorithm for extremum search, optimal in one step. USSR Comput. Math. Math. Phys. **21**(6), 23–39.

A. G. Sukharev (1984): On optimal methods in numerical analysis. Comput. Math., Banach Center Publ. **13**, 575–587. Warsaw.

A. G. Sukharev (1985): The equality of errors in classes of passive and sequential algorithms. USSR Comput. Math. Math. Phys. **25**(1), 193–195.

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

A. V. Suldin (1959/60): Wiener measure and its application to approximation methods I, II. Izv. Yyssh. Uchebn. Zaved. Mat. 6(13), 145–158 and 5(18), 165–179.

G. Talenti (ed.) (1986): Inverse problems. Lecture Notes in Math. **1225**.

V. N. Temlyakov (1985): On the approximate reconstruction of periodic functions of several variables. Soviet Math. Dokl. **31**(1), 246–249.

V. M. Tikhomirov (1960): Widths of sets in functional spaces and the theory of best approximations. Russ. Math. Surv. **15**(3), 75–112.

V. M. Tikhomirov (1969): Best methods of approximation and interpolation of differentiable functions in the space $C[0,1]$. Math. USSR Sbornik **9**(2), 275–289.

V. M. Tikhomirov (1976): Some problems of approximation theory. *(Russian)*. Izdvo, MGU, Moscow.

A. N. Tikhonov, V. Y. Arsenin (1979): Methods for solving ill-posed problems. Wiley, New York.

J. F. Traub, H. Woźniakowski (1980): A general theory of optimal algorithms. Academic Press, New York.

J. F. Traub, G. W. Wasilkowski, H. Woźniakowski (1983): Information, uncertainty, complexity. Addison–Wesley, Reading.

J. F. Traub, G. W. Wasilkowski, H. Woźniakowski (1984): Average case optimality for linear problems. Th. Comp. Sci. **29**, 1–25.

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

J. F. Traub, H. Woźniakowski (1984): Information and computation. In: Advances in Computers, vol. **23**, Academic Press, New York.

N. N. Vakhania (1981): Probability distributions on linear spaces. North Holland, New York.

A. L. Volberg, S. V. Konyagin (1984): There is a homogeneous measure on any compact subset in $\mathbf{R}^n$. Soviet Math. Dokl. **30**, 453–456.

S. M. Voronin, V. I. Skalyga (1984): On quadrature formulas. Soviet Math. Dokl. **29** (3), 616–619.

S. M. Voronin, N. Temirgaliev (1986): Application of Banach measure to quadrature formulas. Math. Notes **39**, 30–34.

D. H. Wagner (1980): Survey of measurable selection theorems: an update. In: Measure theory, Proc. Oberwolfach (D. Kölzow, ed.), Lecture Notes in Math. **794**, 176–216.

G. W. Wasilkowski (1983): Local average errors. Techn. Rep., Dep. Comp. Sci., Columbia University, New York.

G. W. Wasilkowski (1984): Some nonlinear problems are as easy as the approximation problem. Comput. Math. Appl. **10**, 351–363.

G. W. Wasilkowski (1985): Average case optimality. J. Complexity **1**, 107–117.

G. W. Wasilkowski (1986a): Optimal algorithms for linear problems with Gaussian measures. Rocky Mount. J. Math. **16**, 727–749.

G. W. Wasilkowski (1986b): Information of varying cardinality. J. Complexity **2**, 204–228.

G. W. Wasilkowski (1987): Randomization for continuous problems. Preprint, Dep. of Comp. Sci., Columbia University, New York.

G. W. Wasilkowski, H. Woźniakowski (1984): Can adaption help on the average? Numer. Math. **44**, 169–190.

A. G. Werschulz (1985a): Complexity of differential and integral equations. J. Complexity **1**, 232–255.

A. G. Werschulz (1985b): What is the complexity of elliptic systems? J. Approx. Theory **45**, 69–89.

A. G. Werschulz (1986): What is the complexity of related elliptic, parabolic and hyperbolic problems? Math. of Comp. **47**, 461–472.

A. G. Werschulz (1987a): What is the complexity of ill-posed problems? Numer. Funct. Anal. and Optimiz. **9**, 945–967.

A. G. Werschulz (1987b): An information-based approach to ill-posed problems. J. Complexity **3**, 270–301.

A. G. Werschulz, H. Woźniakowski (1986): Are linear algorithms always good for linear problems? Aequationes Math. **31**, 202–211.

H. Woźniakowski (1985): A survey of information-based complexity. J. Complexity **1**, 11–44.

H. Woźniakowski (1986): Information-based complexity. Ann. Rev. Comput. Sci. **1**, 319–380.

N. F. Zaliznyak, A. A. Ligun (1978): On optimum strategy in search of global maximum of function. USSR Comput. Math. Math. Phys. **18**(2), 31–38.

S. K. Zaremba (1968): The mathematical basis of Monte Carlo and Quasi Monte Carlo methods. SIAM Rev. **10**, 303–314.

R. Zielinski (1978): Erzeugung von Zufallszahlen. Deutsch Verlag, Thun.

R. Zieliński, P. Neumann (1983): Stochastische Verfahren zur Suche nach dem Minimum einer Funktion. Akademie Verlag, Berlin.

A. A. Zhensykbaev (1976): On the best quadrature formula on the class $W^r L_p$. Sov. Math. Dokl. **17**, 377–380.

A. A. Zhensykbaev (1981): Monosplines of minimal norm and the best quadrature formulas. Russ. Math. Surv. **36**(4), 121–180.

S. Zubrzycki (1963/64): Some approximate integration formulas of statistical interest. Coll. Math. **11**, 123–135.

A. K. Zvonkin, L. A. Levin (1970): The complexity of finite objects and the development of the concepts of information and randomness by means of the theory of algorithms. Russ. Math. Surv. **15**(6), 83–124.

Notations

$S : F \to M$	the given problem, 1.1.1
$\widetilde{S} : F \to M$	an approximation of S, 1.1.1
$\Delta : M \times F \to \mathbf{R}_0^+$	the error function, 1.1.1
$\Delta(\widetilde{S}(f), f)$	the error of $\widetilde{S}$ for $f \in F$, 1.1.1
Δ_{max}	the maximal error of $\widetilde{S}$, 1.1.1
A	a given set of deterministic methods, 1.1.2
$A(N)$	set of approximations which use an information N, 1.1.2
$r(N) = r_{max}(N)$	(maximal) radius of the information N, 1.1.2
I_n^{ad}, I_n	sets of information operators, 1.1.3
A_n^{ad}, A_n	sets of approximations, 1.1.3
$e_n^{(ad)}(F,S)$	worst case error bounds, 1.1.3
App	approximation problem, 1.1.4
Opt, Opt^*	optimization problems, 1.1.4
Int	integration problem, 1.1.4
λ, λ^s	Lebesgue measure, 1.1.4
$B(X)$	Banach space of bounded real-valued functions on X, 1.1.4
$d_n(F)$	n-width of F in the space $B(X)$, 1.2.1
$d_n^L(F)$	linear n-width of F in $B(X)$, 1.2.6
$rad(D)$	Chebyshev radius of a set, 1.3.1
$diam(D)$	diameter of a set, 1.3.1
$c(M)$	Jung constant of a normed space, 1.3.1
$\alpha(M)$	adaption constant of a normed space, 1.3.1
N_g	a certain nonadaptive information, 1.3.1
F_ω	a Lipschitz class of functions, 1.3.6
F_s	a Lipschitz class of functions, 1.3.6
c_n	n-th covering constant of a metric space, 1.3.6
$a_n \approx b_n$	strong equivalence of sequences, 1.3.6
p_n	n-th packing constant of a metric space, 1.3.7
$\mathrm{Var}(f)$	total variation of a function, 1.3.8
F_{Var}	a set of functions of bounded variation, 1.3.8
$C_s^{k,\alpha}$	Hölder classes of functions, 1.3.9
$a_n \asymp b_n$	weak equivalence of sequences, 1.3.9
$W_\infty^k, W_p^{k,s}$	Sobolev classes, 1.3.10 and 1.3.11
$P(k,s)$	polynomials in s variables with degree less than k, 1.3.12
$\|\mu\|$	total variation of a signed measure, 1.3.12

$Q = (Q(\omega))_{\omega\in\Omega}$	a (generalized) Monte Carlo method, 2.1.2
${}^*C(A)$	set of generalized Monte Carlo methods, 2.1.2
$\Delta(Q,f)$	error of Q for $f \in F$, 2.1.2
$\Delta_{max}(Q)$	error of Q, 2.1.2
${}^*\sigma_n^{(ad)}(F,S)$	error bounds for generalized Monte Carlo methods, 2.1.2
$\widetilde{\Delta}(Q,f)$, $\widetilde{\Delta}_{max}(Q)$	dispersion of a Monte Carlo method, 2.1.2
${}^*\tilde{\sigma}_n^{(ad)}(F,S)$	bounds for the dispersion of Monte Carlo methods, 2.1.2
$C(A)$	set of (restricted) Monte Carlo methods, 2.1.4
$\sigma_n^{(ad)}(F,S)$	error bounds for Monte Carlo methods, 2.1.4
$\tilde{\sigma}_n^{(ad)}(F,S)$	bounds for the dispersion of Monte Carlo methods, 2.1.4
$\Delta_\mu(\widetilde{S})$	average error of $\widetilde{S}$, 2.1.9, 3.1.2
$\int^* h\,d\mu$	upper integral of h with respect to μ, 2.1.9
$e_n^{(ad)}(F,S,\mu)$	bounds for the average error, 2.1.9, 3.1.2
ε_f	Dirac measure, 2.1.9
$C(A_n^{var})$	class of Monte Carlo methods with varying cardinality, 2.1.10
$\sigma_n^{var}(F,S)$	error bounds for Monte Carlo methods, 2.1.10
$\Delta_\mu(Q)$	average error of Q, 3.1.3
$\sigma_n^{(ad)}(F,S,\mu)$	bounds for the average error of M.C. methods, 3.1.3
F_1^l	a class of Lipschitz functions, 3.1.4
$cost_\mu(\widetilde{S})$	average cardinality of a method, 3.1.10
$\alpha(x)$	local error of a method $\widetilde{S}$, 3.2.1
$\Delta_{aver}(\phi,N)$	average a posteriori error, 3.2.2
$a_n(F,S)$	bounds for the average a posteriori error, 3.2.2
$r_{aver}(N)$	average case radius of N, 3.2.2
$\alpha(x_1,\ldots,x_m)$	certain means of the local error, 3.2.7
$\Delta_{aver}^{ad}(\phi,N)$	average a posteriori error of adaptive methods, 3.2.7
$a_n^{ad}(F,S)$	bounds for the average a posteriori error, 3.2.7
$\widetilde{W}_p^k$	a class of periodic Sobolev functions, A.2
$rad^p(m)$	p-radius of a measure, A.4
Δ_μ^p	p-average error of a method, A.5
$r_\mu^p(N)$	p-average radius of an information, A.5

Index