ON THE POWER OF ADAPTION*

ERICH NOVAK†

Abstract. Optimal error bounds for adaptive and nonadaptive numerical methods are compared. Since the class of adaptive methods is much larger, a well-chosen adaptive method might seem to be better than any nonadaptive method. Nevertheless there are several results saying that under natural assumptions adaptive methods are not better than nonadaptive ones. There are also other results, however, saying that adaptive methods can be significantly better than nonadaptive ones as well as bounds on how much better they can be. It turns out that the answer to the "adaption problem" depends very much on the a priori knowledge about the problem in question; even a seemingly small change of the assumptions can lead to a different answer.

Key words. adaptive methods, numerical integration, optimal recovery, zero finding, global optimization, worst case error bounds, average case error bounds, error bounds for randomized (Monte Carlo) methods, complexity of numerical problems, information-based complexity

AMS subject classifications. 41A25, 41A46, 65C05, 65D, 65H05, 65Y20, 68Q25, 90C60

1. The adaption problem. One of the more controversial issues in numerical analysis concerns adaptive algorithms. The use of such algorithms is widespread and many people believe that well-chosen adaptive algorithms are much better than nonadaptive methods in most situations. Such a belief is usually based on numerical experimentation. In this paper we survey what is known theoretically regarding the power of adaption. We will present some results which state that under natural assumptions adaptive methods are not better than nonadaptive ones. There are also other results, however, saying that adaptive methods can be significantly superior to nonadaptive ones. As we will see, the power of adaption is critically dependent on our a priori knowledge concerning the problem being studied; even a seemingly small change in the assumptions can lead to a different answer.

Let us begin with some well-known examples. The bisection method and

*This work was supported by a Heisenberg scholarship of the German Research Council (DFG).

†Mathematisches Institut, Universität Erlangen-Nürnberg, Bismarckstr. 1 1/2, 91054 Erlangen, Germany (novak@mi.uni-erlangen.de).

the Newton method for zero finding of a function are adaptive, since they compute a sequence $(x_n)_n$ of knots that depends on the function. The Gauss formula for numerical integration is nonadaptive since its knots and weights do not depend on the function.

A nonadaptive method provides an immediate decomposition for parallel computation. If adaptive information is superior to nonadaptive information, then an analysis of the tradeoff between using adaptive or nonadaptive information on a parallel computer should be carried out.

To formulate the adaption problem precisely, we need some definitions and notations. Many problems of numerical analysis can be described as computing an approximation of the value $S(f)$ of an operator

$$S : X \to G$$

for $f \in F$, where $F \subset X$. Here we assume that X is a normed space of functions and G is also a normed space. The operator S describes the solution of a mathematical problem, for example the solution of a boundary value problem or an integral equation. Also numerical integration (with $G = \mathbf{R}$) and the recovery of functions (with an imbedding $S = id : X \to L_p$, where $X \subset L_p$) can be stated in this way. In many cases the space X is infinite dimensional and therefore $f \in X$ cannot directly be an input of a computation. We usually replace S by a discretization method, for example, given by a finite element method. Accordingly, numerical methods are often of the form

$$S_n(f) = \varphi(L_1(f), L_2(f), \dots, L_n(f)) = \varphi(N(f)) \tag{1}$$

with linear functionals $L_k : X \to \mathbf{R}$ and a (linear or nonlinear) mapping $\varphi : \mathbf{R}^n \to G$. Hence numerical methods only use partial information $N(f)$ of $f \in X$. The most important example is

$$N(f) = (f(x_1), f(x_2), \dots, f(x_n)),$$

but other functionals are common as well. Examples of such functionals include weighted integrals, Fourier coefficients, wavelet coefficients, or values of a derivative of f.

A method is called *nonadaptive*, if the functionals L_k are fixed in advance and do not depend on f.

One might hope that it is possible to learn about f during the computation of $L_1(f), \dots, L_{k-1}(f)$ in such a way that one can choose the next functional L_k suitably to reduce the error. Therefore one studies *adaptive* methods, where the choice of L_k may depend on the (already computed) values $L_1(f), \dots, L_{k-1}(f)$. In the case $L_k(f) = f(x_k)$, for instance, the knot x_k depends on the known function values via

$$x_k = \psi_k(f(x_1), \dots, f(x_{k-1})),$$

where ψ_k is a function of $(k-1)$ variables. In mathematical statistics, adaptive information is known as sequential design and nonadaptive information is known as non-sequential design.

In (1) we define methods with *fixed cardinality.* For more general methods we also use an *adaptive stopping rule* and obtain methods with *varying cardinality,* where $n = n(f)$ depends on f. After the computation of $L_k(f)$ one decides, on the basis of the computed information $L_1(f), \dots, L_k(f)$, whether additional information is used or not. See [119] or [125] for the exact definitions. We will see in Section 5 that varying cardinality is important for zero finding.

Remark on nonlinear approximation. It should be stressed that the functionals L_k in (1) cannot depend on f in an arbitrary way but only via the already computed values of $L_1(f), \dots, L_{k-1}(f)$. This is because we are interested in feasible computations with small cost, including the cost to obtain information on f. Sometimes it is interesting to even allow the L_k to depend on f in a more general way. This is the core, for example, in efficient data compression. The resulting number n of values $y_1, \dots, y_n$, that allow a recovery of $S(f)$ to within some error, should be small. Important examples include rational approximation, approximation by splines with free knots, and wavelet compression. See [7], [12], [15], [25], [27], and [28]. A general problem is the approximation of $f \in X$ by an arbitrary expression of the form

$$g = \sum_{k=1}^{n} y_k g_{i_k},$$

where $(g_i)_i$ is a given sequence of functions, and the coefficients y_k and the indices i_k may depend on f, see [55]. A good n-term approximation g might be difficult to find, but if it is available then it can easily be transmitted and evaluated. Nonlinear approximation is strongly related to Bernstein and local widths, see [26], [27], [76], [77], and also to recovering infinite dimensional objects from given noisy data, see [31] and [32].

Different settings. There are several problems in numerical analysis, for which all algorithms to find an approximation of the solution are very expensive in the worst case setting. This is the case, for example, for integration or optimization of poorly-behaved functions of several variables. For such problems it is important to know whether they can be solved with justifiable cost at least for most functions and therefore we study the average case and/or randomized methods. Are various stochastic error bounds much better than worst case error bounds? This question is strongly related to the adaption problem because it may happen that adaption does not help in the worst case setting but helps significantly with respect to other settings.

Summary. We summarize the main results on adaption. Most of them will be explained in this paper.

1. On adaptive stopping rules.

1.1. Adaptive stopping rules are not better than nonadaptive ones for many problems, see [70], [119], and [125]. This is always true in the worst case setting. For many natural linear problems, this is also true in the average case setting. Hence we usually only consider methods (1) with a nonadaptive stopping rule.

1.2. For some linear problems adaptive stopping rules are superior, see [91]. An average case analysis of the problem of zero finding shows that adaptive stopping rules are much better, see [80] and [83].

2. For some natural problems adaption does not help.

2.1. This is true for linear problems on symmetric convex sets $F \subset X$ in the worst case setting and also in the average case setting for a Gaussian measure on X, see [5], [35], [58], [59], [70], [119], [120], [130], and [133].

2.2. With respect to the worst case error of deterministic methods, adaption does not help also for integration of monotone or convex functions and also for the problem of global optimization, see [56], [70], [73], [78], and [124].

2.3. For certain problems in nonparametric regression, where the information is disturbed by white noise, adaption does not help for the recovery of functions, see [39].

3. For some natural problems adaption helps significantly.

3.1. If the set F is convex but nonsymmetric then adaption may help a lot even for linear problems in the worst case setting, see [60], [74], [76], [77], [107], [108], and [109].

3.2. In some cases, such as integration of monotone or convex functions, the superiority of adaptive methods can only be seen if randomized methods or average case errors are studied, see [56], [73], and [78]. Adaption only helps on the average for global optimization. The advantage of adaptive methods can not be seen by studying the worst case error of deterministic or randomized methods, see [16], [70], [124] and [126].

3.3. Adaption significantly helps for certain classes of functions with singularities, see [51], [84], [129] and [134].

3.4. Adaption significantly helps for the solution of ordinary differential equations, see [52] and [53].

3.5. For many problems of numerical integration based on function values disturbed by white noise, adaption significantly helps, see [93] and [93a].

Conclusions. Till about 10 years ago, most theoretical results concerning the adaption problem stated that under certain assumptions adaption does not help. In recent years more general assumptions were studied and it turned out that the power of adaption critically depends on our a priori knowledge concerning the problem. This is confirmed by numerical experience.

In [84] the authors present results concerning numerical integration of unimodal peak functions, see also Section 3. For a specific class of integrands an adaptive algorithm is presented that uses 90 function values and is better (in a worst case sense with respect to the logarithmic error criterion) than any nonadaptive method that uses up to 800000 function values.

So a complexity analysis may lead to new efficient methods. Other examples include high-dimensional integration or optimal recovery of functions. An earlier example is the invention of multigrid methods. Prototypes of these methods were first introduced by Fedorenko [34] and Bakhvalov [4] for a complexity analysis. A lot of further work was necessary to transform the early theoretical methods into methods that are useful in practice, see [13] and [47]. Now multigrid is in wide use.

A pure worst case analysis can lead to an inadequate rating of different methods. See [11] for results on the simplex method. An average case analysis can lead to new methods that could not be found otherwise.

Contents. In the rest of this section we discuss the relation between error bounds and complexity bounds. In Section 2 we discuss results for arbitrary linear problems defined on a symmetric convex set. In Section 3 we discuss linear problems on general convex sets. In Section 4 and 5 we study some nonlinear problems: global optimization and zero finding.

Error bounds and complexity results. Assume that a linear operator $S : X \to G$ is to be approximated on the unit ball F of X. For a method S_n of the form (1) we define the worst case error by

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

Let ε_n denote the error

$$\Delta_{\max}(S_n^*) = \inf_{S_n} \Delta_{\max}(S_n) =: \varepsilon_n$$

of the optimal method S_n^* using n function values. In many cases we know that S_n^* can be chosen as linear, so that the cost of computing $S_n^*(f)$ is proportional to n. This cost is mainly the cost of computing the information

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

In this case it would be enough to study the error bounds ε_n. However, there are problems where the cost of computing $S_n(f) = \varphi(N(f))$ from $N(f)$ cannot be neglected because φ is complicated, see [18], [64], [85], and [131]. Also the cost of computing $N(f)$ may be large if N is adaptive and an L_k depends in a complicated way on $L_1(f), \ldots, L_{k-1}(f)$.

To study the cost of such methods one needs a well-defined model of computation. In numerical analysis we usually consider the real number model, where one assumes that arithmetic operations with real numbers and comparisons can be done with unit cost. We often deal with partial information consisting of function values or Fourier coefficients, because a digital computer can only handle finite sets of numbers instead of functions. In information-based complexity it is assumed that certain functionals can be evaluated by an oracle and each call of the oracle costs $\mathbf{c}$, where $\mathbf{c} > 0$. This model of computation is described more carefully in [75], see also [8], [92], [105], [119] and [121].

There is another reason why we want to study the cost of algorithms, as well as the sequence $(\varepsilon_n)_n$ of error bounds. In applications the error level ε often is not fixed and one defines *uniform algorithms* that on input $\varepsilon > 0$ produce an ε-approximation to $S(f)$. In such a case one cannot use a fixed *precomputed* method S_n but must compute the number n of knots and the knots itself during the computation. To study the cost of computing $n = n(\varepsilon)$ and suitable knots we clearly need a model of computation. One particular uniform problem for numerical integration is discussed in [75].

In this paper, however, we consider the adaption problem only in the case where the error level ε is fixed, so that n and the method S_n can be precomputed.

2. Linear problems on symmetric convex sets. In this section we assume that $S : X \to G$ is a linear operator. One important example is the problem of numerical integration where

$$S(f) = \int_\Omega f(x)\,dx, \quad G = \mathbf{R},$$

with some given $\Omega \subset \mathbf{R}^d$.

One often considers error estimates that depend on the norm of $f \in X$. If the method S_n is linear, i.e., of the form

$$S_n(f) = \sum_{k=1}^{n} L_k(f) \cdot g_k$$

with certain $g_k \in G$, then one wants estimates of the form

$$\|S(f) - S_n(f)\| \le c \cdot \|f\|, \quad \forall\, f \in X,$$

with c as small as possible. This corresponds to the worst case analysis with the error

$$\Delta_{\max}(S_n) = \sup_{f \in F} \|S(f) - S_n(f)\|, \tag{2}$$

where F is the unit ball of X. Of course we can use the definition (2) also for adaptive methods and, later, also for sets F different from the unit ball.

The first general result concerning the adaption problem is from Bakhvalov (1971), where S is a linear functional and the L_k are special linear functionals, for instance function evaluations, $L_k(f) = f(x_k)$. Bakhvalov proved that then adaption does not help. A result of Smolyak states that, under the same assumptions, linear methods are optimal in the class of all nonadaptive methods. Smolyak's result was not published in a journal; it is generally known also through Bakhvalov's paper [5]. We formulate the results of Smolyak and Bakhvalov as follows.

THEOREM 1. *Assume that $S : X \to \mathbf{R}$ is a linear functional and the error of S_n is defined by (2), where F is a symmetric convex subset of X.*

Assume that S_n is an arbitrary method of the form (1) *using, for $f = 0 \in F$, the functionals*

$$N^0(f) = (L_1^0(f), L_2^0(f), \dots, L_n^0(f)).$$

Then there is a linear nonadaptive method

$$S_n^*(f) = \sum_{k=1}^n a_k\, L_k^0(f)$$

such that

$$\Delta_{\max}(S_n^*) \le \Delta_{\max}(S_n).$$

Proof. Let S_n be any (adaptive or nonadaptive) method of the form (1). We denote by $L_1^0, L_2^0, \dots, L_n^0$ the (fixed) functionals used by S_n for the zero function $0 \in F$. It is clear that

$$S_n(f) = S_n(0)$$

for all $f \in F$ with

$$N^0(f) = (L_1^0(f), L_2^0(f), \dots, L_n^0(f)) = 0.$$

We have

$$\Delta_{\max}(S_n) \ge \sup_{f \in A} |S(f) - S_n(f)|,$$

where

$$A = \{f \in F \mid N^0(f) = 0\}.$$

Since A is symmetric ($f \in A$ implies $-f \in A$) we have

$$|S(f)| \le \frac{1}{2}|S(f) - S_n(0)| + \frac{1}{2}|S(-f) - S_n(0)|$$

and

$$\sup_{f \in A} |S(f)| \le \sup_{f \in A} |S(f) - S_n(0)|.$$

As a consequence we obtain the important fact that

$$\Delta_{\max}(S_n) \ge \sup\{S(f) \mid f \in F,\ N^0(f) = 0\} =: r.$$

We prove that there is a linear method $S_n^* = \varphi \circ N^0$ with $\Delta_{\max}(S_n^*) = r$. Without loss of generality we may assume that $r < \infty$. Define a convex set by

$$M = \{(S(f), L_1^0(f), \dots, L_n^0(f)) \mid f \in F\} \subset \mathbf{R}^{n+1}$$

and consider a supporting hyperplane H through a boundary point y of M of the form

$$y = (r, 0, \dots, 0)$$

with $r \geq 0$. We obtain $a_k \in \mathbf{R}$ such that

$$S(f) - \sum_{k=1}^{n} a_k \, L_k^0(f) \leq r$$

for all $f \in F$. Due to the symmetry of F, we also obtain that this sum is at least $-r$. Hence we have found that the linear method

$$S_n^*(f) = \sum_{k=1}^{n} a_k \, L_k^0(f)$$

satisfies $\Delta_{\max}(S_n^*) = r \leq \Delta_{\max}(S_n)$ using the same nonadaptive information N^0 that is used by S_n for the function $f = 0$. $\square$

Remarks. a) Intuitively one might say that for zero information we do not have any chance to adjust the next knot in a special way to f to decrease the error. The reader may want to use this argument for the example $S(f) = \int_0^1 f(x)\,dx$ and

$$F = \{f \in C[0,1] \mid |f(x) - f(y)| \leq |x - y|\}.$$

It turns out the midpoint rule S_n is optimal for this particular class of functions with $\Delta_{\max}(S_n) = 1/(2n)$.

b) The adaption problem is more complicated if we consider arbitrary linear operators instead of functionals. It has been known since 1980 that nonadaptive methods are optimal up to a factor 2, see [35] and [120], and it is known from [58] and [59] that there are examples where adaption helps slightly. See also [19].

THEOREM 2. *Assume that $S : X \to G$ is a linear operator and the error of S_n is defined by (2), where F is a symmetric convex subset of X. Assume that S_n is an arbitrary method of the form (1) using, for $f = 0 \in F$, the functionals*

$$N^0(f) = (L_1^0(f), L_2^0(f), \dots, L_n^0(f)).$$

Then there is a nonadaptive method of the form

$$S_n^*(f) = \varphi(N^0(f))$$

such that

$$\Delta_{\max}(S_n^*) \le 2\,\Delta_{\max}(S_n).$$

Hence adaptive methods can only by a factor of at most 2 be better than nonadaptive methods. There are examples where adaptive methods are (slightly) better than nonadaptive ones.

Remarks. a) The results so far did not show any significant superiority of adaptive methods. Nevertheless adaptive methods are often used in practice. An important application of adaptive methods is in finite element computations, see [2] and [33]. A thorough discussion of the above results and their application to the solution of boundary value problems for elliptic partial differential equations is given in the book of Werschulz [133].

b) Of course it should be stressed that these results assume that we have a linear operator S and a convex and symmetric set F. The set F reflects the *a priori knowledge* concerning the problem; often it is known that f has a certain smoothness and this knowledge may be expressed by $f \in F$. If our a priori knowledge about the problem leads to a set F that is either nonsymmetric or nonconvex (or both) then we certainly cannot apply Theorems 1 and 2 and it is possible that adaption is significantly better, see Section 3.

c) The idea behind Theorems 1 and 2 is that nonadaptive information that is good for the zero function $0 \in F$ is also good for any other $f \in F$. This is true for *any linear problem with any norm.* However, these results do *not* automatically lead to good nonadaptive methods. In particular we do not claim that the optimal nonadaptive knots are somehow uniformly distributed or equidistant. There are important examples where regular grid points are rather bad and the optimal (nonadaptive) points are more complicated. We stress this fact because we noticed that some authors compare poor nonadaptive methods based, for example, on a regular grid with sophisticated adaptive methods and (wrongly) conclude that adaptive methods are superior.

d) Many test results for the difficult problem of computing high dimensional integrals can be found in [82] and [104]. These results do not show a general superiority of adaptive algorithms. It seems that global smoothness properties of the integrand can be used better by nonadaptive methods. The results of course strongly depend on the family of integrands, see also [23], [37], [38], [54], [110], and [138].

Example 1. Assume that $X \subset G = L_p([0,1]^d)$ is a normed space. Then we can study the imbedding $S = id : X \to L_p$ on the unit ball F of X. Although this looks like a very special example, it is important because it can often be considered as a component in more general problems. Good knots x_k for an approximation of $S = id$ by a linear method

$$S_n(f) = \sum_{k=1}^{n} f(x_k) \cdot g_k$$

with $g_k \in G$ depend strongly on the space X.

a) Let X be a classical Hölder or Sobolev space, such as

$$X = W_p^r([0,1]^d),$$

with the imbedding condition $pr > d$. Then *regular grid points* together with classical algorithms such as *interpolation by piecewise polynomial functions* lead to the optimal rate of convergence, given by $n^{-r/d}$. See, for example, [20].

b) If X is a space of functions with a bounded mixed derivative such as

$$X = W_p^{(r,r,\dots,r)}([0,1]^d) = \{f \in L_p \mid D^\alpha f \in L_p \text{ for all } \alpha \in \mathbf{N}_0^d \text{ with } \alpha_i \le r\}$$

then regular grid points still give the rate $n^{-r/d}$. This is far from being optimal, because there are methods that yield an order

$$n^{-r} \cdot (\log n)^{(r+1)\cdot(d-1)}.$$

This bound can be achieved by *Smolyak's algorithm* which was introduced in 1963, see [106], and leads to almost optimal methods in any tensor product case. The sample points are *hyperbolic cross points* and the order of convergence depends only weakly on the dimension d. Explicit cost bounds for Smolyak's algorithm were proved by Wasilkowski and Woźniakowski in [132].

Smolyak's algorithm has been developed independently in many papers for specific problems, see [82]. These methods are also called *sparse grid methods* or *Boolean methods* or *discrete blending methods*, see also [24], [44], [81], [113], [114], and [115]. The hyperbolic cross points consist of a finite union of grids. In the simplest case the meshsize in each variable is $h_i = 2^{-k_i}$ with $k_i \in \mathbf{N}$ and we require that

$$\prod_{i=1}^{d} h_i \ge \alpha$$

for a positive parameter α that determines the number of knots.

Example 2. Another important example concerns the numerical computation of weighted integrals

$$S_g(f) = \int_{[0,1]^d} f(x) \cdot g(x)\, dx$$

for functions f from a Sobolev space and integrable weight functions g. As a consequence of Theorem 1 it is enough to consider linear nonadaptive methods of the form

$$S_n(f) = \sum_{k=1}^{n} a_k f(x_k).$$

This is true for any fixed g but the optimal knots x_k in a quadrature formula S_n depend strongly on the weight function. There exists a general method to obtain knots that yield an optimal rate of convergence, see [7] and [70]. It turns out, of course, that uniformly distributed knots are generally not optimal.

Average case analysis. One may argue that adaption does not help because we consider the *worst case setting*. Looking at the proofs of Theorems 1 and 2 one might hope that adaptive methods are much better with respect to an average case analysis. For such a *Bayesian approach to numerical analysis* we need a probability measure P on X. A typical choice of P is a Gaussian measure, see [29], [97], and [119]. The choice of the measure plays a role similar to the choice of the norm (or the set F) in the worst case setting. The following result from [130] reflects the strong symmetry properties of Gaussian measures.

THEOREM 3. *Assume that $S : X \to G$ is linear and the error of S_n is measured by*

$$\Delta_{\mathrm{aver}}(S_n) = \left(\int_X \|S(f) - S_n(f)\|^2 \, dP(f) \right)^{1/2}$$

with a centered Gaussian measure P on X. Then adaption does not help and linear nonadaptive (spline-) algorithms are optimal.

Again this result does not solve the *design problem* of finding optimal sample points. For some important univariate problems, such as weighted integration and optimal recovery, almost optimal sample points are known, see [17], [68], [96], and [100]. Similar problems for the multivariate case were recently solved for the Wiener sheet measure and for the isotropic Wiener measure, see [87], [98], [128], [135] and [136].

Randomized (or Monte Carlo) methods. Does adaption help for randomized methods? Here we study methods of the form

$$S_n^\omega(f) = \varphi^\omega(L_1^\omega(f), L_2^\omega(f), \dots, L_n^\omega(f)) = \varphi^\omega(N^\omega(f)), \tag{3}$$

where the ω indicates that φ and the functionals L_k are random variables. In the adaptive case these random variables also may depend on the already computed information. One may even allow that the number $n = n(f, \omega)$ of functionals depends on ω. In this case we write S_n^ω for any method for which the expected number of functionals satisfies

$$E(n(f, \cdot\,)) \le n$$

for all $f \in F$, where F is a given class of functions. A method S_n^ω of the form (3) is called nonadaptive if the number of functionals is fixed and if all random variables φ and L_k are fixed, i.e., do not depend on the computed values. For a randomized method the error $\|S(f) - S_n^\omega(f)\|$ also depends on ω and we define the worst case error of a randomized method by

$$\Delta_{\max}(S_n^\omega) = \sup_{f \in F} E(\|S(f) - S_n^\omega(f)\|),$$

where E is the mean value over the different ω. See [49], [65], [70], [119], and [126] for recent results and surveys on randomized methods.

It is *not* known whether there is a general result on adaption such as Theorem 1 or Theorem 2 for randomized methods. On the other hand, we do not know an example where adaption helps significantly for a linear operator $S : X \to G$ over a symmetric convex set $F \subset X$. For some linear problems we know that randomized methods are much better than deterministic methods. We mention one particular result for the problem of numerical integration,

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

Here the information is of the form $L_k^\omega(f) = f(x_k^\omega)$ with adaptively chosen

$$x_k^\omega = \psi_k^\omega(f(x_1^\omega), \dots, f(x_{k-1}^\omega)).$$

A special class of methods is the linear (nonadaptive) methods of the form

$$S_n^\omega(f) = \sum_{k=1}^n a_k^\omega\, f(x_k^\omega) \tag{4}$$

with fixed random variables x_k^ω and a_k^ω.

The simplest Monte Carlo method is well-known: the weights $a_k = 1/n$ are equal and do not depend on ω, while the knots x_k are independent and uniformly distributed in $[0,1]^d$. The error of this method is of order $n^{-1/2}$ for many functions, including all continuous functions. There are faster methods, if we assume certain smoothness properties of the integrands, see [10], [46], and [102]. Therefore this crude Monte Carlo method can only be recommended if we do not have any such properties.

Almost optimal randomized quadrature formulas are known for classical function spaces such as Sobolev or Hölder classes. They can be found by a combination of stochastic elements with classical quadrature formulas, based on piecewise polynomial interpolation. This technique is described and further developed in [50]. For many classes we achieve a better order of convergence using randomized methods, but in all these classical examples it is enough to take linear methods of the form (4). The following result is from [3] and [71].

Theorem 4. *Consider the class*

$$F = \{f \in C^k([0,1]^d) \mid \|f^{(\alpha)}\|_\infty \le 1 \text{ for all derivatives of order } |\alpha| = k\}.$$

For deterministic methods the optimal order of convergence is given by

$$\inf_{S_n} \Delta_{\max}(S_n) \asymp n^{-k/d},$$

while randomized methods have optimal order

$$\inf_{S_n^\omega} \Delta_{\max}(S_n^\omega) \asymp n^{-1/2-k/d}.$$

In both cases we allow nonadaptive as well as adaptive methods, but the optimal order is achieved by linear nonadaptive methods.

Open Problem. There are many linear problems where we know, as in Theorem 4, that adaption does not help for randomized methods. It would be interesting to know whether a general result, such as Theorem 2, is true for randomized methods as well.

3. Linear problems on arbitrary convex sets. In this section we still assume that $S : X \to G$ is a linear operator.

In Section 2 we studied the worst case error on symmetric convex sets $F \subset X$. We already observed that this means that $f \in F$ is used as our a priori knowledge about the specific problem. One may want to use as much a priori information as possible. This is important, in particular, if the problem is difficult to solve. Consider, for example, integral equations of the first kind. Here we want to compute $S(f) = u$ given by

$$\int_a^b k(x,y)u(y)\,dy = f(x)$$

with known kernel function k. These equations are important for many applications and are usually ill-posed, i.e., the solution does not depend continuously on the right side f. Integral equations of the first kind and other ill-posed problems are studied in [6], [22], [30], [40], [45], [48], [61], [122], [123], and [133]. In many applications we know that u is a nonnegative (density) function. In other applications we know that f and/or u are monotone or convex. This yields further knowledge about f. It may happen that a problem can be solved with this geometric information but cannot be solved reasonably without it. Hence one might study classes F, such as

$$F = \{f \in C^r[a,b] \mid \|f^{(k)}\|_\infty \le 1,\ f^{(r)} \ge 0\}, \quad k \le r. \tag{5}$$

A class of the form (5) is still convex, but not symmetric. Up to now, knowledge such as "$S(f)$ is monotone (or convex)" has been used for the construction of a good φ to deal with given information N, see [36], [40], [101], and [123] for examples. However, it has usually not been used for the construction of (almost) optimal information N.

We want to study the question whether adaption helps for linear problems defined on a convex but nonsymmetric set F. The first result is due to Kiefer [57] and concerns the problem of numerical integration of monotone functions.[1]

THEOREM 5. *Assume that $S : F_{\mathrm{mon}} \to \mathbf{R}$ is the integration problem*

$$S(f) = \int_0^1 f(x)\,dx$$

[1] Of course there are also important problems where F is not convex. Examples are classes of functions with certain singularities, see [51], [129], and [134]. For nonconvex sets F the advantage of adaptive methods can be very large, even exponential, but there seems to be no general theory. We will see, however, that such a theory is possible in the convex case.

on the class

$$F_{\text{mon}} = \{f : [0,1] \to \mathbf{R} \mid f \textit{ monotone},\ f(0) = 0,\ f(1) = 1\}.$$

Then the nonadaptive trapezoidal rule

$$S_n^*(f) = \frac{1}{2n+2} + \frac{1}{n+1} \sum_{k=1}^{n} f\left(\frac{k}{n+1}\right)$$

is optimal in the class of all adaptive methods.

Proof. Let S_n be any (adaptive or nonadaptive) method that uses n knots. Let $x_1^0, \ldots, x_n^0$ be the knots that are used by S_n for the function $id \in F_{\text{mon}}$, where $id(x) = x$. Here we assume that $x_k^0 \leq x_{k+1}^0$ and we also put $x_0^0 = 0$ and $x_{n+1}^0 = 1$. It is clear that

$$S_n(f) = S_n(id)$$

for all $f \in F_{\text{mon}}$ with

$$N^0(f) := (f(x_1^0), \ldots, f(x_n^0)) = (x_1^0, \ldots, x_n^0) =: y^0.$$

As in the proof of Theorem 1 we obtain

$$\Delta_{\max}(S_n) \geq \frac{1}{2}\left(\sup_{f \in A} S(f) - \inf_{f \in A} S(f)\right) = \frac{1}{2} \sum_{k=0}^{n} (x_{k+1}^0 - x_k^0)^2,$$

where $A = \{f \in F_{\text{mon}} \mid N^0(f) = y^0\}$. It follows that

$$\Delta_{\max}(S_n) \geq \frac{1}{2n+2}.$$

This estimate is optimal since the error of the (nonadaptive) method S_n^* equals $1/(2n+2)$. □

Remarks. a) The proof of Theorem 5 is very similar to that of Theorem 1 since a “worst function” (which is always 0 in the symmetric case) can be identified. There are other classes that can be studied in the same way. Consider, for example, the class

$$F = \{f \in C^1[0,1] \mid \|f'\|_\infty \leq 1,\ f \text{ convex}\}. \tag{6}$$

Then adaption does not help for the problem of numerical integration if we study the worst case error for deterministic methods, see [74]. This time the worst function is given by $f(x) = (x - 1/2)^2$. The optimal method is

$$S_n^*(f) = \frac{1}{8n^2} + \sum_{k=1}^{n} f\left(\frac{2k-1}{2n}\right)$$

with $\Delta_{\max}(S_n^*) = 1/(8n^2)$. Compared to Theorem 5 the proof is more difficult because there is no smallest convex function with some given function values. Equidistant knots are optimal for F_{mon} and also for the class F given by (6). The class

$$F_{\mathrm{con}} = \{f \in C[0,1] \mid \|f\|_\infty \le 1,\ f \text{ convex}\}$$

probably is more interesting. Equidistant knots are far from being optimal for F_{con} and a good quadrature formula uses more knots near the endpoints of the interval. The optimal order of convergence is n^{-2} for both adaptive and nonadaptive methods, see [14] and [78].

b) Up to now we did not present an example where adaptive methods are much better than nonadaptive methods. Such examples were studied in [60], [74], [76], [99], [107], [108], and [109]; we present two of them here.

Example 1. Consider the recovery problem $S = id : F \to L_\infty([0,1])$ for the class

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

with $0 < \alpha < 1$, using function evaluations. This problem can be solved adaptively using the bisection method, while nonadaptive methods are worse. Indeed, Korneichuk [60] proved that

$$\inf_{S_n^{\mathrm{non}}} \Delta_{\max}(S_n^{\mathrm{non}}) \asymp n^{-\alpha} \qquad \text{while} \qquad \inf_{S_n^{\mathrm{ad}}} \Delta_{\max}(S_n^{\mathrm{ad}}) \asymp n^{-1}\log n.$$

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

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

Assume that we want to recover $x \in F$, i.e., $S = id$. We measure the error in the l_∞-norm and (adaptively or nonadaptively) use arbitrary linear functionals as information. One can use deep results of Kashin on the Gelfand-numbers of octohedra, see [89], to prove a lower bound for the error of optimal non-adaptive methods. There exists a positive constant c such that, see [76],

$$\inf_{S_n^{\text{non}}} \Delta_{\max}(S_n^{\text{non}}) \geq \frac{c}{\sqrt{n}\log n}, \qquad \forall\, n.$$

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

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

and

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

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

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

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

$$x_{j_{k+1}} \leq x_{j_k}.$$

We consider the following adaptive information

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

Assume that l is any coordinate not contained in $\{l_1, l_2, \dots, l_{2k-1}\}$. Then one can see that $x_l \leq x_j$ for at least k distinct values of j, and hence

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

Thus for $n = 2m-1$, we get $x_l \leq 1/(m+1) = 2/(n+3)$ for all l different from $l_1, \dots, l_n$. Hence we can recover x from the information N_n with error at most $1/(2m+2) = 1/(n+3)$, i.e., we have a method with

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

So adaptive methods are much better than nonadaptive ones.

Consider now the same problem $S = id : F \to l_\infty$, but with the condition that we only allow methods that use adaptive or nonadaptive "function evaluations", i.e., $L_k(x) = x_{i_k}$. We already constructed an adaptive method with an error bounded by $1/(n+3)$. It is not difficult to see that the nonadaptive information

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

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

$$\inf_{S_n^{\mathrm{non}}} \Delta_{\max}(S_n^{\mathrm{non}}) \asymp \frac{1}{\log(n+1)}.$$

This example shows that adaptive methods can be exponentially better than nonadaptive ones. To guarantee an error of 10^{-3} we need about 1000 function evaluations in the adaptive case and about 2^{1000} in the nonadaptive case. □

Can adaptive methods be arbitrarily better than nonadaptive methods or is there a bound on how much they can be better? This problem was studied in [77]. One can use new inequalities between Gelfand widths and Bernstein widths, and relations between these widths and optimal error bounds for adaptive and nonadaptive methods, respectively. Inequalities between different widths are known in the symmetric case, see [64], [66], [88] and [89], but nonsymmetric sets have not been studied much in approximation theory.

We assume that $S : X \to G$ is a continuous linear mapping that is to be approximated on a convex set $F \subset X$ by a method of the form $\varphi \circ N$, where

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

with arbitrary continuous linear functionals $L_k : X \to \mathbf{R}$. The following result is proved in [77].

Theorem 6.

$$\inf_{S_n^{\mathrm{non}}} \Delta_{\max}(S_n^{\mathrm{non}}) \le 4(n+1)^2 \inf_{S_n^{\mathrm{ad}}} \Delta_{\max}(S_n^{\mathrm{ad}}).$$

This means that there is a universal bound on how much adaptive methods can be better for linear problems on convex sets. The practical implication depends on the speed of convergence of the sequence

$$\alpha_n := \inf_{S_n^{\mathrm{non}}} \Delta_{\max}(S_n^{\mathrm{non}}).$$

If $(\alpha_n)_n$ tends to zero very fast then the improvement by adaptive methods is relatively small. If, however, $(\alpha_n)_n$ tends to zero as n^{-3} or even slower than the improvement by adaption can be huge.

Average case analysis. We saw in Theorem 5 that adaption does not help for the integration of monotone functions. This is a worst case result for deterministic methods.

One can hope that adaptive methods are much better than nonadaptive methods on the average. Assume, for example, that we know $f \in F_{\text{mon}}$ and $f(1/2) = \gamma$. If $\gamma < 1/2$ then the value of $f(3/4)$ gives more information about the integral of f than $f(1/4)$. If, however, $\gamma > 1/2$ then the value of $f(1/4)$ gives more information and it might be better to compute $f(1/4)$ next.

The average case can be studied with the Dubins-Freedman or Ulam measure P on F_{mon}, see [41] for the construction and detailed analysis. The function value $f(1/2)$ is uniformly distributed on $[0,1]$, and a P-random function can be constructed on the dyadic rationals inductively. First put $f(0) = 0$ and $f(1) = 1$. Knowing $f(i/2^k)$ for $i = 0, \dots, 2^k$, choose $f((2i+1)/2^{k+1})$ for $i = 0, \dots, 2^k - 1$ independently according to uniform distributions on the intervals $[f(i/2^k), f((i+1)/2^k)]$.

Nonadaptive quadrature formulas are studied in [42], adaptive methods in [73]. It is proved that the trapezoidal rule has optimal order in the class of nonadaptive methods while there is a "greedy" adaptive method that is much better on the average.

THEOREM 7. *For nonadaptive quadrature formulas using n function values the optimal order is*

$$\inf_{S_n} \Delta_{\text{aver}}(S_n) \asymp n^{-\log 6/\log 4} \asymp n^{-1.29248\dots}.$$

There is an adaptive method with

$$\Delta_{\text{aver}}(S_n^{\text{ad}}) \le 5 \cdot n^{-3/2}.$$

Randomized methods. We compare worst case results for deterministic and randomized methods. The following result, from [73] and [78], says that "adaption may significantly help for convex classes and randomized methods".

THEOREM 8. *Consider the problem of numerical integration on F_{mon} and F_{con}. The optimal order of convergence is n^{-1} for F_{mon} and adaptive or* *nonadaptive deterministic methods. It is also n^{-1} for nonadaptive randomized methods but $n^{-3/2}$ for adaptive randomized methods.*

The optimal order of convergence is n^{-2} for F_{con} and adaptive or nonadaptive deterministic methods. It is also n^{-2} for nonadaptive randomized methods but $n^{-5/2}$ for adaptive randomized methods.

Remarks. a) Analogous classes of monotone and convex functions in the multivariate case are studied in [56] and [86]. Tight lower bounds for nonadaptive Monte Carlo methods are not known in the multivariate case.

b) Theorems 7 and 8 are interesting for the following reason: Classical error bounds in numerical analysis are worst case error bounds for deterministic methods. With respect to such error bounds adaptive quadrature methods are not much better than nonadaptive ones for F_{mon} or F_{con}. Nevertheless there is an advantage of adaptive methods. This advantage of adaptive methods can only be proved via stochastic arguments, however. This means that stochastic error bounds can be used to identify new efficient methods that could not be found by the classical approach, i.e., by a worst case analysis of deterministic methods.

The logarithmic error. Up to now we always considered the *absolute error criterion*, given, in the worst case setting, by (2). This is the error criterion used in most theoretical studies of numerical analysis. In many applications the relative or logarithmic error is more relevant. See [119] for results concerning the relative error in the symmetric case. The integration problem $S(f) = \int_0^1 f(x)\,dx$ was studied for classes of the type

$$F_{\text{mon}}^{\delta} = \{f : [0,1] \to \mathbf{R} \mid f \text{ monotone},\ f(0) = 0,\ f(1) = 1,\ S(f) \geq \delta\}$$

in [84]. One can define the *logarithmic error criterion* in the worst case setting by

$$\Delta_{\log}(S_n) = \sup_{f \in F} |\log S(f) - \log S_n(f)|. \tag{7}$$

Let

$$a(y) = \inf\{S(f) \mid f \in F,\ N(f) = y\} > 0$$

and

$$b(y) = \sup\{S(f) \mid f \in F,\ N(f) = y\}.$$

Given information $N(f) = y$, these numbers are easy to compute. With respect to the logarithmic error, the optimal estimate of $S(f)$ is then the geometric mean $\varphi(y) = \sqrt{a(y)b(y)}$, with error

$$\Delta_{\log} = \frac{\log b - \log a}{2}.$$

It turns out that again nonadaptive methods are optimal. The optimal knots are not equidistant but are in a "geometrical order". One of the results of [84] is as follows.

THEOREM 9. *Consider the integration problem on $F^\delta_{\rm mon}$ with $0 < \delta < 1$. Adaption does not help (for deterministic methods) with respect to the logarithmic error* (7). *Assume that $\delta \neq n/(2n+2)$. Then the optimal nonadaptive knots $x_1, \dots, x_n$ are given by $1 - x_n = \alpha$ and*

$$x_{k+1} - x_k = \alpha \beta^{n-k}$$

with

$$\delta = \frac{n\beta^{n+1} - (n+1)\beta^n + 1}{(\beta^{n+1} - 1)^2}$$

and $\alpha = (\beta - 1)/(\beta^{n+1} - 1)$. In the case $\delta = n/(2n+2)$ the optimal knots are equidistant. For small δ and large n we have

$$\beta \approx 1 - \frac{\log \delta}{n}, \qquad \Delta_{\log}(S_n^*) \approx -\frac{\log \delta}{2n}.$$

Remarks. a) For the optimal method with equidistant knots we only have

$$\Delta_{\log}(S_n) \approx \frac{1}{2n\delta}$$

and therefore the optimal knots are much better than equidistant ones if δ is small.

b) In [84] classes of unimodal functions are also studied. Assume that F contains all $f : [0,1] \to [0,1]$ that are unimodal with $S(f) \geq \delta$. This means that there is an $x^* \in [0,1]$ such that f is nondecreasing in $[0, x^*]$ and nonincreasing in $[x^*, 1]$. This property of f cannot be used efficiently by an adaptive method. Indeed, it turns out that nonadaptive methods based on equidistant knots are almost optimal in the class of all adaptive methods. This is true for the absolute error and for the logarithmic error.

If we consider only those functions that are strictly unimodal, in the sense that they are (strictly) increasing in an interval $[0, x^*]$ and (strictly) decreasing in $[x^*, 1]$, then the results are quite different: Adaption still does not help for the absolute error criterion while adaptive methods are much better than nonadaptive ones for the logarithmic error. In [84] an algorithm is presented that estimates the integral for any positive function f that is unimodal in the strict sense. The method uses Fibonacci-search to approximate the maximum of the function and then uses a "greedy strategy" to determine further knots that lead to an approximation of the integral.

These results show that a seemingly small change in the class F may lead to quite different results. Also the error criterion can be crucial for the adaption problem.

4. Global Optimization. Error bounds for the nonlinear problem of global optimization are closely related to those for the linear problem of optimal recovery in the L_∞-norm. Let F be a symmetric convex set of bounded functions defined on a set Ω. Then we consider methods based on function evaluations for the linear problem $id : F \to B(\Omega)$. The maximal error of an approximation S_n is given by

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

From the basic results of Smolyak and Bakhvalov, see Theorem 1, we know that adaption does not help and that linear methods are optimal. Hence we have the error bound

$$\begin{aligned}\inf_{S_n} \Delta_{\max}(S_n) &= \inf_{x_k} \inf_{g_k \in B(\Omega)} \sup_{f \in F} \|f - \sum_{k=1}^{n} f(x_k) g_k\|_\infty \\ &= \inf_{x_k} \sup\{\|f\|_\infty \mid f \in F,\ f(x_k) = 0,\ k = 1, \ldots, n\} =: \alpha_n(F).\end{aligned}$$

Now we consider the problem of global optimization, defined by $S(f) = \inf f$. Again we are interested in bounds for the error defined by

$$\Delta_{\max}(S_n) = \sup_{f \in F} |\inf f - S_n(f)|.$$

The following result is from [124], a proof can be found also in [70, 1.3.3].

THEOREM 10. *Adaption can help for the problem of global optimization by at most a factor of 2; the error bounds*

$$\tfrac{1}{2}\, \alpha_n(F) \le \inf_{S_n} \Delta_{\max}(S_n) \le \alpha_n(F)$$

are valid for nonadaptive and also for adaptive methods for any symmetric convex $F \subset B(\Omega)$.

Optimal error bounds for global optimization and for optimal recovery are almost equal in the worst case setting if $F \subset B(\Omega)$ is convex and symmetric. Under the same conditions adaptive methods are at most slightly better than nonadaptive methods. The frequent use of adaptive methods for the problem of global optimization cannot be justified by a worst case analysis of deterministic methods.

The sequence $(\alpha_n(F))_n$ provides a bound for the (worst case) complexity of finding the global minimum of $f \in F$. Any method which yields an error of at most $\alpha_n(F)$ for all functions $f \in F$ must use at least n function values for some functions $f \in F$. The following result is well known, see [69], [70], and [119] for this and more general results. By $D^r f$ with $r = (r_1, \ldots, r_d)$ we denote a partial derivative and we put $|r| = \sum r_i$, as usual.

THEOREM 11. *Let* F *be a Hölder class of the form*

$$F = \{f : [0,1]^d \to \mathbf{R} \mid |D^r f(x) - D^r f(y)| \le \|x-y\|^\alpha \text{ for all } r \text{ with } |r| \le k\},$$

where $0 < \alpha \le 1$ *and* $k \in \mathbf{N}_0$. *For the problem of global optimization there is a constant* $c > 0$ *such that*

$$\inf_{S_n} \Delta_{\max}(S_n) \ge c \cdot n^{-(k+\alpha)/d}$$

for any $n \in \mathbf{N}$. *This lower bound is sharp in the order sense and is achieved by a nonadaptive algorithm using function values from a regular grid.*

Remarks. This is a negative result. No clever adaptive algorithm can be better than a nonadaptive method that is based on a regular grid. We stress that the lower bound is still valid if we allow information consisting of derivatives of order at most k. For Lipschitz optimization we have $k = 0$ and $\alpha = 1$. Hence this problem is very difficult even for moderate values of d. In this and some other cases it is possible to specify the optimal constant c. It is known that the complexity result of Theorem 11 is also true if we allow randomized methods, see [70] or [126] for details. This may be surprising because stochastic search methods are often used in practice. How can we avoid this negative result?

First, we can look for "smaller" classes F that allow smaller errors and still contain many interesting functions. We often deal with "nonisotropic" situations where some directions in $\mathbf{R}^d$ are more important than others. Many interesting functions are partially separable, i.e., they can be written as the sum of functions depending on only a few of the variables, see [21] and [116]. Assume, for instance, that f is of the form

$$f(x) = \sum_{i,j=1}^{d} f_{i,j}(x_i, x_j), \tag{8}$$

i.e., f is the sum of functions $f_{i,j}$ depending only on two variables. Let us also assume that $f_{i,j} \in C^2([0,1]^2)$ for all pairs (i,j). By classical results, such as Theorem 11, we can use methods based on a regular grid that guarantee an error of order $n^{-2/d}$. If grid points are used then the order $n^{-2/d}$ is optimal also on the smaller class of functions of the form (8). However, a function of the form (8) is in $C^2([0,1]^d)$ but, in addition, all derivatives of the form $D^r f$ with $r_i \le 1$ for every i exist. We define $W_\infty^{(1,\dots,1)}([0,1]^d)$ by

$$W_\infty^{(1,\dots,1)}([0,1]^d) = \{ f : [0,1]^d \to \mathbf{R} \mid \|D^r f\|_\infty \le 1 \text{ for all } r \text{ with } r_i \le 1 \}.$$

Using results of Temlyakov [113], see also [132] and [136], one can show that there are nonadaptive methods for $W_\infty^{(1,\dots,1)}([0,1]^d)$ with an error bound

$$\Delta_{\max}(S_n) \le c \cdot n^{-1} (\log n)^{2(d-1)}.$$

Now the order of convergence depends only weakly on d. However, this bound cannot be achieved using grid points – grid points only yield the poor bound $n^{-1/d}$. Instead one can use methods based on hyperbolic cross points, see Example 1b of Section 2 and [81].

Second, one can try to define efficient adaptive algorithms that are distinguished not by their worst case performance, since we know from Theorem 10 that adaption can not help in the worst case, but by some other criterion. For instance, Sukharev introduced the concept of sequentially optimal algorithms as algorithms making in each step the best use (with respect to the worst-case information in the following steps) of the information given by $f(x_1), \dots, f(x_{k-1})$. A survey of this concept can be found in [111] or [112]. This approach leads to a new and promising class of algorithms.

Thirdly, instead of a worst case analysis, we can opt for an average case analysis in the hope of finding a method that is good for "most" $f \in F$. To compute the average error one can take the classical Wiener measure on $C([0,1])$ for $d = 1$ and the Wiener sheet measure if $d > 1$. Error bounds are only known for the case $d = 1$, see [94] and [127]. The average error of optimal nonadaptive methods is of order $n^{-1/2}$. Recently Calvin [16] constructed adaptive Monte Carlo methods which yield errors of order $n^{-(1-\delta)}$ for any $\delta > 0$. Thus adaption turns out to be very powerful in the average case setting. The order $n^{-(1-\delta)}$ can also be obtained by a deterministic adaptive method, since it is known that Monte Carlo algorithms cannot be superior to deterministic algorithms in an average case sense, see [70, p. 67]. We stress, however, that such deterministic methods are not known.

There exist several other algorithms that are based on the Bayesian approach, see [9], [67], [117], and [137]. One can define, for example, "greedy methods", i.e., methods that are optimal "in one step". Average error bounds for these methods are, however, yet to be found.

5. Zero Finding. We survey recent results for the solution of the nonlinear equation $f(x) = 0$ of one variable, see also [80]. Zero finding is a classical problem of numerical analysis, and most of the results deal with the asymptotic setting, see [118]. The order of convergence and the efficiency or complexity index are studied, often under the assumption that a good initial approximation to a root is given.

In contrast to the asymptotic setting, error bounds which hold for any $f \in F$ after a fixed number of steps are studied in the worst case setting, see [103]. We also consider the average case setting, where the expected error and cost with respect to a probability measure on F are investigated. In this setting we also study methods with an *adaptive stopping rule*. Actually we will see that methods with such a varying cardinality are much better than methods with a fixed number of knots.

It is well known in practice that methods which usually work fast sometimes fail for specific hard functions f. This is confirmed by a comparison of results in three settings: asymptotic, worst case, and average case. We give such a comparison for the typical class

$$F = \{f \in C^2([0,1]) \mid f(0) < 0 < f(1),\ \ f'(x^*) \neq 0 \text{ if } f(x^*) = 0\}.$$

The order of the Newton method is 2. This means that the error converges quadratically for $f \in F$ if a good starting point x_0 is chosen. It is known that the order 2 can even be achieved by a globally convergent method which uses only function values.

We study methods of the form (1), based only on the knowledge $f \in F$ and the knowledge that is given by the computed values. We do not assume additional knowledge such as "a good starting point for the Newton method". To measure the error of S_n we use the root criterion

$$d(f^{-1}(0), S_n(f)) = \inf\{|x - S_n(f)| \mid f(x) = 0\}$$

and the worst case error

$$\Delta_{\max}(S_n) = \sup_{f \in F} d(f^{-1}(0), S_n(f)).$$

We recall the bisection method S_n^b. Let $a_0 = 0$ and $b_0 = 1$. We set $x_k = (a_{k-1} + b_{k-1})/2$ for $k \in \mathbf{N}$ and

$$a_k = x_k,\ b_k = b_{k-1} \text{ if } f(x_k) \le 0 \quad \text{and} \quad a_k = a_{k-1},\ b_k = x_k \text{ if } f(x_k) > 0.$$

Further we define

$$S_n^b(f) = x_{n+1}(f).$$

It is well known that the worst case error of the bisection method S_n^b satisfies

$$\Delta_{\max}(S_n^b) = 2^{-n-1}.$$

This holds for F and for many other classes. The bisection method is optimal with respect to the worst case error for many classes. The optimal worst case error bounds do not depend on the degree of smoothness, bisection is optimal even for the class of C^∞-functions with a simple zero. Bisection is not optimal, however, if we know an upper bound on $\|f''\|_\infty$ and a positive lower bound on $|f'(x^*)|$ for each zero x^*, see [63].

The bisection method clearly is an adaptive method. The optimal nonadaptive method S_n^* uses equidistant knots, the error is

$$\Delta_{\max}(S_n^*) = \frac{1}{2n+2}.$$

Adaptive methods are much better than nonadaptive methods for this nonlinear problem.

Now we consider average case results. So far Gaussian measures have been used in most of the papers dealing with the average case on infinite dimensional function spaces. For zero finding Gaussian measures and Ulam measures on classes of monotone functions have been used. It is known that the bisection method S_n^b is not optimal in the class of methods with fixed cardinality n but it is almost optimal: A lower bound of the form

$$\int_F d(f^{-1}(0), S_n(f))\, dP(f) \ge \alpha^n, \qquad \text{with } \alpha \in (0, 1/2) \tag{9}$$

holds for all such methods, see [43], [72], [79], and [95].

In the remaining part of this section we discuss results for suitable Gaussian measures on a class

$$F_r = \{f \in C^r[0,1] \mid f(0) < 0 < f(1)\},$$

where $r \in \mathbf{N}_0$. We study zero finding methods $\widetilde{S}$ which are based on function values $f(x_i)$ or derivatives $f^{(k_i)}(x_i)$ at adaptively chosen knots x_i. Of course we assume that $k_i \leq r$. The Gaussian measures which have been used to analyze zero finding are derived from the Wiener measure by r-fold integration and translation by suitable polynomials to match boundary conditions at the endpoints 0 and 1. The following results are from [83] and [95], the Brownian bridge ($r = 0$) was studied earlier in [72] and [79].

A lower bound of the form (9) also holds for the class F_r. Therefore the number of bisection steps, necessary to guarantee a worst case error ε, differs from the number of function evaluations, necessary to obtain an average error ε, at most by a multiplicative constant. This is remarkable for $r \geq 2$ since there are iterative methods which converge superlinearily for all $f \in F$ and, therefore, for almost all $f \in F_r$.

Are *adaptive stopping rules* better than nonadaptive ones? A worst case analysis cannot justify the use of methods with varying cardinality. Due to [125] varying cardinality also does not help much for many linear problems, if the average error and the average cost are defined with respect to a Gaussian measure; but see also [91]. For zero finding, however, varying cardinality is very powerful on the average. In practice the number of evaluations is often determined by an adaptive stopping rule: We stop a method in $n(f)$-th step if

$$d(f^{-1}(0), S_{n(f)}(f)) \leq \varepsilon$$

can be guaranteed.

We present a special algorithm S_ε from [83] with $\Delta_{\max}(S_\varepsilon) \leq \varepsilon$ for $r \geq 0$. This strong error requirement implies that only enclosing methods can be used. See [1] for an analysis of enclosing methods in the asymptotic setting. Numerical results for the method S_ε are presented in [77a].

$i := 0$;
$[s_0, t_0] := [0, 1]$; compute $f(0),\ f(1)$;
do
  for $j := 1, 2$ do (regula falsi)

$i := i + 1;$
$x_i :=$ SEC(s_{i-1}, t_{i-1}); compute $f(x_i), [s_i, t_i]$;
od;
repeat (secant)
$q :=$ SEC(x_i, x_{i-1});
if $q \notin \,]s_i, t_i[$ then break fi;
$i := i + 1;$
$x_i := q$; compute $f(x_i)$, $[s_i, t_i]$;
until $t_i - s_i > (t_{i-3} - s_{i-3})/2$;
$i := i + 1;$ (bisection)
$x_i := (s_i + t_i)/2$; compute $f(x_i)$, $[s_i, t_i]$;
od;

After each computation of a function value $f(x_i)$ we compute the new enclosing interval $[s_i, t_i]$. By SEC(s, t) we denote the "secant point"

$$\mathrm{SEC}(s,t) := \frac{f(t) \cdot s - f(s) \cdot t}{f(t) - f(s)}.$$

This point is only used if it is well-defined and within the current interval $[s_i, t_i]$ with a change of sign. We stop if $t_i - s_i \leq 2\varepsilon$. In that case we return $(s_i + t_i)/2$.

We mention some properties of the method S_ε that are also important for the proof of average case results. The method uses steps of the regula falsi (R), the secant method (S), and the bisection method (B). A typical pattern is R R S ... S B R R S B R R S S S S S S ... We always have a length reduction $t_i - s_i \leq (t_{i-4} - s_{i-4})/2$ of the interval with a guaranteed zero. The error of the method is bounded by ε for each $f \in F_0$. The computational cost is proportional to the number of function evaluations.

The following result holds for r-fold Wiener measures and $r \geq 2$. The constant c in (10) depends on the measure P, the proof in [83] gives a suitable c that can be computed numerically.

THEOREM 12. *Let $0 < \varepsilon < 1/2$. For the hybrid secant-bisection method S_ε,*

$$\int_{F_r} n(f)\, dP(f) \leq \frac{1}{\log \beta} \cdot \log\log(1/\varepsilon) + c, \tag{10}$$

where $\beta = \frac{1+\sqrt{5}}{2}$.

We only sketch the basic idea of the rather technical and long proof. For $\gamma_i > 0$ we define a subset G of F_r by

$$G = \{g \in F_r \mid |g''(x) - g''(y)| \leq \gamma_1 \, |x-y|^{1/3}, \\ |g''(z)| \geq \gamma_2, \, |g'(z)| \geq \gamma_3 \text{ if } g(z) = 0\}.$$

Let $g \in G$. After $k = k(G)$ steps we can guarantee the following: If there is a further bisection step then this is the last one and we obtain the pattern B R R S S S S ... It is possible to give explicit estimates of $k(G)$. It is also possible to estimate the probability of G and so, finally, the result can be proved.

One should note that such an average case analysis includes a worst case analysis for many different subsets G of F_r. It is crucial, of course, that a membership $f \in G$ can not be checked by any algorithm that is based on finite information. Hence the constants γ_i cannot be used in a program.

The algorithm S_ε is almost optimal in a strong sense. Actually one can prove a lower bound for very general methods: instead of function evaluation at any knot we also allow the evaluation of derivatives; instead of enclosing methods with a guaranteed error ε we only require that the average error is bounded by ε. Nevertheless, the average number of knots of any such algorithm cannot be much smaller than the average number of knots of our algorithm S_ε. Again we consider r-fold Wiener measures with $r \geq 2$.

THEOREM 13. *Let $\varepsilon > 0$ and assume that $\widetilde{S}$ is a method with*

$$\int_{F_r} d(f^{-1}(0), \widetilde{S}(f)) \, dP(f) \leq \varepsilon.$$

Then

$$\int_{F_r} n(f) \, dP(f) \geq \frac{1}{\log \alpha} \cdot \log\log(1/\varepsilon) + c_\alpha,$$

for any α with $\alpha > r + 1/2$.

The constant c_α depends on the measure, but is of course independent of $\widetilde{S}$. We believe that the bound $r+1/2$ on α is optimal. Together with Theorem 12 we obtain that the method S_ε is almost optimal if $r \geq 2$. Possibly, this particular method is also optimal for $r = 0$ and $r = 1$ but this is not known so far.

The stopping rule "$t_i - s_i \leq 2\varepsilon$" is adaptive since the number $n(f)$ of function evaluations depends on $f \in F_r$. We have seen that such an adaptive stopping rule is crucial.

Note the huge difference between the orders $\log(1/\varepsilon)$ and $\log\log(1/\varepsilon)$ of the worst and average case complexity. This difference is due to the fact that we switch from maximal $n = n(f)$ to average $n(f)$, the difference between worst case errors and average case errors turns out to be insignificant for this particular problem.

Acknowledgments. The author would like to express his sincere thanks to Klaus Ritter, Joe Traub, Greg Wasilkowski, Art Werschulz, and Henryk Woźniakowski for many helpful discussions on adaption and this paper.

REFERENCES

[1] G. Alefeld and F. A. Potra, *On enclosing simple roots of nonlinear equations*, Math. of Comp., 61 (1993), pp. 733–744.

[2] I. Babuška and W. C. Rheinboldt, *Error estimates for adaptive finite element computations*, SIAM J. Numer. Anal., 15 (1978), pp. 736–754.

[3] N. S. Bakhvalov, *On approximate computation of integrals*, Vestnik MGU, Ser. Math. Mech. Astron. Phys. Chem., 4 (1959), pp. 3–18. [In Russian]

[4] N. S. Bakhvalov, *On the convergence of a relaxation method with natural constraints on the elliptic operator*, USSR Comput. Maths. Math. Phys., 6(5) (1966), pp. 101–135.

[5] N. S. Bakhvalov, *On the optimality of linear methods for operator approximation in convex classes of functions*, USSR Comput. Maths. Math. Phys., 11 (1971), pp. 244–249.

[6] J. Baumeister, *Stable Solution of Inverse Problems*, Vieweg, Braunschweig, 1987.

[7] M. Š. Birman and M. Z. Solomjak, *Piecewise-polynomial approximations of functions of the class* W_p^α, Math. USSR Sbornik, 2 (1967), pp. 295–317.

[8] L. Blum, M. Shub, and S. Smale, *On a theory of computation and complexity over the real numbers: NP completeness, recursive functions and universal machines*, Bull. Amer. Math. Soc., 21 (1989), pp. 1–46.

[9] C. G. E. Boender and H. E. Romeijn, *Stochastic methods*, In: Handbook of Global Optimization, R. Horst, P. M. Pardalos, eds., Kluwer, Dordrecht, 1995, pp. 829–869.

[10] K. Bogues, C. R. Morrow, and T. N. L. Patterson, *An implementation of the method of Ermakov and Zolotukhin for multidimensional integration and interpolation*, Numer. Math., 37 (1981), pp. 49–60.

[11] K. H. Borgwardt, *The Simplex Method*, Springer-Verlag, Berlin, 1987.

[12] D. Braess, *Nonlinear Approximation Theory*, Springer-Verlag, Berlin, 1986.

[13] J. H. Bramble, *On the development of multigrid methods and their analysis*, Proc. of Symp. Appl. Math., 48 (1994), pp. 5–19.

[14] H. Brass, *Zur Quadraturtheorie konvexer Funktionen*, in Numerical Integration, G. Hämmerlin, ed., ISNM 57, Birkhäuser, 1982, pp. 34–47.

[15] Yu. A. Brudnyi, *Adaptive approximation of functions with singularities*, Trans. Moscow Math. Soc. 55 (1994), pp. 123–186.

[16] J. M. Calvin, *Average performance of a class of adaptive algorithms for global optimization*, preprint, Georgia Institute of Technology, 1995.

[17] S. Cambanis, *Sampling designs for time series*, In: Time series in time domain, Handbook of Statistics 5, E. J. Hannan, P. R. Krishnaiah, and M. M. Rao, eds., North-Holland, Amsterdam, 1985, pp. 337–362.

[18] M. Chu, *There exists a problem whose computational complexity is any given function of the information complexity*, J. Complexity, 10 (1994), pp. 445–450.

[19] O. R. Chugan and A. G. Sukharev, *On adaptive and nonadaptive stochastic and deterministic algorithms*, J. Complexity, 6 (1990), pp. 119–127.

[20] P. G. Ciarlet, *The Finite Element Method for Elliptic Problems*, North-Holland, Amsterdam, 1978.

[21] A. R. Conn, N. I. M. Gould, Ph. L. Toint, *Lancelot, a Fortran Package for Large-Scale Nonlinear Optimization*, Springer-Verlag, Berlin, 1992.

[22] I. Csiszár, *Why least squares and maximum entropy? An axiomatic approach to inference for linear inverse problems*, Annals of Statistics, 19 (1991), pp. 2032–2066.

[23] P. J. Davis and P. Rabinowitz, *Methods of Numerical Integration*, Second edition, Academic Press, Orlando, Florida, 1984.

[24] F.-J. Delvos and W. Schempp, *Boolean Methods in Interpolation and Approximation*, Pitman Research Notes in Math. 230, Longman, Essex, 1989.

[25] R. A. DeVore, *Degree of nonlinear approximation*, Approximation Theory VI: Vol. 1, C. K. Chui, L. L. Schumaker and J. D. Ward (eds.), Academic Press, 1989, pp. 175–201.

[26] R. A. DeVore, R. Howard, and C. Micchelli, *Optimal nonlinear approximation*, Manuscripta Math., 63 (1989), pp. 469–478.

[27] R. A. DeVore, G. Kyriazis, D. Leviatan and V. M. Tikhomirov, *Wavelet compression and nonlinear n-widths*, Adv. in Comput. Math., 1 (1993), pp. 197–214.

[28] R. A. DeVore, G. G. Lorentz, *Constructive Approximation*, Springer-Verlag, Berlin, 1993.

[29] P. Diaconis, *Bayesian numerical analysis*, Statistical decision theory and related topics IV, Papers 4th Purdue Symp., West Lafayette, Indiana, 1988, Vol. 1, pp. 163–175.

[30] Nho Hào Dinh, *A mollification method for ill-posed problems*, Numer. Math., 68 (1994), pp. 469–506.

[31] D. L. Donoho, *Nonlinear wavelet methods for recovery of signals, densities, and spectra from indirect and noisy data*, Proc. of Symp. in Applied Math., 47 (1993), pp. 173–205.

[32] D. L. Donoho, I. M. Johnstone, G. Kerkyacharian, and D. Picard, *Wavelet shrinkage: asymptopia?*, J. R. Statist. Soc. B, 57 (1995), pp. 301–369.

[33] K. Eriksson, D. Estep, P. Hansbo and C. Johnson, *Introduction to adaptive methods for differential equations*, Acta Numerica 1995, Cambridge University

Press, pp. 105–158.

[34] R. P. Fedorenko, *The speed of convergence of one iterative process*, USSR Comput. Maths. Math. Phys., 4(3) (1964), pp. 227–235.

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

[36] C. Gellrich and B. Hofmann, *A study of regularization by monotonicity*, Computing, 50 (1993), pp. 105–125.

[37] A. C. Genz, *Subregion adaptive algorithms for multiple integrals*, Contemporary Mathematics 115 (1991), pp. 23–31.

[38] A. C. Genz and A. A. Malik, *An adaptive algorithm for numerical integration over an N-dimensional rectangular region*, J. of Comp. and Appl. Maths., 6 (1980), 295–302.

[39] G. K. Golubev, *Sequential design of an experiment for nonparametric estimation of smooth regression functions*, Problems Inform. Transmission, 28 (1992), pp. 265–268 and 395.

[40] R. Gorenflo and S. Vessella, *Abel Integral Equations: Analysis and Applications*, Lecture Notes in Mathematics 1461, Springer-Verlag, New York, 1991.

[41] S. Graf, R. D. Mauldin, and S. C. Williams, *Random homeomorphisms*, Advances in Math. 60 (1986), pp. 239–359.

[42] S. Graf and E. Novak *The average error of quadrature formulas for functions of bounded variation*, Rocky Mt. J. Math., 20 (1990), pp. 707–716.

[43] S. Graf, E. Novak, and A. Papageorgiou, *Bisection is not optimal on the average*, Numer. Math., 55 (1989), pp. 481–491.

[44] M. Griebel, M. Schneider, and Ch. Zenger, *A combination technique for the solution of sparse grid problems*, In Iterative Methods in Linear Algebra, R. Beauwens and P. de Groen, eds., Elsevier Science Publishers, North-Holland, 1992, pp. 263–281.

[45] C. W. Groetsch, *Inverse Problems in the Mathematical Sciences*, Vieweg, Braunschweig, 1993.

[46] S. Haber, *Numerical evaluation of multiple integrals*, SIAM Rev., 12 (1970), pp. 481–526.

[47] W. Hackbusch, *Multi-grid methods and applications*, Springer-Verlag, New York, 1985.

[48] P. C. Hansen, *Numerical tools for analysis and solution of Fredholm integral equations of the first kind*, Inverse Problems, 8 (1992), pp. 849–872.

[49] S. Heinrich, *Random approximation in numerical analysis*, In Functional Analysis, K. D. Bierstedt et al., eds., Lecture Notes in Pure and Applied Mathematics 150, Marcel Dekker, New York, 1993, pp. 123–171.

[50] S. Heinrich and P. Mathé, *The Monte Carlo complexity of Fredholm integral equations*, Math. Comp., 60 (1993), pp. 257–278.

[51] I. P. Huerta, *Adaption helps for some nonconvex classes*, J. Complexity, 4 (1986), pp. 333–352.

[52] B. Z. Kacewicz, *How to increase the order to get minimal-error algorithms for systems of ODE*, Numer. Math., 45 (1984), pp. 93–104.

[53] B. Z. Kacewicz, *On sequential and parallel solution of initial value problems*, J. Complexity, 6 (1990), pp. 136–148.

[54] D. K. Kahaner, *A survey of existing multidimensional quadrature routines*, Contemporary Mathematics 115 (1991), pp. 9–22.

[55] B. S. Kashin and V. N. Temlyakov, *On the best m-term approximation and the entropy of sets in L_1*, Mat. Sametki, 56(5) (1994), pp. 57–86. [In Russian]

[56] C. Katscher, E. Novak, and K. Petras, *Quadrature formulas for multivariate convex functions*, J. Complexity, to appear.

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

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

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

[60] N. P. Korneichuk, *Optimization of active algorithms for recovery of monotonic functions from Hölder's class*, J. Complexity, 10 (1994), pp. 265–269.

[61] A. K. Louis, *Inverse und schlecht gestellte Probleme*, Teubner, Stuttgart, 1989.

[62] Z. Luo and G. Wahba, *Hybrid adaptive splines*, Dep. of Statistics, University of Wisconsin, Madison, WI, Technical Report 947, 1995.

[63] G. D. Maistrowskii, *On the optimality of Newton's method*, Soviet Math. Dokl., 13 (1972), pp. 838–840.

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

[65] P. Mathé, *Approximation Theory of Stochastic Numerical Methods*, Habilitationsschrift, Berlin, 1994.

[66] B. S. Mityagin and G. M. Henkin, *Inequalities between n-diameters*, In Proc. of the Seminar on Func. Anal. Voronezh, 7 (1963). [In Russian]

[67] J. Mockus, *Bayesian Approach to Global Optimization, Theory and Applications*, Kluwer, Dordrecht, 1989.

[68] T. Müller-Gronbach, *Optimal designs for approximating the path of a stochastic process*, preprint, FU Berlin, 1993.

[69] A. S. Nemirovsky and D. B. Yudin, *Problem Complexity and Method Efficiency on Optimization*, John Wiley and Sons, Chicester, 1983.

[70] E. Novak, *Deterministic and Stochastic Error Bounds in Numerical Analysis*, Lecture Notes in Math. 1349 (1988), Springer-Verlag, Berlin.

[71] E. Novak, *Stochastic properties of quadrature formulas*, Numer. Math., 53 (1988), pp. 609–620.

[72] E. Novak, *Average case results for zero finding*, J. Complexity, 5 (1989), pp. 489–501.

[73] E. Novak, *Quadrature formulas for monotone functions*, Proc. of the AMS, 115 (1992), pp. 59–68.

[74] E. Novak, *Quadrature formulas for convex classes of functions*, In Numerical Integration IV, H. Braß and G. Hämmerlin, eds., ISNM 112 (1993), Birkhäuser, Basel, pp. 283–296.

[75] E. Novak, *The real number model in numerical analysis*, J. Complexity, 11 (1995), pp. 57–73.

[76] E. Novak, *Optimal recovery and n-widths for convex classes of functions*, J. Approx. Th., 80 (1995), pp. 390–408.

[77] E. Novak, *The adaption problem for nonsymmetric convex sets*, J. Approx. Th., 82 (1995), pp. 123–134.

[77a] E. Novak, *The Bayesian approach to numerical problems: results for zero finding*, In: Proc. of the IMACS-GAMM Int. Symp. on Num. Methods and Error Bounds, J. Herzberger, ed., Akademie Verlag, Berlin, 1996, pp. 164–171.

[78] E. Novak and K. Petras, *Optimal stochastic quadrature formulas for convex functions*, BIT, 34 (1994), pp. 288–294.

[79] E. Novak and K. Ritter, *Average errors for zero finding: lower bounds*, Math. Z., 211 (1992), pp. 671–686.

[80] E. Novak and K. Ritter, *Some complexity results for zero finding for univariate functions*, J. Complexity, 9 (1993), pp. 15–40.

[81] E. Novak and K. Ritter, *Global optimization using hyperbolic cross points*, In: State of the art in global optimization: computational methods and applications. C. A. Floudas, P. M. Pardalos, eds., Kluwer, Dordrecht, 1996, pp. 19–33.

[82] E. Novak and K. Ritter, *High dimensional integration of smooth functions over cubes*, preprint, Erlangen, 1995, to appear in Numer. Math.

[83] E. Novak, K. Ritter, and H. Woźniakowski, *Average case optimality of a hybrid secant-bisection method*, Math. Comp., 64 (1995), pp. 1517–1539.

[84] E. Novak and I. Roschmann, *Numerical integration of peak functions*, to appear.

[85] E. W. Packel, *Do linear problems have linear optimal algorithms?*, SIAM Rev., 30 (1988), pp. 388–403.

[86] A. Papageorgiou, *Integration of monotone functions of several variables*, J. Complexity, 9 (1993), pp. 252–268.

[87] S. H. Paskov, *Average case complexity of multivariate integration for smooth functions*, J. Complexity, 9 (1993), pp. 291–312.

[88] A. Pietsch, *Eigenvalues and s-Numbers*, Cambridge University Press, Cambridge, 1987.

[89] A. Pinkus, *n-Widths in Approximation Theory*, Springer, Berlin, 1985.

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

[91] L. Plaskota, *A note on varying cardinality in the average case setting*, J. Complexity, 9 (1993), pp. 458–470.

[92] L. Plaskota, *Noisy Information and Computational Complexity*, Cambridge University Press, 1996.

[93] L. Plaskota, *On adaptive designs in statistical estimation. Or, how to benefit from noise?*, preprint, Warsaw, 1995.

[93a] L. Plaskota, *Worst case complexity of problems with random information noise*, preprint, Warsaw, 1996.

[94] K. Ritter, *Approximation and optimization on the Wiener space*, J. Complexity, 6 (1990), pp. 337–364.

[95] K. Ritter, *Average errors for zero finding: lower bounds for smooth or monotone functions*, Aequationes Math., 48 (1994), pp. 194–219.

[96] K. Ritter, *Asymptotic optimality of regular sequence designs*, preprint, Erlangen, 1995.

[97] K. Ritter, *Average case analysis of numerical problems*, Habilitationsschrift, 1996, Erlangen.

[98] K. Ritter, G. W. Wasilkowski, and H. Woźniakowski, *Multivariate integration and approximation for random fields satisfying the Sacks-Ylvisaker conditions*, Annals of Appl. Prob., to appear.

[99] G. Rote, *The convergence rate of the sandwich algorithm for approximating convex functions*, Computing, 48 (1992), pp. 337–361.

[100] J. Sacks and D. Ylvisaker, *Statistical design and integral approximation*, In: Proc. 12th Bienn. Semin. Can. Math. Congr., R. Pyke, ed., Can. Math. Soc., Montreal 1970, pp. 115–136.

[101] H. Schwetlick and V. Kunert, *Spline smoothing under constraints on derivatives*, BIT, 33 (1993), pp. 512–528.

[102] A. F. Siegel and F. O'Brian, *Unbiased Monte Carlo integration methods with exactness for low order polynomials*, SIAM J. Sci. Stat. Comput., 6 (1985), pp. 169–181.

[103] K. Sikorski, *Optimal solution of nonlinear equations*, J. Complexity, 1 (1985), pp. 197–209.

[104] I. H. Sloan and S. Joe, *Lattice Methods for Multiple Integration*, Clarendon Press, Oxford, 1994.

[105] S. Smale, *Some remarks on the foundations of numerical analysis*, SIAM Rev., 32 (1990), pp. 211–220.

[106] S. A. Smolyak, *Quadrature and interpolation formulas for tensor products of certain classes of functions*, Math. USSR Doklady, 4 (1963), pp. 240–243.

[107] Gy. Sonnevend, *On the optimization of adaptive numerical algorithms of approximation*, Methods of Operations Research, 31 (1979), Verlag A. Hain, Königstein, pp. 581–595.

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

[109] Gy. Sonnevend, *Sequential algorithms of optimal order for the uniform recovery of functions with monotone* $(r-1)$ *derivatives*, Analysis Math., 10 (1984), pp. 311–335.

[110] J. Spanier and E. H. Maize, *Quasi-random methods for estimating integrals using relatively small samples*, SIAM Rev., 36 (1994), pp. 18–44.

[111] A. G. Sukharev, *The concept of sequential optimality for problems in numerical analysis*, J. Complexity, 3 (1987), pp. 347–357.

[112] A. G. Sukharev, *Min-max Algorithms in Problems of Numerical Analysis*, Nauka, Moscow, 1989. [In Russian]

[113] V. N. Temlyakov, *Approximate recovery of periodic functions of several variables*, Math. USSR Sbornik, 56 (1987), pp. 249–261.

[114] V. N. Temlyakov, *Approximation of Functions with Bounded Mixed Derivative*, Proc. Steklov Inst. Math. 178 (1989).

[115] V. N. Temlyakov, *On approximate recovery of functions with bounded mixed derivative*, J. Complexity, 9 (1993), pp. 41–59.

[116] Ph. L. Toint and A. Griewank, *Numerical experiments with partially separable optimization problems*, Lecture Notes in Math. 1066, Springer-Verlag, Berlin, 1984, pp. 203–220.

[117] A. Törn and A. Žilinskas, *Global Optimization*, Lecture Notes in Computer Science 350 (1989), Springer.

[118] J. F. Traub, *Iterative Methods for the Solution of Equations*, Englewood Cliffs, N. J., 1964. Reissued: Chelsea Press, New York, 1982.

[119] J. F. Traub, G. W. Wasilkowski, and H. Woźniakowski, *Information-Based Complexity*, Academic Press, Boston, 1988.

[120] J. F. Traub and H. Woźniakowski, *A General Theory of Optimal Algorithms*, Academic Press, New York, 1980.

[121] J. F. Traub and H. Woźniakowski, *Information-based complexity: new questions for mathematicians*, Math. Intell., 13 (1991), pp. 34–43.

[122] Y. Vardi and D. Lee, *From image deblurring to optimal investments: maximum likelihood solutions for positive linear inverse problems*, J. R. Statist. Soc. B, 55 (1993), pp. 569–612.

[123] G. Wahba, *Spline Models for Observational Data*, CBMS-NSF Regional Conference Series in Applied Math. 59, SIAM, Philadelphia, 1990.

[124] G. W. Wasilkowski, *Some nonlinear problems are as easy as the approximation problem*, Comput. Math. Appl., 10 (1984), pp. 351–363.

[125] G. W. Wasilkowski, *Information of varying cardinality*, J. Complexity, 2 (1986), pp. 204–228.

[126] G. W. Wasilkowski, *Randomization for continuous problems*, J. Complexity, 5 (1989), pp. 195–218.

[127] G. W. Wasilkowski, *On average complexity of global optimization problems*, Math. Programming, 57 (1992), pp. 313–324.

[128] G. W. Wasilkowski, *Integration and approximation of multivariate functions: average case complexity with isotropic Wiener measure*, J. Approx. Th., 77 (1994), pp. 212–227.

[129] G. W. Wasilkowski and F. Gao, *On the power of adaptive information for functions with singularities*, Math. Comp., 58 (1992), pp. 285–304.

[130] G. W. Wasilkowski, H. Woźniakowski, *Can adaption help on the average?*, Numer. Math., 44 (1984), pp. 169–190.

[131] G. W. Wasilkowski, H. Woźniakowski, *There exists a linear problem with infinite combinatory complexity*, J. Complexity, 9 (1993), pp. 326–337.

[132] G. W. Wasilkowski and H. Woźniakowski, *Explicit cost bounds of algorithms for multivariate tensor product problems*, J. Complexity, 11 (1995), pp. 1–56.

[133] A. G. Werschulz, *The Computational Complexity of Differential and Integral Equations*, Oxford University Press, Oxford, 1991.

[134] A. G. Werschulz, *The complexity of two-point boundary-value problems with piecewise analytic data*, J. Complexity, 10 (1994), pp. 367–383.

[135] H. Woźniakowski, *Average case complexity of linear multivariate problems. I. Theory*, J. Complexity, 8 (1992), pp. 337–372.

[136] H. Woźniakowski, *Average case complexity of linear multivariate problems. II. Applications*, J. Complexity, 8 (1992), pp. 373–392.

[137] A. G. Zhilinskas, *Single-step Bayesian search method for an extremum of functions of a single variable*, Cybernetics, 11 (1975), pp. 160–166.

[138] D. Zwillinger, *Handbook of Integration*, Jones and Bartlett Publishers, Boston, 1992.