Some Complexity Results for Zero Finding For Univariate Functions

Erich Novak* and Klaus Ritter†

Mathematisches Institut, Bismarckstraße 1 1/2, D-8520 Erlangen, Germany

January 1993

Dedicated to Professor J. F. Traub on the occasion of his sixtieth birthday.

Abstract

This paper gives a survey of complexity results for zero finding. We focus on zero finding for univariate, at least continuous functions with a change of sign on some interval. One section, however, deals with zero finding for complex polynomials. We present asymptotic, worst case, and average case results.

1. Introduction

This paper gives a survey of complexity results for zero finding. We focus on zero finding for univariate, at least continuous functions with a change of sign on some interval. One section deals with zero finding for complex polynomials, where we only mention some recent results. The interested reader should consult the respective papers of Pan, Renegar, Shub, Smale, and others, see section 2.

For the framework which is used here (real number model, information based complexity) we refer to Traub, Woźniakowski (1980), Traub, Wasilkowski, Woźniakowski (1988), and Smale (1990). Generally spoken, this framework can be used to analyze the complexity of problems of numerical analysis. We are mainly interested in average case results, and therefore present worst case and asymptotic results merely for the purpose of a comparison. A more complete survey on worst case results is Sikorski (1985), for results in the asymptotic setting see Traub et al. (1980, 1988).

A zero finding problem is defined by the following ingredients. First a class F of functions is specified, meaning that we wish to approximate zeros of functions $f \in F$. The error of an approximation is defined either in the root or in the residual sense. Besides the a priori information that $f \in F$, we clearly need some additional information to approximate a root of f. For classes F of polynomials of bounded degree it is reasonable to assume that $f \in F$

* supported by a Heisenberg scholarship of the Deutsche Forschungsgemeinschaft (DFG)

† supported by a postdoc grant of the Deutsche Forschungsgemeinschaft (DFG)

is given by its coefficients, i.e., the information about f is complete. See section 2 for some results about this case.

In sections 3–7 we assume that F is infinitely dimensional and that some incomplete information $N(f)$ about f is obtainable. We assume that $N(f)$ consists of function or derivative values at sequentially chosen knots and that the informational cost is proportional to the number of knots. In general the information is only partial, i.e., there are functions in F which yield the same information while having different zeros. However, we always assume that the information is exact, i.e., it is not effected by some noise.

Using the information $N(f)$ an approximation to a zero of f is computed. By choosing a model of computation we determine the kind of approximations which are computable and the cost which is caused by the computation. As it is often done in numerical analysis, we do not use Turing machines and a bit number model but prefer to work with a real number model with infinite precison. A survey of results in the bit number model is Schönhage (1987), see Pan (1987) for recent results.

Zero finding is a classical problem of numerical analysis, and most of the results deal with the asymptotic setting. The order of convergence and the efficiency or complexity index are studied under the assumption that a good initial approximation to a root is given. The computational cost is often neglected, and error bounds are not obtained. The investigation of best possible orders of convergence was initiated by Traub (1964), and subsequently various results in this field have been obtained, see section 3.

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. 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 which use a varying number of knots, depending on $f \in F$. Actually we will see in section 7 that methods with such a varying cardinality are much better than methods with a fixed number of knots. It is well known by practitioners that methods which usually work fast do sometimes fail or yield comparatively large errors for some hard functions f. This is confirmed by a comparison of results in three settings: asymptotic, worst case, and average case.

In the rest of this introduction we give such a comparison in one typical case. Let

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

It is well known that the Newton method has 2 as its order of convergence. This means that the error converges quadratically for $f \in F$ if a good starting point x_0 is chosen. We may ask whether the order 2 is optimal among all methods that compute $f(x_i)$ and $f'(x_i)$ in every chosen knot x_i. We also can ask whether there are respective globally convergent methods, i.e., methods that do not rely on a well chosen starting point. It is known that the optimal order is 2 for the class F, but this order can even be achieved by a globally convergent method which uses only function values. The optimal order is 3 if we use $f(x_i)$ and $f'(x_i)$ for each knot x_i but only consider C^3-functions with a simple zero. Known methods with optimal efficiency are rather complicated, as they have a large computational cost. Therefore the construction of simple methods with a high efficiency is interesting, see section 3 for more details.

The order of convergence is not directly related to error estimates for a method after n steps. For the bisection method and the root criterion we clearly have the worst case error $1/2^{n+1}$ after n steps, while there is no similar result for the Newton method. Actually it is not difficult to prove that bisection is optimal for F. This worst case optimality holds for many other classes, too, see section 4. One feels, however, that other methods are better on the average. This question is studied in sections 6 and 7, in section 5 we describe the probability measures that are used.

Suitable measures on F are the conditional 2-folded Wiener measures. With respect to these measures it turns out that any method with an average error ϵ needs at least $\log(1/\epsilon)$ knots for some functions, see section 6. In section 7, however, we describe a method with varying cardinality such that the expected number of knots is $O(\log\log(1/\epsilon))$ while the error is bounded by ϵ for every $f \in F$.

2. Worst case and probabilistic results for polynomials

In this section we discuss the zero finding problem on classes

$$F_d = \left\{ z^d + \sum_{i=0}^{d-1} a_i z^i \mid a_i \in \mathbb{C},\ |a_i| \leq 1 \right\}$$

of complex polynomials for $d \geq 2$. Observe that arbitrary polynomials of degree d may be transformed to elements from F_d by a dilatation and a scalar multiplication. Since problems over $\mathbb{R}$ are considered in the following sections, we identify $\mathbb{C}$ with $\mathbb{R}^2$ and use a real number model of computation here, too. It seems natural to assume that the a priori information about $f \in F_d$ is complete, i.e., the coefficients of f are the input of any zero finding method. These methods may then use real constants and the arithmetic operations $+, -, \cdot, /$ over $\mathbb{R}$, as well as branching based on comparisons over $\mathbb{R}$.

The computational cost of a method could be defined using different models for the computation over $\mathbb{R}$, such as the real number RAM (random access machine), algebraic computation trees, or machines over $\mathbb{R}$, see Blum, Shub, Smale (1989). For the present paper the differences between these models seem to be minor and we do not discuss them. We consider zero finding methods in a real number RAM and define $\mathrm{cost}_C(\phi, f)$, the computational cost of the method ϕ applied to f, by the number of arithmetic operations and branches, which are performed with the coefficients of f as input. Let us note that this real number model is equivalent to a complex number RAM with arithmetic operations and conjugation over $\mathbb{C}$ and with branching as above, and that the cost in both models are equivalent.

Let $B_r(z)$ denote the closed ball with center z and radius r, and let S_r denote the sphere with center 0 and radius r.

A worst case analysis for approximating all zeros of f is due to Pan (1987) and Renegar (1987), see Pan (1992) for a survey. The following problem is studied. Given $\epsilon > 0$, find points $z_1, \ldots, z_d \in \mathbb{C}$ such that $f(z) = \prod_{i=1}^{d}(z - z_i^*)$ with $|z_i - z_i^*| \leq \epsilon$. In the worst case analysis it is required that a method ϕ solves the problem for any $f \in F_d$. Moreover the cost of ϕ is defined by

$$\mathrm{cost}_{\max}(\phi) = \sup_{f \in F_d} \mathrm{cost}_C(\phi, f),$$

and the worst case ϵ-complexity $\mathrm{comp}_{\max}(\epsilon)$ is the minimal cost which is needed to solve the problem with the given accuracy ϵ. Pan (1987, 1992) describes an algorithm which is based on the Newton iteration and exploits Weyl's geometric construction for search and exclusion. This yields the upper bound

$$\mathrm{comp}_{\max}(\epsilon) = O(d^2 \log d \cdot \log\log(1/\epsilon)).$$

A lower bound is $d + \log\log(1/\epsilon)$, in particular

$$\mathrm{comp}_{\max}(\epsilon) \asymp \log\log(1/\epsilon)$$

holds for fixed $d \geq 2$.

For zero finding using parallel processors see Pan (1987, 1992), Kim, Sutherland (1991), Neff (1991), and Renegar (1991). Zero finding for systems of polynomials is studied in the worst case in Renegar (1989). The results mentioned so far depend on the fact that a uniform bound on the degree of the polynomials is given. Several zero finding problems for classes of polynomials of unbounded degree turn out to be unsolvable in the worst case, see Wasilkowski (1983), Saari (1987), and Sikorski (1989).

The probabilistic analysis of polynomial zero finding is initiated by Smale (1981). Here nondeterministic methods are analyzed, and the normalized Lebesgue measure P_d on (the set of coefficients corresponding to) F_d is used to derive results which do not hold for any $f \in F_d$ but with a large probability or on the average. Observe that almost every $f \in F_d$ has only simple zeros. In the following we give a sample of probabilistic results for a single polynomial. Systems of polynomials are studied in Renegar (1987a) and in several papers of Shub, Smale (1992).

Consider the Newton iteration with a starting point which is chosen at random. Since the iteration is not generally convergent, it is important to know the probability of convergence. For a polynomial of degree 2, convergence occurs with almost any starting point from $\mathbb{C}$. Therefore

$$\inf_{f \in F_2} \lambda(\Omega_1(f)) = 1,$$

where λ denotes the normalized Lebesgue measure on $B_2(0)$ and $\Omega_1(f)$ denotes the set of starting points from $B_2(0)$ yielding convergence to a zero of f. However,

$$\inf_{f \in F_d} \lambda(\Omega_1(f)) < 1$$

holds for $d \geq 3$, see Smale (1985) for these results. Friedman (1989) shows that

$$(1/d)^{c \cdot d^2 \cdot \log d} \leq \inf_{f \in F_d} \lambda(\Omega_1(f))$$

for any $d \geq 3$ with a constant $c > 0$, hereby confirming a conjecture of Smale (1985) saying that the infimum is always positive.

Suppose we wish to find a point $z \in \mathbb{C}$ with $|f(z)| \leq \epsilon$ for given $f \in F_d$ and $\epsilon > 0$. Smale (1985) constructs and analyzes a method ϕ of the following form. Let n be a suitable function of d and ϵ with $n = O(d + \log(1/\epsilon))$. Choose $z_0 \in S_3$ at random and perform a modified Newton iteration. If the problem is not solved within n steps, start the iteration again with a new starting point. More precisely, the choice of starting points occurs independently according to the uniform distribution on S_3. Clearly the real number model of computation has to be augmented by some random number generator. For any $f \in F_d$ and any $\epsilon > 0$ the expected number of starts of the iteration is bounded by 6. Consequently,

$$\sup_{f \in F_d} E(\mathrm{cost}_C(\phi, f)) = O(d^2 + d \cdot \log(1/\epsilon))$$

holds for the expected computational cost $E(\mathrm{cost}_C(\phi, f))$. See Shub, Smale (1985, 1986) for related results.

Smale (1987) introduces a version of the notion of approximate zeros to study sets of starting points for the Newton iteration which yield fast convergence. A point $z_0 \in \mathbb{C}$ is called approximate zero of a polynomial f, if the Newton iteration, starting with z_0, is well defined and

$$|z_{k+1} - z_k| \leq (1/2)^{2^k - 1} \cdot |z_1 - z_0|$$

holds for $k \in \mathbb{N}_0$. This property implies the error estimate

$$|z_k - z^*| \leq (1/2)^{2^k - 3} \cdot |z_0 - z^*|,$$

where $f(z^*) = 0$, see Renegar (1987). Let $\Omega_2(f)$ denote the set of all points $z_0 \in B_1(0)$ which satisfy this estimate, and let λ denote the Lebesgue measure. Then

$$\int_{F_d} \lambda(\Omega_2(f)) \, dP_d(f) \geq c$$

holds for the average area of these sets with a constant $c > 0$ not depending on d. Moreover, let

$$\alpha(z, f) = |f(z)/f'(z)| \cdot \max_{k \geq 2} \left| f^{(k)}(z)/(k! \cdot f'(z)) \right|^{1/(k-1)}.$$

Then

$$\alpha(z, f) < 1/8$$

implies that z is an approximate zero of f. These results are due to Smale (1987), where also a condition involving only z, $f(z)$, $f'(z)$, and d is given. The above condition is used to construct a Monte Carlo method for zero finding.

Finally we consider purely iterative zero finding methods. These iterative methods are defined by $x_{k+1} = T(x_k, f)$, where $T : \mathbb{C} \cup \{\infty\} \times G_d \to \mathbb{C} \cup \{\infty\}$ is such that $T(z, f)$ can be formed by the complex arithmetic operations $+, -, \cdot, /$ from the coefficients of f and from z. Here G_d denotes the space of all complex polynomials of degree at most d. An iteration is called generally convergent, if there is an open set $U \subset \mathbb{C} \cup \{\infty\} \times G_d$ of full Lebesgue measure such that $(z, f) \in U$ implies that the iteration, started with z, converges to a zero of f. For instance, the Newton iteration is purely iterative and generally convergent for $d = 2$. While there is a generally convergent purely iterative algorithm for $d = 3$, too, there is no such algorithm for any $d \geq 4$, see McMullen (1985). However, if the complex arithmetic operations and conjugation are permissible to form T, then generally convergent iterations do exist for any degree d, see Shub, Smale (1986a).

3. Asymptotic statements

Consider the equation $f(x) = 0$ for a continuous real valued function f. If f is from some space with uncountably infinite dimension, then it cannot be involved directly in the approximate computation of a root x^* of f. Rather some data on f, like function values, are computed first, and based on this information an approximation to x^* is constructed. Henceforth we consider functions $f \in C^r([0, 1])$, where $r \in \mathbb{N}_0$ or $r = \infty$, and we assume that the data consist of function values and values of certain derivatives of f. The sequential computation of these data starts at a knot $x_1 \in [0, 1]$, and the selection of all following knots may depend on the previously computed data. This is described by mappings $\psi_2, \psi_3, \ldots$ with values in $[0, 1]$. In the i-th step the knot

$$x_i = \psi_i(f[x_1], \ldots, f[x_{i-1}])$$

is used. Here $f[x_j]$ denotes the information that is computed at the knot x_j, in the case of Hermite information we have

$$f[x_j] = (f(x_j), \ldots, f^{(k_j)}(x_j)),$$

where k_j may depend on j. More generally we assume that certain derivatives $f^{(l)}(x_j)$ for $l \in \mathbb{N}_0$ with $l \leq r$ are computed. A natural restriction on the selection of the knots is to require that x_{i+1} may only depend on the data found at $x_i, \ldots, x_{\max(i-m,1)}$. In this case $m \in \mathbb{N}_0$ is called the memory of the method, which is said to be without memory if $m = 0$.

The mappings ψ_i and $x_i \mapsto f[x_i]$ give rise to an iterative method. Of course we are interested in those iterative methods for which the sequence of knots converges to a root x^*. This is a very old problem and a great variety of such iteration schemes have been discovered. The order of convergence of the scheme is the infimum of the order of (x_i) over all functions f from some class F and all starting points that are closely enough to a zero x^* of f. Here the order of a particular sequence is defined by

$$\lambda = \liminf_{i \to \infty} (-\log |x_i - x^*|)^{1/i}.$$

Often it is assumed that f is sufficiently smooth and x^* is a simple zero of f. In Traub (1964), Brent (1973), Bus, Dekker (1975), Woźniakowski (1976), and Sikorski, Trojan (1990) also multiple zeros are studied.

In the following we assume that F is the class of analytic functions with a simple zero in $]0,1[$. It is well known, for example, that the Newton method has order 2 even on the class of functions from $C^2([0,1])$ having only simple zeros. We may ask whether there are different methods with a better order of convergence on F.

Traub (1964) is the first who studies upper bounds for the order of convergence. He proves that any iteration scheme using Hermite information with $k_i = k$ without memory can generate a method with maximal order $k+1$. He also conjectures that no method with memory can have an order greater than $k+2$. This conjecture is proved by Brent, Winograd, Wolfe (1973). This bound is best possible, for it is attained by certain schemes based on polynomial interpolation.

The order of convergence should be related to the number of function and derivative evaluations that are used to compute the knots, see Traub (1964), Ostrowski (1973), and Traub, Woźniakowski (1976). We assume that n_i such evaluations are necessary to compute $x_2, \dots, x_i$. For example $n_i = (i-1)\cdot(k+1)$ in the case of Hermite information with $k_i = k$ for all i. Then the efficiency index θ of the method is defined by

$$\theta = \liminf_{i\to\infty}(-\log|x_i - x^*|)^{1/n_i},$$

and $1/\theta$ is called complexity index. We conjecture that $\theta \le 2$ for any method and note that this follows in many special cases from known results, for example if every $f[x_i]$ is of Hermite type.

Methods with a high efficiency or order of convergence tend to be very complicated, however. Sometimes the functions ψ_i are different for each i, i.e., the method is nonstationary, and they also cannot be given in a constructive way using only the basic arithmetic operations. Therefore many authors study 'simple' methods with a 'good' efficiency index, we just mention King (1973) and Bus, Dekker (1975).

It is interesting to study systematically such methods where the memory is bounded in a prescribed way. An important class of such methods is the class of multipoint iterations without memory. Kung and Traub (1974) study such methods that only use n function or derivative evaluations per step. Here one step may include several of the x_i's. They conjecture a bound for the order of convergence of such methods. This conjecture is proved by Kung and Traub (1976) for $n = 2$ and by Woźniakowski (1976) if each $f[x_i]$ consists of Hermite information. Wasilkowski (1981) can prove still more general cases of the Kung-Traub conjecture.

Let us stress that the statements of this section are basically asymptotic and local, we assume that the starting points are sufficiently close to x^* and we get results on the order λ, not on the errors $|x_i - x^*|$ itself. It is known, however, that the order of optimal local methods cannot be better than the order of optimal global methods. Here a method is said to be global if it converges for any function $f \in C([0,1])$ with $f(0)\cdot f(1) < 0$.

4. Worst case results for continuous or smooth functions

In the remaining part of this paper classes $F \subset F^*$ of infinite dimension are considered, where

$$F^* = \{f \in C([0,1]) \mid f(0) < 0 < f(1)\}.$$

As motivated in section 3, we study methods which first compute data on $f \in F$ at sequentially chosen knots and then construct an approximation to a root of f. For simplicity we assume that the data consist of Hermite information

$$f[x_i] = (f(x_i), \ldots, f^{(k)}(x_i))$$

of order $k \in \mathbb{N}_0$. Clearly only those k with $F \subset C^k([0,1])$ are admissible.

The formal description of a zero finding method is given by a knot $x_1 \in [0,1]$ and by mappings $\psi_i^k : \mathbb{R}^{(i-1)\cdot(k+1)} \to [0,1]$ for $i = 2, \ldots, n$ and $\phi_n^k : \mathbb{R}^{n\cdot(k+1)} \to [0,1]$. In the i-th step the knot

$$x_i = \psi_i^k(f[x_1], \ldots, f[x_{i-1}])$$

is chosen and the data

$$N_i^k(f) = (f[x_1], \ldots, f[x_i])$$

are known after this step. After n evaluations the approximation

$$S_n^k(f) = \phi_n^k(N_n^k(f))$$

to a root of f is constructed. Observe that the number of evaluations does not depend on f.

It should be stressed that any zero finding method is only based on the knowledge $f \in F$ and the knowledge that is given by N_n^k. We do not assume additional knowledge such as 'a good starting point for the Newton method'. If such an additional knowledge is given it must be translated in our language; it usually means that we have to consider a subset $\widetilde{F} \subset F$ instead of F.

Now we define the cost and the error of S_n^k for a function $f \in F$. To compute $S_n^k(f)$ we need $n \cdot (k+1)$ function values and values of derivatives, i.e., a certain subroutine or oracle is called n times. We assume that we have to pay c_k each time, where

$$0 < c_0 \le c_1 \le \ldots,$$

which causes the informational cost $n \cdot c_k$. The total cost amounts to

$$\mathrm{cost}(S_n^k, f) = \mathrm{cost}_I(S_n^k, f) + \mathrm{cost}_C(S_n^k, f) = n \cdot c_k + \mathrm{cost}_C(S_n^k, f),$$

where $\mathrm{cost}_I(S_n^k, f)$ denotes the informational cost and $\mathrm{cost}_C(S_n^k, f)$ is the computational cost defined similarly as in section 2. The error of S_n^k for f is defined by the root criterion

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

in this section, later we will also use the residual criterion.

In this section we study the maximal error

$$\Delta_{\max}(S_n^k) = \sup_{f \in F} \Delta(S_n^k, f)$$

and the maximal cost

$$\mathrm{cost}_{\max}(S_n^k) = \sup_{f \in F} \mathrm{cost}(S_n^k, f),$$

and we define the (worst case) ϵ-complexity of the problem by

$$\mathrm{comp}_{\max}(\epsilon) = \inf\{\mathrm{cost}_{\max}(S_n^k) \mid \Delta_{\max}(S_n^k) \leq \epsilon\}.$$

Hence $\mathrm{comp}_{\max}(\epsilon)$ is the minimal worst case cost, which is needed to compute a zero with an error of at most ϵ for any $f \in F$. We might also allow methods which use a varying number of evaluations in the definition of the complexity, see section 7. However, this can effect the worst case complexity at most by a multiplicative constant.

In complexity theory it is, as a rule, very difficult to prove tight lower bounds. For continuous problems on a infinitely dimensional set F one often only studies bounds for the informational cost. It is remarkable that one can prove rather sharp complexity results in this way for many problems like numerical integration or zero finding. In the worst case analysis of zero finding the estimate

$$\mathrm{comp}_{\max}(\epsilon) \geq \inf\{n \cdot c_k \mid \Delta_{\max}(S_n^k) \leq \epsilon\}$$

is used and the numbers

$$\inf_{S_n^k} \Delta_{\max}(S_n^k)$$

are studied for $n \in \mathbb{N}$ and admissible $k \in \mathbb{N}_0$. We obtain tight bounds if there are respective methods with a small computational cost.

The bisection method B_n^0 on F^* is formally defined by the knot $x_1 = \psi_1^0 = 1/2$ and by

$$\psi_{i+1}^0(y_1, \ldots, y_i) = \psi_i^0(y_1, \ldots, y_{i-1}) - \mathrm{sgn}\,(y_i)/2^{i+1}$$

and $\phi_n^0 = \psi_{n+1}^0$. It can be proved that bisection is optimal for many different classes $F \subset F^*$ and under different assumptions for the allowed methods, see Kung (1976) and Sikorski (1982, 1989).

For instance, let F be a class such that

$$\{f \in C^\infty([0,1]) \mid f(0) = -1,\ f(1) = 1,\ f' \geq 1\} \subset F \subset F^*.$$

Moreover let $S_n^k = \phi_n^k \circ N_n^k$ be an arbitrary zero finding method. Then there are two functions $f_1, f_2 \in F$ with $N_n^k(f_1) = N_n^k(f_2)$ but $d(f_1^{-1}(0), f_2^{-1}(0))$ close to $(1/2)^n$. Since $S_n^k(f_1) = S_n^k(f_2)$, we conclude that

$$\inf_{S_n^k} \Delta_{\max}(S_n^k) = \Delta_{\max}(B_n^0) = (1/2)^{n+1}$$

for any $n \in \mathbb{N}$ and any admissible k, i.e., B_n^0 is worst case optimal among all methods which use n knots. Here even methods which use derivatives of arbitrarily high order could be permitted. The computational cost of the bisection method satisfies

$$\sup_{f \in F} \mathrm{cost}_C(B_n^0, f) \asymp n$$

and therefore

$$\mathrm{comp}_{\max}(\epsilon) \asymp \log(1/\epsilon)$$

holds. Hence we have a rather large worst case complexity even for classes of smooth functions having simple zeros. The same complexity bound holds on the class

$$F = \{f \in C^1([0,1]) \cap F^* \mid \gamma_1 \leq f' \leq \gamma_2\}$$

with $0 < \gamma_1 < \gamma_2$, see Kung (1976), although bisection is no longer worst case optimal, see also Micchelli, Miranker (1975).

These results may be surprising, as they show that modifications of the Newton method (the Newton method itself is not necessarily defined on F) are not essentially better than bisection in the worst case on the respective classes F. But of course there are other classes such that bisection is far from being optimal. For instance, consider

$$F = \{f \in C^2([0,1]) \cap F^* \mid |f'(x^*)| \geq \gamma_1 \text{ if } f(x^*) = 0,\ \|f''\|_\infty \leq \gamma_2\}$$

with $\gamma_i > 0$. In this case the ϵ-complexity is significantly smaller than $\log(1/\epsilon)$, since

$$\mathrm{comp}_{\max}(\epsilon) \asymp \log\log(1/\epsilon)$$

holds. The lower bound is due to Maistrowskii (1972), and the following methods yield the upper bound. First carry out $n_0 = \lceil \log(\gamma_2/\gamma_1) \rceil$ bisection steps to obtain a subinterval with a change of sign and length at most γ_1/γ_2. Then start the Newton iteration with the center of this subinterval. After $n_0 + n$ steps the root error is bounded by $(1/2)^{(2^n + n_0)}$ for any $f \in F$, see Kantorovich (1948) and Traub, Woźniakowski (1979).

5. On some measures in function spaces

For problems of numerical linear algebra or linear programming the problem elements, e.g., the matrices and vectors defining a system of linear equations, belong to finite dimensional spaces. Problems of numerical analysis, however, are usually studied on infinite dimensional classes of functions. For the zero finding problem the class F may be either of finite dimension, e.g., a class of polynomials having fixed degree, or infinite dimensional, e.g., defined by regularity and boundary conditions.

This difference is important, if we want to study the average performance of zero finding methods with respect to a probability measure on F. While normalized Lebesgue measures can be used on classes of polynomials, see section 2, there is no analogue to the Lebesgue measure on infinite dimensional spaces.

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 as well as a particular type of non-Gaussian measures have been used. The non-Gaussian measures are the Dubins-Freedman or Ulam measures on classes of continuous, monotone functions, see Graf, Mauldin, Williams (1986) for the construction and a detailed analysis. The Ulam measure P with boundary values $a_0 < 0 < b_0$ is a probability measure on the class

$$F = \{f \in C([0,1]) \mid f(0) = a_0,\ f(1) = b_0,\ f \text{ strictly increasing}\},$$

equipped with the topology of uniform convergence and with the respective Borel σ-algebra. The function value $f(1/2)$ is uniformly distributed on $[a_0, b_0]$, and a P-random function can be constructed on the dyadic rationals inductively. First put $f(0) = a_0$ and $f(1) = b_0$. 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 $[f(i/2^k), f((i+1)/2^k)]$. Plots of these random functions are given in Graf, Mauldin, Williams (1986). It is proven there that $P(A) > 0$ holds for any open neighborhood A of any $f \in F$.

Graf, Novak (1990) and Novak (1992) use the Ulam measure for an average case analysis of numerical integration of monotone functions. For the zero finding problem this measure is used by Graf, Novak, Papageorgiou (1989) and Ritter (1992).

In the remaining part of this section we discuss Gaussian measures, see Billingsley (1968), Hida (1980), Adler (1981), and Vakhania, Tarieladze, Chobanyan (1987). Let $C^r([0,1])$ be equipped with the norm $\|f\| = \max(\|f\|_\infty, \dots, \|f^{(r)}\|_\infty)$ and with the respective Borel σ-algebra. A probability measure on a measurable subset of $C^r([0,1])$ is called Gaussian, if the joint distribution of any finite number of function evaluations $f(t_1), \dots, f(t_n)$ is a normal distribution.

Suldin (1959) is the first who analyzes average errors for numerical problems on infinite dimensional spaces. He uses the Wiener measure on $C([0,1])$ to define the average error of quadrature formulas. Gaussian measures on classes of univariate functions with higher regularity are considered by Sacks, Ylvisaker (1970), Speckman (1979), and Lee, Wasilkowski (1986), e.g. In particular, the r-folded Wiener measure on $C^r([0,1])$ is used. This measure is obtained as the image of the Wiener measure under the transformation $g \mapsto f$, where

$$f(t) = (r-1)!^{-1} \int_0^t g(u) \cdot (t-u)^{r-1}\, du,$$

i.e., under r-fold integration of the functions $g \in C([0,1])$.

Wasilkowski, Gao (1992) use superpositions of Gaussian measures to analyze problems for functions having singularities. Two different generalizations of the Wiener measure, the Wiener sheet measure and the isotropic Wiener measure, are used to study problems for multivariate functions. As in the univariate case related Gaussian measures with higher regularity are considered, too. See, e.g., Papageorgiou, Wasilkowski (1990), Woźniakowski (1991), Wasilkowski (1992), Woźniakowski (1992), Paskov (1993), and Ritter, Wasilkowski, Woźniakowski (1993). Gaussian measures on periodic Sobolev spaces are used in Traub, Wasilkowski, Woźniakowski (1988) and Heinrich (1990), e.g. The average case analysis with respect to finitely additive Gaussian measures on Hilbert spaces dates back to Larkin (1972).

The Gaussian measures which are used to analyze zero finding are derived from the Wiener measure by r-fold integration and translation by suitable polynomials. For $r = 0$ Novak (1989) and Novak, Ritter (1992) use the Brownian bridge with boundary values $a_0 < 0 < b_0$ on the class

$$F = \{f \in C([0,1]) \mid f(0) = a_0,\ f(1) = b_0\}.$$

This measure is the image of the Wiener measure under the translation $g \mapsto f$, where

$$f(t) = g(t) + (1-t) \cdot (a_0 - g(0)) + t \cdot (b_0 - g(1)).$$

Zero finding for functions with higher regularity is studied by Ritter (1992) and Novak, Ritter, Woźniakowski (1993). They use a generalization of the above construction, and consider the images of the r-folded Wiener measure under the translation defined by

$$f = g + p(a - (g(0), \dots, g^{(r)}(0)), b - (g(1), \dots, g^{(r)}(1))).$$

Here $a, b \in \mathbb{R}^{r+1}$, and $p(u,v)$ denotes the polynomial of degree at most $2r+1$ with

$$p(u,v)^{(i)}(0) = u_i \qquad \text{and} \qquad p(u,v)^{(i)}(1) = v_i$$

for $i = 0, \dots, r$. Obviously f fulfills the boundary conditions $f^{(i)}(0) = a_i$ and $f^{(i)}(1) = b_i$. If $a = 0$, then the image measures are the conditional r-folded Wiener measures given $g^{(i)}(1) = b_i$, see Lee, Wasilkowski (1986). We will use this as a name for any of these measures. Hence the conditional r-folded Wiener measure with boundary values a, b is a Gaussian measure on the class

$$F = \{f \in C^r([0,1]) \mid f^{(i)}(0) = a_i,\ f^{(i)}(1) = b_i \text{ for } i = 0, \dots, r\},$$

and for zero finding $a_0 < 0 < b_0$ is required.

We summarize some properties of the conditional r-folded Wiener measures. If $r \geq 1$, then almost every function has only simple zeros and the number of zeros is almost surely finite, see Novak, Ritter, Woźniakowski (1992). The mean of the measure P with boundary values a, b is the polynomial $p(a,b)$ which interpolates the boundary values, i.e.,

$$\int_F f(t)\, dP(f) = p(a,b)(t)$$

holds for $t \in [0,1]$. Since P is Gaussian, it is symmetric with respect to the mean, saying that $f \mapsto 2 \cdot p(a,b) - f$ leaves P invariant. Furthermore the covariances of function evaluations do not depend on the boundary values but only on r. A time-inversion property holds, since the covariances of $f(s)$ and $f(t)$ and of $f(1-s)$ and $f(1-t)$ coincide. In particular the variances are given by

$$\int_F (f(t) - p(a,b)(t))^2 \, dP(f) = c_r \cdot (t \cdot (1-t))^{2r+1}$$

with a constant $c_r > 0$. Any open neighborhood A of any $f \in F$ satisfies $P(A) > 0$, and therefore F is the support of P. The $i+1$-dimensional process $(f(t), \ldots, f^{(i)}(t))$ has the Markov property iff $i = r$.

Functions with fixed values and derivatives up to the order r at the points 0 and 1 are considered, if the average is studied with respect to a conditional r-folded Wiener measure. The a priori knowledge of high order derivatives may be unrealistic, and therefore stochastic models with nonconstant boundary values are interesting in particular for large r. The superposition of conditional r-folded Wiener measures is a suitable way to construct such models. Moreover, a Wiener measure with covariance kernel $\sigma \cdot \min(s,t)$ can be used to construct the conditional r-folded measures. Hereby the covariances can be amplified suitably.

Finally we give some plots of random functions with respect to conditional r-folded Wiener measures for $r = 1$ and 3. These functions are constructed in the following way. Brownian paths are approximated according to Lévy's construction, see Hida (1980). By r-fold integration and translation as indicated we obtain the random functions. The respective mean is given as a dotted line, and each of the two pictures contains 7 functions. We have used software of Frank Nestel as well as the Erlangen graphics system.

Figure 1: $r = 1$

Figure 2: $r = 3$

6. Average case results for methods with a fixed number of knots

The remaining two sections deal with the average case on the function classes

$$F = \{f \in C([0,1]) \mid f(0) = a_0,\ f(1) = b_0,\ f \text{ strictly increasing}\}$$

equipped with the respective Ulam measure P, and

$$F = \{f \in C^r([0,1]) \mid f^{(i)}(0) = a_i,\ f^{(i)}(1) = b_i \text{ for } i = 0, \ldots, r\}$$

equipped with the respective conditional r-folded Wiener measure P. Henceforth

$$a_0 < 0 < b_0$$

is assumed. In the case $a_0 \cdot b_0 > 0$ the following decision problem would be interesting for conditional r-folded Wiener measures. Decide with a probability of failure bounded by $\delta > 0$ whether $f \in F$ has a zero or not. We do not discuss this problem here.

In this section we study zero finding methods $S_n^k = \phi_n^k \circ N_n^k$ of the form which was already discussed in section 4. These methods use Hermite information $N_n^k(f)$ of order k at n sequentially chosen knots, where n does not depend on $f \in F$, i.e., the number of knots is fixed. Methods with a varying number of knots are studied in the following section. The definition of $\mathrm{cost}(S_n^k, f)$, the cost of S_n^k for the function f, is given in section 4. The error $\Delta(S_n^k, f)$ of S_n^k for f, however, is either defined by the root criterion

$$\Delta(S_n^k, f) = \Delta^{\mathrm{ro}}(S_n^k, f) = d(S_n^k(f), f^{-1}(0))$$

as in section 4 or by the residual criterion

$$\Delta(S_n^k, f) = \Delta^{\mathrm{re}}(S_n^k, f) = |f(S_n^k(f))|.$$

The expectation

$$\Delta_{\mathrm{avg}}(S_n^k) = \int_F \Delta(S_n^k, f)\, dP(f)$$

of the root or residual error with respect to P is called average error of S_n^k. The integral is well defined, i.e., the integrand is measurable, if the mappings ψ_i^k, which describe the selection of the knots, and the mapping ϕ_n^k are measurable. This will be always assumed in the following. Analogously to the worst case we study the numbers

$$\inf_{S_n^k} \Delta_{\mathrm{avg}}(S_n^k)$$

for $n \in \mathbb{N}$ and admissible $k \in \mathbb{N}_0$.

Many results concerning the average error are known for linear problems, like numerical integration or reconstruction (approximation) of functions, if Gaussian measures are used to define the average. See section 5 for references. For nonlinear problems, like zero finding or global optimization, however, only few results are known. In particular, lower bounds were obtained only recently, see Novak, Ritter (1992), Ritter (1992) for zero finding and Ritter (1990) for global optimization using fixed knots.

The bisection method is worst case optimal for any of the classes F mentioned above, if the root criterion is used, since

$$\inf_{S_n^k} \Delta_{\max}^{\mathrm{ro}}(S_n^k) = (1/2)^{n+1}$$

holds. On the average smaller errors are obtainable as the upper bound

$$\inf_{S_n^0} \Delta_{\mathrm{avg}}^{\mathrm{ro}}(S_n^0) \leq \gamma_1 \cdot \alpha^n$$

with constants $\gamma_1 > 0$ and

$$\alpha < 1/2$$

shows. The constants may depend on F, equipped with the measure as above, and the bound is proven with respect to an Ulam measure, see Graf, Novak, Papageorgiou (1989), and with respect to a Brownian bridge, see Novak (1989). For the conditional r-folded Wiener measures with $r \geq 1$ a proof could be given similarly.

We sketch some observations which lead to the proof in the case of a Brownian bridge. Suppose that $f(s) < 0 < f(t)$ are computed and that no function value is known inside of $]s,t[$. A bisection step reduces the length of the subinterval, necessarily containing a zero, almost surely by the factor $1/2$. Let x be the knot belonging to a regula falsi step. Then the conditional distribution of $f(x)$ is normal with mean zero, and therefore the expected factor of reduction is $1/2$, too, independently of $f(s)$ and $f(t)$. The one step optimal choice regarding the expected factor of reduction lies between $(s+t)/2$ and x, and the zero finding methods S_n^0, which are defined hereby, satisfy $\Delta_{\mathrm{avg}}^{\mathrm{ro}}(S_n^0) \leq \gamma_1 \cdot \alpha^n$.

The question arises whether there are methods S_n^k with average errors converging to zero superlinearily. A negative answer is given by Novak, Ritter (1992) for Brownian bridges and by Ritter (1992) for the other measures. Here it is enough to consider the case $k = r$, in particular, this means $k = 0$ in the case of an Ulam measure or a Brownian bridge. According to these papers there exist constants $\beta > 0$ and $\gamma_2 > 0$, which only depend on the respective measure, such that

$$\inf_{S_n^r} \Delta_{\mathrm{avg}}(S_n^r) \geq \gamma_2 \cdot \beta^n$$

holds for any n. This result holds for the root and the residual criterion. Therefore the number of bisection steps, necessary to guarantee a maximal root error ϵ, differs from the number of function or derivative evaluations, necessary to obtain an average error ϵ, at most by a multiplicative constant. This seems to be remarkable especially for conditional r-folded Wiener measures with $r \geq 2$, since there are iterative methods which converge superlinearily on a set of functions with full measure, see section 3 and section 5.

In the lower bound the number β can be chosen independently from the special boundary values, i.e., we can take the same β for all Ulam measures and we can take the same $\beta = \beta_r$ for all conditional r-folded Wiener measures. We even conjecture that the optimal β is the same for all Ulam measures and for all conditional r-folded Wiener measures, respectively.

Now we want to explain some ideas that are related to the proof of the lower bounds. We also indicate why the proof of such lower bounds is quite different for linear and nonlinear problems. Assume that a method $S_n^k = \phi_n^k \circ N_n^k$ is given. Using conditional probability measures, the average error can be written in the form

$$\Delta_{\text{avg}}(S_n^k) = \int_{N_n^k(F)} \int_F \Delta(S_n^k, f) \, dP(f \mid N_n^k = y) \, dN_n^k P(y).$$

The inner integral

$$\Delta_{\text{avg}}(S_n^k, y) = \int_F \Delta(S_n^k, f) \, dP(f \mid N_n^k = y)$$

is called local average error. For linear problems with a Gaussian measure we cannot have too small local average errors because of

$$\inf_{S_n^k} \inf_y \Delta_{\text{avg}}(S_n^k, y) = \inf_{S_n^k} \Delta_{\text{avg}}(S_n^k),$$

see Traub, Wasilkowski, Woźniakowski (1988). This also means that a method S_n^k is optimal iff $\Delta_{\text{avg}}(S_n^k, y)$ is almost everywhere the constant defined by this infimum.

In a typical nonlinear problem the local average errors $\Delta_{\text{avg}}(S_n^k, y)$ are distributed quite differently. These numbers may be arbitrarily small for some favorable data y and may be large for other y. For the zero finding problem and all the measures mentioned above the following is true. For each method S_n^k there is a set of unfavorable data $M \subset \mathbb{R}^{n\cdot(k+1)}$ such that

$$\inf_{y \in M} \Delta_{\text{avg}}(S_n^k, y) \geq \gamma_2 \cdot \beta_1^n$$

and

$$N_n^k P(M) \geq \beta_2^n.$$

A proof of this result uses induction and is based on the scaling and quasi Markov properties of the Ulam and the conditional r-folded Wiener measures.

It can be seen that such a result does not lead to nontrivial lower bounds for large n if we also consider methods with a varying number of knots. Nevertheless, in the case of conditional r-folded Wiener measures one can use a statement of a similar form to prove such lower bounds even then. In the forthcoming paper Novak, Ritter, Woźniakowski (1993) the following is proved. For each method S_n^k using a fixed number n of knots and for any set $M \subset \mathbb{R}^{n\cdot(k+1)}$ with $P(\{f \in F_0 \mid N_n^k(f) \in M\}) \geq 1/(2n)$ we have the estimate

$$\int_M \Delta_{avg}(S_n^k, y) \, dN_n^k P(y) \geq \alpha^{\beta^n}.$$

Here $\alpha > 0$ and $\beta > 1$ do not depend on n, and F_0 is a fixed set of functions with $P(F_0) = 1/2$. Taking into account the upper bound from section 7 this implies that the average case ϵ-complexity is of the order $\log\log(1/\epsilon)$ if $r \geq 2$.

7. Average case results for methods with a varying number of knots

So far we have discussed worst case and average case results for zero finding methods which use a fixed number of knots. In practice, however, the number of evaluations is often determined by a stopping rule, like $|f(x_i)| \leq \epsilon$, and may therefore vary over the class F. Methods of the latter type are said to be of varying cardinality.

As previously we assume that the data on $f \in F$ consist of Hermite information of order k at sequentially chosen knots. In the case of varying cardinality a decision to stop or to compute additional data is made after each step. This decision may be based on all data which are known at that time. If $n(f)$ is the total number of knots which are used for f, then the approximation to a root of f is of the form

$$S^k(f) = \phi^k_{n(f)}(N^k_{n(f)}(f)),$$

where $\phi^k_i : \mathbb{R}^{i\cdot(k+1)} \to [0, 1]$. See Wasilkowski (1986) for a formal definition of methods with varying cardinality.

The model of computation from section 4 is used, such that the cost of S^k for f is given by

$$\text{cost}(S^k, f) = n(f) \cdot c_k + \text{cost}_C(S^k, f).$$

In contrast to section 6 we consider the average cost

$$\text{cost}_{\text{avg}}(S^k) = \int_F \text{cost}(S^k, f)\, dP(f)$$

of S^k, which may be much lower than the maximal cost. Analogously to the worst case ϵ-complexity the average case ϵ-complexity of the problem is defined as

$$\text{comp}_{\text{avg}}(\epsilon) = \inf\{\text{cost}_{\text{avg}}(S^k) \mid \Delta_{\text{avg}}(S^k) \leq \epsilon\},$$

where

$$\Delta_{\text{avg}}(S^k) = \int_F \Delta(S^k, f)\, dP(f)$$

is the average error of S^k.

A worst case analysis cannot justify the use of methods with varying cardinality for a numerical problem. Due to Wasilkowski (1986) varying cardinality also does not help essentially for many linear problems, if the average error and the average cost are defined with respect to a Gaussian measure. For zero finding, however, varying cardinality is very powerful. This will be proven in Novak, Ritter, Woźniakowski (1993), and we discuss some of their results here.

We study a method which uses the function value $f(x_i)$ and the first derivative $f'(x_i)$ for each knot x_i and is defined for each $f \in C^1([0,1]) \cap F^*$. The method is easy to describe and also can easily be implemented on a computer. Moreover we can perform an average case analysis of that method using well known facts about the Newton method. This method is almost optimal, see the remarks at the end of section 6, but other such hybrid methods may be even slightly better, see, e.g., Bus, Dekker (1975). Methods of this form are still considered as "one of the best computer algorithms that we have for finding a real zero", see Kahaner, Moler, Nash (1989). We have taken our algorithm just for the sake of simplicity.

First we define $[s_0, t_0] = [0, 1]$ and evaluate f and f' at $x_{-1} = 0$ and $x_0 = 1$. Now we define the knot x_i under the condition that s_{i-1} and t_{i-1} and also the respective values of f and f' are known. Assume that i is odd, and define $p \in \{s_{i-1}, t_{i-1}\}$ by

$$|f(p)| = \min(|f(s_{i-1})|,\ |f(t_{i-1})|)$$

and $p = s_{i-1}$ if $|f(s_{i-1})| = |f(t_{i-1})|$. Then we try to perform a Newton step beginning from p, i.e., we try to define

$$q = p - f(p)/f'(p).$$

We distinguish two cases. If the resulting point q is well defined and satisfies $q \in [s_{i-1}, t_{i-1}]$, then we put $x_i = q$. If the resulting point q is not defined (because of zero division) or is not in $[s_{i-1}, t_{i-1}]$, then we put $x_i = x_{i-1}$, i.e., we actually do not perform an evaluation in this step. Now we assume that i is even. Then we perform a bisection step, i.e., we put $x_i = (s_{i-1} + t_{i-1})/2$. For arbitrary i the new subinterval is given by $[s_i, t_i] = [s_{i-1}, x_i]$ if $f(x_i) \geq 0$ and $[s_i, t_i] = [x_i, t_{i-1}]$ otherwise.

If the residual criterion is used then we apply the described method together with the simple stopping rule $|f(x_i)| \leq \epsilon$. In the case of the root criterion we do not have such a simple stopping rule, actually we compute function values in some additional knots and then also define a stopping rule that guarantees

$$\Delta(S^1, f) \leq \epsilon$$

for all $f \in C^1([0,1]) \cap F^*$.

Now assume that $f \in C^2([0,1]) \cap F^*$ with $\|f''\|_\infty \leq \gamma_1$ and $|f'(x^*)| \geq \gamma_2$ for any x^* with $f(x^*) = 0$. Then we can determine a number i_0 such that the subsequence

$$x_{i_0} \to x_{i_0+2} \to x_{i_0+4} \to \dots$$

is defined by the Newton method. We therefore get an error bound using known results of Kantorovich type on the Newton method. For both stopping rules we can prove

$$\int_F n(f)\, dP(f) \leq \gamma \cdot \log\log(1/\epsilon)$$

with repect to any conditional r-folded Wiener measure with $r \geq 2$. Since the average cost of these methods is of the same order, the upper bound

$$\mathrm{comp}_{\mathrm{avg}}(\epsilon) = O(\log\log(1/\epsilon))$$

holds. Observe that this bound is much better than the respective bound for methods with fixed cardinality, although we can guarantee that the error is at most ϵ for all functions from the respective class F.

We just remark that this bound can even be achieved by methods which only use function values. We have to replace the Newton steps by suitably defined steps of the secant method.

Acknowledgments. We thank Mike Shub, Greg Wasilkowski, and Henryk Woźniakowski for valuable comments. Our contributions to the average case analysis for zero finding were stimulated by Joe Traub. Not only for this we express our gratitude.

References

Adler, R. J. (1981), "The Geometry of Random Fields", Wiley, New York.

Billingsley, P. (1968), "Convergence of Probability Measures", Wiley, New York.

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

Brent, R. P. (1973), "Algorithms for Minimization without Derivatives", Prentice-Hall, New York.

Brent, R. P., Winograd, S., and Wolfe, P. (1973), Optimal iterative processes for root-finding, Numer. Math. **20**, 327–341.

Bus, J. C. P., and Dekker, T. J. (1975), Two efficient algorithms with guaranteed convergence for finding a zero of a function, ACM Trans. Math. Software **1**, 330–345.

Friedman, J. (1989), On the convergence of Newton's method, J. Complexity **5**, 12–33.

Graf, S., Mauldin, R. D., and Williams, S. C. (1986), Random homeomorphisms, Adv. Math. **60**, 239–359.

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

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

Heinrich, S. (1990), Invertibility of random Fredholm operators, Stoch. Anal. Appl. **8**, 1–59.

HIDA, T. (1980), "Brownian Motion", Springer-Verlag, New York.

KAHANER, D., MOLER, C., AND NASH, S. (1989), "Numerical Methods and Software", Prentice-Hall, Englewood Cliffs.

KANTOROVICH, L. V. (1948), Functional analysis and applied mathematics (in Russian), Uspehi Mat. Nauk **3**, 89–185.

KIM, M., AND SUTHERLAND, S. (1991), Families of parallel root-finding algorithms, Preprint.

KING, R. F. (1973), An improved Pegasus method for root finding, BIT **13**, 423–427.

KUNG, H. T. (1976), The complexity of obtaining starting points for solving operator equations by Newton's method, in "Analytic Computational Complexity" (Traub, J. F., ed.), 35–57, Academic Press, New York.

KUNG, H. T., AND TRAUB, J. F. (1974), Optimal order of one-point and multipoint iterations, J. Assoc. Comput. Mach. **21**, 643–651.

KUNG, H. T., AND TRAUB, J. F. (1976), Optimal order and efficiency for iterations with two evaluations, SIAM J. Numer. Anal. **13**, 84–99.

LARKIN, F. M. (1972), Gaussian measure in Hilbert space and applications in numerical analysis, Rocky Mountain J. Math. **2**, 379–421.

LEE, D., AND WASILKOWSKI, G. W. (1986), Approximation of linear functionals on a Banach space with a Gaussian measure, J. Complexity **2**, 12–43.

MAISTROWSKII, G. D. (1972), On the optimality of Newton's method, Soviet Math. Dokl. **13**, 838–840.

MCMULLEN, C. (1985), Families of rational maps and iterative root-finding algorithms, Ph.D. thesis, Harvard University.

MICCHELLI, C. A., AND MIRANKER, W. L. (1975), High order search methods for finding roots, J. Assoc. Comput. Mach. **22**, 52–60.

NEFF, A. C. (1991), Specified precision polynomial root isolation is NC, J. Comput. System Sci. , to appear.

NOVAK, E. (1989), Average-case results for zero finding, J. Complexity **5**, 489–501.

NOVAK, E. (1992), Quadrature formulas for monotone functions, Proc. Amer. Math. Soc. **115**, 59–68.

NOVAK, E., AND RITTER, K. (1992), Average errors for zero finding: lower bounds, Math. Z., to appear.

NOVAK, E., RITTER, K., AND WOŹNIAKOWSKI, H. (1993), Average case optimality of a hybrid Newton-bisection method with a stopping rule, tentative title, in preparation.

OSTROWSKI, A. M. (1973), "Solution of Equations in Euclidean and Banach Spaces", Academic Press, New York.

PAN, V. (1987), Sequential and parallel complexity of approximate evaluation of polynomial zeros, Comput. Math. Appl. **14**, 591–622.

PAN, V. (1992), Complexity of computations with matrices and polynomials, SIAM Rev. **34**, 225–262.

PAPAGEORGIOU, A., AND WASILKOWSKI, G. W. (1990), On the average complexity of multivariate problems, J. Complexity **6**, 1–23.

PASKOV, S. (1993), Average case complexity of multivariate integration for smooth functions, J. Complexity, to appear.

RENEGAR, J. (1987), On the worst-case arithmetic complexity of approximating zeros of polynomials, J. Complexity **3**, 90–113.

RENEGAR, J. (1987a), On the efficiency of Newton's method in approximating all zeros of a system of complex polynomials, Math. Oper. Res. **12**, 121–148.

RENEGAR, J. (1989), On the worst-case arithmetic complexity of approximating zeros of systems of polynomials, SIAM J. Computing **18**, 350–370.

RENEGAR, J. (1991), Computational complexity of solving real algebraic formulae, in "Proc. Int. Congress of Math., Kyoto, 1990" (Satake, I., ed.), Vol. II, 1595–1606, Springer Verlag, Tokyo.

RITTER, K. (1990), Approximation and optimization on the Wiener space, J. Complexity **6**, 337–364.

RITTER, K. (1992), Average errors for zero finding: lower bounds for smooth or monotone functions, Techn. Rep. No. 209–92, Department of Computer Science, Univ. Kentucky, Lexington.

RITTER, K., WASILKOWSKI, G. W., AND WOŹNIAKOWSKI, H. (1993), On multivariate integration for stochastic processes, Preprint.

SAARI, D. G. (1987), Some informational requirements for convergence, J. Complexity **3**, 302–311.

SACKS, J., AND YLVISAKER, D. (1970), Statistical designs and integral approximation, in "Proc. 12th Bienn. Seminar Can. Math. Congr. 1969" (Pyke, R., ed.), 115–136, Can. Math. Soc., Montreal.

SCHÖNHAGE, A. (1987), Equation solving in trems of computational complexity, in "Proc. Int. Congress of Math., Berkeley, 1986" (Gleason, A. M., ed.), Vol. I, 131–153, Amer. Math. Soc., Providence.

SHUB, M., AND SMALE, S. (1985), Computational complexity: on the geometrie of polynomials and a theory of cost, I, Ann. Sci. École Norm. Sup. **18**, 107–142.

SHUB, M., AND SMALE, S. (1986), Computational complexity: on the geometrie of polynomials and a theory of cost, II, SIAM J. Computing **15**, 145–161.

SHUB, M., AND SMALE, S. (1986a), On the existence of generally convergent algorithms, J. Complexity **2**, 2–11.

SHUB, M., AND SMALE, S. (1992), Complexity of Bezout's theorem I, II, III, Preprint.

SIKORSKI, K. (1982), Bisection is optimal, Numer. Math. **40**, 111–117.

SIKORSKI, K. (1985), Optimal solution of nonlinear equations, J. Complexity **1**, 197–209.

SIKORSKI, K. (1989), Study of linear information for classes of polynomial equations, Aequationes Math. **37**, 1–14.

SIKORSKI, K. , AND TROJAN, G. M. (1990), Asymptotic near optimality of the bisection method, Numer. Math. **57**, 421–433.

SMALE, S. (1981), The fundamental theorem of algebra and complexity theory, Bull. Amer. Math. Soc. **4**, 1–36.

SMALE, S. (1985), On the efficiency of algorithms of analysis, Bull. Amer. Math. Soc. **13**, 87–121.

SMALE, S. (1987), Algorithms for solving equations, in "Proc. Int. Congress of Math., Berkeley, 1986" (Gleason, A. M., ed.), Vol. I, 172–195, Amer. Math. Soc., Providence.

SMALE, S. (1990), Some remarks on the foundations of numerical analysis, SIAM Rev. **32**, 211–220.

SPECKMAN, P. (1979), L_p approximation of autoregressive Gaussian processes, Techn. Rep., Department of Statistics, Univ. Oregon, Eugene.

SULDIN, A. V. (1959), Wiener measure and its applications to approximation methods I (in Russian), Izv. Vyssh. Uchebn. Zaved. Mat. **13**, 145–158.

TRAUB, J. F. (1964), "Iterative Methods for the Solution of Equations", Prentice-Hall, Englewood Cliffs, reissued: Chelsea Press, New York, 1982.

TRAUB, J. F., WASILKOWSKI, G. W., AND WOŹNIAKOWSKI, H. (1988), "Information-Based Complexity", Academic Press, New York.

TRAUB, J. F., AND WOŹNIAKOWSKI, H. (1976), Strict lower and upper bounds on iterative computational complexity, in "Analytic Computational Complexity" (Traub, J. F., ed.), 15–34, Academic Press, New York.

TRAUB, J. F., AND WOŹNIAKOWSKI, H. (1979), Convergence and complexity of Newton iteration for operator equations, J. Assoc. Comput. Mach. **26**, 250–258.

TRAUB, J. F., AND WOŹNIAKOWSKI, H. (1980), "A General Theory of Optimal Algorithms", Academic Press, New York.

Vakhania, N. N., Tarieladze, V. I., and Chobanyan, S. A. (1987), "Probability Distributions on Banach Spaces", Reidel, Dordrecht.

Wasilkowski, G. W. (1981), n-evaluation conjecture for multipoint iterations for the solution of scalar nonlinear equations, J. Assoc. Comput. Mach. **28**, 71–80.

Wasilkowski, G. W. (1983), Any iteration for polynomial equations using linear information has infinite complexity, J. Theor. Comput. Sci. **22**, 195–208.

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

Wasilkowski, G. W. (1992), Integration and approximation of multivariate functions: average case complexity with isotropic Wiener measure, Bull. Amer. Math. Soc., to appear.

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

Woźniakowski, H. (1976), Maximal order of multipoint iterations using n evaluations, in "Analytic Computational Complexity" (Traub, J. F., ed.), 75–107, Academic Press, New York.

Woźniakowski, H. (1991), Average case complexity of multivariate integration, Bull. Amer. Math. Soc. **24**, 185–194.

Woźniakowski, H. (1992), Average case complexity of linear multivariate problems, I. Theory, II. Applications, J. Complexity **8**, 337–372, 373–392.