AVERAGE-CASE OPTIMALITY
OF A HYBRID SECANT-BISECTION METHOD

ERICH NOVAK, KLAUS RITTER, AND HENRYK WOŹNIAKOWSKI

Abstract. We present an average-case complexity analysis for the zero-finding problem for functions from $C^r([0, 1])$, $r \geq 2$, which change sign at the endpoints. This class of functions is equipped with a conditional $r$-folded Wiener measure. We prove that the average-case complexity of computing an $\varepsilon$-approximation is of order $\log \log (1/\varepsilon)$, and that a hybrid secant-bisection method with a suitable adaptive stopping rule is almost optimal. This method uses only function evaluations.

We stress that the adaptive stopping rule is crucial. If one uses a nonadaptive stopping rule, then the cost has to be of order $\log (1/\varepsilon)$. Hence, the adaptive stopping rule is exponentially more powerful than arbitrary nonadaptive stopping rules.

Our algorithm is a slightly simplified version of the hybrid methods proposed by Dekker in 1969 and Bus and Dekker in 1975. These algorithms are still considered as “the best algorithms for zero-finding” by Kahaner, Moler, and Nash in their book on numerical methods.

1. Introduction

Zero-finding is a classical problem of numerical analysis. Most of the results have been obtained for the asymptotic setting in which the order of convergence and the efficiency index are studied for methods consisting of an infinite number of steps. Convergence is usually guaranteed by assuming that a good initial approximation to a root is given, see Traub [13], Ortega and Rheinboldt [8], and Ostrowski [9].

In contrast to the asymptotic setting, (global) error bounds which hold after a fixed number of steps are studied in the worst-case setting. Usually this is done without assuming a good initial approximation. Instead, error bounds are derived for some classes of functions $F$. The class $F$ is chosen in such a way that the error bounds tend to zero as the number of steps goes to infinity.

The complexity of zero-finding in the worst-case setting is understood as the minimal worst-case cost of computing an approximation with error at most $\varepsilon$ for any function from the class $F$. A survey of worst-case complexity results can be found in Sikorski [12]. In particular, for a number of classes $F$, it is known that the bisection method is optimal and the worst-case complexity is of...
order $\log(1/\varepsilon)$. This holds, for instance, for the class
$$F = \{ f \in C^r([0,1]) : f(0) \cdot f(1) < 0 \},$$
and even for the subclass of $C^\infty$-functions having only simple zeros. One
may hope, however, that other methods, such as modifications of the secant or
Newton method, should be significantly better than bisection "on the average".
A survey of average-case results may be found in Novak and Ritter [7].

In this paper we prove that, indeed, a proper modification of the secant
method is not only much better than bisection, but that it is almost optimal
on the average. We deal with the average-case setting in which the error and
cost are defined on the average with respect to a probability measure on the
class $F$ with $r \geq 2$ and with fixed boundary conditions at the endpoints. The
assumption $r \geq 2$ is needed since we use an error formula for the secant method
which depends on the behavior of the second derivatives. The case $r < 2$ is
open.

As the probability measure we choose a conditional $r$-folded Wiener measure.
This measure is obtained from the classical Wiener measure by $r$-fold integra-
tion of the paths and translation by suitable polynomials to fit the boundary
conditions. For such a measure, the set of functions with only simple roots has
full measure. However, we still have many "ill-conditioned" functions in $F$
since the probability of
$$\left\{ f \in F : \frac{\|f''\|_\infty}{\min_{x^* \in f^{-1}(0)} |f'(x^*)|} > u \right\}$$
is strictly positive for arbitrarily large $u$. For our analysis it is crucial to estab-
lish upper bounds for this probability.

We study methods which use function or derivative evaluations at sequen-
tially chosen knots, the number of which is determined by a stopping rule. The
stopping rule is nonadaptive if the number of knots is the same for all func-
tions, and it is adaptive otherwise. The goal of the method is to compute an
approximation $x$ to a root of $f \in F$ with error at most $\varepsilon$. Here the error is
understood either in the root or in the residual sense. The root sense means
that we take the distance between $x$ and the nearest root of the function $f$,
whereas the residual sense means that we take $|f(x)|$.

As is often done in numerical analysis, we do not use Turing machines or a bit
model, but prefer to work with a real-number model having infinite precision.
The average cost of a method is defined as the average number of function and
derivative evaluations plus arithmetic operations and comparisons used by the
method. Thus, the average cost is at least proportional to the average number
of knots at which function or derivative evaluations are computed. As we shall
see, for the hybrid methods presented in the paper the average cost is indeed
proportional to the average number of knots. The complexity of zerofinding in
the average-case setting is understood as the minimal average cost of computing
an approximation with average error at most $\varepsilon$.

We prove that the average-case complexity is of order $\log \log(1/\varepsilon)$ and a
hybrid secant-bisection method with a suitable adaptive stopping rule is almost
optimal. The same optimality result also holds for a hybrid Newton-bisection
method.
We stress that adaptive stopping rules do not play a role in the asymptotic or worst-case setting. In the average-case setting, they are also not important for many linear problems, see Wasilkowski [14]. For zero-finding, which is a nonlinear problem, adaptive stopping rules are very important. It is known, see Ritter [10], that any method with an average error at most $\varepsilon$ needs at least about $\log(1/\varepsilon)$ evaluations for some functions $f$ from $F$. That means that methods with nonadaptive stopping rules must use about $\log(1/\varepsilon)$ evaluations. To achieve the cost of order $\log\log(1/\varepsilon)$, optimal methods must use adaptive stopping rules. 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 indeed confirmed by the average case analysis. This also proves that adaptive stopping rules are exponentially more powerful than nonadaptive ones.

The adaptive stopping rules of the hybrid secant-bisection and Newton-bisection method are very simple. We just check whether the function value or the length of the interval which contains a zero is comparable to $\varepsilon$. These adaptive stopping rules have one additional property. Namely, they guarantee that the error is at most $\varepsilon$ for every function $f \in F$. This means that optimality of the hybrid methods is preserved even if the error is defined in the worst-case sense and with the cost defined in the average-case sense.

2. Problem formulation and main results

We study zero-finding for a class $F$ of functions $f : [0, 1] \rightarrow \mathbb{R}$ which depends on certain parameters. We denote the smoothness of functions $f$ by $r$ and assume that $r \geq 2$. We need to fix boundary values of $f$ and its derivatives in order to equip $F$ with a proper variant of the Wiener measure. We denote these boundary values by $a_i$ and $b_i$, and define

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

To guarantee that $f$ has a root, it is enough to assume that $f(0) \cdot f(1) \leq 0$, i.e., $a_0 \cdot b_0 \leq 0$. To exclude the trivial case, we assume that

$$f(0) = a_0 < 0 \quad \text{and} \quad f(1) = b_0 > 0.$$ 

Let the space $C^r([0, 1])$ be equipped with the norm

$$\| f \| = \max\{ \| f \|_\infty, \ldots, \| f^{(r)} \|_\infty \},$$

where $\| \cdot \|_\infty$ denotes the supremum norm, and consider the Borel $\sigma$-algebra on the closed subspace $F$ of $C^r([0, 1])$. The average-case setting is defined with respect to a Gaussian measure $P$ on $F$. This measure $P$, which is called a conditional $r$-folded Wiener measure, is constructed from the classical Wiener measure in the following way. By $r$-fold integration of the Brownian paths we get a Gaussian measure, called an $r$-folded Wiener measure, on the class of functions $g \in C^r([0, 1])$ with $g(0) = \cdots = g^{(r)}(0) = 0$. Let $p(g)$ denote the polynomial of degree at most $2r + 1$ such that $g - p(g) \in F$, i.e., $g - p(g)$ satisfies the respective boundary conditions. By the translation $g \mapsto g - p(g)$ of the $r$-fold integrated Brownian paths we obtain the conditional $r$-folded
Wiener measure $P$ on $F$. Basic properties of $P$, plots of $P$-random functions as well as references to applications of $r$-folded Wiener measures to problems of numerical analysis are given in Novak and Ritter [7].

We now describe zero-finding methods. As usual in numerical analysis, an approximation to a root of $f$ is obtained by computing function and possibly derivative values of $f$ at sequentially chosen knots $x_1, \ldots, x_{\nu(f)}$. The total number $\nu(f)$ of knots is determined by a stopping rule. The stopping rule is called nonadaptive if $\nu(f)$ takes the same value for all $f \in F$; otherwise the stopping rule is called adaptive. For simplicity, we assume that at all knots $x_i$ we compute the Hermite information of order $k \leq r$,

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

Here, $k$ is fixed for a particular zero-finding method. The sequential computation of the Hermite information starts at a knot

$$x_1 \in [0, 1],$$

and the selection of the remaining knots may depend on the previously computed data. This is formally described by mappings

$$\psi_i^k : \mathbb{R}^{(i-1) \times (k+1)} \to [0, 1], \quad i \geq 2.$$

In the $i$th step the knot

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

is selected and $f[x_i]$ is computed. Thus, the information

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

is known at the $i$th step. A decision to stop or to compute additional information is made after each step. This is formally described by mappings

$$\chi_i^k : \mathbb{R}^{i \times (k+1)} \to \{0, 1\}, \quad i \geq 1.$$

The total number of knots which is used for the function $f$ is given by

$$\nu(f) = \min\{i \in \mathbb{N} : \chi_i^k(N_i^k(f)) = 1\}.$$

Obviously, only the case $\nu(f) < \infty$ for all $f \in F$ is of practical interest, although in the average-case setting, we may have $\nu(f) = \infty$ for $f$ belonging to a set of measure zero. After $\nu = \nu(f)$ knots have been computed, the approximation

$$S_\nu^k(f) = \phi_\nu^k(N_\nu^k(f))$$

to a root of $f$ is constructed, where

$$\phi_i^k : \mathbb{R}^{i \times (k+1)} \to [0, 1], \quad i \geq 1.$$

Summarizing, a zero-finding method $S_\nu^k$ is formally given by a starting point $x_1$, and by the mappings $\psi_i^k$, $\chi_i^k$, and $\phi_i^k$. For simplicity, we assume that these mappings are defined over the whole space.
For lower bounds we want to allow very general methods. The only restriction on the mappings \( \psi_i^k, x_i^k, \) and \( \phi_i^k \) is Borel measurability. For upper bounds we present relatively simple methods and prove their optimality.

The error of a zero-finding method \( S_0^k \) for a function \( f \in F \) is defined either in the root sense

\[
\Delta^\text{ro}(S_0^k, f) = \inf \{|S_0^k(f) - x^*| : x^* \in f^{-1}(0)\}
\]

or in the residual sense

\[
\Delta^\text{re}(S_0^k, f) = |f(S_0^k(f))|.
\]

We simply write \( \Delta \) in the statements which hold for \( \Delta^\text{ro} \) and \( \Delta^\text{re} \).

In §3 we analyze a hybrid secant-bisection method. This method combines bisection and secant steps in a suitable way, and its computational cost is proportional to \( \nu(f) \). In particular, we prove the following theorem.

**Theorem A.** There exists a constant \( \alpha \), depending only on the regularity \( r \geq 2 \) and the boundary values \( a_i \) and \( b_i \), with the following property. For any \( \varepsilon \in (0, 1/2) \), there exists a hybrid secant-bisection method \( S_0^k \) with worst-case error

\[
\sup_{f \in F} \Delta(S_0^k, f) \leq \varepsilon
\]

whose average number of knots satisfies

\[
\int_F \nu(f) \, dP(f) \leq 1/\log((1 + \sqrt{5})/2) \cdot \log \log(1/\varepsilon) + \alpha.
\]

In §4 we prove that this hybrid method is almost optimal, even if we do not require the error bound for all \( f \in F \) but only on the average.

**Theorem B.** Let \( \beta > r + 1/2 \). There exists a constant \( \gamma \), depending only on \( \beta \), the regularity \( r \geq 2 \) and the boundary values \( a_i \) and \( b_i \), with the following property. For any \( \varepsilon \in (0, 1/2) \) and any zero-finding method \( S_0^k \) with average error

\[
\int_F \Delta(S_0^k, f) \, dP(f) \leq \varepsilon,
\]

the average number of its knots satisfies

\[
\int_F \nu(f) \, dP(f) \geq 1/\log \beta \cdot \log \log(1/\varepsilon) + \gamma.
\]

As already mentioned, the use of the average number of knots is critical here. For any method \( S_0^k \) with average error at most \( \varepsilon \), the essential supremum of \( \nu \) with respect to \( P \) is of order \( \log(1/\varepsilon) \), see Ritter [10].

**3. Hybrid methods and proof of the upper bounds**

We first describe some ideas which will allow us to find a hybrid method that is almost optimal in the average-case setting.

To explain these ideas, we restrict ourselves for a moment to the error in the residual sense and consider a hybrid Newton-bisection method. This method consists of Newton and bisection steps performed alternatively. If the Newton step cannot be performed, or if it gives an approximation outside the last
interval which contains a zero, we replace this Newton step by the bisection step.

More precisely, the hybrid method is defined as follows. Let \([s_0, t_0] = [0, 1]\). Assume inductively that \(s_{i-1}\) and \(t_{i-1}\), with \(s_{i-1} \leq t_{i-1}\), as well as the values of \(f\) and \(f'\) at these knots have been already computed. We now define \(x_i, s_i,\) and \(t_i\). Assume that \(i\) is odd, and take \(p \in \{s_{i-1}, t_{i-1}\}\) with

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

If \(q = p - f(p)/f'(p)\) is well defined and satisfies \(q \in (s_{i-1}, t_{i-1})\), then perform the Newton step \(x_i = q\). Otherwise, and for any even \(i\), perform a bisection step \(x_i = (s_{i-1} + t_{i-1})/2\). For arbitrary \(i\), the new subinterval \([s_i, t_i]\) is given by \([s_i, t_i] = [s_{i-1}, t_{i-1}]\) if \(f(x_i) > 0\) and by \([s_i, t_i] = [x_i, t_{i-1}]\) if \(f(x_i) < 0\). The stopping rule is simple. We stop if \(|f(x_i)| \leq \varepsilon\), and the output is \(x_i\). Clearly, this stopping rule is adaptive. It is well known that the error of the Newton method depends on the ratio between \(\|f''\|\infty\) and the minimum of \(|f'(x^*)|\) over the roots \(x^*\) of \(f\). For \(u > 0\), let

\[
F(u) = \left\{f \in F : \frac{\|f''\|\infty}{\min_{x^* \in f^{-1}(0)} |f'(x^*)|} \leq u \right\}.
\]

Obviously, for large \(u\) we may need to perform many bisection steps to guarantee good properties of the Newton steps. It turns out that there exists an index \(i = i(u)\) such that the points \(x_i, x_{i+2}, x_{i+4}, \ldots\) are computed by the Newton method and are not affected by the intermediate bisection steps, i.e.,

\[
x_{i+2(j+1)} = x_{i+2j} - f(x_{i+2j})/f'(x_{i+2j}), \quad j = 0, 1, \ldots.
\]

The index \(i(u)\) can be estimated by using a well-known result of Kantorovich [5]. Obviously, \(i(u) \to +\infty\) as \(u \to +\infty\). It turns out that the total number of evaluations needed for an \(\varepsilon\)-residual error is of order \(\log\log(1/\varepsilon) + i(u)\). We stress that it is not possible to determine such an index \(i(u)\) during the computation.

To estimate the average number of evaluations, it is necessary to estimate the average value of \(i(u)\). This can be done by estimating the probability of the set \(F(u)\). For the conditional \(r\)-folded Wiener measure \(P\), see §2, with \(r \geq 2\) we have

\[
P(F(u)) \geq 1 - c(\log u)^{1/2}/u \quad \forall u \geq 2,
\]

with a positive constant \(c\). (This follows from Lemmas 3.1 and 3.2 which will be proven later.)

From this estimate of \(P(F(u))\), it can be shown that the average number of evaluations of the hybrid Newton-bisection method to compute an \(\varepsilon\)-approximation in the residual sense is of order \(\log\log(1/\varepsilon)\). From Theorem B we conclude that this method is almost optimal.

We do not present a proof of this result (though this was the starting point of our paper) because we are able to analyze an improved hybrid method which takes care of some drawbacks of the presented hybrid Newton-bisection method. We list these drawbacks.

(a) We want to avoid the computation of the derivative and therefore want to replace the Newton steps by suitable secant steps.
(b) The number of bisection steps is too large. It seems reasonable to require that bisection steps are performed only at the beginning. That is, for each function there exists an index \( k \) such that bisection steps are \textit{not} performed after \( k \) steps.

(c) For the root criterion, we want to guarantee that the error is small. Hence it is not enough to find a function value which is small—instead we really need two function values \( f(x_i) \) and \( f(x_j) \) such that \( f(x_i) \cdot f(x_j) \leq 0 \) and \( |x_i - x_j| \leq 2e \). See Alefeld, Potra, Shi [1] for a study of this error criterion in the asymptotic setting. The Newton method—as well as many other methods—does not have this property, even if the function \( f \) is very smooth. To achieve a small guaranteed error in the root sense by the Newton method, it seems necessary to compute the function values at some extra “auxiliary” knots. Although this preserves the optimal order \( \log\log(1/e) \), the multiplicative constant is larger.

We now describe and analyze a hybrid method which uses the ideas presented above and is free of the drawbacks (a), (b) and (c).

Our method uses only function values, that is, the parameter \( k \) of §2 is now zero. The method is well defined on each class \( F \) of (2.1), even with regularity \( r = 0 \) or 1, although the analysis requires that \( r \geq 2 \). The method is easy to describe and also can easily be implemented on a computer. Moreover, an average-case analysis of this method uses well-known facts about the secant method.

The method is a hybrid secant-bisection method which consists of a combination of secant and bisection steps. The method computes points \( x_i \in [0, 1] \) at which \( f \) is evaluated, and subintervals \([s_i, t_i]\) \( \subset [0, 1] \). One always has \( x_i \in [s_i, t_i] \subset [0, 1, x_1, \ldots, x_i] \) and \( f(s_i) \leq 0 \leq f(t_i) \) with strict inequalities if \( s_i < t_i \). Moreover, \( x_i \in [s_{i-1}, t_{i-1}] \) and \([s_i, t_i] \subset [s_{i-1}, t_{i-1}] \).

We give a formal definition of the method in a pseudocode together with some comments. The method depends on the error criterion and on the required accuracy \( \varepsilon > 0 \) via its adaptive stopping rule \( \text{DONE}_i \) and its output \( \text{APPROX}_i \). These are very simple and given by

\[
\text{DONE}_i := ((f(x_i) - s_i) \leq 2e) \quad \text{and} \quad \text{APPROX}_i := (s_i + t_i)/2
\]

if we consider the error in the root sense, and by

\[
\text{DONE}_i := (|f(x_i)| \leq \varepsilon) \quad \text{and} \quad \text{APPROX}_i := x_i
\]

if we consider the error in the residual sense.

For the secant step, we compute

\[
\text{SEC}(x, y) := \begin{cases} 
  x - (x - y)/(f(x) - f(y)) \cdot f(x) & \text{if } f(x) \neq f(y), \\
  \text{undefined} & \text{otherwise}
\end{cases}
\]

with \( x, y \in [0, 1] \), and for the bisection step, we compute

\[
\text{BIS}_i := (s_{i-1} + t_{i-1})/2.
\]

In each step, a new subinterval

\[
\text{SUB}_i := \begin{cases} 
  [s_{i-1}, x_i] & \text{if } f(x_i) > 0, \\
  [x_i, t_{i-1}] & \text{if } f(x_i) < 0, \\
  [x_i, x_i] & \text{otherwise}
\end{cases}
\]
is constructed. The hybrid secant-bisection method is defined as follows:

\[
i := 0;
\]

\[\text{if } -a_0 < b_0 \text{ then } x_0 := 0 \text{ else } x_0 := 1 \text{ fi;}
\]

\[[s_0, t_0] := [0, 1] ;\]

\[\text{while not DONE}_i \text{ do}
\]

\[\text{for } j := 1, 2 \text{ do}
\]

\[i := i + 1 ;
\]

\[x_i := \text{SEC}(s_{i-1}, t_{i-1}) ;
\]

\[[s_i, t_i] := \text{SUB}_i ;
\]

\[\text{if DONE}_i \text{ then return APPROX}_i \text{ fi;}
\]

\[\text{od;}
\]

\[\text{repeat}
\]

\[q := \text{SEC}(x_i, x_{i-1}) ;
\]

\[\text{if } q = \text{undefined or } q \notin (s_i, t_i) \text{ then break fi;}
\]

\[i := i + 1 ;
\]

\[x_i := q ;
\]

\.[s_i, t_i] := \text{SUB}_i ;
\]

\[\text{if DONE}_i \text{ then return APPROX}_i \text{ fi;}
\]

\[\text{until } t_i - s_i > (t_{i-3} - s_{i-3})/2 ;
\]

\[i := i + 1 ;
\]

\[x_i := \text{BIS}_i ;
\]

\[[s_i, t_i] := \text{SUB}_i ;
\]

\[\text{od;}
\]

\[\text{return APPROX}_i ;
\]

Here "break" causes the method to leave the repeat-until loop. An evaluation of \( f \) only occurs when \( \text{SUB}_i \) is computed. The function values which are used in \( \text{SEC} \) are already known at that time. Each evaluation is preceded by a check of the stopping rule. Disregarding the stopping rule, the method works as follows. At the beginning and after each bisection step we perform two steps of the regula falsi starting from the endpoints of the respective subinterval \([s_{i-1}, t_{i-1}]\). Then we perform secant steps as long as they are well defined, lead into the current subinterval, and reduce the length of the subinterval by a factor at least \( 1/2 \) in every three steps. A bisection step is made if one of these conditions is violated.

Note that this hybrid method could also be defined in terms of the mappings \( \psi^0_i, \chi^0_i, \) and \( \phi^0_i \), see §2. Clearly, \( \chi^0_i \), and therefore also \( \nu \), depend on the accuracy \( \varepsilon \) and the error criterion. Without loss of generality we may assume that \( \varepsilon < \min\{1/2, -a_0, b_0\} \), which guarantees that \( f \) has to be evaluated at least once. Given \( \varepsilon \) and the error criterion, we denote the resulting hybrid secant-bisection method by \( S^0 \).

Since we halve the length of the subinterval \([s_i, t_i]\) at least in every fourth step, the hybrid method terminates for any \( f \in F \).
we have
\[ \sup_{f \in F} \nu(f) \leq 4 \cdot [1/ \log 2 \cdot \log(1/\varepsilon)], \]
while the number of steps is no longer uniformly bounded over any class \( F \) in case of the residual criterion. However \( \nu(f) = O(\log(1/\varepsilon)) \) for any \( f \in F \).

Clearly, the error of the hybrid method is at most \( \varepsilon \) for both error criteria and every \( f \in F \).

We show that the average value of \( \nu(f) \) is of order \( \log \log(1/\varepsilon) \). The proof relies on a worst-case analysis for subsets \( F_\delta \subset F \), the subsets \( F_\delta \) will be defined later, with the following properties:

(i) the probability of \( F_\delta \) is large, \( P(F_\delta) \geq 1 - \delta \),
(ii) the number of bisection steps of the method \( S_\delta \) is bounded by \( c \log(1/\delta) \) for each \( f \in F_\delta \). The constant \( c \) does not depend on the accuracy \( \varepsilon \) or on \( \delta \).
(iii) the number of secant steps is also small for \( f \in F_\delta \); it is bounded by
\[ 1/\log((1 + \sqrt{5})/2) \cdot \log \log(1/\varepsilon) + c \cdot \log(1/\delta) \]
with the same constant \( c \).

To prove this, we need a number of properties of the conditional \( r \)-folded Wiener measure \( P \) which is defined on the class \( F \), see §2. The constants \( c_j \) and \( d_j \) which will appear in our analysis will depend only on the regularity parameter \( r \) and the boundary values \( a_i, b_i \), see (2.1).

Owing to the construction of \( P \), its mean \( p \) is the polynomial of minimal degree that interpolates the boundary values. The covariance kernel \( K \) of \( P \) is of the form
\[ K(s, t) = \int_0^1 (s-u)_+^r \cdot (t-u)_+^r \, du + L(s, t) \]
for \( s, t \in [0, 1] \), where \( L \) is a polynomial of degree at most \( r + 1 \) in each variable, see Sacks and Ylvisaker [11, p. 2060].

**Lemma 3.1.** There exist constants \( c_1 \geq 1 \) and \( c_2 > 0 \) such that
\[ P \left( \left\{ f \in F : \sup_{s, t \in [0, 1]} \left| f''(s) - f''(t) \right|/|s-t|^{1/3} > u \right\} \right) \leq c_1 \cdot \exp(-c_2 \cdot u^2) \quad \forall u \geq 0. \]

Here \( 0/0 = 0 \).

**Proof.** It is well known that for any exponent \( 0 < d < 1/2 \), with probability one the Brownian paths are Hölder continuous with exponent \( d \). Let \( B \) denote the Banach space of continuous functions on \([0, 1]\) with Hölder exponent \( 1/3 \). The construction of \( P \) guarantees that \( f'' \in B \) for \( f \in F \) with probability one. Hence, the estimate follows from Ledoux and Talagrand [6, Lemma 3.1], who give an upper bound for the Gaussian measure of a ball in a Banach space. \( \square \)

**Lemma 3.2.** Let
\[ q(t) = K(t, t) \cdot K^{(2, 2)}(t, t) - K^{(2, 0)}(t, t)^2, \]
see (3.1), and
\[ c_3 = \sup_{0 < t < 1} \frac{1}{\pi \cdot q(t)^{1/2}} \cdot \exp \left( -\frac{p(t)^2}{2K(t, t)} \right). \]
Then $c_3 < \infty$ and
\[
P \left( \left\{ f \in F : \inf_{x^* \in f^{-1}(0)} |f''(x^*)| < u, \|f'\|_{\infty} \leq v \right\} \right) \leq c_3 \cdot u \cdot v \quad \forall u, v \geq 0.
\]
Let
\[
c_4 = \sup_{0 < t < 1} \frac{1}{(2\pi \cdot K(t, t))^{1/2}} \cdot \exp \left( -\frac{p(t)^2}{2K(t, t)} \right).
\]
Then $c_4 < \infty$ and
\[
P \left( \left\{ f \in F : \inf_{x^* \in f^{-1}(0)} |f'(x^*)| < u \right\} \right) \leq c_4 \cdot u \quad \forall u \geq 0.
\]

Proof. Consider the joint distribution of $f(t)$ and $f''(t)$. Its covariance matrix $Q(t)$ and mean $m(t)$ are given by
\[
Q(t) = \begin{pmatrix} K(t, t) & K^{(2, 0)}(t, t) \\ K^{(2, 0)}(t, t) & K^{(2, 2)}(t, t) \end{pmatrix}
\]
and
\[
m(t) = (p(t), p''(t))^T.
\]
Note that $q(t)$ is the determinant of $Q(t)$. Since $P$ has full support in $F$, we have $q(t) > 0$ for $0 < t < 1$. Clearly, $q$ is a polynomial, and therefore
\[
q(t) \geq \min\{t, 1 - t\}^{d_1}
\]
for a suitable $d_1 \in \mathbb{N}$. Moreover, there is a positive constant $d_2$ such that
\[
\lambda(t) \leq K(t, t) + K^{(2, 2)}(t, t) \leq d_2 \cdot \min\{t, 1 - t\},
\]
where $\lambda(t)$ is the larger eigenvalue of $Q(t)$.

For $0 < t < 1$, the density function $h(t, \cdot, \cdot)$ of the joint distribution of $f(t)$ and $f''(t)$ is given by
\[
h(t, \zeta, \eta) = \frac{1}{2\pi \cdot q(t)^{1/2}} \cdot \exp \left( -\frac{\langle (\zeta, \eta)^T - m(t), Q(t)^{-1}(\zeta, \eta)^T - m(t) \rangle}{2} \right).
\]
We claim that $h$, extended by
\[
h(0, \zeta, \eta) = h(1, \zeta, \eta) = 0,
\]
is continuous on the set
\[
A = [0, 1] \times [-\min\{-a_0, b_0\}/2, \min\{-a_0, b_0\}/2] \times \mathbb{R}.
\]
Since $Q(t)$ and $m(t)$ depend continuously on $t$, the density $h$ is continuous on $(0, 1) \times \mathbb{R}^2$. If $(t_i, \zeta_i, \eta_i) \in A$ tend to $(0, \zeta, \eta) \in A$, then $\lim_{t \to 0} \|Q(t_i) - m(t_i)\|_2 \geq |\zeta - a_0| \geq |a_0|/2$ and
\[
0 \leq \lim_{i \to \infty} h(t_i, \zeta_i, \eta_i) \leq \lim_{i \to \infty} \frac{1}{2\pi \cdot (q(t_i))^{1/2}} \cdot \exp \left( -\frac{|a_0|^2}{8\lambda(t_i)} \right) = 0.
\]
Similarly, we conclude the continuity of $h$ at $(1, \zeta, \eta) \in A$.

Define
\[
F_1 = \left\{ f \in F : \inf_{x^* \in f^{-1}(0)} |f''(x^*)| < u, \|f'\|_{\infty} \leq v \right\}
\]
and
\[ C_i = \bigcup_{k=1}^{i-1} \{ f \in F : |f''(k/i)| < u, |f(k/i)| \leq v/(2i) \}. \]

Then continuity of \( f'' \) and the inequality \( |f(x)| \leq v|x-x^*| \) for \( f \in F_1 \) yield
\[ F_1 \subset \bigcup_{j=1}^{\infty} \bigcap_{i=j}^{\infty} C_i, \]
which implies
\[ P(F_1) \leq \lim_{i \to \infty} \inf P(C_i) \leq \lim_{i \to \infty} \inf (i-1) \cdot 2u \cdot v/i \cdot \sup_{0 < i < 1} \sup_{|\zeta| \leq v/(2i)} \sup_{|\eta| < u} \sup h(t, \zeta, \eta) \]
\[ \leq 2uv \cdot \sup_{0 < i < 1} \sup_{\eta \in \mathbb{R}} h(t, 0, \eta). \]

For a fixed \( 0 < t < 1 \), we directly compute
\[ \sup_{\eta \in \mathbb{R}} h(t, 0, \eta) = \frac{1}{2\pi \cdot q(t)^{1/2}} \cdot \exp \left( -\frac{p(t)^2}{2K(t, t)} \right). \]

Since \( K(t, t) \leq \lambda(t) \) and since \( p(0) = a_0 \) and \( p(1) = b_0 \) are nonzero, we obtain
\[ \lim_{t \to 0} \sup_{\eta \in \mathbb{R}} h(t, 0, \eta) = \lim_{t \to 1} \sup_{\eta \in \mathbb{R}} h(t, 0, \eta) = 0 \]
and therefore
\[ c_3 = 2 \cdot \sup_{0 < t < 1} \sup_{\eta \in \mathbb{R}} h(t, 0, \eta) < \infty. \]

This completes the proof of the first estimate.

Define
\[ F_2 = \left\{ f \in F : \inf_{x^* \in f^{-1}(0)} |f'(x^*)| < u \right\} \]
and
\[ D_i = \bigcup_{k=1}^{i-1} \{ f \in F : |f(k/i)| \leq u/(2i) \}. \]

Arguing as in the first part of the proof, we find that
\[ P(F_2) \leq \lim_{i \to \infty} \inf P(D_i) \leq \sup_{0 < t < 1} \frac{1}{(2\pi \cdot K(t, t))^{1/2}} \cdot \exp \left( -\frac{p(t)^2}{2K(t, t)} \right) \cdot u < \infty. \]

Remark 3.1. As a consequence of Lemma 3.2, the following properties hold
with probability one on \( F \). A function \( f \in F \) has only simple zeros, and therefore the number of zeros of \( f \) is finite. A function \( f \) is strictly convex or
strictly concave in small neighborhoods of its zeros.

Now we define the sets \( F_\delta \) which will be considered in a worst-case analysis.

Let
\[ F(u, v, w) = \left\{ f \in F : \sup_{s, t \in [0, 1]} \frac{|f''(s) - f''(t)|}{|s - t|^{1/3}} \leq u, \inf_{x^* \in f^{-1}(0)} |f'(x^*)| \geq v, \inf_{x^* \in f^{-1}(0)} |f''(x^*)| \geq w \right\}. \]

Define
\[ F_\delta = F(u_\delta, v_\delta, w_\delta), \]
where
\[ u_\delta = \left(1/c_2 \cdot \log(3c_1/\delta)\right)^{1/2}, \quad v_\delta = \delta/(3c_4), \quad w_\delta = \delta/(3c_3 \cdot (|a_1| + |a_2| + u_\delta)) \]
with the constants from Lemmas 3.1 and 3.2.

**Lemma 3.3.** If \( 0 < \delta \leq 1 \), then
\[ P(F_\delta) \geq 1 - \delta. \]

**Proof.** Observe that
\[ \sup_{t, s \in [0, 1]} \frac{|f''(s) - f''(t)|}{|s - t|^{1/3}} \leq u \]
implies
\[ ||f'||_\infty \leq |a_1| + |a_2| + u. \]
Therefore,
\[ P(F(u, v, w)) \geq 1 - P \left( \left\{ f \in F : \sup_{s, t \in [0, 1]} \frac{|f''(s) - f''(t)|}{|s - t|^{1/3}} > u \right\} \right) \]
\[ = 1 - P \left( \left\{ f \in F : \inf_{x^* \in f^{-1}(0)} |f'(x^*)| < v \right\} \right) \]
\[ = 1 - P \left( \left\{ f \in F : \inf_{x^* \in f^{-1}(0)} |f''(x^*)| < w, \|f'||_\infty \leq |a_1| + |a_2| + u \right\} \right) \]
\[ \geq 1 - c_1 \cdot \exp(-c_2 \cdot u^2) - c_4 \cdot v - c_3 \cdot w \cdot (|a_1| + |a_2| + u) \]
by Lemmas 3.1 and 3.2, and \( P(F_\delta) \geq 1 - \delta \) follows. \( \square \)

We are ready to analyze the behavior of the hybrid method \( S_0^\delta \) on the sets \( F_\delta \) as well as on the whole class \( F \). We present a theorem whose part includes Theorem A.

**Theorem 3.1.** There exist constants \( c_5 \) and \( \alpha \), depending only on the regularity \( r \geq 2 \) and the boundary values \( a_i \) and \( b_i \), with the following property. For any \( \varepsilon \in (0, \min\{1/2, -a_0, b_0\}) \) and for any \( 0 < \delta \leq 1/2 \), the hybrid secant-bisection method \( S_0^\delta \) satisfies
\[ \sup_{f \in F_\delta} \nu(f) \leq 1/\log((1 + \sqrt{5})/2) \cdot \log(1/\varepsilon) + c_5 \cdot \log(1/\delta) \]
and
\[ \int_F \nu(f) \, dP(f) \leq 1/\log((1 + \sqrt{5})/2) \cdot \log(1/\varepsilon) + \alpha. \]

**Proof.** Let \( f \in F_\delta \), and disregard the stopping rule for a moment. Consider the infinite sequences of points \( x_i, s_i, \) and \( t_i \) which are defined by the hybrid method. Without loss of generality we assume that \( f(x_i) \neq 0 \) for all \( i \). Let
\[ n = n_\delta = 4 \max \left\{ \left[ 3 \log_2 \frac{u_\delta}{w_\delta} \right] + 1, \left[ \log_2 \frac{2(|a_2| + u_\delta)}{v_\delta} \right] \right\} \]
and \( l = t_n - s_n > 0 \), and observe that
\[ \nu_\delta \leq 4 \cdot \log(1/\delta) \]
with a suitable constant \( d_1 > 0 \). Since \( t_i - s_i \leq (t_{i-4} - s_{i-4})/2 \) for any \( i \geq 4 \), the length \( l \) of the interval \([s_n, t_n]\) is at most

\[
l \leq 2^{-n/4} \leq \frac{1}{2(\|a_2\| + u_\delta)}.
\]

Let \( x^* \in [s_n, t_n] \) be a zero of \( f \). For \( x \in [s_n, t_n] \) and a suitable \( \zeta \in [s_n, t_n] \), we have

\[
|f'(x)| = \left| f'(x^*) + f''(0)(x - x^*) + \frac{f''(\zeta) - f''(0)}{\zeta^{1/3}}(x - x^*) \right|
\geq \frac{1}{2} - (\|a_2\| + u_\delta)l \geq \frac{1}{2l} > 0.
\]

From \( f(s_n) < f(t_n) \) we conclude that \( f' \) is strictly positive on \([s_n, t_n]\). In particular, this means that \( x^* \) is the unique zero of \( f \) in this interval. Similarly,

\[
|f''(x)| = \left| f''(0) + \frac{f''(x) - f''(0)}{x^{1/3}}x^{1/3} \right| \leq |a_2| + u_\delta,
\]

\[
|f''(x)| = \left| f''(x^*) + \frac{f''(x) - f''(x^*)}{(x - x^*)^{1/3}}(x - x^*)^{1/3} \right|
\geq u_\delta - u_\delta l^{1/3} \geq \frac{1}{2} - 2^{-n/12} > 0,
\]

where the strict inequality is due to the definition of \( n \). Moreover,

\[
\sup_{\zeta_1, \zeta_2 \in [s_n, t_n]} \frac{|f''(\zeta_1)|}{|f''(\zeta_2)|} \leq \frac{1}{2l}.
\]

In what follows we assume that \( f'' \) is strictly positive on \([s_n, t_n]\). The case \( f'' < 0 \) can be studied analogously.

For any two distinct points \( x, y \in [s_n, t_n] \) the secant step

\[
q = \text{SEC}(x, y) = x - (x - y)/(f(x) - f(y)) \cdot f(x)
\]
is well defined. The known error formula for the secant method yields

\[
q - x^* = C(x, y)(x - x^*)(y - x^*),
\]

where \( C(x, y) = f''(\zeta_1)/(2f'(\zeta_2)) \) for some \( \zeta_1, \zeta_2 \in [s_n, t_n] \). Observe that

\[
0 < C(x, y) \leq \frac{1}{2l}
\]
and

\[
|q - x^*| \leq \frac{1}{2} \min\{|x - x^*|, |y - x^*|\}.
\]

Define \( e_i = |x_i - x^*| \) and \( l_i = t_i - s_i \), and let \( \sigma_i \) denote the sign of \( x_i - x^* \).

**Claim 1.** Assume that \( x_{i-2} = \text{BIS}_{i-2} \) or \( x_i = \text{SEC}(x_{i-1}, x_{i-2}) \) with \( i \geq n + 2 \). Then \( \text{SEC}(x_i, x_{i-1}) \in (s_i, t_i) \).

Let \( q = \text{SEC}(x_i, x_{i-1}) \) and assume that \( x_{i-2} = \text{BIS}_{i-2} \) at first. Then \( x_{i-1} \) and \( x_i \) are computed by the regula falsi and \( (\sigma_{i-1}, \sigma_i) = (-, -) \) by (3.4). Therefore, \( t_i = t_{i-2} \) and \( q > x^* \), and (3.5) yields \( q - x^* \leq (t_i - x^*)/8 \). We conclude that \( q \in (x_i, t_i) \).

Now assume that \( x_i = \text{SEC}(x_{i-1}, x_{i-2}) \). If \( (\sigma_{i-2}, \sigma_{i-1}) = (-, +) \), then \( x_i, q < x^* \) and \( x^* - q \leq e_i/2 = (x^* - s_i)/2 \). Hence, \( q \in (s_i, t_i) \). In the cases \( (\sigma_{i-2}, \sigma_{i-1}) = (-, -), (+, -), \) or \( (+, +) \), Claim 1 may be verified similarly.
During the computation each secant step $x_i = \text{SEC}(x_{i-1}, x_{i-2})$ is followed by a check of the condition $l_i \leq l_{i-3}/2$. We study this condition for the following two cases: four consecutive secant steps or six steps with one bisection step.

**Claim 2.** Assume that $x_i = \text{SEC}(x_{i-j}, x_{i-j-2})$ for $j = 0, \ldots, 3$ and $(\sigma_{i-1}, \sigma_i) \in \{(-, -), (+, -), (-, +)\}$ with $i \geq n + 5$. Then $l_i \leq l_{i-3}/2$.

If $(\sigma_{i-1}, \sigma_i) = (-, -)$, then $(\sigma_{i-5}, \ldots, \sigma_{i-2}) = (+, -, -, +)$ by (3.4). Therefore, $[s_i, t_i] = [x_i, x_{i-2}]$ and $[s_{i-3}, t_{i-3}] = [x_{i-3}, x_{i-5}]$, and (3.5) implies

$$l_i = e_i + e_{i-2} \leq (e_{i-3} + e_{i-5})/2 = l_{i-3}/2.$$  

In the cases $(\sigma_{i-1}, \sigma_i) = (+, -)$ or $(-, +)$, Claim 2 may be verified similarly.

**Claim 3.** Assume that $x_i = \text{SEC}(x_{i-1}, x_{i-2})$ and $x_{i-j} = \text{BIS}_{i-j}$ for $j = 3, 4$, or 5 with $i-j \geq n$. Then $l_i \leq l_{i-3}/2$.

If $x_{i-3} = \text{BIS}_{i-3}$, then we already know that $(\sigma_{i-2}, \sigma_{i-1}, \sigma_i) = (-, -, +)$ and $t_{i-1} = t_{i-3}$. Therefore, $l_{i-3} \geq e_{i-2} + e_{i-3} - x^*$ and (3.5) yields

$$l_i = e_i + e_{i-1} \leq (t_{i-3} - x^*)/8 + e_{i-2}/2 \leq l_{i-3}/2.$$  

A similar argument works for $j = 4$ or 5.

As a consequence of Claims 1-3 we get the following. If $x_{i_0} = \text{BIS}_{i_0}$ for $i_0 \geq n$, then no further bisection step will be made. Indeed, the next two steps $i_0 + 1$ and $i_0 + 2$ must be performed by the regula falsi. Then the step $i_0 + 3$ will be done by the secant method because of Claim 1. The steps $i_0 + 4, i_0 + 5$, and $i_0 + 6$ are secant steps by Claims 1 and 3, and $(\sigma_{i_0+5}, \sigma_{i_0+6}) = (-, +)$. Finally, the step $i_0 + 7$ and all the remaining steps are secant steps by Claims 1 and 2.

Note that the lack of bisection steps for $i > i_0$ implies that the sequence $(\sigma_{i_0+1}, \sigma_{i_0+2}, \ldots)$ does not contain three consecutive + or three consecutive −. Hence, $l_{i+2} \leq e_{i+2} + e_i$, and (3.5) implies

$$l_{i+2} \leq 5/4e_i \quad \forall i \geq i_0 + 1.$$  

To find an index $i$ such that $l_i \leq 2\epsilon$, we consider signs of $(\sigma_n, \sigma_{n+1})$. Let $M = (1 + \sqrt{5})/2$, and assume first that

$$(\sigma_n, \sigma_{n+1}) \in \{(-, -), (+, -), (-, +)\}.$$  

If one of the knots $x_n, \ldots, x_{n+5}$ is computed by bisection, then (3.6) holds with $i_0 = n + 5$. Otherwise, $x_i = \text{SEC}(x_{i-1}, x_{i-2})$ for $i \geq n + 2$ by Claims 1 and 2, and the sequence $\sigma_n, \sigma_{n+1}, \ldots$ does not contain three consecutive + or three consecutive −. Hence, (3.6) holds with $i_0 = n + 5$ in both cases, and

$$e_i \leq e_{i-1}e_{i-2}/(2l) \quad \forall i \geq n + 8,$$  

follows from (3.4). Clearly, $e_{n+6}, e_{n+7} \leq l$. Using the well-known direct estimate for $e_i$, we have

$$\min\{i \in \mathbb{N} : e_i \leq \epsilon, i \geq n + 6\} \leq 1/\log M \cdot \log \log(1/\epsilon) + n + d_2$$  

with a positive constant $d_2$. From (3.6) we obtain

$$\min\{i \in \mathbb{N} : l_i \leq 2\epsilon\} \leq 1/\log M \cdot \log \log(1/\epsilon) + n + d_2 + 2.$$
Now we assume that
\[(\sigma_n, \sigma_{n+1}) = (+, +).\]
In this case there exists \(i_0 \geq n + 1\) such that \(x_{i_0} = \text{BIS}_{i_0}\), and no further bisection step is made. Clearly,
\[e_i \leq e_{i-1}e_{i-2}/(2l) \quad \forall n + 2 \leq i < i_0,
\]
while we only know that \(e_{i_0} \leq l/2\). However,
\[e_n \geq e_{n+1} \geq \cdots \geq e_{i_0-1} \geq \min\{e_{i_0-1}, e_{i_0}\} \geq e_{i_0+1} \geq e_{i_0+2} \geq \cdots,
\]
which implies
\[e_{i_0+3} \leq e_{i_0+2}e_{i_0+1}/(2l) \leq e_{i_0-1}e_{i_0-2}/(2l)\]
if \(i_0 > n + 1\) and
\[e_{i_0+4} \leq e_{i_0+3}e_{i_0+2}/(2l) \leq e_{i_0+3}e_{i_0-1}/(2l).
\]
Finally,
\[e_i \leq e_{i-1}e_{i-2}/(2l) \quad \forall i \geq i_0 + 5.
\]
The estimates above show that we lose at most three secant steps because of the bisection step. Therefore,
\[(3.9) \quad \min\{i \in \mathbb{N} : e_i \leq \epsilon, i \geq n\} \leq 1/(\log M \cdot \log \log(1/\epsilon)) + n + d_3
\]
with a constant \(d_3 > 0\).
Let \(\lambda = x^*-s_n\). Observe that \(l_i \leq \lambda + e_i\) holds for \(n \leq i \leq i_0\) and \(i \geq i_0 + 3\).
If \(\lambda \leq \epsilon\), then \(e_i \leq \epsilon\) with \(i \geq n\) implies \(l_{i+2} \leq 2\epsilon\), and (3.9) yields
\[(3.10) \quad \min\{i \in \mathbb{N} : l_i \leq 2\epsilon\} \leq 1/(\log M \cdot \log \log(1/\epsilon)) + n + d_3 + 2.
\]
Consider the case \(\lambda > \epsilon\). If \(n + 3 \leq i < i_0\) and \(e_{i-3} \leq \epsilon\), then
\[\frac{l_i}{l_{i-3}} = \frac{\lambda + e_i}{\lambda + e_{i-3}} > \frac{\epsilon}{\epsilon + e_{i-3}} \geq 1/2.
\]
Hence, (3.9) shows that
\[i_0 \leq 1/(\log M \cdot \log \log(1/\epsilon)) + n + d_3 + 4,
\]
and we get
\[(3.11) \quad \min\{i \in \mathbb{N} : l_i \leq 2\epsilon\} \leq 1/(\log M \cdot \log \log(1/\epsilon)) + n + d_3 + 7,
\]
using (3.6).
Combining (3.3), (3.8), (3.10), and (3.11), we obtain
\[\min\{i \in \mathbb{N} : l_i \leq 2\epsilon\} \leq 1/(\log M \cdot \log \log(1/\epsilon) + d_4 \cdot \log(1/\delta))
\]
for any \(f \in F_\delta\) with a suitable constant \(d_4\). For the root criterion, this yields the needed worst-case estimate on \(\sup_{f \in F_\delta} \nu(f)\). For the residual criterion, (3.2) yields
\[|f(x_i)| \leq \|f'\|_{\infty} \cdot e_i \leq (|a_1| + |a_2| + (1/c_2 \log(3c_1/\delta))^{1/2}) \cdot e_i.
\]
Thus, we terminate for the residual criterion if \(e_i = O(\epsilon/\sqrt{\log 1/\delta})\). We get the needed estimate from the root criterion with \(\epsilon\) replaced by \(O(\epsilon/\sqrt{\log 1/\delta})\).
We now estimate the average value of \(\nu(f)\) over the class \(F\). We know that
\[\sup_{f \in F_\delta} \nu(f) \leq A + c_5 \log(1/\delta)
\]

for any \( 0 < \delta \leq 1/2 \), with a positive constant \( c_5 \) and \( A = 1/ \log M \cdot \log \log (1/\epsilon) \). Let \( B = [A + 1 + c_5 \log 2] \). Assume that \( n \geq B \) and define

\[
\delta = \exp(-(n - A - 1)/c_5).
\]

Then \( \delta \leq 1/2 \), and \( \nu(f) \geq n \) implies \( f \notin F_\delta \). Hence, \( P(\nu(f) \geq n) \leq \delta \) follows from Lemma 3.3, and we finally obtain

\[
\int_F \nu(f) \, dP(f) = \sum_{n=1}^{\infty} P(\nu(f) \geq n) \leq B - 1 + \sum_{n=B}^{\infty} \exp(-(n - A - 1)/c_5) \leq A + \alpha
\]

with a suitable constant \( \alpha \). \( \square \)

4. PROOF OF THE LOWER BOUNDS

In this section we prove Theorem B. We start with a lemma concerning interpolation by polynomials. As in §2, for \( g \in C^r([0, 1]) \) denote

\[
g[s] = (g(s), \ldots, g^{(r)}(s)).
\]

For \( 0 \leq s < t \leq 1 \), let \( p(g; s, t) \) be the polynomial of degree at most \( 2r + 1 \) which satisfies

\[
p(g; s, t)[s] = g[s] \quad \text{and} \quad p(g; s, t)[t] = g[t].
\]

**Lemma 4.1.** Let \( r \geq 1 \). There exists a constant \( L = L(r) > 0 \) such that

\[
\sup_{s \leq x \leq t} |g'(x) - p'(g; s, t)(x)| \leq L \cdot (t - s)^{r-1} \sup_{s \leq x \leq t} |g^{(r)}(x)|
\]

for any \( g \in C^r([0, 1]) \) and \( 0 \leq s < t \leq 1 \).

**Proof.** The operator \( g \mapsto p(g; 0, 1) \) is linear and bounded on \( C^r([0, 1]) \). Hence, there is a constant \( L > 0 \) such that

\[
\|g^{(r)} - p^{(r)}(g; 0, 1)\|_{\infty} \leq L/2 \cdot \|g^{(r)}\|_{\infty}
\]

for any \( g \in C^r([0, 1]) \) with \( g[0] = 0 \). For an arbitrary \( g \in C^r([0, 1]) \), let \( \tilde{g} \) denote the Taylor polynomial of degree at most \( r \) with \( \tilde{g}[0] = g[0] \). Then

\[
\|g^{(r)} - p^{(r)}(g; 0, 1)\|_{\infty} = \|(g - \tilde{g})^{(r)} - p^{(r)}(g - \tilde{g}; 0, 1)\|_{\infty} \leq L/2 \cdot \|(g - \tilde{g})^{(r)}\|_{\infty} \leq L \cdot \|g^{(r)}\|_{\infty}.
\]

For \( 0 \leq s < t \leq 1 \), let \( T = t - s \), and define \( \overline{g}(z) = g(s + z \cdot T)/T^r \), \( z \in [0, 1] \). Then

\[
\overline{g}^{(j)}(z) = g^{(j)}(s + z \cdot T) \cdot T^{j-r}, \quad j = 0, 1, \ldots, r.
\]

Therefore, \( \overline{p}(g; s, t) = p(\overline{g}; 0, 1) \). Hence,

\[
\sup_{s \leq x \leq t} |g^{(r)}(x) - p^{(r)}(g; s, t)(x)| = \sup_{0 \leq z \leq 1} |\overline{g}^{(r)}(z) - \overline{p}^{(r)}(g; s, t)(z)| = \|\overline{g}^{(r)} - p^{(r)}(\overline{g}; 0, 1)\|_{\infty} \leq L \cdot \|\overline{g}^{(r)}\|_{\infty} = L \cdot \sup_{s \leq x \leq t} |g^{(r)}(x)|.
\]
Since \( g' - p'(g; s, t) \) has a root of multiplicity \( r \) at \( s \), we have

\[
\sup_{s \leq x \leq t} |g'(x) - p'(g; s, t)(x)| \leq \frac{T^{r-1}}{(r-1)!} \cdot \sup_{s \leq x \leq t} |g^{(r)}(x) - p^{(r)}(g; s, t)(x)|
\]

\[
\leq L \cdot T^{r-1} \cdot \sup_{s \leq x \leq t} |g^{(r)}(x)|,
\]

which completes the proof. \( \square \)

We also need the following lemma on the absolute value of a Gaussian random variable.

**Lemma 4.2.** There exists a positive constant \( K_1 \) such that

\[
E(|Z|) \geq K_1 \cdot \max\{|m|, \sigma\}
\]

for any Gaussian random variable \( Z \) with mean \( m \) and variance \( \sigma^2 \).

If additionally \( m \neq 0 \) or \( \sigma \neq 0 \), then there exists a positive constant \( K_2 \) such that

\[
\text{Prob}(\{ |Z| \leq u \cdot E(|Z|) \}) \leq K_2 \cdot u \quad \forall u \geq 0.
\]

**Proof.** Since \( E(|Z|) \) is a convex and even function of the mean \( m \), we get \( E(|Z|) \geq K_1 \cdot \sigma \), where \( K_1 = (2\pi)^{-1/2} \int_{-\infty}^{\infty} |v| \cdot \exp(-v^2/2) \, dv \leq 1 \). Moreover, \( E(|Z|) \geq |E(Z)| = |m| \), and the first estimate follows.

It is sufficient to show the second estimate for \( 0 \leq u \leq 1/2 \). The estimate is obvious for \( \sigma = 0 \), and it remains to consider the case \( \sigma = 1 \) and \( m > 0 \). If \( m \leq 1 \), then \( E(|Z|) \leq m + K_1 \leq 2 \), and

\[
\text{Prob}(\{ |Z| \leq u \cdot E(|Z|) \}) \leq 4(2\pi)^{-1/2} \cdot u.
\]

If \( m \geq 1 \), then \( E(|Z|) \leq 2m \), and \( \text{Prob}(\{ |Z| \leq u \cdot 2m \}) \) is maximal for \( m = 1 \). Therefore, the above estimate also holds in this case. \( \square \)

For fixed regularity \( r \geq 2 \) and boundary values \( a_i, b_i \) we consider the conditional \( r \)-folded Wiener measure \( P \) on the class \( F \) defined in §2. We also fix \( \beta > r + 1/2 \). In what follows, we use some constants \( c_j \) and \( d_j \) which only depend on the parameters \( a_i, b_i, r, \) and \( \beta \).

We can choose a constant \( c_1 > 0 \) such that the class

\[
G = \{ g \in F : \|g\| \leq c_1 \}
\]

satisfies

\[
\frac{\log \beta + \log(r + 1/2)}{2 \log \beta} \leq P(G)^2 < 1.
\]

Indeed, since \( P(G) \to 1 \) for \( c_1 \to \infty \) and since the left-hand side of (4.1) is smaller than 1, we can satisfy (4.1) for large \( c_1 \). Recall that by \( \|g\| \) we mean here \( \max_{0 \leq j \leq r} \|g^{(j)}\|_\infty \).

We first consider information operators \( N^r_f \) that use Hermite information of order \( r \), see (2.2). Without loss of generality we assume that \( N^r_f \) uses different knots from \((0, 1)\) for any function from \( F \). Since the knots usually do not depend monotonically on the step number, we let

\[
0 < z_{i,1}(y) < \cdots < z_{i,r}(y) < 1
\]

as the increasing rearrangement of the knots used for \( f \in F \), where \( N^r_f(y) = y \). We also set \( z_{i,0}(y) = 0 \) and \( z_{i,i+1}(y) = 1 \). Observe that the knots \( z_{i,j}(y) \) may
only depend on $N_{i-1}(f)$, i.e., on the data which are computed in the first $i - 1$
steps.

Let $f \in F$ be any function with $N_i'(f) = y$. It is well known that the
conditional measures $P(\cdot | N_i' = y)$ are Gaussian with the following prop-
erties. On each of the subintervals $[s, t] = [z_{i,j-1}(y), z_{i,j}(y)]$, the mean of
$P(\cdot | N_i' = y)$ is given by the polynomial $p(f; s, t)$. Function evaluation at
$x \in [s, t]$ has variance

$$
\int_F (f(x) - p(f; s, t)(x))^2 dP(f|N_i' = y)
\tag{4.2}
$$

If $x, v \in [0, 1]$ lie in different subintervals, then function evaluation at $x$ and
$v$ are independent with respect to $P(\cdot | N_i' = y)$.

Consider a zero-finding method $S_i = \phi_i' \circ N_i$ that uses a fixed number $i$
of
knots for any function from $F$. Given the data $N_i' = y$, the method determines
a point $x$ from $[0, 1]$ as an approximation to a root, and the average residual
error is the mean of the absolute value of a Gaussian random variable. If
$x \in [s, t] = [z_{i,j-1}(y), z_{i,j}(y)]$, the average residual error is bounded from
below by

$$
e_{i,j}(y) = \inf_{s \leq x \leq t} \int_F |f(x)| dP(f|N_i' = y).
$$

Since $x$ is in some subinterval, the average residual error is bounded from
below by

$$
e_i(y) = \min_{j = 1, \ldots, i+1} e_{i,j}(y).
$$

The following lemma relates a lower bound on $e_{i,j}(y)$ in terms of the boundary
values of the function at the respective subinterval.

**Lemma 4.3.** There exists a positive constant $c_2$ such that for any information
operator $N_i'$ and any $g \in G$ we have

$$
e_{i,j}(y) \geq c_2^{r+1/2} \cdot \min\{|g(z_{i,j-1}(y))|, |g(z_{i,j}(y))|\}^{r+1/2},
$$

where $y = N_i'(g)$.

**Proof.** Let $[s, t] = [z_{i,j-1}(y), z_{i,j}(y)]$ and $p = p(g; s, t)$. Assume that $x \in
[s, t]$. The conditional mean and variance of $f(x)$, given $N_i' = y$, are equal
to $p(x)$ and $\sigma^2(x)$, the latter defined by (4.2). Clearly,

$$
\sigma(x) \geq d_1^{1/2} \cdot \min\{x - s, t - x\}^{r+1/2}
$$

with $d_1 = (2r + 1)^{-1/2} (r!)^{-2}$. Lemma 4.1 and $\|g\| \leq c_1$ imply that

$$
\sup_{s \leq x \leq t} |p'(x)| \leq d_2
$$

with $d_2 = c_1 \cdot (L + 1)$.

Let $M = \min\{|g(s)|, |g(t)|\}$. If

$$
\min\{x - s, t - x\} \leq M/(2d_2),
$$

then

$$
|p(x)| \geq M - M/(2d_2) \cdot \sup_{s \leq x \leq t} |p'(x)| \geq M/2,
\tag{4.3}
$$
and otherwise

\[ \sigma(x) \geq d_1^{1/2} d_2^{-r-1/2} 2^{-r-1/2} \cdot M^{r+1/2}. \]  

From Lemma 4.2 we know that

\[ \int_F |f(x)| \, dP(f|N_i' = y) \geq K_1 \cdot \max\{|p(x)|, \sigma(x)\}. \]

Hence, \( M \leq \|g\|_\infty \leq c_1 \), and (4.3) and (4.4) imply

\[ \int_F |f(x)| \, dP(f|N'_i = y) \geq c_2^{r+1/2} \cdot M^{r+1/2} \]

for a suitable constant \( c_2 > 0 \). \( \square \)

The distribution of \( e_i(N'_i(g)) \) on the class \( G \) is studied in the following lemma.

**Lemma 4.4.** Let \( \beta(i) = \sum_{j=1}^i (r+1/2)^j \). There exists a positive constant \( c_3 \) such that for any information operator \( N'_i \) we have

\[ P\{ e_i(N'_i(g)) \leq u \} \cap G \leq c_3 \cdot i \cdot u \quad \forall u \geq 0. \]

**Proof.** We prove the lemma by induction on \( i \). Lemma 4.3 and \( N'_i(f) = f[x_1] \) imply that

\[ P\{ e_1(N'_i(g)) \leq u^{r+1/2} \} \cap G \leq P\{ c_2 \cdot \min\{|a_0|, |f(x_1)|, |b_0|\} \leq u \} \cap G. \]

Clearly,

\[ d_1 = \inf_{0 \leq x \leq 1} \int_F |f(x)| \, dP(f) > 0, \]

and Lemma 4.2 implies

\[ P\{ |f(x_1)| \leq u/c_2 \} \leq K_2/(c_2d_1) \cdot u. \]

Hence,

\[ P\{ e_1(N'_i(g)) \leq u^{1/2} \} \cap G \leq d_2 \cdot u \]

for any \( u \geq 0 \) with a suitable constant \( d_2 > 0 \).

Let \( y = N'_i(g) \) for \( g \in G \). In the next step \( i + 1 \), we evaluate \( \rho = g[\xi] \), where

\[ \xi = \psi_{i+1}'(y) \in (z_{i,j-1}(y), z_{i,j}(y)) = (s, t). \]

Clearly,

\[ \inf_{0 \leq x \leq 1} \int_F |f(x)| \, dP(f|N'_{i+1} = (y, \rho)) \geq e_i(y). \]

By Lemma 4.3, we have

\[ \inf_{s \leq x \leq t} \int_F |f(x)| \, dP(f|N'_{i+1} = (y, \rho)) \geq c_2^{r+1/2} \cdot \min\{|g(s)|, |g(t)|\}^{r+1/2} \]

and

\[ \inf_{\xi \leq x \leq t} \int_F |f(x)| \, dP(f|N'_{i+1} = (y, \rho)) \geq c_2^{r+1/2} \cdot \min\{|g(\xi)|, |g(t)|\}^{r+1/2}. \]
This implies that
\[ e_{i+1}(N_{i+1}^r(g)) = e_{i+1}(y, \rho) \geq \min\{e_i(y), c_2^{r+1/2} \cdot \min\{|g(s)|, |g(t)|, |g(\xi)|\}^{r+1/2}\}, \]
and therefore
\[
P\{\{e_{i+1}(N_{i+1}^r(g)) \geq (u \cdot e_i(y))^{r+1/2}\} \cap G|N_i^r = y) \leq P\{\{e_i(y) \leq (u \cdot e_i(y))^{r+1/2}\} \cap G|N_i^r = y) + P\{\{\min\{|g(s)|, |g(t)|\} \leq u \cdot e_i(y)/c_2\} \cap G|N_i^r = y) + P\{\{|g(\xi)| \leq u \cdot e_i(y)/c_2\} \cap G|N_i^r = y)\}.
\]
Assume that \(e_i(y) > 0\), and let \(h(y) = \int_F |f(\xi)| dP(f|N_i^r = y)\). Clearly,
\[e_i(y) \leq \min\{|a_0|, |b_0|, |g(s)|, |g(t)|, h(y)|. \]
If \(0 < u < \min\{c_2, \min\{|a_0|, |b_0|\}^{(2r+1)/(2r+1)}\}\), then Lemma 4.2 yields
\[
P\{\{e_{i+1}(N_{i+1}^r(g)) \geq (u \cdot e^r(y))^{r+1/2}\} \cap G|N_i^r = y) \leq P\{\{|h(\xi)| \leq u \cdot h(y)/c_2\} \cap G|N_i^r = y) \leq K_2/c_2 \cdot u.
\]
Hence,
\[
P\{\{e_{i+1}(N_{i+1}^r(g)) \geq (u \cdot e^r(y))^{r+1/2}\} \cap G|N_i^r = y) \leq d_3 \cdot u
\]
for any \(u \geq 0\) with a suitable constant \(d_3 > 0\).
Define \(c_3 = \max\{d_2, d_3\}\), see (4.5). From (4.6) we get
\[
P\{\{e_{i+1}(N_{i+1}^r(g)) > (u \cdot u^{\beta(i)})^{r+1/2}\} \cap G) = \int_{N_i^r(F)} P\{\{e_{i+1}(N_{i+1}^r(g)) > (u \cdot u^{\beta(i)})^{r+1/2}\} \cap G|N_i^r = y) dN_i^r P(y)
\]
\[
\geq \int_{\{e_i(y) > u^{\beta(i)}\}} P\{\{e_{i+1}(N_{i+1}^r(g)) > (u \cdot e_i(y))^{r+1/2}\} \cap G|N_i^r = y) dN_i^r P(y)
\]
\[
\geq \int_{\{e_i(y) > u^{\beta(i)}\}} (P(G|N_i^r = y) - c_3 \cdot u) dN_i^r P(y)
\]
\[
\geq P\{\{e_i(N_{i+1}^r(g)) > u^{\beta(i)}\} \cap G) - c_3 \cdot u.
\]
Since \(\beta(i+1) = (1 + \beta(i)) \cdot (r + 1/2)\), we obtain by induction
\[
P\{\{e_{i+1}(N_{i+1}^r(g)) > u^{\beta(i+1)}\} \cap G) \geq P\{\{e_1(N_{i+1}^r(g)) > u^{\beta(1)}\} \cap G) - c_3 \cdot i \cdot u
\]
\[
\geq 1 - c_3 \cdot u - c_3 \cdot i \cdot u = 1 - c_3 \cdot (i + 1) \cdot u.
\]
This completes the proof. \(\Box\)

We are ready to prove Theorem B. It is enough to prove that for any \(\beta > r + 1/2\) there is a positive constant \(\lambda\) such that for any zerofinding method \(S_i^\beta\) with
\[
(4.7) \quad \int_F \nu(f) dP(f) \leq n
\]
we have

\[ (4.8) \int_F \Delta(S^k, f) \, dP(f) \geq \lambda^\beta. \]

We first consider the error \( \Delta_{re} \) in the residual sense. Obviously, it is sufficient to consider zero-finding methods \( S'_j(f) = \phi_{\nu(f)}(N'_j(f)) \) which use Hermite information of order \( r \). Let

\[ A_i = \{ f : \nu(f) = i \} = (N'_i)^{-1}(B_i) \]

with a Borel measurable set \( B_i \subset \mathbb{R}^{i \cdot (r+1)} \). Then (4.7) yields

\[ \sum_{j=1}^{\infty} j \cdot P(A_j \cap G) = \int_G \nu(g) \, dP(g) \leq n. \]

Take \( m = \lceil n/P(G)^2 \rceil \) and \( p = P(G) \cdot (1 - P(G))/m \). There exists an index \( i \in \{1, \ldots, m\} \) such that

\[ P(A_i \cap G) \geq p. \]

Indeed, if this is not the case, we have

\[ \sum_{j=1}^{\infty} \nu(g) \, dP(g) \leq \sum_{j=1}^{m} \nu(g) \, dP(g) \leq P(G) \cdot (1 - P(G))/m < n, \]

which is a contradiction.

For such an index \( i \), let

\[ w = (P(A_i \cap G)/(c_3 \cdot i))^{\beta(i)}, \]

where \( c_3 \) and \( \beta(i) \) are chosen as in Lemma 4.4. Then, using Lemma 4.4, we have

\[ \int_F \Delta_{re}(S'_j, f) \, dP(f) \geq \int_{B_i} e_i(y) \, dN'_i P(y) \geq \int_{A_i \cap G} e_i(N'_i(g)) \, dP(g) \]

\[ = \int_0^{\infty} P(\{e_i(N'_i(g)) > v\} \cap A_i \cap G) \, dv \]

\[ \geq \int_0^{w} (P(A_i \cap G) - c_3 \cdot i \cdot v^{1/\beta(i)}) \, dv \]

\[ = P(A_i \cap G)/(1 + \beta(i)) \cdot (P(A_i \cap G)/(c_3 \cdot i))^{\beta(i)}. \]

Observe that for suitable positive constants \( d_1 \) and \( d_2 \) we have

\[ i \leq d_1 \cdot n, \]

\[ P(A_i \cap G) \geq d_2/n, \]

\[ \beta(i) \leq (2r + 1)/(2r - 1) \cdot (r + 1/2)^i \leq 5/3 \cdot (r + 1/2)^i. \]

Therefore,

\[ \int_F \Delta_{re}(S'_j, f) \, dP(f) \geq d_2/n \cdot 3/10(2/(2r + 1))^{d_1 \cdot n} \cdot (d_2/(d_1 c_3))/n^2)^{\beta(i)}. \]
Thus, it remains to show that

\[(4.9) \quad \beta(i) \cdot \log n \leq d_3 \cdot \beta^n \]

for some positive constant \(d_3 > 0\).

Using (4.1), we can find a positive constant \(d_4\) such that

\[
\log \beta(i) \leq i \cdot \log(r + 1/2) + \log(5/3) \leq n \cdot \log(r + 1/2)/P(G)^2 + d_4
\]

\[
\leq n \cdot \log \beta \cdot 2 \log(r + 1/2)/(\log \beta + \log(r + 1/2)) + d_4.
\]

Clearly, for some positive \(d_5\) we have for any integer \(n\),

\[
\log \log n \leq n \cdot \log \beta \cdot (\log \beta - \log(r + 1/2))/(\log \beta + \log(r + 1/2)) + d_5.
\]

Hence,

\[
\log \beta(i) + \log \log n \leq n \cdot \log \beta + d_4 + d_5
\]

and (4.9) follows. This completes the proof of (4.8) for the error \(\Delta^{re}\).

We now consider the error \(\Delta^{ro}\) in the root sense. Consider an arbitrary method \(S^k_f\) which satisfies (4.7). We relate the average error of \(S^k_f\) in the root sense to the average error of a method \(\tilde{S}^k_{v+1}\) in the residual sense. Let \(z = 0\) if \(|a_0| < |b_0|\), and \(z = 1\) if \(|a_0| \geq |b_0|\). Then \(\tilde{S}^k_{v+1}\) is defined by

\[
\tilde{S}^k_{v+1}(f) = \begin{cases} 
S^k_f(f) & \text{if } |f(S^k_f(f))| \leq \min\{|a_0|, |b_0|\}, \\
z & \text{otherwise}.
\end{cases}
\]

We have

\[
\Delta^{re}(\tilde{S}^k_{v+1}, f) = \min\{|a_0|, |b_0|\} \quad \forall f \in F,
\]

and

\[
\Delta^{ro}(\tilde{S}^k_{v+1}, f) \leq u \cdot \Delta^{ro}(S^k_f, f) \quad \forall f \in F : \|f'\|_\infty \leq u.
\]

Since \(\tilde{S}^k_{v+1}\) uses at most \(n + 1\) knots, the proof for the error in the residual sense yields

\[
\int_F \Delta^{re}(\tilde{S}^k_{v+1}, f) \, dP(f) \geq \lambda^{\beta^{v+1}}.
\]

Therefore,

\[
\int_F \Delta^{ro}(S^k_f, f) \, dP(f) \geq 1/u \int_{\{\|f'\|_\infty \leq u\}} \Delta^{re}(\tilde{S}^k_{v+1}, f) \, dP(f)
\]

\[
\geq 1/u \cdot \lambda^{\beta^{v+1}} - 1/u \cdot P(\{f \in F : \|f'\|_\infty > u\}) \cdot \min\{|a_0|, |b_0|\}.
\]

Lemma 3.1 of Ledoux and Talagrand [6] states that there exist positive constants \(d_6\) and \(d_7 < 1\) such that

\[
P(\{f \in F : \|f'\|_\infty > u\}) \leq d_6 \cdot d_7^{u^2} \quad \forall u \geq 0.
\]

We choose

\[
u = ((\beta^{n+1} \log(1/\lambda) + \log(2d_6 \min\{|a_0|, |b_0|\}))/\log(1/d_7))^{1/2}
\]

to obtain

\[
\int_F \Delta^{ro}(S^k_f, f) \, dP(f) \geq 1/(2u) \cdot \lambda^{\beta^{v+1}} \geq \lambda^\tilde{\lambda}^{\beta^n}
\]

with a suitable constant \(\tilde{\lambda} > 0\). This completes the proof of (4.8) and of Theorem B. \(\square\)
ACKNOWLEDGMENTS

We thank T. Linder, J. F. Traub, G. W. Wasilkowski, and A. G. Werschulz for valuable comments.

The first author was supported by a Heisenberg scholarship of the Deutsche Forschungsgemeinschaft (DFG). The second author was supported by a postdoctoral grant of the Deutsche Forschungsgemeinschaft (DFG). The third author was supported in part by the National Science Foundation and the Air Force Office of Scientific Research.

BIBLIOGRAPHY


Mathematisches Institut, Universität Erlangen-Nürnberg, Bismarckstr. 1 ½, 91054 Erlangen, Germany
E-mail address: novak@mi.uni-erlangen.de
ritter@mi.uni-erlangen.de

Department of Computer Science, Columbia University, New York, NY 10027, USA,
and Institute of Applied Mathematics, University of Warsaw, ul. Banacha 2, 02-097 Warszawa, Poland
E-mail address: henryk@ground.cs.columbia.edu