ON THE POWER OF ADAPTIVE INFORMATION
FOR FUNCTIONS WITH SINGULARITIES

G. W. WASILKOWSKI AND F. GAO

ABSTRACT. We study from a probabilistic viewpoint the problem of locating singularities of functions using function evaluations. We show that, under the assumption of a Wiener-like probability distribution on the class of singular functions, an adaptive algorithm can locate a singular point accurately with only a small probability of failure. As an application, we show that an integration algorithm that adaptively locates a singular point is probabilistically superior to nonadaptive algorithms.

1. INTRODUCTION

We study the problem of locating singular points of functions. More specifically, let $f: [0, 1] \to \mathbb{R}$ be such that for some point $z = z_f \in (0, 1)$ the function f restricted to $[0, z)$ and to $(z, 1]$ is in C^k, while over the interval $[0, 1]$, f is in C^{s-1} and $f^{(s)}$ does not exist at z_f. Here s and k are integers, $0 \le s < k$. We assume that the position of z_f is unknown. Instead, we want to locate z_f through evaluating f at a (presumably small) number of points.

Locating singular points is an important problem by itself. It is also a key subproblem in a number of numerical problems including adaptive integration.

To be more specific, let us consider the integration problem for functions with k large relative to s. If the singular point were known, one could approximate the desired integral by approximating $\int_0^{z_f} f(x)\,dx$ and $\int_{z_f}^1 f(x)\,dx$ separately, with small error using relatively few function evaluations, by taking advantage of high regularity of the integrand in the two subintervals. The same would be possible if one were able to locate z_f with high accuracy. However, without locating z_f, one would need many more evaluations to guarantee small error because of the overall low regularity of the integrand. Therefore, not surprisingly, many integration codes contain adaptive schemes for locating singular points. These schemes differ from one another, but most of them are based on the following intuitive approach: evaluate f at a number of points in $[0, 1]$ (a partition of $[0, 1]$), and find out where the computed values of

f (or divided differences) change most drastically; if such a change in some interval is noticed, then restrict attention to this interval, partition it, and carry out the same procedure there; otherwise, take another partition of $[0, 1]$ by introducing extra points. This is repeated until a sufficiently small interval is determined to include possibly the singularity. The idea behind this approach is that the singularity at z_f is likely to cause f to behave more drastically near z_f.

It is well known that unless more properties of f are known and exploited, for any code there are always functions that will fool it. This implies that in the worst case, adaption does not help for the integration problem. In fact, in the worst case one cannot approximate singular points with error less than $\frac{1}{2}$ of the length of the interval (see §2).

Adaptive integration rules are common in practice. Although they do not always work, they do work satisfactorily *"most of the time,"* at least empirically.

This paper presents a mathematical model for studying this problem. By assuming the existence of a reasonable probability distribution on functions with a singularity in their sth derivatives, we give an algorithm based on the above intuitive approach, which locates z to high accuracy in a small number of steps, provided one permits a very small probability of failure. For a more rigorous statement and the result, see §5. Next, we apply our result to the integration problem. The conclusion is that adaptive integration rules are much superior to nonadaptive rules for all functions except a set of small measure.

Of course, the power of adaption depends very much on the underlying probability measure. If one, for instance, endows the class of functions with a Gaussian measure, then adaption essentially does *not* help (see, e.g., [8]). (Roughly speaking, this is because Gaussian measures are concentrated on functions with no singularities.) On the other hand, one could endow the class with a measure that conveys much information about a position of singular points. But this would make the problem trivial and not interesting from a practical point of view. Therefore, in this paper, we propose a new probability measure, which we call *Wiener-like* measure (see §3). It has all the important properties of k-fold Wiener measures, and yet, unlike Wiener measures, it is concentrated on functions with singularities. Some probability measures concentrated on singular functions have been studied in the statistical literature, see, e.g., [7], however they are more restrictive than our measure. We are not aware of any previous use of the Wiener-like measure in the literature.

The implication of our study is twofold. First, it gives a quantitative formulation of what the numerical analyst observes and believes in, namely a good code works well most of the time. Second, it represents a new methodology for tackling this problem: by using a probabilistic assumption that characterizes the practical situation, the design and analysis of an algorithm can now be done with mathematical rigor.

2. WORST-CASE ANALYSIS

Recall that we consider the following class of functions $f: [0, 1] \to \mathbb{R}$. For a given positive integer k, a function f has continuous kth derivative everywhere except at the *singular point* z_f, i.e., $f|_{[0, z_f)} \in C^k$ and $f|_{(z_f, 1]} \in C^k$; at z_f, $f^{(s)}$ does not exist but $f^{(s-1)}$ is absolutely continuous and $\|f^{(s)}\|_\infty$ is

bounded. Here, $s = s_f$ is an integer, $0 \le s < k$; $\| \cdot \|_\infty$ is the L_∞-norm.

As stated in the Introduction, we would like to locate z_f using function evaluations. To motivate our probabilistic analysis, we begin with a simple observation concerning the worst case.

Proposition 1. *For every n, any algorithm that uses n function evaluations has worst-case absolute error not smaller than $\frac{1}{2}$.*

Sketch of Proof. Consider any algorithm that uses n function evaluations at some points $x_1, \ldots, x_n$. (There is no assumption on the position of the points; they can be chosen in any adaptive way.) Consider now two functions f_1 and f_2 from our class that have the following properties:

(i) they attain the same values at the points x_i, $f_1(x_i) = f_2(x_i)$ for $i = 1, \ldots, n$,

(ii) $z_{f_1} \approx 0$ and $z_{f_2} \approx 1$.

They are easy to construct. Because of (i), the algorithm cannot distinguish between f_1 and f_2, and thus produces the same approximation for both of them. Hence, (ii) completes the proof. $\square$

This result states that no matter what algorithm we use, there are always functions that cause the algorithm to fail. However, endowing the class of functions with a probability measure, we will demonstrate that for some algorithms the probability of failure is very small.

3. WIENER-LIKE DISTRIBUTION

From now on, we shall endow the class of functions with a probability measure and analyze the performance of algorithms from a probabilistic perspective. Of course, the conclusions of a probabilistic analysis depend on the underlying probability distribution. Care must be exercised to choose an assumption that is relevant to practice. As a very first step, we would like to use a more conventional probability measure. To make it more practical, one can further tailor the probabilistic assumption to suit specific practical situations.

Among the most frequently used measures on function spaces are the Wiener and k-fold Wiener measures. However, these measures are concentrated on functions with no singularities, and hence do not allow analysis over a class of functions with singular points. We propose new measures that have all important properties of k-fold Wiener measures and yet are concentrated on the class of singular functions.

Recall that the k-fold Wiener measure w_k is the Gaussian measure on $C^k[0, 1]$ with zero mean and correlation given by

$$E_{w_k}(f(x)f(y)) = \int_0^1 \frac{(x - t)_+^k (y - t)_+^k}{k! k!} \, dt, \quad \text{where } (x - t)_+^k = \max\{(x - t)^k, 0\}.$$

Equivalently, f distributed according to w_k can be viewed as the Gaussian stochastic process with zero mean and correlation given above.

To define our *Wiener-like measures*, we first define the *conditional* probability with $z_f = z$ and s fixed. To define this conditional probability, denoted by Prob_z, it is equivalent to specify its corresponding random functions (stochastic

processes) f. They are given as follows. For fixed z and s $(s < k)$,

$$f(x) = \int_0^1 f^{(s)}(t)\, d(-(x-t)_+^s/s!)$$

with

$$f^{(s)}(x) = \begin{cases} g_1(1-x) & \text{if } x \le z, \\ g_2(x) & \text{if } x > z, \end{cases}$$

where g_1, g_2 are independent and identically distributed according to w_{k-s}. Equivalently,

$$f(x) = \int_0^1 ((z-t)_+^0 g_1(1-t) + (t-z)_+^0 g_2(t))\, d(-(x-t)_+^s/s!).$$

Obviously, $f^{(s-1)}$ is absolutely continuous, $f^{(s)}(z)$ does not exist, and $f|_{[0,z)}$ and $f|_{(z,1]}$ are from C^k almost surely.

After defining conditional probability Prob_z, one could put some Borel probability measure ρ on z to get a corresponding *Wiener-like* probability measure μ_ρ,

$$\mu_\rho(A) = \int \text{Prob}_z(A)\, \rho(dz),$$

on the class of functions with singularities. (Of course, one could also model functions with more than one singular point; see Remark 7.)

The results we are going to present are robust with respect to distributions on z. Therefore, instead of choosing a specific ρ, we shall study the worst case with respect to these distributions (or equivalently, the worst case with respect to z). Similarly, our algorithm does not assume knowledge of s, and the value of s will be left as a parameter in the probability estimates. For simplicity (see Remarks 1 and 4), we shall assume that

$$z \in [a, b] \subset [0, 1] \quad \text{with sufficiently large } a \text{ and } 1 - b.$$

We end this section by providing some simple facts (without proofs) concerning the random functions f.

Given $z_f = z$, let

$$f_1(x) = \int_0^z g_1(1-t)\, d(-(x-t)_+^s/s!)$$

and

$$f_2(x) = \int_z^1 g_2(t)\, d(-(x-t)_+^s/s!).$$

Obviously, $f(x) = f_1(x) + f_2(x)$, and f_1 and f_2 are independent. The function f_1 is a Gaussian stochastic process with zero mean and correlation given by

$$\tag{1} E(f_1(x_j)f_1(x_p)) = \int_0^1 \left[\int_0^z \frac{(x_j-t)_+^{s-1}(1-t-u)_+^{k-s}}{(s-1)!(k-s)!}\, dt \right] \times \left[\int_0^z \frac{(x_p-v)_+^{s-1}(1-v-u)_+^{k-s}}{(s-1)!(k-s)!}\, dv \right] du.$$

In particular, for $x > z$, f_1 is a polynomial of degree $s - 1$,

(2)
$$f_1(x) = (-1)^s(g(1 - z) - g(1))$$
$$+ \sum_{j=1}^{s-1}(-1)^{s+j}\left(g^{(j)}(1 - z)\frac{(x - z)^j}{j!} - g^{(j)}(1)\frac{x^j}{j!}\right)$$

with g distributed according to w_k. The function f_2 vanishes on $[0, z]$, and for $x > z$ it is a Gaussian stochastic process with zero mean and correlation given by

(3)
$$E(f_2(x_j)f_2(x_p)) = \int_0^1 \left[\int_z^{x_j} \frac{(x_j - t)^{s-1}(t - u)_+^{k-s}}{(s - 1)!(k - s)!} dt\right]$$
$$\times \left[\int_z^{x_p} \frac{(x_p - v)^{s-1}(v - u)_+^{k-s}}{(s - 1)!(k - s)!} dv\right] du.$$

4. ALGORITHM

To present the algorithm, we need some notation. Given h, and $n + 1$ equally spaced points $x_0, \ldots, x_n \in [0, 1]$ ($h = x_{i+1} - x_i$), by $X_l = X_l(f; h)$ we mean the $(k + 1)$st backward difference of f at the point x_l, i.e.,

$$X_l = \sum_{j=0}^{k+1}\binom{k + 1}{j}(-1)^j f(x_{l-j}) \quad \text{for } l = k + 1, \ldots, n.$$

We say that the points $\{x_l\}_{l=0}^n$ form a *semipartition* of an interval $[a, b]$ if $x_k = a$ and $x_{n-k} = b$.

Consider now a function T such that $T(h)$ converges to zero slower than $h^{k+1/2}$, i.e.,

$$\lim_{h \to 0} \frac{T(h)}{h^{k+1/2}} = +\infty.$$

Recall that we are interested in locating z_f in a given interval $[a_0, b_0] = [a, b]$ that is a proper subset of $[0, 1]$. For given n_0, $n_0 > 4k + 2$, the algorithm will perform a number of steps, in each producing a new subinterval $[a_{i+1}, b_{i+1}]$ that with large probability contains z_f and has a diameter significantly smaller than the diameter of the previously constructed interval $[a_i, b_i]$. Hence, with large probability, after the rth step, z_f will be located in a small interval $[a_{r+1}, b_{r+1}]$, and $(a_{r+1} + b_{r+1})/2$ will be a very accurate approximation to z_f. More precisely, the algorithm consists of the following steps:

Step 0. Set $[a_0, b_0] = [a, b]$. Choose $n_0 + 1$ equally spaced points $x_0, \ldots, x_{n_0}$ with $x_k = a_0$ and $x_{n_0-k} = b_0$. Obviously, $h_0 = x_{i+1} - x_i = (b_0 - a_0)/(n_0 - 2k)$.

Compute the backward differences X_j for $j = k + 1, \ldots, n_0$, and find an index l with $|X_l| \geq T(h_0)$. If such an l exists, then

$$[a_1, b_1] := [x_{l-k-1}, x_l] \cap [a_0, b_0].$$

Otherwise, $[a_1, b_1] := [a_0, b_0]$.

Step i $(i \geq 1)$. If $[a_i, b_i] = [a_{i-1}, b_{i-1}]$, then form a new semipartition of $[a_i, b_i]$ with $h_i = h_{i-1}/2$ and the corresponding $n_i = 2n_{i-1} - 2k$. This can

be done by deleting x_i (with $i < \lceil k/2 \rceil$ and $i > n_i - \lceil k/2 \rceil$) from the old partition, and adding extra $n_{i-1} - 2k + 2\lceil k/2 \rceil$ points between the old points.

Else, take $n_i = n_0$ and a new semipartition for $[a_i, b_i]$. In this case, $b_i - a_i \leq h_{i-1}(k+1) = (b_{i-1} - a_{i-1})(k+1)/(n_{i-1} - 2k)$ and $h_i \leq h_{i-1}(k+1)/(n_0 - 2k)$.

Next, we compute the corresponding backward differences, construct $[a_{i+1}, b_{i+1}]$ (in the same way as in Step 0), and go to the next $(i+1)$st step.

Remark 1. For simplicity, we assume that the a priori bounds a and b on z_f are not too close to 0 and 1, respectively, so that a semipartition $x_0, \ldots, x_{n_0}$ can be constructed. Formally, this means that $a \geq kh_0$ and $1 - b \geq kh_0$. If this assumption does not hold, then in each step of the algorithm we consider only those points x_j that are in $[0, 1]$.

This modification does not alter the algorithm performance if $z_f = \Omega(h_0)$ and $1 - z_f = \Omega(h_0)$. In general, if z_f is not known to be far away from 0 and 1, we need another modification which is discussed in Remark 4.

5. PROPERTIES OF THE ALGORITHM

We present a probabilistic analysis of the algorithm of §4, assuming that the underlying probability measure is an arbitrary Wiener-like measure μ_ρ, as defined in §3, with $z_f \in [a, b]$ for any f.

Sections 5.1 and 5.2 deal with the algorithm's performance in one step. Note that if an uncertainty interval $[a_i, b_i]$ is reduced at the ith step, then

$$b_{i+1} - a_{i+1} \leq h_i(k+1) \leq \frac{(b_i - a_i)(k+1)}{n_0 - 2k}.$$

Hence, the new uncertainty interval is at least $(n_0 - 2k)/(k+1)$ times smaller. Although $[a_{i+1}, b_{i+1}]$ need not contain the singular point z (we refer to this as *a wrong decision*), this can happen only with small probability, as will be proven in §5.1. We would also like to know whether the reduction of the uncertainty interval occurs frequently. This is indeed the case, as will be proven in §5.2.

Using these results, we will prove in §5.3 that with a relatively large probability, the singular point z_f can be located very accurately in a relatively small number of steps.

5.1. **Probability of wrong decision.** Consider one step of the algorithm with spacing $h = h_i$ and the number of points $n + 1 = n_i + 1$. Then we have the following lemma.

Lemma 1. *The probability that $|X_l| \geq T(h)$ for some l, but z does not belong to $[x_{l-k-1}, x_l]$ is bounded from above by*

$$\sqrt{\frac{2}{\pi}} \frac{h^{k+1/2}(n - 2k)}{T(h)} \exp\left(\frac{-(T(h))^2}{2h^{2k+1}}\right).$$

Proof. Since $\mu_\rho(A) = \int_{a_0}^{b_0} \mathrm{Prob}_z(A)\rho(dz) \leq \sup_{a_0 < z < b_0} \mathrm{Prob}_z(A)$ for an arbitrary measurable set A, it is enough to show that the expression given in the lemma is an upper bound on $\mathrm{Prob}_z(|X_l| \geq T(h))$ for every $z \in [a_0, b_0]$.

Consider $x_l \leq z$. Then X_l is the backward difference of f_1. Since X_l is a linear combination of $f_1(x_{l-j})$'s, the fact that the $f_1(x_{l-j})$'s are Gaussian implies that X_l has a normal distribution with zero mean. Letting $\sigma_l =$

$E((X_l(f_1\,;h))^2)$ be the covariance of X_l, we have

$$\text{Prob}_z(|X_l| \geq T(h)) = \frac{1}{\sqrt{2\pi\sigma_l}} \int_{|x| \geq T(h)} e^{-x^2/(2\sigma_l)}\, dx$$
$$\leq \sqrt{\frac{2}{\pi}} \frac{\sqrt{\sigma_l}}{T(h)} \exp\left(\frac{-(T(h))^2}{2\sigma_l}\right).$$

We need to estimate σ_l. It is easy to see that for x_j, $x_p \leq z$, each inner integral in (1) as a function of x_j (or x_p) is a sum of a polynomial of degree k and the term $(1 - u - x_j)_+^k/k!$. Therefore,

$$\sigma_l = \int_0^1 \left[\sum_{j=0}^{k+1} \binom{k+1}{j}(-1)^j \frac{(1 - x_{l-j} - u)_+^k}{k!}\right]^2 du.$$

Hence, σ_l does not depend on l, $\sigma_l = \sigma$. Furthermore, it is well known (see, e.g., [4]) that

$$\sigma_l = \sigma \leq h^{2k+1}.$$

Consider now $z < x_{l-k-1}$. Then $X_l = X_l(f_1\,;h) + X_l(f_2\,;h) = X_l(f_2\,;h)$ because of (2). Also for this case we have $\sigma_l = \sigma \leq h^{2k+1}$. Indeed, each inner integral in (3) as a function of x_j (or x_p) is a polynomial of degree k if $u \leq z$. Hence, for such u, the backward difference is zero. For $u > z$, these integrals reduce to $(x - u)_+^k/k!$ and that is why $\sigma_l = \sigma \leq h^{2k+1}$.

Therefore, for any l with $z \notin [x_{l-k-1}, x_l]$, we have

$$\text{Prob}_z(|X_l| \geq T(h)) \leq \sqrt{\frac{2}{\pi}} \frac{h^{k+1/2}}{T(h)} \exp\left(\frac{-(T(h))^2}{2h^{2k+1}}\right).$$

Since there are $n - 2k$ such indices l, the proof is complete. $\square$

5.2. Probability of not reducing the uncertainty interval.
Consider now one step with spacing $h = h_i$ and $z_f \in [a_i, b_i]$.

Lemma 2. *The probability $P(h) = \mu_p([a_i, b_i] = [a_{i+1}, b_{i+1}])$ that the uncertainty interval will not be reduced is bounded by*

(4) $$P(h) \leq \beta_{k,s}(T(h)h^{-(k+s)/2})^{k-s+1}$$

with

$$\beta_{k,s} = \frac{2^{(k-s+1)(k+s+2)/2}(\prod_{i=s}^k i!)(\prod_{i=1}^{k-s+1}(k - s + i)!)^{1/2}}{\pi^{(k-s+1)/2}((2(k - s) - 1)!!)^s(\prod_{i=1}^{k-s} i!)^{3/2}}.$$

Proof. As in the proof of Lemma 1, we consider the conditional probability Prob_z.

Given z, let x_l be the partition point such that $x_{l-1} \leq z \leq x_l$. Then for every p, the probability $P(h)$ is bounded from above by

$$P(h) \leq \sup_{x_{l-1} \leq z \leq x_l} \text{Prob}_z(\{|X_j(f\,;h)| \leq T(h) : \forall j = l, \dots, k + l\}).$$

Since

$$X_j(f\,;h) = X_j(f_1\,;h) + X_j(f_2\,;h) \quad \text{for every } l \leq j \leq k + l$$

and

$$[X_l(f_i\,;h), \dots, X_{k+l}(f_i\,;h)]^T \quad (i = 1, 2)$$

are independent Gaussian random vectors, each with zero mean, Anderson's inequality (see [1]) implies that

(5)
$$\text{Prob}_z(\{|X_j(f; h)| \le T(h): \forall j = l, \ldots, k + l\})$$
$$\le \text{Prob}_z(\{|X_j(f_i; h)| \le T(h): \forall j = l, \ldots, k + l\}) \quad \forall i = 1, 2.$$

Assume first that

(6)
$$z \ge \tfrac{1}{2} \quad \text{and} \quad x_l - z \ge h/2.$$

The right-hand side of (5) can be estimated by considering $i = 2$ and $l \le j \le k - s + l$ instead of $l \le j \le k + l$, i.e.,

$$\text{Prob}_z(\{|X_j(f_2; h)| \le T(h): \forall j = l, \ldots, k + l\})$$
$$\le \text{Prob}_z(\{|X_j(f_2; h)| \le T(h): \forall j = l, \ldots, k - s + l\}) =: \widetilde{P}_z.$$

We need to estimate $\widetilde{P}_z$. With $y_j = X_j(f_2; h)$, the random vector

$$[y_l, \ldots, y_{k-s+l}]^\mathsf{T}$$

has a normal distribution with zero mean and $(k - s + 1) \times (k - s + 1)$ covariance matrix $C = (c_{i,j})_{i,j=l}^{k-s+l}$ with

$$c_{i,j} = E(X_i(f_2; h), X_j(f_2; h)).$$

Thus,

(7)
$$\widetilde{P}_z = \frac{1}{(2\pi)^{(k-s+1)/2}\sqrt{\det(C)}} \int_{\|\vec{y}\|_\infty \le T(h)} \exp(-\langle C^{-1}\vec{y}, \vec{y}\rangle/2)\, d\vec{y}$$
$$\le \frac{(T(h))^{k-s+1} 2^{(k-s+1)/2}}{\pi^{(k-s+1)/2}\sqrt{\det(C)}}.$$

Hence, we need to estimate from below the determinant of the matrix C.
Note that

$$c_{i,j} = \int_0^1 g_i(u) g_j(u)\, du$$

with

$$g_i(u) = \sum_{p=0}^{k+1} \binom{k+1}{p} (-1)^p \int_z^{x_{i-p}} \frac{(x_{i-p} - t)_+^{s-1}(t - u)_+^{k-s}}{(s-1)!(k-s)!}\, dt$$
$$= \sum_{p=0}^{i-l} \binom{k+1}{p} (-1)^p \int_z^{x_{i-p}} \frac{(x_{i-p} - t)_+^{s-1}(t - u)_+^{k-s}}{(s-1)!(k-s)!}\, dt.$$

Thus,

$$c_{i,j} = \int_0^1 g_i(u) g_j(u)\, du = \tilde{c}_{i,j} + \hat{c}_{i,j},$$

where

$$\tilde{c}_{i,j} = \int_0^z g_i(u) g_j(u)\, du \quad \text{and} \quad \hat{c}_{i,j} = \int_z^1 g_i(u) g_j(u)\, du.$$

Since $\widetilde{C} = (\tilde{c}_{i,j})$ and $\widehat{C} = (\hat{c}_{i,j})$ are symmetric and nonnegative definite, $\det(C) \ge \det(\widetilde{C})$.

We now estimate the determinant of $\widetilde{C}$. For $u \leq z$,

$$\int_z^{x_i} \frac{(x_i - t)^{s-1}(t - u)_+^{k-s}}{(s-1)!(k-s)!} \, dt = \sum_{q=s}^{k} \frac{(x_i - z)^q (z - u)^{k-q}}{q!(k-q)!}.$$

Therefore,

$$\tilde{c}_{i,j} = \sum_{q_1,q_2=s}^{k} b_{q_1,q_2} \zeta_{i,q_1} \zeta_{j,q_2},$$

where

$$b_{q_1,q_2} = \frac{z^{2k-q_1-q_2+1}}{(2k - q_1 - q_2 + 1)q_1!q_2!(k - q_1)!(k - q_2)!}$$

and

$$\zeta_{i,q} = \sum_{p=0}^{i-l} \binom{k+1}{p} (-1)^p (x_{i-p} - z)^q.$$

Denoting $B = (b_{p,q})_{p,q=s}^{k}$ and $\vec{\zeta}_i = [\zeta_{i,s}, \ldots, \zeta_{i,k}]^{\mathrm{T}}$, we can write $\widetilde{C} = Z^{\mathrm{T}} B Z$ with the matrix Z given by $Z = [\vec{\zeta}_l, \ldots, \vec{\zeta}_{k-s+l}]$. Obviously,

$$\det(\widetilde{C}) = \det(B)(\det(Z))^2.$$

Furthermore, the determinant of Z equals the determinant of the matrix

$$V = \begin{pmatrix} (x_l - z)^s & \cdots & (x_{k-s+l} - z)^s \\ \vdots & \cdots & \vdots \\ (x_l - z)^k & \cdots & (x_{k-s+l} - z)^k \end{pmatrix}.$$

However, $V = \widetilde{V}D$, where D is a diagonal matrix, $D = \operatorname{diag}((x_i - z)^s)_{i=l}^{k-s+l}$ and $\widetilde{V}$ is the following Vandermonde matrix:

$$\widetilde{V} = \begin{pmatrix} 1 & \cdots & 1 \\ (x_l - z) & \cdots & (x_{k-s+l} - z) \\ \vdots & \cdots & \vdots \\ (x_l - z)^{k-s} & \cdots & (x_{k-s+l} - z)^{k-s} \end{pmatrix}.$$

The determinant of D is obviously equal to

$$\det(D) = \prod_{i=l}^{k-s+l} (x_i - z)^s.$$

It is known that

$$\det(\widetilde{V}) = \prod_{i=l}^{k-s+l-1} \prod_{j=i+1}^{k-s+l} (x_i - x_j).$$

Hence,

$$(8) \quad \det(C) \geq \det(\widetilde{C}) = \det(B) \left(\prod_{i=l}^{k-s+l} (x_i - z)^s \right)^2 \left(\prod_{i=l}^{k-s+l-1} \prod_{j=i+1}^{k-s+l} (x_i - x_j) \right)^2.$$

In view of (6), we have

$$\prod_{i=l}^{k-s+l} (x_i - z)^s \geq h^{s(k-s+1)} \left(\frac{(2(k-s)-1)!!}{2^{k-s+1}} \right)^s$$

and

$$\prod_{i=l}^{k-s+l-1} \prod_{j=i+1}^{k-s+l} |x_i - x_j| = \prod_{i=l}^{k-s+l-1} h^{k-s+l-i}(k-s+l-i)!$$

$$= h^{(k-s)(k-s+1)/2} \prod_{i=1}^{k-s} i!.$$

Thus,

$$\sqrt{\det(C)} \geq h^{(k+s)(k-s+1)/2} \alpha_{k,s},$$

where

$$\alpha_{k,s} = \left(\frac{(2(k-s)-1)!!}{2^{k-s+1}} \right)^s \sqrt{\det(B)} \prod_{i=1}^{k-s} i!.$$

Hence, by (7),

$$\widetilde{P}_z \leq (T(h)h^{-(k+s)/2})^{k-s+1} \beta_{k,s}(z)$$

with

$$\beta_{k,s}(z) = \frac{2^{(k-s+1)(2s+1)/2}}{\pi^{(k-s+1)/2}((2(k-s)-1)!!)^s \sqrt{\det(B)} \prod_{i=1}^{k-s} i!}.$$

To estimate the determinant of B, note that

$$\det(B) = z^{(k-s+1)^2} \left(\prod_{p=s}^{k} p!(k-p)! \right)^{-2} \det(H_{k-s+1}),$$

where H_{k-s+1} is the $(k-s+1) \times (k-s+1)$ Hilbert matrix. It is known that

$$\det(H_{k-s+1}) = \frac{\prod_{i=1}^{k-s}(i!)^3}{\prod_{i=1}^{k-s+1}(k-s+i)!}.$$

Hence,

$$\beta_{k,s}(z) = \frac{2^{(k-s+1)(2s+1)/2}(\prod_{i=s}^{k} i!)(\prod_{i=1}^{k-s+1}(k-s+i)!)^{1/2}}{\pi^{(k-s+1)/2} z^{(k-s+1)^2/2}((2(k-s)-1)!!)^s (\prod_{i=1}^{k-s} i!)^{3/2}}.$$

Since $\beta_{k,s}(z)$ is decreasing, $\beta_{k,s}(z) < \beta_{k,s}(a)$. This completes the proof under the assumption (6).

Suppose that (6) does not hold. If $z \geq \frac{1}{2}$ (i.e., $z - x_l < h/2$), then instead of $X_l, \ldots, X_{k-s+l}$ consider $X_{k+l}, \ldots, X_{s+l}$ in estimating $P(h)$ by $\sup_z \widetilde{P}_z$. As before, the problem reduces to estimating the determinant of a corresponding matrix $\widetilde{C}$. However, for $i = s+l, \ldots, k+l$, it is easy to see that

$$g_i(u) = -\sum_{p=i-l+1}^{k+1} \binom{k+1}{p} (-1)^p \int_{z}^{x_{i-p}} \frac{(x_{i-p}-t)^{s-1}(t-u)_+^{k-s}}{(s-1)!(k-s)!} dt.$$

Hence, the matrix $\widetilde{C}$ has the same structure as before, with the only difference that instead of $x_l, \ldots, x_{k-s+l}$ we have $x_{l-1}, \ldots, x_{l-k+s-1}$ with $|z - x_{l-1}| \geq h/2$. Hence the determinant of $\widetilde{C}$ is bounded exactly as in the case when $x_l - z \geq h/2$.

If $z < \frac{1}{2}$ then we use (5) with $i = 1$. Depending on the sign of $x_l - z \geq \frac{1}{2}$, we consider $X_l, \ldots, X_{k-s+l}$ or $X_{k+l}, \ldots, X_{s+l}$ in estimating $P(h)$ by $\sup_z \widetilde{P}_z$. In order to estimate the determinant of the corresponding matrix $\widetilde{C}$, we use the following observation: If $x_j \leq z$, then

$$\int_0^z \frac{(x_j - t)_+^{s-1}(1 - t - u)_+^{k-s}}{(s-1)!(k-s)!} = \int_0^z \frac{(x_j - t)^{s-1}(1 - t - u)_+^{k-s}}{(s-1)!(k-s)!}$$
$$- \int_{x_j}^z \frac{(x_j - t)^{s-1}(1 - t - u)_+^{k-s}}{(s-1)!(k-s)!}.$$

Therefore, the corresponding function $g_j(u)$ equals

$$g_j(u) = -\sum_{p=i+l}^{k+1} \binom{k+1}{p}(-1)^p \int_{x_{l-p}}^z \frac{(x_{i-p} - t)^{s-1}(1 - t - u)_+^{k-s}}{(s-1)!(k-s)!} \, dt.$$

Hence, the matrix $\widetilde{C}$ has the same structure as before with the only difference being that the entries b_{q_1, q_2} of the matrix B have $(1-z)^{2k-q_1-q_2+1}$ instead of $z^{2k-q_1-q_2+1}$. Since $z < \frac{1}{2}$ implies $1 - z > \frac{1}{2}$, this completes the proof. □

Remark 2. As we shall see, Lemma 2 plays a crucial role in our analysis. Although (4) is not sharp for relatively large h, we believe that it provides a sharp estimate of $P(h)$ for sufficiently small h. Furthermore, even for moderate values of h, it gives a nontrivial bound on $P(h)$.

To see this, consider $T(h) = h^k$. The estimate (4) is a product of the constant $\beta_{k,s}$ and the spacing h raised to the power $\alpha = (k - s + 1)(k - s)/2$. The constant $\beta_{k,s}$ might be very large. However, when multiplied by h^α, the whole expression attains a small value for even a moderate value of h. For instance, consider $k = 4$ and $s = 0$. Then $\beta_{4,0} \approx 2.3 \times 10^{13}$ and $\alpha = 10$, and therefore the estimate (4) implies $P(h) \leq 2300(10h)^{10}$.

For moderate values of h, one could consider a similar algorithm with the only change being that the backward differences of order $(k + 1)$ are replaced by the backward differences of order $(\hat{k} + 1)$ with $s \leq \hat{k} < k$. Then the corresponding estimate would take the same form as in Lemma 2 with α replaced by $\hat{\alpha} = (\hat{k} - s)(\hat{k} - s + 1)/2$, which is smaller than α, but with the constant $\hat{\beta}_{k,s}$ smaller than $\beta_{k,s}$.

For simplicity, we shall assume from now on that h is small enough; modifications of the algorithm will be reported elsewhere (see Remark 6).

5.3. Main theorem.

Consider $T(h) = \sqrt{2}h^{k+\gamma}$ with $\gamma \in [0, \frac{1}{2})$ and $n_0 > 4k+2$. Let

$$\alpha = \frac{(k - s + 2\gamma)(k - s + 1)}{2} \quad \text{and} \quad \nu = \nu(n_0, k) = \frac{k + 1}{n_0 - 2k}.$$

Then the function $P(h)$ from Lemma 2 is bounded by $P(h) \leq \widetilde{\beta}_{k,s} h^\alpha$ with $\widetilde{\beta}_{k,s} = 2^{(k-s+1)/2} \beta_{k,s}$.

For $r \geq 1$, let $[a_r, b_r]$ denote the uncertainty interval obtained after r steps. Let $N_r = N_r(f)$ denote the total number of function evaluations performed in the steps $0, \ldots, r-1$. As explained earlier, $[a_r, b_r]$ need not contain z_f. Of course, we would like this to happen only with a small probability; this probability is bounded from above in part (i) of the following theorem. Otherwise, if $z_f \in [a_r, b_r]$, we would like the diameter $b_r - a_r$ and the number N_r of function evaluations to be small with large probability. Part (ii) of the following theorem estimates this probability from below. Furthermore, it provides upper bounds on the expected values of $b_r - a_r$ and N_r, respectively.

Theorem 1. (i) *The uncertainty interval* $[a_r, b_r]$ *does not contain* z_f *with probability*

$$
\mu_p(z_f \notin [a_r, b_r])
$$
$$
\leq \frac{(b_0 - a_0)^{1/2-\gamma}(n_0 - 2k)^{1/2+\gamma}}{\sqrt{\pi}} \sum_{i=0}^{r-1} 2^{i(1/2+\gamma)} \exp\left(-\left(\frac{2^i(n_0 - 2k)}{b_0 - a_0}\right)^{1-2\gamma}\right)
$$
$$
= (1 + o(1)) \frac{(b_0 - a_0)^{1/2-\gamma}(n_0 - 2k)^{1/2+\gamma}}{\sqrt{\pi}} \exp\left(-\left(\frac{n_0 - 2k}{b_0 - a_0}\right)^{1-2\gamma}\right).
$$

(ii) *Let* $z_f \in [a_r, b_r]$. *Then for every* $p = 0, \ldots, r-1$,

$$
\mu_p((b_r - a_r \leq c_p) \wedge (N_r(f) \leq d_p))
$$
$$
(9) \qquad\qquad \geq 1 - \widetilde{\beta}_{k,s}\left(\frac{b_0 - a_0}{2^p(n_0 - 2k)}\right)^\alpha \frac{1 - 2^{-\alpha(r-p)}}{1 - 2^{-\alpha}},
$$

where $c_p = (b_0 - a_0)\nu^{r-p}/2^p$, *and*

$$
d_p = \begin{cases} r(n_0 + 1) & \text{if } p = 0, 1, \\ (2^p - 1)(n_0 - 2k) + p2\lceil k/2 \rceil + (r - p)(n_0 + 1) & \text{otherwise.} \end{cases}
$$

The expected value of $b_r - a_r$ *is bounded from above by*

$$
(b_0 - a_0)\left(\nu^r + \widetilde{\beta}_{k,s}\left(\frac{b_0 - a_0}{n_0 - 2k}\right)^\alpha (\kappa_1 + 2^{\alpha(1-r)})\right)
$$

with

$$
\kappa_1 = \begin{cases} \nu^{r-1}\dfrac{r-1}{2} & \text{if } 2^{\alpha+1}\nu = 1, \\[2mm] \nu^{r-1}\dfrac{1 - (2^{\alpha+1}\nu)^{1-r}}{2 - (2^\alpha\nu)^{-1}} & \text{otherwise} \end{cases}
$$
$$
\leq \frac{r-1}{2}\max\{\nu^{r-1}, 2^{(\alpha+1)(1-r)}\},
$$

and the expected value of $N_r(f)$ *is bounded from above by*

$$
r(n_0 + 1) + \widetilde{\beta}_{k,s}\left(\frac{b_0 - a_0}{2(n_0 - 2k)}\right)^\alpha \kappa_2
$$

with

$$\kappa_2 = 2(n_0 - 2k) \left(\frac{1 - 2^{(\alpha-1)(2-r)}}{1 - 2^{(1-\alpha)}} - \frac{1 - 2^{\alpha(2-r)}}{1 - 2^{-\alpha}} \right)$$

$$+ (2\lceil k/2 \rceil - 2k - 1) \frac{1 - 2^{\alpha(2-r)}}{1 - 2^{-\alpha}}$$

$$- (n_0 + 1 - 2\lceil k/2 \rceil) \left(\frac{1 - 2^{\alpha(1-r)}}{(1 - 2^{-\alpha})^2} - \frac{(r-1)2^{\alpha(2-r)}}{1 - 2^{-\alpha}} \right)$$

$$\leq \frac{n_0 - 2k}{(1 - 2^{-\alpha})(2^{\alpha-1} - 1)} + (n_0 + 1 - 2\lceil k/2 \rceil) \frac{(r-1)2^{\alpha(2-r)}}{1 - 2^{-\alpha}}.$$

Proof. The first part follows from Lemma 1 and the fact that $h_i \leq h_0 2^{-i}$ in the ith step.

We now prove the second part. Let R_i be the set of functions for which the uncertainty is reduced in the ith step, i.e., $[a_{i+1}, b_{i+1}] \neq [a_i, b_i]$. Let F_i be the complement of R_i. Consider the following families of sets:

$$G_0 = \bigcap_{j=0}^{r-2} R_j, \qquad G_p = F_{p-1} \cap \bigcap_{j=p}^{r-2} R_j \quad \text{for } p = 1, \ldots, r-2, \qquad G_{r-1} = F_{r-2},$$

and

$$H_p = G_p \cap R_{r-1} \quad \text{for } p = 0, \ldots, r-1, \qquad H_r = F_{r-1}.$$

Obviously, $H_p \subseteq G_p$ ($p < r$). Furthermore, $\{G_p\}_{p=0}^{r-1}$ and $\{H_p\}_{p=0}^{r}$ form partitions of our class of functions. Note that

$$(10) \qquad \text{Prob}_z(H_p) \leq \text{Prob}_z(G_p) \leq \text{Prob}_z(F_{p-1}) \leq \widetilde{\beta}_{k,s} h_{p-1}^\alpha \leq \widetilde{\beta}_{k,s} (h_0 2^{1-p})^\alpha$$

for $1 \leq p \leq r-1$, and $\text{Prob}_z(H_r) \leq \widetilde{\beta}_{k,s} (h_0 2^{1-r})^\alpha$.

We prove that the numbers c_p and d_p defined in the theorem satisfy

$$(11) \qquad c_p \geq \max_{f \in H_p}(b_r - a_r), \qquad d_p \geq \max_{f \in G_p} N_r(f) \geq \max_{f \in H_p} N_r(f) \quad \forall p \leq r-1.$$

Indeed, for $f \in H_p$ the algorithm reduces the uncertainty interval in all steps from the pth through $(r-1)$st. Hence, $b_r - a_r \leq h_p(k+1)\nu^{r-p} \leq c_p$, since $h_p \leq 2^{-p} h_0$. To show the bound on $\max_{f \in G_p} N_r(f)$, let $m_j(f)$ denote the number of function evaluations performed in the jth step. Obviously, if $j = 0$ or $f \in R_{j-1}$, then $m_j(f) \leq n_0 + 1$. Otherwise, if $f \in F_j$, then $m_{j+1} = n_j - 2k + 2\lceil k/2 \rceil$. Since $n_j \leq 2(n_{j-1} - k)$, we have $n_j - 2k \leq 2^j(n_0 - 2k)$. Therefore, $N_r(f) = n_0 + 1 + \sum_{j=1}^{r-1} m_j(f) \leq d_p$ for any $f \in G_p$, as claimed in (11).

Note that (9) follows from (10), since $((b_r - a_r > c_p) \vee (N_r(f) > d_p))$ implies that $f \in \bigcup_{j=p+1}^{r} H_j$.

To estimate the expected value of $b_r - a_r$, observe that

$$\int (b_r - a_r) \text{Prob}_z(df) = \sum_{p=0}^{r} \int_{H_p} (b_r - a_r) \text{Prob}_z(df),$$

since the family $\{H_p\}_{p=0}^{r}$ is a partition. Hence, $b_r - a_r \leq b_0 - a_0$, and (11) imply

$$\int (b_r - a_r) \text{Prob}_z(df) \leq \sum_{p=0}^{r-1} c_p \text{Prob}_z(H_p) + (b_0 - a_0) \text{Prob}_z(H_r).$$

With the help of this and (10), it is easy to establish the bound on the expectation of $b_r - a_r$.

To estimate the expected value of N_r, we consider the partition $\{G_p\}_{p=0}^{r-1}$. Then

$$\int N_r(f)\operatorname{Prob}_z(df) = \sum_{p=0}^{r-1}\int_{G_p} N_r(f)\operatorname{Prob}_z(df)$$

$$\le d_0\operatorname{Prob}_z(G_0 \cup G_1) + \sum_{p=2}^{r-1} d_p\operatorname{Prob}_z(G_p).$$

Since $d_p \ge d_0$ for $p \ge 2$, from (10) we get

$$\int N_r(f)\operatorname{Prob}_z(df) \le d_0 + \sum_{p=2}^{r-1}(d_p - d_0)\widetilde{\beta}_{k,s}(h_0 2^{1-p})^\alpha$$

$$\le r(n_0 + 1) + \widetilde{\beta}_{k,s}\left(\frac{b_0 - a_0}{n_0 - 2k}\right)^\alpha \kappa.$$

This completes the proof. □

Remark 3. The proof of Theorem 1(ii) relies heavily on the use of the estimate of $P(h)$ from Lemma 2. Even though this estimate seems to be sharp (we assume, of course, that h is small), we believe that the bound (9) is not sharp. A better estimate than (9) would be of interest. Nevertheless, even with this estimate, the superiority of adaptive function evaluations for the singularity approximation problem is apparent. The same is valid for the integration problem discussed in the next section.

Remark 4. We show how to modify the algorithm in the case of $a \approx 0$ or $b \approx 1$. This includes $a = 0$ and $b = 0$ where no a priori bound on z_f is known and/or z_f can be very close to either 0 or 1. As we shall see, the modified algorithm performs almost as well as the original algorithm with given a and b that are not too close to 0 and 1.

This algorithm depends on parameters v and q_i $(i = 0, \ldots, v)$ whose choice we shall discuss later.

For simplicity, we first outline the modified algorithm, assuming that $a = 0$ and $b \le (n_0 - k)/n_0$. Initially, $i = 0$ and $b^{(0)} = b$.

A1: Apply q_i steps of the algorithm of §4 for the interval $[0, b^{(i)}]$ (in the first such step we use $h_0^{(i)} = b^{(i)}/(n_0 - k)$ and $x_j^{(i)} = jh_0^{(i)}$, $0 \le j \le n_0$, as a partition). If a new uncertainty interval is constructed, then we apply $r - q_i$ additional steps and terminate. Otherwise, go to A2.

A2: If $i = v - 1$, then we terminate. Otherwise, $i := i + 1$, $b^{(i)} := h^{(i-1)}k2^{1-q_{i-1}}$, and A1 is repeated.

For $a \ge k/n_0$ and $b = 1$, the algorithm is defined in a similar way with the only change being that instead of the interval $[0, b^{(i)}]$ we use $[a^{(i)}, 1]$, where $a^{(i)} = 1 - h^{(i-1)}k2^{1-q_{i-1}}$.

Consider now $a = 0$ and $b = 0$. Perform A1 and/or A2 for $i = 0$ $(a^{(0)} = 0, b^{(0)} = 1$, and $h^{(0)} = 1/n_0)$. The algorithm will construct a new uncertainty interval or, after q_0 unsuccessful steps, will set $i := 1$, $b^{(1)} := h^{(0)}k2^{1-q_0}$, and

$a^{(1)} := 1 - h^{(0)}k2^{1-q_0}$. Next, for $i \geq 1$, **A1** and **A2** are performed in parallel for two subintervals $[0, b^{(i)}]$ and $[a^{(i)}, 1]$.

We now discuss the performance of the modified algorithm. Let PW stand for the probability of wrong decision. Obviously, the probability that the algorithm will choose a wrong subinterval during **A1** ($0 \leq i \leq v - 1$) is bounded as in Theorem 1(i) with $a = 0$ and $b = 1$. By Lemma 2, the probability that in the $(i + 1)$st step the algorithm will search for z_f in $[0, b^{(i+1)}) \cup (a^{(i+1)}, 1]$ whereas $b^{(i+1)} \leq z_f \leq a^{(i+1)}$ is bounded by $\beta_{k,s}(2^{1-q_i}h^{(i)})^\alpha$. Hence,

$$PW \leq (1 + o(1))\frac{(n_0 - 2k)^{1/2+\gamma}}{\sqrt{\pi}} \exp(-(n_0 - 2k)^{1-2\gamma})$$
$$+ \beta_{k,s} \sum_{i=0}^{v-2} (2^{1-q_i}h^{(i)})^\alpha.$$

Obviously,

$$h^{(i)} \leq \frac{b^{(i)}}{n_0 - k} \leq \frac{h^{(i-1)}2^{1-q_{i-1}}k}{n_0 - k} \leq h^{(0)}2^{c_i}\left(\frac{k}{n_0 - k}\right)^i$$

with $c_i = i - \sum_{j=0}^{i-1} q_i$.

Consider now

(12) $q_i = 1, \qquad i = 1, 2, \ldots.$

Then the probability of wrong decision is bounded by

$$PW \leq (1 + o(1))\left(\frac{(n_0 - 2k)^{1/2+\gamma}}{\sqrt{\pi}} \exp(-(n_0 - 2k)^{1-2\gamma})\right.$$
$$\left. + \beta_{k,s}\left(\frac{1}{2^{q_0-1}(n_0 - k)}\right)^\alpha\right).$$

Furthermore, for $v \geq \max\{\alpha/(k - s), r - p\}$ and $q_0 = p$, one can show that (9) of Theorem 1 holds with c_p defined as before and d_p equal to the old d_p plus $(2^p + 2(v - 1))n_0$. (This follows from already established results and the simple fact that $E(X_l X_l) \geq \Theta(|x_l - z|^{2s})$ for $|x_l - z| < (k + 1)h$.) This shows that the modified algorithm (even for $a = 0$ and $b = 0$) works almost as well as the algorithm of §4 (for a and b not too close to 0 and 1, respectively).

6. INTEGRATION PROBLEM

We now apply the results to the integration problem where one is interested in approximating $\int_0^1 f(x)\,dx$ to within a given accuracy ε by performing as few function evaluations as possible.

More specifically, suppose that we want to construct an approximation $A(f)$ to $\int_0^1 f(x)\,dx$ with the error

$$\left|\int_0^1 f(x)\,dx - A(f)\right| \leq \varepsilon \max\{\|f^{(s)}\|_\infty, \|f^{(k)}\|_\infty\}$$

for functions f from the class considered in this paper. Here, for simplicity, we assume that s is fixed and given (see Remark 7).

It can be easily shown that in order to produce such approximations, any algorithm would require at least $\Omega(1/\varepsilon^{1/(s+1)})$ function evaluations in the worst case. This lower bound is achieved by algorithms that use *nonadaptive* function evaluations. Hence, from the worst-case point of view, adaption does not help.

Now relax the worst-case requirement by allowing the algorithm to fail with a small probability, say with probability not exceeding a given small number δ. If *nonadaptive* evaluations are performed, then it can be shown that $\Omega(1/\varepsilon^{1/(s+1)})$ of them are still needed. However, one can do much better by using adaptive function evaluations.

Consider the following scheme. Choose $n_0 \geq 4k + 2$ and start locating the singular point until reaching $[a_r, b_r]$ with

$$(13) \qquad \left(\frac{b_r - a_r}{n_0 - 2k}\right)^{s+1} (n_0 - 2k) \leq \frac{\varepsilon}{2}.$$

Once such an interval is found, performing extra $n_0 - 2k$ function evaluations at equally spaced points in the interval $[a_r, b_r]$, one will be able to approximate $\int_{a_r}^{b_r} f(x)\,dx$ with the error not exceeding $\varepsilon\|f^{(s)}\|_\infty/2$. The integral over $[0, a_r] \cup [b_r, 1]$ can be approximated to within $\varepsilon\|f^{(k)}\|_\infty/2$ by using function values already computed plus at most $\hat{n} = \varepsilon^{-1/k}$ additional evaluations at points outside the interval $[a_r, b_r]$.

Obviously, once we are able to construct the interval $[a_r, b_r]$ containing z_f, this algorithm will approximate the integral of f with the error $\leq \varepsilon\max\{\|f^{(s)}\|_\infty, \|f^{(k)}\|_\infty\}$. The total number of function evaluations performed does not exceed

$$(14) \qquad \hat{n} + (n_0 - 2k) + N_r,$$

where N_r is the number of evaluations needed to construct $[a_r, b_r]$.

Now, given ε and δ, we can apply Theorem 1 to choose n_0 so that the total number of function evaluations is small. To this end, recall that

$$\alpha = \frac{(k - s + 2\gamma)(k - s + 1)}{2} \quad \left(\gamma \in \left[0, \frac{1}{2}\right)\right), \qquad \nu = \frac{k+1}{n_0 - 2k} \leq \frac{1}{2}.$$

Note that $N_r + (n_0 - 2k) \leq d_p + (n_0 - 2k) = 2^p(n_0 - 2k) + 2p\lceil k/2 \rceil + (r - p)(n_0 + 1)$, where n_0, r, and $p < r$ are to satisfy

$$(15) \qquad \mu_p(z \notin [a_r, b_r]) + \widetilde{\beta}_{k,s}\left(\frac{b_0 - a_0}{(n_0 - 2k)2^p}\right)^\alpha \frac{1}{1 - 2^{-\alpha}} \leq \delta$$

(the condition that with probability $\leq \delta$, a wrong decision has been made, or $b_r - a_r > c_p$, or $N_r > d_p$) and

$$(16) \qquad \frac{b_r - a_r}{n_0 - 2k} \leq \frac{c_p}{n_0 - 2k} = \frac{b_0 - a_0}{2^p(n_0 - 2k)}\nu^{r-p} \leq \left(\frac{\varepsilon}{2(n_0 - 2k)}\right)^{1/(s+1)}$$

(to guarantee that (13) holds).

In what follows, we assume for simplicity that δ is sufficiently small. Then for $n_0 - 2k \approx \ln(\delta^{-1}\ln\delta^{-1})$ and $2^p(n_0 - 2k) \approx (b_0 - a_0)(2\widetilde{\beta}_{k,s}/\delta)^{1/\alpha}$ the inequality (15) holds. Using this, we get that (16) holds for

$$r - p \approx (\ln(2(n_0 - 2k)/\varepsilon)^{1/(s+1)} - \ln(2\widetilde{\beta}_{k,s}/\delta)^{1/\alpha})/\ln(1/\nu).$$

For such values of n_0, r, and p, we have

$$N_r \leq (b_0 - a_0) \left(\frac{2\widetilde{\beta}_{k,s}}{\delta} \right)^{1/\alpha} + \frac{\ln \varepsilon^{-1}}{s+1} \left(\frac{\ln \delta^{-1} \ln \delta^{-1}}{\ln \ln \delta^{-1}} + 1 \right) (1 + o(1)).$$

Therefore, the total number of function evaluations is bounded by

$$\hat{n} + (n_0 - k) + N_r \leq \varepsilon^{-1/k} + (b_0 - a_0) \left(\frac{2\widetilde{\beta}_{k,s}}{\delta} \right)^{1/\alpha}$$

(17)
$$+ \frac{\ln \varepsilon^{-1}}{s+1} \left(\frac{\ln \delta^{-1} \ln \delta^{-1}}{\ln \ln \delta^{-1}} + 1 \right) (1 + o(1))$$

$$\leq \varepsilon^{-1/k}(1 + o(1)) + (b_0 - a_0) \left(\frac{2\widetilde{\beta}_{k,s}}{\delta} \right)^{1/\alpha}$$

for small ε and δ. Thus, for small ε and δ, the adaptive integration rule is much superior to any nonadaptive rule, especially when $s \ll k$.

For fixed δ, and ε tending to zero, this adaptive integration can be improved as follows. Suppose that in the adaptive scheme presented above, the integrals over $[0, a_r]$ and $[b_r, 1]$ are approximated by the integrals of natural splines of degree $2k + 1$ that interpolate f at equally spaced points. Since $f|_{[0, a_r]}$ is distributed according to w_k, the results from [6, p. 366] for probabilistic integration with normalized error criterion state that using

$$\tilde{n}_1 = \tilde{n}_1(\varepsilon, \delta) = \left(\frac{\sqrt{\ln \delta^{-1}}}{\varepsilon} \right)^{(1+o(1))/(k+1)} \quad \text{evaluations},$$

the error of approximating $\int_0^{a_r} f(x)\,dx$ does not exceed $\varepsilon \|f^{(k)}\|_\infty / 4$ with probability at least $1 - \delta/4$. The same can be proven for the integral over $[b_r, 1]$. Hence, using

$$\tilde{n} = \tilde{n}(\varepsilon, \delta) = \left(\frac{\sqrt{\ln \delta^{-1}}}{\varepsilon} \right)^{(1+o(1))/(k+1)} \quad \text{evaluations},$$

the error of approximating the integral over $[0, a_r] \cup [b_r, 1]$ does not exceed $\varepsilon \|f^{(k)}\|_\infty / 2$ with probability at least $1 - \delta/2$. Replacing the right-hand side δ of (15) by $\delta/2$, we get that with probability at least $1 - \delta$, the error of approximating the integral $\int_0^1 f(x)\,dx$ does not exceed $\varepsilon \max\{\|f^{(s)}\|_\infty, \|f^{(k)}\|_\infty\}$ when using

(18)
$$\tilde{n} + (n_0 - k) + N_r \leq \left(\frac{\sqrt{\ln \delta^{-1}}}{\varepsilon} \right)^{(1+o(1))/(k+1)}$$

$$+ (b_0 - a_0) \left(\frac{4\widetilde{\beta}_{k,s}}{\delta} \right)^{1/\alpha} \quad \text{evaluations}.$$

We do not know if this adaptive rule is optimal with respect to δ. However, if one considers fixed (even very small) δ, and ε tending to zero, then the rule is (modulo constants) optimal. Indeed, for small ε the number of evaluations is proportional to $\tilde{n}(\varepsilon, \delta)$. Even if the singular points were given explicitly for

every function f, one would still need $\tilde{n}(\varepsilon, \delta)$ function evaluations, which has been proven in [6, p. 366].

For δ small relative to ε, the bound on the number of function evaluations depends on δ essentially through $(2\widetilde{\beta}_{k,s}/\delta)^{1/\alpha}$. We believe that the actual number of function evaluations is smaller than the bound presented above. In order to get such a smaller bound, one would need to improve the estimate (9) (see Remarks 3 and 6).

We end this section with the following remark.

Remark 5. Our adaptive information scheme is based on computing backward differences and comparing them to h^k. Although in some adaptive quadrature rules there is no explicit use of backward differences, the backward differences are often used implicitly. To see this, let us very briefly consider the adaptive Simpson's rule (see, e.g., [2, 3]), where the decision about partition is made based on whether $|\widehat{S}_i - S_i|/15$ is small. Here, $\widehat{S}_i$ is the (composite) Simpson's rule based on $f(x_i)$, $f(x_i + h)$, ..., $f(x_i + 4h)$, and S_i is the Simpson's rule based on $f(x_i)$, $f(x_i + 2h)$, $f(x_i + 4h)$. A simple calculation gives us $(\widehat{S}_i - S_i)/15 = -h(\Delta_{i+4}^4 f)/45$, where $\Delta_{i+4}^4 f$ is the 4th-order backward difference of f at x_{i+4}. This means that new evaluation points are chosen based on the size of a corresponding backward difference, which is the approach taken in our algorithm.

7. CONCLUDING REMARKS

We end the paper with a number of remarks outlining our research plans.

Remark 6. We would like to improve the estimates obtained in Theorem 1. Note that small improvements in (9) can be easily obtained. For instance, we could take $T(h)$ that is closer to $h^{k+1/2}$ than $T(h) = \sqrt{2}h^{k+\gamma}$ $(\gamma < \frac{1}{2})$, say $T(h) \approx h^{k+1/2}/\ln h^{-1}$, or we could take (in initial steps) backward differences of order smaller than $k + 1$ (see Remark 2). However, significant improvements might require much more elaborate analysis and/or modification of our algorithm.

One might improve the algorithm by introducing a *backtracking* technique. For instance, if the uncertainty interval is not reduced after several steps, the algorithm will backtrack to an earlier partition and find a new interval to work on. This should reduce the probability of making wrong decisions.

Remark 7. In this paper we consider functions f with exactly one singular point z_f. Hence, each f consists of two pieces of k-regular functions such that $f^{(s-1)}$ is absolutely continuous. We plan to extend our analysis for classes of functions f that consist of an unknown (bounded) number of regular pieces (i.e., f might have a number of singular points), each piece of different regularity k_i and each singular point of different regularity s_i. The parameters k_i and s_i need not be known, and some k_i's might equal infinity. Some singular points z_i might be *poles* of $f^{(s_i)}$. However, poles are easier to approximate than z_i's with bounded $\|f^{(s_i)}\|_\infty$, as assumed in this paper.

We plan to analyze the problem of approximating the singular points for such a generalized class of functions. Here, by approximating all singular points of f we mean constructing a set (a union of intervals) containing the singular points and having small Lebesgue measure.

If we assume that an upper bound on the number of singular points is given, then a relatively simple modification of our algorithm approximates the singular points well with high probability. This modification is based on the following observation. Let $X_x^p = X_x^p(f)$ be the backward difference of order p of f at a point x. Consider x such that $f|_{(z,x]} \in C^k$. Then for every $p \leq s$, the expected value of $(X_x^p)^2$ is proportional to h^{2p}. For $s < p \leq k$, it is proportional to h^{2s} or to h^{2p} depending on whether x is *close* to z or not. This, together with the analysis, will be reported in the future. We also plan to analyze the problem of estimating the regularity parameters k_i and s_i along with the points z_i.

Remark 8. In §6 we used an integration problem only for the purpose of illustration. Some practically important questions were left untouched. One such question concerns getting an a posteriori error estimate. To be more specific, recall that in the integration problem of §6 we wanted to approximate $\int_0^1 f(x)\,dx$ with an error not exceeding $\varepsilon \max\{\|f^{(s)}\|_\infty, \|f^{(k)}\|_\infty\}$. Adaptive rules try to approximate integrals with small errors, but in addition to that, they attempt to provide the user with an a posteriori bound on the error. For instance, in the adaptive Simpson's rule (see Remark 5), $(\widehat{S}_i - S_i)/15$ is used as an estimate of $\int_{x_i}^{x_{i+4}} f(x)\,dx - \widehat{S}_i$.

We want to analyze classical techniques for getting a posteriori bounds, which we expect to be reliable in a probabilistic sense (of course, in the worst case, getting nontrivial reliable bounds is impossible). In particular, we would like to know the probability that

$$\text{small } \frac{|\widehat{S}_i - S_i|}{15} \quad \text{implies} \quad \text{small } \left| \int_{x_i}^{x_{i+4}} f(x)\,dx - \widehat{S}_i \right|.$$

This is in the spirit of [5], where for functions with *no* singularities probabilistic a posteriori estimates have been studied. We plan to investigate this problem for functions with a number of singular points (see Remark 7) and we want to study adaptive quadrature rules that are efficient (require a relatively small number of function evaluations) and provide probabilistically reliable error estimates.

BIBLIOGRAPHY

1. T. W. Anderson, *The integral of a symmetric unimodal function over a symmetric convex set and some probability inequalities*, Proc. Amer. Math. Soc. **6** (1955), 170–176.

2. S. D. Conte and C. de Boor, *Elementary numerical analysis—an algorithmic approach*, 3rd ed., McGraw-Hill, New York, 1980.

3. P. J. Davis and P. Rabinowitz, *Methods of numerical integration*, 2nd ed., Academic Press, Orlando, FL, 1984.

4. C. de Boor, *Splines as linear combinations of B-splines. A survey*, Approximation Theory II (G. G. Lorentz, C. K. Chui, and L. L. Schumaker, eds.), Academic Press, New York, 1976, pp. 1–47.

5. F. Gao, *A probabilistic theory for error estimation in automatic integration*, Numer. Math. **56** (1989), 309–329.

6. J. F. Traub, G. W. Wasilkowski, and H. Woźniakowski, *Information-based complexity*, Academic Press, New York, 1988.

7. G. Wahba, *Improper priors, spline smoothing and the problem of guarding against model errors in regression* J. Roy. Statist. Soc. Ser. B **40** (1978), 364–372.

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

DEPARTMENT OF COMPUTER SCIENCE, UNIVERSITY OF KENTUCKY, LEXINGTON, KENTUCKY 40506
E-mail address: greg@ms.uky.edu

DEPARTMENT OF COMPUTER SCIENCE, UNIVERSITY OF BRITISH COLUMBIA, VANCOUVER, BRITISH COLUMBIA V6T 1W5, CANADA
E-mail address: gao@cs.ubc.ca