On Detecting Regularity of Functions: A Probabilistic Analysis

by
F. Gao
and
G. W. Wasilkowski

Technical Report 91-28
November 1991

Department of Computer Science
University of British Columbia
Rm 333 - 6356 Agricultural Road
Vancouver, B.C.
CANADA V6T 1Z2

On Detecting Regularity of Functions: A Probabilistic Analysis

F. Gao*
Department of Computer Science
University of British Columbia
Vancouver, B.C., V6T 1W5, CANADA

and

G. W. Wasilkowski†
Department of Computer Science
University of Kentucky
Lexington, KY 40506, USA

Abstract: We study the problem of detecting the regularity degree $deg(f) = \max\{k : k \leq r, f \in C^k\}$ of functions based on a finite number of function evaluations. Since it is impossible to find $deg(f)$ for any function f, we analyze this problem from a probabilistic perspective. We prove that when the class of considered functions is equipped with a Wiener-type probability measure, one can compute $deg(f)$ exactly with super exponentially small probability of failure. That is, we propose an algorithm which, given n function values at equally spaced points, might propose a value different than $deg(f)$ only with probability $O\left((n^{-1}\ln n)^{(n-r)/4}\right)$. Hence, regularity detection is easy in the probabilistic setting even though it is unsolvable in the worst case setting.

*Research supported in part by Natural Sciences and Engineering Research Council of Canada under Grant OGP0041639.

†Research supported in part by the National Science Foundation under Grant CCR-91-14042.

1 Introduction

We study the problem of detecting the degree of regularity of a function f based on a finite number of its values $f(x_i)$ $(1 \leq i \leq n)$. That is, for a continuous function $f : [0,1] \to \mathbb{R}$. we would like to compute the maximal integer $k = k(f)$ such that $f \in C^k = C^k[0,1]$. Since to approximate $f^{(k)}$ (should it exist) at least $k+1$ function values are needed, deciding whether $f \in C^k$ for $k \geq n$ is impossible. Therefore we study the following modified problem: *Given a positive integer* r *(*$r \ll n$*); compute* $\deg(f) \equiv \max\{k : k \leq r, f \in C^k\}$.

Detecting regularity of functions is an important problem from a theoretical point of view. The theoretical optimality of many algorithms for various problems, such as integration or function approximation (see, e.g., [5, 6, 9, 10] and papers cited therein), depends on the regularity of the underlying class of functions, and in general different classes (say C^{k_1} and C^{k_2}) poses different optimal algorithms. Thus, if the regularity of functions under considerations is unknown, optimality results that assume known regularity need not be applicable. The knowledge of the regularity degree is also very helpful from a practical point of view since, without it, good decisions concerning an appropriate choice of algorithms as well as termination of an algorithm are difficult to make. To see it, consider briefly the integration problem where we want to approximate $S(f) = \int_0^1 f(x)\,dx$. Suppose our algorithm uses a quadrature Q_k that relies on the assumption $f \in C^k$. If $k \neq \deg(f)$ then two scenarios could happen. In case of underestimating $\deg(f)$, Q_k is much less efficient than $Q_{\deg(f)}$. Furthermore, since bounded (small) $\|f^{(\deg(f))}\|$ need not imply small $\|f^{(k)}\|$, the error of Q_k could be very large. In case of overestimating $\deg(f)$, the rule Q_k is, modulo a multiplicative constant, as good as $Q_{\deg(f)}$ (see e.g., [11]). However, the constant grows exponentially with $k-\deg(f)$. Moreover, since the actual error of Q_k behaves asymptotically as the error of $Q_{\deg(f)}$, its converges is significantly slower than the anticipated convergence of Q_k when applied to k-times differentiable functions. This could result in premature termination of the algorithm.

Thus, the precise knowledge of $\deg(f)$ is of a great deal of interest. However, as it is well known, without some very restrictive assumptions on the class of functions, computing $\deg(f)$ for all functions in the class is intractable in the worst case setting. This is why we address this problem from a probabilistic perspective. More precisely, by assuming the existence of a reasonable probability measure on the underlying class of functions, we give an algorithm that computes the correct $\deg(f)$ with a very high probability. This algorithm makes the decision based on the behavior of forward differences of function values. Since decisions on termination of algorithms in numerical quadratures are often based (at least implicitly) on some form of differences, this result can also be viewed as a theoretical basis for these numerical techniques. When the regularity of the functions is known, a probabilistic analysis of numerical integration algorithms that use divided differences as termination criteria has been pursued in [3, 4].

The main result of the paper states that one can compute *deg*(f) exactly with super exponentially small probability of failure. That is, our algorithm might propose a value different than *deg*(f) only with probability $O\left((n^{-1}\ln n)^{(n-r)/4}\right)$. (Recall that n stands for the number of function values and r is the bound on the regularity degree.) Hence the regularity detection is an easy problem from the probabilistic complexity point of view, whereas it is unsolvable in the worst case setting.

We stress that functions arising in practice are more complicated than those studied in this paper. Indeed, in this paper we concentrate on functions that have the same regularity degree $\deg(f)$ in the whole domain $[0,1]$. However, in practice we often deal with functions that are piecewise regular. That is, f consists of a number of regular pieces, $f(x) = f_i(x)$ for $x \in I_i = (z_{i-1}, z_i)$. In each subinterval, the degree of regularity $\deg_i = \deg(f_i)$ can be different. Furthermore, at each singular point $z_i = z_i(f)$ that is unknown and varies with f, the function f can have a different (and unknown) degree s_i of smoothness. Hence, for theoretical results to have an impact on practical applications, one needs to consider the more general regularity detection problem where f is piecewise regular, as described above, and the task is to approximate the intervals $I_i = (z_{i-1}, z_i)$ together with computing the regularity degree $\deg_i$ of f restricted to I_i.

This paper does not address the general problem. It only constitutes another step in this direction. It is a continuation of our previous paper [12], where the problem of approximating singular points $z(f)$ of piecewise regular functions has been studied. We proved there that, with a very high probability, one can approximate singular points very accurately with relatively low cost. Based on that and the results presented in this paper, in a forthcoming paper [2] we will show that under some relatively nonrestrictive assumptions (such as uniformly bounded and not too large number of pieces f_i) the general problem is tractable in the probabilistic case.

2 Problem formulation and basic definitions

Let $F = C^0$ be the class of continuous functions defined on the interval $[0,1]$. Let r be a positive integer, and let the regularity degree of f be defined by

$$\deg(f) = \max\{k : k \leq r, f \in C^k\}.$$

Here, $C^k = \{f : [0,1] \to \mathbb{R} : f^{(k)} \text{ continuous}\}$. Since $F = \bigcup_{k=0}^r C^k$ and $C^{k+1} \subset C^k$, the space F can be endowed with the following Borel probability measure Prob:

$$\text{Prob}(C^r) = \alpha_r \qquad \text{and} \qquad \text{Prob}(C^k \setminus C^{k+1}) = \alpha_k \quad \forall\, k \in \{0, \ldots, r\},$$

and the conditional probability

$$\text{Prob}(f \in A | \deg(f) = k) = w_k(A) \quad \forall \text{ Borel set } A \subseteq C^k.$$

Here α_k are nonnegative numbers such that $\sum_{k=0}^r \alpha_k = 1$ and w_k is the k-fold Wiener measure. That is, when $\deg(f) = k$, the function f is distributed according to the k-fold Wiener measure w_k, and the probability that $\deg(f) = k$ equals α_k.

The problem studied in this paper is to compute $\deg(f)$ for all functions $f \in F$ but a set of small probability Prob. More specifically, given $n \gg r$ and

$$N_n(f) = [f(x_0), \ldots, f(x_n)],$$

we want an algorithm $\psi : \mathbb{R}^n \to \{0, \ldots, r\}$ for which

$$\text{Prob}(\{f \in F : \psi(N_n(f)) \neq \deg(f)\}) \quad \text{is small.}$$

In this paper, we assume that the points x_i are equally spaced in some interval $[a,b] \subseteq [0,1]$, i.e.,

$$x_i = a + ih \qquad \text{with} \qquad h = \frac{b-a}{n}.$$

Before presenting an algorithm that computes $\deg(f)$ with very high probability (see Section 3), we comment on the choice of the probability Prob. The numbers α_k are needed only for a precise definition of Prob. In this paper, we do not assume that they are known. Instead, we present an algorithm that works very well regardless of their values. The choice of k-fold Wiener measures w_k is somewhat arbitrary. However, due to their interesting mathematical properties they are one of the most frequently used measures on function spaces in a variety of fields (e.g., numerical analysis, operation research, physics, statistics). Furthermore, their properties provide an adequate model for what $\deg(f)$ means. A more detailed discussion of Wiener measures can be found in e.g., [1, 7, 9]. Here we briefly recall their basic properties which will be used in this paper.

For $k=0$, w_0 (the classical Wiener measure) is a zero-mean Gaussian measure with the covariance kernel

$$K_0(x,y) := \int_{C^0} f(x)f(y)\, w_0(df) = \min\{x,y\}.$$

Equivalently, f is a zero-mean Gaussian stochastic process with the autocorrelation $K_0(x,y)$ given above. For $k \geq 1$, f distributed according to w_k can be viewed as a k-fold integrated Wiener process g, i.e., $f(x) = \int_0^1 g(t)(x-t)_+^{k-1}/(k-1)!\, dt$ with g is distributed according to w_0. Hence, w_r is a zero-mean Gaussian measure with the the covariance kernel

$$K_k(x,y) := \int_{C^k} f(x)f(y)\, w_k(df) = \int_0^1 \frac{(x-t)_+^k (y-t)_+^k}{k!\, k!}.$$

We end this section by the following remarks

Remark 1 Note that w_k concentrates on functions with $f^{(j)}(0) = 0$ for $0 \leq j \leq k$. This (a rather peculiar) property of w_k could easily be removed by taking $f(x) = f_1(x) + f_2(1-x)$ with independent f_1, f_2, both distributed according to w_k. It is possible to show, that our algorithm works as well for such a modified probability distribution; for the modified distribution, its probability of failure differs from $\mathrm{Prob}(\{f : \deg(f) \neq \psi(N_n(f))\})$ only by a multiplicative constant of the order of unity. For piecewise regular functions that we will study in [2], their distribution will be specified by a distribution of pieces f_i, which in turn will be equal to $f_i(x) = g_i(x)$ or $f_i(x) = g_i(1-x)$ (for x close to zero) with g_i being a $\deg_i$-fold Wiener process. Thus, f will not vanish at 0.

Remark 2 Since our measure concentrates on functions with the same regularity everywhere in $[0,1]$, $r+1$ function values at points close to one another suffice to detect the correct $\deg(f)$ with a high probability. In fact, when the property that f vanishes at zero together with its first k derivatives is utilized, we could use only one value $f(h)$ for a small h. The smaller h is the higher the probability of success. Hence the regularity detection problem as posed in this paper would be trivial when arbitrary function values are allowed. However, as stated in Introduction our eventual goal is to tackle the problem for piecewise regular functions. For piecewise regular functions, that is no longer the case. Because of varying regularity in unknown subintervals I_i, one needs to use points across the whole interval. This is why we have chosen $N_n(f)$ consisting of function values at equally spaced points. The fact that the points we choose are equally spaced is not very crucial to our estimates. Similar conclusions can be drawn when values at $0 \le x_1 \le \ldots \le x_n \le 1$ are used with $\min_i(x_i - x_{i-1}) \approx \max_i(x_i - x_{i-1}) \approx h$.

3 An algorithm

We propose the following algorithm for deciding the regularity of f.

For $k = 1, \ldots, r+1$ and $i \le n-k$, let $X_{k,i} = X_{k,i}(f)$ be the kth forward difference of f at $x_i = a + ih$. Obviously,

$$X_{k,i} = \sum_{j=0}^{k} \binom{k}{j} (-1)^j f(x_{i+j}) \quad \text{and} \quad X_{k,i} = X_{k-1,i+1} - X_{k-1,i} \text{ with } X_{0,i} = f(x_i).$$

For $k = 1, \ldots, r+1$, define

$$X_k = X_k(f) := \max_{0 \le i \le n-k} |X_{k,i}(f)|.$$

The decision algorithm is

$$\psi(N_n(f)) = \max\left\{j : j \le r \text{ and } X_{j+1}(f) \le b_j h^j\right\}$$

if such a k exists, and

$$\psi(N_n(f)) = 0$$

otherwise. Here $b_j = b_j(n)$ are positive reals whose choice will be addressed in Section 4.2.

The error of ψ is defined by

$$E(\psi) = \text{Prob}\left(\{f \in F : \deg(f) \ne \psi(N_n(f))\}\right).$$

4 Estimating $E(\psi)$

Let A_k be the set of functions with $\deg(f) = k$ for which ψ delivers incorrect value. Then

$$E(\psi) \le \sum_{k=0}^{r} \alpha_k w_k(A_k).$$

Denoting by A_k^- and A_k^+ the subsets of A_k on which ψ underestimates and overestimates the degree k, respectively, we have $w_k(A_k) = w_k(A_k^-) + w_k(A_k^+)$. Of course, $A_r^+ = A_0^- = \emptyset$ and thus

$$E(\psi) \le \sum_{k=1}^{r} \alpha_k w_k(A_k^-) + \sum_{k=0}^{r-1} \alpha_k w_k(A_k^+). \tag{1}$$

The sets

$$A_k^- = \left\{f : \deg(f) = k, \forall j \ge k : X_{j+1} > b_j h^j\right\} \qquad \text{for } k \ge 1.$$

and

$$A_k^+ = \left\{f : \deg(f) = k, \exists j \ge k+1 : X_{j+1} \le b_j h^j)\right\} \qquad \text{for } k \le r-1.$$

4.1 Estimating $w_k(A_k)$

In this subsection $k \in \{0, \ldots, r\}$ and $j \in \{k+1, \ldots, r+1\}$. Since w_k is a Gaussian measure, the random variables $X_{j,l}$ $(0 \leq l \leq n-j)$ restricted to C^k are also Gaussian. Take $l_1, l_2 \leq n-j$. Since

$$\int_{C^k} f(x)f(y)\, w_k(df) = \int_0^1 \frac{(x-t)_+^k}{k!} \frac{(y-t)_+^k}{k!}\, dt,$$

the expected value of $X_{j,l_1}X_{j,l_2}$ (with respect to w_k) is

$$E_k(X_{j,l_1}X_{j,l_2}) = \int_0^1 \Delta_{l_1}^j \left(\frac{(\cdot - t)_+^k}{k!}\right) \Delta_{l_2}^j \left(\frac{(\cdot - t)_+^k}{k!}\right) dt,$$

where Δ_l^j is the jth forward difference operator at x_l. Thus,

$$E_k(X_{j,l_1}X_{j,l_2}) = 0 \quad \text{if } j \geq k+1 \text{ and } |l_1 - l_2| > j. \tag{2}$$

From the well-know properties of B-splines we also conclude that for $j = k+1$, $E_k(X_{k+1,l}X_{k+1,l})$ does not depend on l and is bounded by

$$E_k(X_{k+1,l}X_{k+1,l}) = h^{2k}||N_l^{k+1}||_{L_2} \leq h^{2k+1} = \left(\frac{b-a}{n}\right)^{2k+1}, \tag{3}$$

where N_l^{k+1} is the lth normalized B-spline of degree $k+1$.

Proposition 1 *For every $k \in \{1, \ldots, r\}$,*

$$w_k(A_k^-) \leq \sqrt{\frac{2}{\pi}} \frac{(n-k)\sqrt{h}}{b_k} \exp\left(-b_k^2/(2h)\right).$$

Proof: Let $B_k = w_k(\{f \in A_k^- : X_{k+1} > b_k h^k\})$. Then $w_k(A_k^-) \leq B_k$. Since $X_{k+1}(f) > b_k h^k$ is equivalent to $|X_{k+1,l}(f)| > b_k h^k$ for some $l \in \{0, \ldots, n-k-1\}$, we get

$$B_k \leq \sum_{l=0}^{n-k-1} w_k(\{f \in C^k : |X_{k+1,l}| > b_k h^k\}).$$

Denote $\sigma = E_k(X_{k+1,l}X_{k+1,l})$. Since $X_{k+1,l}$ has a normal $\mathcal{N}(0,\sigma)$ distribution,

$$B_k \leq (n-k)\sqrt{\frac{2}{\pi}} \int_{\beta_{k,n}}^{+\infty} e^{-t^2/2}\, dt \leq (n-k)\sqrt{\frac{2}{\pi}}\, \beta_{k,n}^{-1} \int_{\beta_{k,n}}^{+\infty} t\, e^{-t^2/2}\, dt = (n-k)\sqrt{\frac{2}{\pi}}\, \beta_{k,n}^{-1}\, e^{-\beta_{k,n}^2/2},$$

where $\beta_{k,n} = b_k h^k/\sqrt{\sigma}$. Hence, (3) completes the proof. □

Since $A_k^+ = \bigcup_{j \geq k+1} \{f : \deg(f) = k, X_{j+1} \leq b_j h^j\}$, we estimate $w_k(A_k^+)$ by studying first the probability of $X_{j+1} \leq b_j h^j$, separately for each j.

Proposition 2 *For every $k \in \{0, \ldots, r-1\}$ and $j \geq k+1$,*

$$w_k\left(\left\{\deg(f) = k : X_{j+1} \leq b_j h^j\right\}\right) \leq \frac{1}{\Gamma\left(\frac{j+3}{2}\right)^{n_j+1}} \left(h^{j-k-1/2} b_j \tilde{c}_{k,j}\right)^{(n-j)/2},$$

where

$$n_j = \left\lfloor \frac{n-2j-1}{2(j+1)} \right\rfloor \quad \textit{and} \quad \tilde{c}_{k,j} = \frac{(2j-k-1)!\sqrt{(j+1)(2j+1)(2j-k-1)}}{j! 2\pi^{j-k-1}}.$$

To prove this proposition we need the following two Lemmas.

Lemma 1 *Let j, m, i be integers satisfying $j \geq k+1$, $m \geq 0$, and $1 \leq i \leq n-m-j-1$. Let $S = (s_{l_1,l_2})$ be an $(i \times i)$ matrix with the entries*

$$s_{l_1,l_2} = E_k\left(X_{j+1,l_1} X_{j+1,l_2}\right) \quad m \leq l_1, l_2 \leq m+i-1,$$

which obviously is positive definite. Then

$$h^{-k-1/2} c_{k,i,j} \, ||S^{1/2}\vec{y}||_\infty \geq ||\vec{y}||_2 \quad \forall \vec{y} \in \mathbb{R}^i,$$

where

$$c_{k,i,j} = \frac{(i+j-k-1)!\sqrt{i(i+j)(i+j-k-1)}}{\pi^{j-k-1}(i-1)!\sqrt{2}}.$$

Proof: Due to the equivalence of $||\cdot||_\infty$ and $||\cdot||_2$ on $\mathbb{R}^i$, we have

$$||S^{1/2}\vec{y}||_\infty \geq i^{-1/2}||S^{1/2}\vec{y}||_2 = i^{-1/2}\left\langle S\vec{y}, \vec{y}\right\rangle_2^{1/2} \geq ||\vec{y}||_2\sqrt{i^{-1}\lambda(S)}, \tag{4}$$

where $\langle \cdot, \cdot \rangle_2$ is the standard inner product on $\mathbb{R}^i$ and $\lambda(S)$ is the minimal eigenvalue of S.

To estimate $\lambda(S)$, note that

$$\begin{aligned}
s_{l_1,l_2} &= E_k(X_{j,l_1} X_{j,l_2}) = \int_0^1 \Delta_{l_1}^{j+1}\left(\frac{(\cdot - t)_+^k}{k!}\right) \Delta_{l_2}^{j+1}\left(\frac{(\cdot - t)_+^k}{k!}\right) dt \\
&= \int_0^1 \left(\sum_{p_1=0}^{j+1}(-1)^{p_1}\binom{j+1}{p_1}\frac{(x_{l_1+p_1} - x)_+^k}{k!}\right)\left(\sum_{p_2=0}^{j+1}(-1)^{p_2}\binom{j+1}{p_2}\frac{(x_{l_2+p_2} - x)_+^k}{k!}\right) dx \\
&= \sum_{r_1,r_2=0}^{j-k-1}(-1)^{r_1}\binom{j-k-1}{r_1}(-1)^{r_2}\binom{j-k-1}{r_2} b_{l_1+r_1,l_2+r_2},
\end{aligned} \tag{5}$$

where b_{s_1,s_2} equals

$$\int_0^1 \left(\sum_{p_1=0}^{k+2}(-1)^{p_1}\binom{k+2}{p_1}\frac{(x_{s_1+p_1} - x)_+^{k+1}}{(k+1)!}\right)'\left(\sum_{p_2=0}^{k+2}(-1)^{p_2}\binom{k+2}{p_2}\frac{(x_{s_2+p_2} - x)_+^{k+1}}{(k+1)!}\right)' dx.$$

Hence

$$b_{s_1,s_2} = h^{2k+2}\int_0^1 \left(N_{s_1}^{k+2}(x)\right)'\left(N_{s_2}^{k+2}(x)\right)' dx \tag{6}$$

with N_s^{k+2} being a normalized B-spline of degree $k+2$.

From (5) we get

$$S = VBV^T, \tag{7}$$

where B is a $(i+j-k-1)\times(i+j-k-1)$ matrix whose (s_1,s_2)th entry is given by (6), $m-(j-k)+1 \le s_1, s_2 \le m+i-1$. The matrix V equals

$$V = W_i \cdot W_{i+1} \cdot \ldots \cdot W_{j+i-k-2}$$

where W_l is the following $l \times (l+1)$ matrix

$$W_l = \begin{bmatrix} -1 & 1 & 0 & 0 & . & . & . & 0 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & . & . & . & 0 & 0 & 0 & 0 \\ . & . & . & . & . & . & . & . & . & . & . \\ 0 & 0 & 0 & 0 & . & . & . & 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & 0 & . & . & . & 0 & 0 & -1 & 1 \end{bmatrix}$$

Due to (7),

$$\lambda(S) \ge \lambda(B)\lambda(VV^T), \tag{8}$$

where $\lambda(B)$ and $\lambda(VV^T)$ are the minimal eigenvalues of B and VV^T, respectively.

We first estimate

$$\lambda(B) = \min_{\vec{\alpha}\in\mathbb{R}^{i+j-k-1}} \frac{\langle B\vec{\alpha}, \alpha\rangle_2}{\langle \vec{\alpha}, \alpha\rangle_2}.$$

Letting $g_{\vec{\alpha}}(x) = \sum_{s=m-(j-k)+2}^{m+i} \alpha_s N_s^{k+2}(x)$, (6) implies $\langle B\vec{\alpha}, \vec{\alpha}\rangle_2 = h^{2k+2}||g'_{\vec{\alpha}}||_{L_2}^2$. Since $\mathrm{supp} g_{\vec{\alpha}} \subseteq [x_{m-j}, x_{m+i}]$,

$$||g_{\vec{\alpha}}||_{L_\infty} \le \sqrt{(i+j)h/2}\,||g'_{\vec{\alpha}}||_{L_2}.$$

It is well known, see e.g., [8] (Thm. 4.44), that $||g_{\vec{\alpha}}||_{L_\infty} \ge ||\vec{\alpha}||_\infty(2/\pi)^{k+2}$. Since $||\vec{\alpha}||_\infty/||\vec{\alpha}||_2 \ge (i+j-k-1)^{-1/2}$, we get

$$\sqrt{\lambda(B)} \ge \frac{h^{k+1/2}\sqrt{2}}{\sqrt{(i+j)(i+j-k-1)}}\left(\frac{2}{\pi}\right)^{k+2}. \tag{9}$$

To estimate $\lambda(VV^T)$ note that

$$\lambda(VV^T) \ge \prod_{l=i}^{j+i-k-2} \lambda(W_l W_l^T)$$

Since $W_l W_l^T$ is an $l \times l$ tri-diagonal matrix with diagonal elements equal to 2 and codiagonal elements equal to -1, it is well known that $\lambda(W_l W_l^T) = 4\sin^2(\pi/(2l)) \ge (\pi/l)^2$. Hence

$$\sqrt{\lambda(VV^T)} \ge \frac{\pi^{j-k-1}}{\prod_{l=i}^{j+i-k-2} l}. \tag{10}$$

Hence (4), (8), (9), and (10) imply

$$||S^{1/2}\vec{y}||_\infty \ge ||\vec{y}||_2 h^{k+1/2} \frac{\pi^{j-k-1}\sqrt{2}}{\sqrt{i(i+j)(i+j-k-1)}\prod_{l=i}^{i+j-k-2} l}.$$

This completes the proof. □

Lemma 2 *Let j, m, i be as in Lemma 1. Then*

$$w_k\left(\left\{f \in C^k : \max_{m+1\le l\le m+i} |X_{j+1,l}(f)| \le b_j h^j\right\}\right) \le \left(h^{j-k-1/2} b_j \hat{c}_{k,i,j}\right)^i \frac{1}{\Gamma\left(\frac{i}{2}+1\right)},$$

where $\hat{c}_{k,i,j} = c_{k,i,j}/\sqrt{2}$.

Proof: Let L denotes $w_k\left(\left\{f \in C^k : \max_{m+1\le l\le m+i} |X_{j+1,l}(f)| \le b_j h^j\right\}\right)$. The random vector $[X_{j+1,m+1}, \ldots, X_{j+1,m+i}]^T$ has a normal $\mathcal{N}(0, S)$ distribution with S given in Lemma 1. Hence

$$L = \frac{1}{\sqrt{(2\pi)^i \det(S)}} \int_{\vec{x}\in \mathbb{R}^i, \|\vec{x}\|_\infty \le b_j h^j} \exp\left(-\left\langle S^{-1}\vec{x}, \vec{x}\right\rangle_2 /2\right) d\vec{x}.$$

Changing the variables, $\vec{y} := S^{-1/2}\vec{x}$, we get

$$L = (2\pi)^{-i/2} \int_{\|\vec{y}\|_\infty \le b_j h^j} \exp(-\|\vec{y}\|_2^2/2)\, d\vec{y} \le (2\pi)^{-i/2} \mu_i\left(\left\{\vec{y} \in \mathbb{R}^i : \|S^{1/2}\vec{y}\|_\infty \le b_j h^j\right\}\right),$$

where μ_i is the Lebesgue measure on $\mathbb{R}^i$. Due to Lemma 1, we have

$$L \le (2\pi)^{-i/2} \mu_i\left(\left\{\vec{y} \in \mathbb{R}^i : \|\vec{y}\|_2 \le R\right\}\right) = (2\pi)^{-i/2} \frac{R^i \pi^{i/2}}{\Gamma(i/2+1)} = \frac{R^i}{2^{i/2}\Gamma(i/2+1)}$$

with $R = h^{j-k-1/2} b_j / c_{k,i,j}$. This completes the proof. □

Proof of Proposition 2: Recall that for given k and $j \ge k+1$, $n_j = \lfloor (n-2j-1)/(2(j+1)) \rfloor$. Define

$$Y_{j+1,s}(f) = \max_{2s(j+1)\le l\le 2s(j+1)+j} |X_{j+1,l}(f)| \quad \text{for } s = 0, \ldots, n_j$$

and

$$Y_{j+1,n_j+1}(f) = \max_{n-j\le l\le 2(n_j+1)(j+1)} |X_{j+1,l}(f)|$$

with the convention that $Y_{j+1,n_j+1} \equiv 0$ if $n - j > 2(n_j+1)(j+1)$. Lemma 2 with $i = j+1$ yields

$$w_k\left(\left\{f \in C^k : Y_{j+1,s} \le b_j h^j\right\}\right) \le \left(h^{j-k-1/2} b_j \hat{c}_{k,j+1,j}\right)^{j+1} \frac{1}{\Gamma\left(\frac{j+3}{2}\right)} \tag{11}$$

for $s = 0, \ldots, n_j$, and with $i^* = (n - j - 2(n_j+1)(j+1))_+$ yields

$$w_k\left(\left\{f \in C^k : Y_{j+1,n_j+1} \le b_j h^j\right\}\right) \le \left(h^{j-k-1/2} b_j \hat{c}_{k,j+1,j}\right)^{i^*} \frac{1}{\Gamma\left(\frac{j+3}{2}\right)^{sgn(i^*)}} \tag{12}$$

since $i^* \le j$ and $(\hat{c}_{k,i^*,j})^{i^*}/\Gamma(i^*/2+1) \le (\hat{c}_{k,j+1,j})^{i^*}/\Gamma((j+3)/2)$. Finally define

$$Y_{j+1}(f) = \max_{0\le s\le n_j+1} |Y_{j+1,s}(f)|.$$

Obviously, $Y_{j+1} \le X_{j+1}$, and therefore

$$w_k\left(\left\{f \in C^k : X_{j+1} \le b_j h^j\right\}\right) \le w_k\left(\left\{f \in C^k : Y_{j+1} \le b_j h^j\right\}\right).$$

Since $X_{j+1,l}$ are Gaussian, (2) implies that $Y_{j+1,s}$ are independent. and therefore

$$w_k\left(\left\{f \in C^k : X_{j+1} \leq b_j h^j\right\}\right) \leq \prod_{s=0}^{n_j+1} w_k\left(\left\{f \in C^k : Y_{j+1,m_s} \leq b_j h^j\right\}\right).$$

Hence, due to (11) and (12),

$$w_k\left(\left\{f \in C^k : X_{j+1} \leq b_j h^j\right\}\right) \leq \frac{1}{\Gamma\left(\frac{j+3}{2}\right)^{n_j+1}} \left(h^{j-k-1/2} b_j \hat{c}_{k,j+1,j}\right)^{(j+1)(n_j+1)+i^*}.$$

Since $(j+1)(n_j+1)+i^* \geq (n-j)/2$, this completes the proof. $\square$

From Proposition 2 we immediately get

Proposition 3 *For* $0 \leq k \leq r-1$,

$$w_k(A_k^+) \leq \sum_{j=k+1}^{r} \left(h^{j-k-1/2} b_j \tilde{c}_{k,j}\right)^{(n-j)/2}.$$

4.2 Main Result

Recall that the probability measure Prob depends on the the values α_k. Therefore, the probability of failure depends on α_k's as well, $E(\psi) = E(\psi; \vec{\alpha})$ with $\vec{\alpha} = [\alpha_0, \ldots, \alpha_r]$. Since from a practical point of view it is unreasonable to assume that these values are known, our main result provides an estimate of the probability of failure for the worst possible choice of α_k's. That is, we estimate

$$\sup_{\vec{\alpha}} E(\psi; \vec{\alpha}) \leq \max_{0 \leq k \leq r} w_k(A_k).$$

Using the upper bounds obtained in Propositions 1 and 3 to estimate $w_k(A_k)$, we obtain an upper bound that is minimized (asymptotically for large n) when the values b_k satisfy:

$$b_k^2 = \ln\left(h^{-d_1} d_2 / \left(\ln(h^{-d_1} d_2)\right)^{d_1}\right) \tag{13}$$

with

$$d_1 = \frac{(n-k+2)h}{2} \quad \text{and} \quad d_2 = \left(\frac{(b-a)4}{k(k+1)(2k+1)}\sqrt{2/\pi}\right)^{2h}.$$

Then we arrive immediately at the following

Theorem 1 *Let* b_k *satisfy (13). Then*

$$\sup_{\vec{\alpha}} E(\psi; \vec{\alpha}) \leq 2\left(\frac{(n-r+2)(b-a)r(r+1)(2r+1)}{8n^2} \ln(n/(b-a))\right)^{(n-r)/4} (1+o(1)).$$

The term $o(1)$ *tends to zero exponentially fast with* n *tending to infinity, and is small even for* n *close to* r.

The theorem shows that the probability of wrong decision is super exponentially small in the number of function evaluations.

References

[1] R. J. Adler, *The Geometry of Random Fields*, John Wiley & Sons, New York, 1981.

[2] F. Gao and G. W. Wasilkowski, in progress.

[3] F. Gao, "Probabilistic analysis of numerical integration algorithms," *Journal of Complexity* **7** (1991) 58-69.

[4] F. Gao, "On the power of a posteriori error estimation for numerical integration and function approximation," *Curves and Surfaces, Proc. AFA International Conference on Curves and Surfaces* (eds. P. J. Laurent, A. L. Méhauté, and L. L. Schumaker) 197-207, Academic Press, San Diego, California, 1991.

[5] C. A. Micchelli and T. Rivlin, "Lectures on optimal recovery," in *Numerical Analysis Lancaster 1984*, (P. R. Turner, ed.), 21-93, Lecture Notes in Mathematics **1129**, Springer-Verlag, Berlin, 1985.

[6] E. Novak, *Deterministic and Stochastic Error Bounds in Numerical Analysis*, Lecture Notes in Mathematics **1349**, Springer-Verlag, Berlin, 1988.

[7] K. R. Parthasarathy, *Probability Measures on Metric Spaces*, Academic Press, New York, 1967.

[8] L. L. Schumaker, *Spline Functions: Basic Theory*, John Wiley & Sons, New York, 1981.

[9] J. F. Traub, G. W. Wasilkowski, H. Woźniakowski, *Information-Based Complexity*, Academic Press, New York, 1988.

[10] G. Wahba, *Spline Models for Observational Data*, SIAM-NSF Regional Conference Series in Appl. Math. **59**, 1990.

[11] G. W. Wasilkowski, "On piecewise-polynomial approximation of functions with a bounded fractional derivative in an L_p-norm," *Journal of Approximation Theory* **62** (1990) 372-380.

[12] G. W. Wasilkowski and F. Gao, "On the power of adaptive information for functions with singularities," *Mathematics of Computation*, 1992, to appear.