ADAPTION MAKES IT EASY TO INTEGRATE FUNCTIONS WITH UNKNOWN SINGULARITIES

LESZEK PLASKOTA AND GRZEGORZ W. WASILKOWSKI

ABSTRACT. We study numerical integration $I(f) = \int_0^T f(x)\,dx$ for functions f with *singularities*. Nonadaptive methods are inefficient in this case, and we show that the problem can be efficiently solved by *adaptive* quadratures at cost similar to that for functions with no singularities.

Consider first a class $\mathcal{F}_r$ of functions whose derivatives of order up to r are continuous and uniformly bounded for any but one singular point. We propose adaptive quadratures Q_n^*, each using at most n function values, whose *worst case* errors $\sup_{f\in\mathcal{F}_r} |I(f) - Q_n^*(f)| = \Theta(n^{-r})$. On the other hand, the worst case error of nonadaptive methods is $\Omega(n^{-1})$.

These worst case results do not extend to the case of functions with two or more singularities; however, adaption shows its power even for such functions in the *asymptotic setting*. That is, let F_r^∞ be the class of r-smooth functions with arbitrary (but finite) number of singularities. Then a generalization of Q_n^* yields adaptive quadratures Q_n^{**} such that $|I(f) - Q_n^{**}(f)| = O(n^{-r})$ for any $f \in F_r^\infty$. In addition, we show that for any sequence of nonadaptive methods there are 'many' functions in F_r^∞ for which the errors converge no faster than n^{-1}.

Results of numerical experiments are also presented.

1. INTRODUCTION

It is well known that scalar integrals $I(f) = \int_0^T f(x)\,dx$ can be efficiently approximated when the integrand is smooth. If f is r-times continuously differentiable then the composite Gauss or Newton-Cotes quadratures based on equispaced division of the interval $[0, T]$ achieve the best possible convergence rate of n^{-r} (where n is the number of function evaluations). On the other hand, those quadratures fail for functions having some singularities; even one discontinuity of f suffices to slow down the convergence to the level n^{-1}. It is intuitively clear that this is because those quadratures use the same (nonadaptive) sampling for each f even though the singularities may be located in any part of the interval. Therefore one may hope to do much better by adjusting the sampling to the underlying function, i.e., by using *adaptive* methods.

Many adaptive methods for functions with or without singularities have already been proposed in the literature and used in practice. They include adaptive Simpson's quadrature [12, 13], CADRE [7], QUADPACK [16], and an adaptive meta-algorithm [19], to mention just a few. For more complete list of excellent adaptive methods we refer

Date: April 7, 2005.

1991 *Mathematics Subject Classification.* 65D30, 65D32.

Key words and phrases. numerical integration, singularities, adaptive quadratures, Simpson's rule.

The authors were partially supported, respectively, by the State Committee for Scientific Research of Poland (KBN) and by the National Science Foundation under Grant CCR-0095709.

to, e.g., [6, 11]. The superiority of adaptive methods has so far been verified mainly empirically. More precisely, extensive tests indicate that adaptive quadratures work very efficiently *"most of the time"*; however, they fail for some integrands. There are theoretical results as well, however they hold for rather restrictive classes of functions or are probabilistic in nature, see, e.g. [9, 14, 20] and [15] for a survey.

On the other hand, there are a number of results stating that, under certain assumptions on the classes of functions, nonadaptive methods are as powerful as adaptive methods for linear problems in the worst case setting. Perhaps the first such result is due to Kiefer [10] who showed in 1957 that the nonadaptive composite Trapezoid rule is optimal in the worst case setting for the class of monotone functions with $f(0) = 0$ and $f(1) = 1$. The most general results for approximating linear functionals (e.g., integrals) is due to Bakhvalov [2] who showed that nonadaptive algorithms are optimal provided the class of functions is balanced and convex. This result has later been extended to approximation of linear operators, see [8, 22]. Similar results hold in the average case and probabilistic settings provided that the probability measure imposed on the class of functions is Gaussian, see e.g., [21, 24], and for problems with noisy information, see [17]. Moreover, if the function class is a ball in a Banach space then adaption does not help also in the asymptotic setting. Recall that in the asymptotic setting one is interested in the rate of convergence for each individual function f rather than in the convergence of the worst case error, see [23] and [21].

Convex classes of functions exclude functions with unknown singularity. So do statistical models that are based on 'regular' probability measures such as Gaussian measures. This is why the theoretical results mentioned above do not contradict the empirical observations. Actually, there have been theoretical results showing in a rigorous way the power of adaption, see e.g., [15]. In particular, [25] proposed a Gaussian-like probability measure on classes of functions with unknown singular points and showed that adaptive quadratures drastically outperform nonadaptive ones in the corresponding probabilistic setting.

In the present paper, we consider the worst case and asymptotic settings. We introduce a rather natural classes of singular integrands for which adaptive methods are much better than nonadaptive ones. Actually, the adaptive methods we propose make integrating functions with singularities almost as easy as integrating functions without singularities.

More specifically, for a given regularity parameter $r \geq 1$, let $\mathcal{F}_r$ be a class of functions $f : [0, T] \to \mathbb{R}$ that have continuous and uniformly bounded derivatives $f^{(j)}$, $0 \leq j \leq r$, for all but (perhaps) one unknown point. (For a complete definition of $\mathcal{F}_r$, see Section 2 for periodic and Section 5 for nonperiodic case.) At that point, $f^{(s)}$ is discontinuous, where s is an unknown integer between 0 and r. The magnitude of the discontinuity jump is also unknown; it can be large or very small. Then any method using n function values at nonadaptively chosen (i.e., a priori fixed) points has the worst case error with respect to $\mathcal{F}_r$ at least proportional to n^{-1}. On the other hand, we propose a special adaptive selection of the sampling points leading to a composite quadrature Q_n^* (based, e.g., on Newton-Cotes or Gauss rules) that has the worst case error of order n^{-r}; hence it is optimal.

The algorithm relies on finding and removing from the domain a small subinterval that possibly contains 'large' singularity, and then applying the composite quadrature on the remaining part of the domain. First, for an initial (nonadaptive) division $0 = x_0 < x_1 < \cdots < x_m = T$, the divided differences of f are evaluated at all consecutive $r+1$ points x_i. If the divided differences do not exceed a given threshold D then the initial division becomes the final one. This is the case when there is no singularity or the discontinuity jumps are small enough for the composite rules to work well. On the other hand, if the largest divided difference, say $|f[x_{i-r},\ldots,x_i]|$, is larger than the threshold then additional r points are introduced in $[x_{i-r},x_i]$ and the divided differences are evaluated in that subinterval. This adaptive procedure continues until the length of the subinterval is of order $m^{-(r+1)}$; see Section 4 for details.

An important feature of the algorithm Q_n^* is that the only parameters of the problem used are T and the regularity r. Moreover, its worst case error is $O(n^{-r})$ independently of the threshold D and bounds L_j on the derivatives $f^{(j)}$, $0\le j\le r$. On the other hand, it is clear that the constant in the O-notation and, more generally, the (exact) error of Q_n^* do depend on those parameters and on the basic rule being used. For instance, if L_r is known then it is wise to set $D = L_r/r!$, because then the adaptive procedure is performed only when the 'large' singularity has been detected. For more, see Remark 3 where computation of an ε-approximation together with special cases is discussed.

Note that the main idea behind our algorithm is similar to that already proposed earlier in the literature. To see this, let $S(\cdot;a,b)$ denote basic Simpson's rule on the interval $[a,b]$, i.e.,

$$S(f;a,b) \;=\; \frac{b-a}{6}\cdot\big(f(a)+4f(c)+f(b)\big) \quad \text{with} \quad c=\frac{a+b}{2}.$$

A classical derivation of adaptive Simpson's rule (see, e.g., such textbooks and monographs as [1, 3, 4, 6, 18]) is based on the observation that

$$E(f;a,a+h) \;:=\; \frac{S(f;a,a+h/2)+S(f;a+h/2,a+h)-S(f;a,a+h)}{15}$$

$$\text{estimates} \quad \int_a^{a+h} f(t)\,dt - \big(S(f;a,a+h/2)+S(f;a+h/2,a+h)\big)$$

very well if the 4th derivative of f does not change too much in $[a,a+h]$. Therefore $[a,a+h]$ is subdivided if the *empirical error* $|E(f;a,a+h)|$ is large. One can verify that

$$E(f;a,a+h) \;=\; \frac{2h^5}{15}\cdot f[a,a+h/4,a+h/2,a+3h/4,a+h],$$

and large values of E correspond exactly to large values of the corresponding divided differences. However, we choose to divide the interval $[a,a+h]$ because, as mentioned earlier and formally shown in Section 3, for our class of integrands, a large value of the divided difference implies that a singular point is there.

So far we have restricted the class of functions to those with only one singularity. It turns out that this assumption is essential for adaption to help in the worst case setting. Indeed, we show in Section 5 that the worst case error of any (adaptive or not) method is $\Omega(n^{-1})$ for classes of integrands with the number of singularities uniformly bounded by q, for any $q\ge 2$. Moreover, for classes with arbitrarily (but finitely) many singularities, there are no methods whose worst case errors converge to zero with increasing n. This negative conclusion is a result of the pessimistic nature of the worst case setting; namely, for any n there are functions for which two or more singularities are so close one to each other that there is no method able to separate them using n function evaluations. Fortunately, adaption regains its superiority in the less demanding *asymptotic setting*.

Recall that, in the worst case setting, for a given number n of function evaluations, we consider the worst case error of a quadrature Q_n, i.e., the error for the 'most difficult' integrand f. This integrand may depend on n, and different values of n may have completely different 'most difficult' integrands. In the asymptotic setting, instead of fixing n and varying f, an integrand f is fixed and then the speed of convergence of $|I(f) - Q_n(f)|$ is analyzed for $n \to \infty$.

In Section 6, we consider such an asymptotic setting for the class F_r^∞ of functions with continuous derivatives of order up to r at all but singular points. Each function may have arbitrary (yet finite) number of singularities. We show that the adaptive quadrature Q_n^* can be generalized yielding an adaptive quadrature Q_n^{**} such that the error $|I(f) - Q_n^{**}(f)|$ converges at least as fast as n^{-r}, and this convergence cannot be improved. Moreover, we show that for any sequence of nonadaptive quadratures there are 'many' functions in our class for which the error is only $\Omega(n^{-1})$.

The adaptive quadratures Q_n^* and Q_n^{**} have been implemented and tested for $r = 2$ and $r = 4$, where we used, respectively, the Midpoint and Trapezoid rules, and Simpson's rule. We stress that both implementations run in $O(n)$ total time. The numerical results perfectly confirm the theory. Some of them are presented in Section 7.

We believe that the idea of getting some *quantitative* information about singularities from analyzing divided differences (see the auxiliary results of Section 3) can be applied to solve some other related problems. For instance, our results can be extended to the problem of L_p-approximation ($1 \le p < \infty$) of functions with singularities. It would also be interesting to consider functions of more than one variable.

2. Problem formulation and main result

For given $T > 0$ and integer $r \ge 1$, let F_r be the set of functions $f : [0, T] \to \mathbb{R}$ that are r-times continuously differentiable at all but perhaps one *singular* point s_f. For each f the singularity s_f is not known a priori. We also assume until Section 5 that f and its derivatives up to rth are periodic. The latter simplifies some technical considerations. In Section 5 we show that this assumption is not crucial and can be easily removed.

More specifically, $f \in F_r$ iff there are a function $g_f \in C^r([0, T])$ and a point $s_f \in [0, T)$ such that

$$f(x) \;=\; \begin{cases} g_f(T + x - s_f) & 0 \le x < s_f, \\ g_f(x - s_f) & s_f \le x \le T. \end{cases}$$

This allows us to view the functions f as T-periodic on $\mathbb{R}$ by setting $f(s_f + kT + x) = g_f(x)$ for all integers k and $x \in [0, T)$, and to consider the values of f at arguments $x \notin [0, T]$.

Our aim is to approximate the integral

$$I(f) \;=\; \int_0^T f(x)\,dx$$

for functions $f \in F_r$. Any method is based on values of f at finitely many points. It could be a *quadrature*

$$Q_n(f) \;=\; \sum_{j=1}^n a_j f(x_j), \tag{1}$$

where $n \ge 1$, $a_j \in \mathbb{R}$, and $x_j \in [0,T)$, $1 \le j \le n$, are independent of f, or more generally, it could be

$$Q_n(f) \;=\; \varphi(f(x_1), \dots, f(x_n)),$$

where x_j are again chosen independently of f, and $\varphi : \mathbb{R}^n \to \mathbb{R}$ is an arbitrary mapping. Such methods (or quadratures) are called *nonadaptive*. (Examples are provided by, e.g., composite Newton-Cotes or Gauss quadratures.) We also allow *adaptive* methods where the choice of successive x_j and/or the number n of them may depend on the previously obtained information about f, i.e., x_1 is fixed but

$$x_j \;=\; x_j\big(f(x_1), \dots, f(x_{j-1})\big) \quad \text{for} \quad j \ge 2.$$

An adaptive method is called an adaptive quadrature if it is of the form (1) when restricted to any collection of functions $f \in F_r$ for which information $f(x_j)$, $1 \le j \le n$, is evaluated at the same points x_j. (For more discussion on adaptive methods we refer to, e.g., [17, 21].)

Given positive L_0 and L_r, define the function class

$$\mathcal{F}_r \;=\; \mathcal{F}_r(L_0, L_r) \;=\; \left\{ f \in F_r : \quad \|g_f\|_C \le L_0,\ \|g_f^{(r)}\|_C \le L_r \right\},$$

where $\|g\|_C = \max_{0 \le t \le T} |g(t)|$.

Remark 1. Note that functions in $\mathcal{F}_r$ have *all* the derivatives (up to rth) uniformly bounded, i.e., there are $L_j > 0$ (dependent on L_0, L_r, and T) such that $\|g_f^{(j)}\|_C \le L_j$, $1 \le j \le r-1$.

In the *worst case setting*, the quality of any method Q_n is measured by its *worst case error* with respect to $\mathcal{F}_r$, i.e.,

$$\mathrm{e}^{\mathrm{wor}}(Q_n; \mathcal{F}_r) \;=\; \sup_{f \in \mathcal{F}_r} |I(f) - Q_n(f)|.$$

If we knew a priori the exact location of s_f, for all $f \in \mathcal{F}_r$, then our problem would be equivalent to integrating functions $g \in C^{(r)}([0,T])$ with $\|g\|_C \le L_0$ and $\|g^{(r)}\|_C \le L_r$. It is well known that in this case adaptive methods are not better than nonadaptive ones. Furthermore, the minimal worst case error of methods that use no more than n function evaluations is $\Theta(n^{-r})$, and this is achieved by, e.g., the composite Newton-Cotes quadratures of degree r that are based on equispaced knots.

Our assumption that the singularity is not known a priori radically changes the picture; namely, we have the following result.

Theorem 1. *Let $r \geq 2$.*
(i) For any nonadaptive method Q_n that uses n function evaluations,

$$\mathrm{e}^{\mathrm{wor}}(Q_n; \mathcal{F}_r) \;\geq\; \frac{T L_0}{n+1}.$$

(ii) There exist adaptive quadratures Q_n^*, each using no more than n function evaluations, such that

$$\mathrm{e}^{\mathrm{wor}}(Q_n^*; \mathcal{F}_r) \;\leq\; C_r T^{r+1} L_r n^{-r}$$

holds for all sufficiently large n. Here C_r is independent of T, L_r, and L_0.

To show (i), suppose that Q_n is a nonadaptive quadrature using the points $0 \leq x_1 < x_2 < \ldots < x_n < T$. Then there exists $1 \leq k \leq n$ such that $x_{k+1} - x_k \geq T/n$ (where, by convention, $x_{n+1} = T + x_1$). Let $h = x_{k+1} - x_k - \delta$, where $0 < \delta < \min\{T, x_{k+1}\} - x_k$, and $c = 2L_0/(T+h)$. Let f be generated by $g_f(x) = cx - L_0$ with $s_f = x_k + \delta$, and $\tilde{f}$ be generated by $g_{\tilde{f}}(x) = c(x+h) - L_0$ with $s_{\tilde{f}} = x_{k+1}$ (or $s_{\tilde{f}} = x_1$ if $k = n$). Obviously $f, \tilde{f} \in \mathcal{F}_r$. Moreover,

$$(\tilde{f} - f)(x) \;=\; \begin{cases} 2L_0 - ch & x \in [x_k + \delta, x_{k+1}), \\ 0 & \text{otherwise.} \end{cases}$$

Since $f(x_k) = \tilde{f}(x_k)$ for all k, we have $Q_n(\tilde{f}) = Q_n(f)$ and $I(\tilde{f}) - I(f) = h(2L_0 - ch)$. Hence

$$\mathrm{e}^{\mathrm{wor}}(Q_n; \mathcal{F}_r) \;\geq\; h(L_0 - ch/2) \;=\; \frac{hL_0T}{T+h}.$$

The fact that h can be arbitrarily close to T/n completes the proof of (i).

The proof of (i) requires some auxiliary results from the next section. It is presented in Section 4 together with the quadrature Q_n^*.

3. Auxiliary lemmas

For a function $f \in F_r$, we denote

$$\Delta_f^{(j)} \;:=\; f^{(j)}(s_f^-) - f^{(j)}(s_f^+) \;=\; g_f^{(j)}(T) - g_f^{(j)}(0), \qquad 0 \leq j \leq r.$$

Let S_r be an arbitrary (nonadaptive) quadrature for approximating $\int_0^1 h(y)\,dy$ that is exact for all polynomials of degree $r-1$,

$$S_r(h) \;=\; \sum_{j=1}^{\overline{r}} a_j h(y_j).$$

For any $a < b$, let $S_r(\cdot; a, b)$ be the quadrature for approximating $\int_a^b h(x)\,dx$ obtained from S_r by

$$S_r(h; a, b) \;=\; (b-a) \cdot \sum_{j=1}^{\overline{r}} a_j h(x_j) \qquad \text{where} \qquad x_j = a + y_j(b-a).$$

For instance, $S_r(\cdot; a, b)$ can be a (basic) Newton-Cotes rule with $\overline{r} \geq r$ for r odd, and $\overline{r} \geq r - 1$ for r even, or (Legendre) Gauss rule with $\overline{r} \geq \lceil r/2 \rceil$.

Lemma 1. *For any $f \in F_r$ with $s_f \in (a,b]$ and $\|g_f^{(r)}\|_C \le L_r$ we have*

$$\left|\int_a^b f(x)\,dx \,-\, S_r(f;a,b)\right| \le \sum_{j=0}^{r-1} \alpha_j\,|\Delta_f^{(j)}|\,(b-a)^{j+1} \,+\, \alpha_r\,L_r\,(b-a)^{r+1}$$

for some α_j's that are independent of a, b, and f.

Proof. We have

$$f(x) \,=\, \begin{cases} f^-(x) & 0 \le x < s_f, \\ f^+(x) & s_f \le x \le T, \end{cases}$$

where

$$f^-(x) \;=\; \sum_{j=0}^{r-1} g^{(j)}(T)(x-s_f)^j/j! \,+\, g^{(r)}(\xi_x)(x-s_f)^r/r!$$

$$f^+(x) \;=\; \sum_{j=0}^{r-1} g^{(j)}(0)(x-s_f)^j/j! \,+\, g^{(r)}(\xi_x)(x-s_f)^r/r!$$

Since $S_r(\cdot;a,b)$ is exact for polynomials of degree $r-1$, the error for f equals the error for $\tilde f$ defined as

$$\tilde f(x) \;=\; f(x) \,-\, \sum_{j=0}^{r-1} g^{(j)}(0)(x-s_f)^j/j!$$

$$(2) \qquad\qquad = \; \sum_{j=0}^{r-1} \Delta_f^{(j)}\,\phi_{j,s}(x) + g^{(r)}(\xi_x)(x-s_f)^r/r!$$

Here and elsewhere

$$\phi_{j,s}(x) \;:=\; \frac{(x-s)_-^j}{j!} \;=\; \frac{(\min\{x-s,0\})^j}{j!}.$$

Changing variables to $y=(x-a)/(b-a)$, we obtain that α_j can be defined as

$$(3) \qquad \alpha_j \;:=\; \max_{0\le s\le 1}\left|\int_0^1 \phi_{j,s}(y)dy - S_r(\phi_{j,s};0,1)\right| \qquad \text{for} \quad j \le r-1,$$

and

$$(4) \qquad \alpha_r \;:=\; \sup_{0\le s\le 1}\,\sup\left\{\left|\int_0^T \psi(x)\,dx - S_r(\psi;0,1)\right| \;:\quad \psi:[0,T]\to\mathbb{R}\,,\right.$$
$$\left. |\psi(x)| \;\le\; |x-s|^r/r! \quad \forall\, x\in[0,1]\right\}.$$

□

Remark 2. Lemma 1 together with Remark 1 yields the following property of non-adaptive quadratures. Let $m \ge 1$. Let $Q_n^C(f) = \sum_{i=1}^m S_r(f;(i-1)h,ih)$ be the corresponding to S_r composite quadrature, where $h=T/m$ and n is the number of function evaluations used. Let $\mathcal{F}_r(s) = \mathcal{F}_r \cap C^{s-1}([0,T])$ with $s \ge 1$. Then

$$\mathrm{e}^{\mathrm{wor}}(Q_n^C;\mathcal{F}_r(s)) \;=\; O(n^{-(\overline{s}+1)}) \quad \text{where} \quad \overline{s} \;=\; \min\{s,r\}.$$

In what follows, we use the following notation: $t_i := ih$ for all (not necessarily integer) i, where $h = T/m$ for some integer $m \ge r$. By $f[t_{i-j}, t_{i-j+1}, \ldots, t_i]$ we denote the divided difference of the function f with respect to the $j+1$ points $t_{i-j}, \ldots, t_i$.

Lemma 2. *Let w_{r-1} be a polynomial of degree at most $r-1$, and let*

$$f(x) \;=\; \begin{cases} w_{r-1}(x) & x < t_{k+1}, \\ 0 & x \ge t_{k+1}. \end{cases}$$

If

$$f[t_{i-r}, \ldots, t_i] \;=\; 0 \qquad k+1 \le i \le k+r,$$

then $w_{r-1} \equiv 0$.

Proof. Since $f(t_j) = 0$ for all $j \ge k+1$, vanishing of the divided difference for $i = k+r$ implies that also $f(t_k) = w_{r-1}(t_k) = 0$. Using the same argument for $i = k+r-1$, we get that also $f(t_{k-1}) = w_{r-1}(t_{k-1}) = 0$, and next that $f(t_{i-r}) = w_{r-1}(t_{i-r}) = 0$ for $i = k+r-2, \ldots, k+1$. However, then w_{r-1} vanishes at r different points t_{i-r} which completes the proof. □

Lemma 3. *There exist $M, E > 0$ (dependent on r) with the following property. For any $D \ge 0$, $h > 0$, and $f \in F_r$ with $s_f \in (t_k, t_{k+1}]$ and $\|g_f^{(r)}\|_C \le L_r$, if the divided differences satisfy*

$$|f[t_{i-r}, t_{i-r+1}, \ldots, t_i]| \;\le\; D \qquad k+1 \le i \le k+r,$$

then

$$|\Delta_f^{(j)}| \;\le\; M(D + EL_r) \cdot h^{r-j} \qquad 0 \le j \le r-1.$$

Proof. The divided differences of f at $r+1$ points equal the corresponding divided differences of $\tilde{f}(x) = f(x) - \sum_{j=0}^{r-1} g^{(j)}(0)(x - s_f)^j/j!$. Letting $\tau = (s_f - t_k)/h$ we have from (2) that

$$\tilde{f}(t_m) \;=\; h^r L_r c_{r,m}(\tau) + \begin{cases} \sum_{j=0}^{r-1} h^j \Delta_f^{(j)} c_{j,m}(\tau) & m \le k, \\ 0 & m \ge k+1, \end{cases}$$

where $c_{j,m}(\tau) = (m-k-\tau)^j/j!$, $0 \le j \le r-1$, and $|c_{r,m}(\tau)| \le |m-k-\tau|^r/r!$ We also have

$$\tilde{f}[t_{i-r}, \ldots, t_i] \;=\; h^{-r} \sum_{m=i-r}^{i} b_{m-i+r} \tilde{f}(t_m),$$

with $b_s = \prod_{s \ne \ell = 0}^{r} (s-\ell)^{-1}$. Hence

$$f[t_{i-r}, \ldots, t_i] \;=\; \sum_{j=0}^{r-1} a_{i,j}(\tau) h^{j-r} \Delta_f^{(j)} + e_i(\tau) \quad k+1 \le i \le k+r, \tag{5}$$

where

$$a_{i,j}(\tau) \;=\; \sum_{m=i-r}^{k} c_{j,m}(\tau) b_{m-i+r} \qquad 0 \le j \le r-1$$

and

$$e_i(\tau) \;=\; L_r \cdot \sum_{m=i-r}^{i} c_{r,m}(\tau) b_{m-i+r}.$$

Let A_τ be the $r \times r$ matrix of the coefficients $a_{i,j}(\tau)$. Formally $\tau \in (0,1]$, but we naturally extend the definition of A_τ to $\tau = 0$. We need to know that A_τ is nonsingular. Indeed, the equality $A_\tau z = 0$, $z = (z_1, \dots, z_{r-1})^T$, means vanishing of the divided differences of the function

$$v(x) = \begin{cases} \sum_{j=0}^{r-1} \Delta^{(j)} (x - t_k - \tau)^j / j! & x < t_{k+1}, \\ 0 & x \geq t_{k+1}, \end{cases}$$

where $\Delta^{(j)} = h^{r-j} z_{j+1}$. By Lemma 2 this implies $z \equiv 0$.

Observe that the coefficients of A_τ and, consequently, the coefficients of A_τ^{-1} depend continuously on τ. Hence the quantity

$$M := \max_{0 \leq \tau \leq 1} \|A_\tau^{-1}\|_\infty \tag{6}$$

is well defined and finite. Writing (5) as a linear system of $r \times r$ equations with the matrix A_τ,

$$d = A_\tau z + e_\tau,$$

we obtain that

$$\|z\|_\infty \leq \|A_\tau^{-1}\|_\infty \left(\|d\|_\infty + \|e_\tau\|_\infty \right).$$

This implies

$$|h^{j-r} \Delta_f^{(j)}| \leq M \cdot (D + E L_r)$$

where

$$E := \max_{0 \leq \tau \leq 1} \max_{k+1 \leq i \leq k+r} \left| \sum_{m=i-r}^{i} c_{r,m}(\tau) b_{m-i+r \cdot} \right|, \tag{7}$$

which completes the proof. □

Lemma 4. *For any integer i we have*

$$|f[t_{i-r}, t_{i-r+1}, \dots, t_i]| \leq \max_{2i-r \leq j \leq 2i} |f[t_{(j-r)/2}, t_{(j-r+1)/2}, \dots, t_{j/2}]|.$$

Proof. Suppose without loss of generality that $f[t_{i-r}, \dots, t_i] > 0$ and that the lemma does not hold. Let w_r be the polynomial of degree at most r interpolating f at t_j, $i - r \leq j \leq i$, and define $\tilde{f} := f - w_r$. Note that divided differences of w_r at any $r+1$ different points are the same and, hence, equal to $f[t_{i-r}, \dots, t_i]$. Therefore,

$$\tilde{f}[t_{(j-r)/2}, t_{(j-r+1)/2}, \dots, t_{j/2}] < 0 \qquad 2i - r \leq j \leq 2i.$$

This implies that the divided differences based on r successive points,

$$\tilde{f}[t_{(j-r+1)/2}, t_{(j-r+2)/2}, \dots, t_{j/2}] \qquad 2i - r - 1 \leq j \leq 2i,$$

decrease with j and, consequently, they change sign at most ones (where hitting zero is also considered a sign change). Proceeding inductively we obtain that the divided differences based on successive $r-1$ points change sign at most twice, those based on $r-2$ differences change sign at most three times, and so on. We eventually conclude that the values $f(t_{j/2})$, $2(i-r) \leq j \leq 2i$, change sign at most r times. This is a contradiction since $\tilde{f}$ vanishes at $r+1$ points t_j for $i - r \leq j \leq i$. □

Note that Lemma 4 holds true for arbitrary (not necessarily equidistant) increasing sequence of knots $t_{j/2}$, $2(i-r) \leq j \leq 2i$.

4. The adaptive quadrature

We are ready to prove (ii) of Theorem 1 by constructing the adaptive quadrature Q_n^*. The quadrature uses divided differences (corresponding to an initial equidistant division of the domain $[0,T]$) to get some information about the size of the discontinuity jumps $\Delta_f^{(j)}$. If these are 'small', the composite quadrature based on S_r is applied on $[0,T]$. Otherwise an adaptive procedure is performed to determine a small interval suspected to contain a 'large' singularity. This interval is then removed and the composite quadrature is applied on the remaining part of the domain.

The following pseudocode describes Q_n^*. In this algorithm, the threshold D is formally an arbitrary real number, and $b > 0$.

```
    THE ALGORITHM
01  begin
02    choose m ≥ r;   h := T/m;
03    for i := 0, 1, ..., m−1  do  d_i := f[(i−r)h, ..., (i−1)h, ih];
04    s := arg max_{0≤i≤m−1} |d_i|;
05    if |d_s| ≤ D  then  Q*_n(f) := Σ_{i=1}^{m} S_r(f; (i−1)h, ih)  else
06    begin
07      u := kh;   δ := h;
08      while rδ ≥ bh^{r+1}  do
09      begin
10        δ := δ/2;
11        for i := 0, 1, ..., r  do  e_i := f[u − δ(i+r), ..., u − δ(i+1), u − δi];
12        s := arg max_{0≤i≤r} |e_i|;
13        u := u − δs
14      end;
15      l_1 := min{i : ih > u};   l_2 := max{i : ih < u + T − bh^{r+1}};
16      Q*_n(f) := S_r(f; u, l_1 h) + S_r(f; l_2 h, u + T − bh^{r+1})
17                   + Σ_{i=l_1+1}^{l_2} S_r(f; (i−1)h, ih)
18    end
19  end.
```

To prove that Q_n^* has the desired properties, we set $t_i = ih\ \forall i$, and assume that the singularity $s_f \in (t_k, t_{k+1}]$. We also introduce β_r to be such that for any $a < b$ and $f \in C^r([a,b])$

$$\left| \int_a^b f(x)\,dx - S_r(f;a,b) \right| \le \beta_r \|f^{(r)}\|_{C([a,b])} (b-a)^{r+1}. \tag{8}$$

(Obviously, one can take $\beta_k = \alpha_k$ from Lemma 1, but better estimates are available; see, e.g., Remark 3.)

We consider two cases dependent on the size of the largest divided difference $|d_s|$ computed in lines 03 and 04.

Case 1: $|d_s| \le D$.

Then by Lemmas 1 and 3 we have

$$\left|\int_{t_k}^{t_{k+1}} f(x)\,dx \;-\; S_r(f;t_k,t_{k+1})\right| \;\le\; h^{r+1}L_r\,(\alpha_r+B_D)$$

where $B_D \;=\; M(D/L_r+E)\sum_{j=0}^{r-1}\alpha_j$, which implies that

$$|I(f)-Q_n^*(f)| \;\le\; m^{-r}T^{r+1}L_r\left(\beta_r+m^{-1}(B_D+\alpha_r-\beta_r)\right).$$

In this case the quadrature uses at most $m(\overline{r}+1)$ points.

Case 2: $|d_s|>D$.

If, in addition, $|d_s| > L_r/r!$ then by Lemma 4 the singularity is in the removed interval. Using (8) we obtain

$$|I(f)-Q_n^*(f)| \;\le\; m^{-r}T^{r+1}L_r(\beta_r+m^{-1}bL_0/L_r).$$

On the other hand, if $|d_s|\le L_r/r!$ (which can be the case only when $D<L_r/r!$) then

$$|I(f)-Q_n^*(f)| \;\le\; m^{-r}T^{r+1}L_r(\beta_r+m^{-1}(B+\alpha_r-\beta_r+bL_0/L_r))$$

where $B=M(1/r!+E)\sum_{j=0}^{r-1}\alpha_j$.

The adaptive procedure 06 to 18 is in this case performed $\lfloor r\log_2 m+\log_2(r/(bT^r))\rfloor+1$ times, and each time r new function values are evaluated. Hence the total number of points used is at most

$$(m+1)(\overline{r}+1)+(\lfloor r\log_2 m+\log_2(r/(bT^r))\rfloor+1)\,r. \tag{9}$$

We now let $C_r>\overline{r}^r\beta_r$ and $m=m(n)$ be chosen such that it is the largest integer for which n is not smaller than (9). Then, for sufficiently large n we obtain

$$\mathrm{e}^{\mathrm{wor}}(Q_n^*;\mathcal{F}_r) \;\le\; C_rT^{r+1}L_rn^{-r},$$

as claimed in Theorem 1 (ii).

Remark 3. Suppose one wants to be sure that the error is at most $\varepsilon>0$. Then the choice of m depends on ε, parameters r, T, L_0, L_r of the problem, parameters D and b of the algorithm, and on the choice of S_r through β_r and α_j, $0\le j\le r$, defined by (8) and (3)-(4). For instance, let $D=L_r/r!$ and $b=(B+\alpha_r-\beta_r)L_r/L_0$. Then $|I(f)-Q_n^*(f)|\le\varepsilon\ \forall f\in\mathcal{F}_r$ if

$$m^{-r}T^{r+1}L_r(\beta_r+\gamma_r/m) \;\le\; \varepsilon,$$

where $\gamma_r=B+\alpha_r-\beta_r$.

Consider now special cases when the Midpoint, Trapezoid, or Simpson's rules are used as the basic quadrature S_r.

For Midpoint and Trapezoid rules the regularity parameter is set to $r=2$. We have $\alpha_0=1/2$, $\alpha_1=1/8$, and $\alpha_2=7/24$, $\beta_2=1/24$ for the Midpoint, and $\alpha_2=5/12$, $\beta_2=1/12$ for the Trapezoid. The constant E given by (7) equals 3/2. To find M given by (6), we check that

$$A_\tau \;=\; \frac{1}{2}\cdot\begin{pmatrix}1 & -\tau+1\\ -1 & \tau\end{pmatrix} \qquad\text{and}\qquad A_\tau^{-1} \;=\; 2\cdot\begin{pmatrix}\tau & \tau-1\\ 1 & 1\end{pmatrix},$$

which leads to $M=4$. Hence $\gamma_2=21/4$ for the Midpoint and $\gamma_2=16/3$ for the Trapezoid rules.

For Simpson's rule the regularity is $r = 4$. We have $\alpha_0 = 1/3$, $\alpha_1 = 1/24$, $\alpha_2 = 1/324$, $\alpha_3 = 1/1152$, $\alpha_4 = 49/2880$, $\beta_4 = 1/2880$, and $E = 17/144$. The constant M is now more difficult to get. However, one can verify that

$$A_\tau = \frac{1}{24} \cdot \begin{pmatrix} 1 & -\tau+1 & \frac{1}{2}\tau^2 - \tau + \frac{1}{2} & -\frac{1}{6}\tau^3 + \frac{1}{2}\tau^2 - \frac{1}{2}\tau + \frac{1}{6} \\ -3 & 3\tau - 2 & -\frac{3}{2}\tau^2 + 2\tau & \frac{1}{2}\tau^3 - \tau^2 + \frac{2}{3} \\ 3 & -3\tau + 1 & \frac{3}{2}\tau^2 - \tau - \frac{1}{2} & -\frac{1}{2}\tau^3 + \frac{1}{2}\tau^2 + \frac{1}{2}\tau + \frac{1}{6} \\ -1 & \tau & -\frac{1}{2}\tau^2 & \frac{1}{6}\tau^3 \end{pmatrix}$$

and

$$A_\tau^{-1} = 24 \cdot \begin{pmatrix} \frac{1}{6}\tau^3 + \frac{1}{2}\tau^2 + \frac{1}{3}\tau & \frac{1}{6}\tau^3 - \frac{1}{6}\tau & \frac{1}{6}\tau^3 - \frac{1}{2}\tau^2 + \frac{1}{3}\tau & \frac{1}{6}\tau^3 - \tau^2 + \frac{11}{6}\tau - 1 \\ \frac{1}{2}\tau^2 + \tau + \frac{1}{3} & \frac{1}{2}\tau^2 - \frac{1}{6} & \frac{1}{2}\tau^2 - \tau + \frac{1}{3} & \frac{1}{2}\tau^2 - 2\tau + \frac{11}{6} \\ \tau + 1 & \tau & \tau - 1 & \tau - 2 \\ 1 & 1 & 1 & 1 \end{pmatrix},$$

which leads to $M = 96$. Hence $\gamma_4 = 5.8273\ldots$

5. Nonperiodic functions with multiple singularities

We now consider the integration problem for functions that are not necessarily periodic and may have more than one singularity.

Specifically, let F_r^∞ be the set of r-smooth functions $f : [0,T) \to \mathbb{R}$ with finitely many singularities. That is, $f \in F_r^\infty$ iff there are an integer $k = k(f) \ge 0$, points

$$0 = s_0 < s_1 < \cdots < s_k < s_{k+1} = T,$$

and functions $g_j \in C^r([s_j, s_{j+1}])$, $0 \le j \le k$, such that $f(x) = g_j(x)$ for all $x \in [s_j, s_{j+1})$, $0 \le j \le k$, and $f(T) = g_k(T)$. Let F_r^q be the set of functions with at most q singularities, i.e.,

$$F_r^q = \{ f \in F_r^\infty : \quad k = k(f) \le q \}.$$

Obviously, $F_r^0 \subset F_r^1 \subset \cdots \subset F_r^q \subset \cdots$ and $F_r^\infty = \cup_{q=0}^\infty F_r^q$.

Given positive L_0 and L_r, we analyze the worst case errors $\mathrm{e}^{\mathrm{wor}}(Q_n; \mathcal{F}_r^q)$ of quadratures Q_n with respect to the classes

$$\mathcal{F}_r^q = \mathcal{F}_r^q(L_0, L_r) = \left\{ f \in F_r^q : \|g_j\|_C \le L_0 \text{ and } \|g_j^{(r)}\|_C \le L_r \text{ for } 0 \le j \le k(f) \right\},$$

$0 \le q \le \infty$. We consider four cases depending on q.

Case 1: $q = 0$.

This corresponds to functions with no singularities. Then nonadaptive quadratures, e.g., composite Newton-Cotes, achieve the minimal error of $\Theta(n^{-r})$.

Case 2: $q = 1$.

This is almost the case considered in Theorem 1. Despite the fact that now the functions need not be periodic, Theorem 1 holds true for $\mathcal{F}_r^1$; however, the adaptive quadrature Q_n^* has to be modified. Indeed, Lemma 3 cannot be applied for the initial equidistant sampling to detect a 'large' singularity at

$$s_1 \in [0, (r-1)h) \cup [T - (r-1)h, T] \qquad (h = T/m).$$

To avoid this difficulty, we just add some new sample points in the vicinities of 0 and T. For instance, we can use sampling of size h on the interval $[2rh, T - 2rh]$, sampling of size $h/2$ on $[rh, 2rh]$ and $[T-2rh, T-rh]$, of size $h/4$ on $[rh/2, rh]$ and $[T-rh, T-rh/2]$, and generally, sampling of size $2^{-j}h$ on $[2^{1-j}rh, 2^{2-j}rh]$ and $[T-2^{2-j}rh, T-2^{1-j}rh]$ for $1 \le j \le k$, where $k = \lceil r\log(1/h)\rceil$. It is easy to verify, using Lemma 3 and the adaptive procedure, that if $s_1 \in [\delta, T-\delta]$, $\delta = 2^{2-k}rh \le 4rh^{r+1}$, and the singularity is 'large' then it can be detected with precision $O(h^{r+1})$. This in turn is sufficient to approximate the integral with the worst case error $O(n^{-r})$.

Case 3: $2 \le q < \infty$.

It is clear that for a fixed value of q, the worst case error of the composite Newton-Cotes quadratures is $\Theta(n^{-1})$ with the constant in the Θ-notation depending on q. It turns out that this error level cannot be improved. Indeed, we have the following proposition.

Proposition 1. *For any $q \in [2,\infty)$, any n, and any (in general adaptive) method Q_n,*

$$e^{\mathrm{wor}}(Q_n;\mathcal{F}_r^q) \;\ge\; L_0\,T\,\max\left\{\frac{\lfloor q/2\rfloor}{n+1}\,,\,1\right\}.$$

Proof. Let $x_1 \le x_2 \le \cdots \le x_n$ be the points at which the function values are computed for $f \equiv 0$. With x_0 and x_{n+1} denoting respectively 0 and T, let K be the set of indices $i \in \{1,\ldots,n+1\}$ for which $x_{i-1} \ne x_i$. Clearly its cardinality $k = |K|$ is at least 1 and $\sum_{i\in K}(x_i - x_{i-1}) = T$. If $q/2 \le k$, let $P \subset K$ be a set of $\lfloor q/2\rfloor$ indices for which

$$\sum_{i\in P}(x_i - x_{i-1}) \;\ge\; T\,\frac{\lfloor q/2\rfloor}{n+1}.$$

If $q/2 > k$, we choose $P = K$. For $0 < \delta < \min_{i\in P}(x_i - x_{i-1})/2$, consider two functions $f_{-1,\delta}$ and $f_{+1,\delta}$ defined as follows:

$$f_{\ell,\delta}(x) \;=\; \begin{cases} \ell L_0 & \text{if } x \in (x_{i-1}+\delta, x_i-\delta) \text{ for } i \in P,\\ 0 & \text{otherwise.}\end{cases}$$

Clearly, for $\ell = \pm 1$, we have that $f_{\ell,\delta} \in \mathcal{F}_r^q$, and $I(f_{\ell,\delta}) = \ell L_0\lfloor q/2\rfloor(T/(n+1) - 2\delta)$ when $q/2 < k$ and $I(f_{\ell,\delta}) = \ell(L_0T - k\delta)$ otherwise. Moreover, since both functions vanish at the x_i's, the method uses those points to evaluate both functions and, as a result of that, $Q_n(f_{-1,\delta}) = Q_n(f_{+1,\delta})$. This and the fact that δ can be arbitrarily small complete the proof. □

Case 4: $q = \infty$.

It follows from Proposition 1 that the worst case error of any method is bounded from below by L_0T regardless of n.

We summarize this is the following theorem.

Theorem 2. *The minimal worst case errors for classes $\mathcal{F}_r^q$ that can be achieved by nonadaptive and adaptive methods using n function values are given as follows:*

class of functions	*nonadaptive methods*	*adaptive methods*
$\mathcal{F}_r^0$	$\Theta(n^{-r})$	$\Theta(n^{-r})$
$\mathcal{F}_r^1$	$\Theta(n^{-1})$	$\Theta(n^{-r})$
$\mathcal{F}_r^q$, $2 \le q < \infty$	$\Theta(n^{-1})$	$\Theta(n^{-1})$
$\mathcal{F}_r^\infty$	$\Theta(1)$	$\Theta(1)$

In particular, adaption helps in the worst case setting only when $q = 1$.

6. The asymptotic setting

In this section we show that the convergence rate of n^{-r} can be preserved even for the class F_r^∞ if we switch from the worst case to the *asymptotic setting.* That is, instead of fixing n and considering the worst function for that n, we now fix a function f and see how fast the error $|I(f) - Q_n(f)|$ goes to zero when $n \to \infty$. We assume that $f \in F_r^\infty$, i.e., the only a priori knowledge used is that f has r continuous derivatives (except at an arbitrary, but finite number of singular points). In particular, the bounds on the derivatives and the number of singularities are unknown.

Theorem 3. *There exist adaptive quadratures* Q_n^{**}, $n \ge 1$, *each using at most n function evaluations, such that for all* $f \in F_r^\infty$ *we have*

$$\limsup_{n\to\infty}\; n^r \cdot |I(f) - Q_n^{**}(f)| \;<\; \infty.$$

Proof. The quadrature Q_n^{**} is constructed based on Q_n^* from Section 4. Let $f \in F_r^\infty$ and s_j, $1 \le j \le k$, be the singular points of f, $0 = s_0 < s_1 < \cdots < s_k < s_{k+1} = T$.

In the preliminary step we choose $m \ge r$ and compute the divided differences $d_i = f[t_{i-r}, \ldots, t_i]$ for $r \le i \le m$ ($t_j = jh\ \forall j$). Observe that if m is large enough so that

$$h \;=\; \frac{T}{m} \;\le\; \frac{1}{2r}\,\min_{0\le j\le k}(s_j - s_{j-1}) \tag{10}$$

then all the divided differences from Lemma 3 corresponding to each singularity are computed and, moreover, these cover different intervals for different s_j.

Then we choose $l = l(m)$ (which will be specified later) and $l^* \le l$ indices $i_1, \ldots, i_{l^*}$ as follows. If the largest divided difference, say $|d_s|$, does not exceed a given threshold D then $l^* = 0$, otherwise $i_1 = s$. Suppose we have defined $i_1, \ldots, i_{j-1}$, $1 \le j \le l$. Consider the set Z_j consisting of all the divided differences that are at least D and use points different from those used by $d_{i_1}, \ldots, d_{i_{j-1}}$. If Z_j is empty then $l^* = j - 1$, otherwise i_j is the index of the largest element in Z_j.

Next we perform the adaptive procedure separately for each of the l^* disjoint intervals $[t_{i_j-r}, t_{i_j}]$ obtaining l^* disjoint intervals (a_j, b_j), $a_1 < b_1 < a_2 < b_2 < \cdots < a_{l^*} < b_{l^*}$ where $(b_j - a_j) = O(h^{r+1})\ \forall j$.

Finally, $Q_n^{**}(f)$ is obtained by applying the composite quadratures (based on S_r) on the $l^* + 1$ intervals $[0, a_1], [b_1, a_2], \ldots, [b_{l^*-1}, a_{l^*}], [b_{l^*}, T]$.

Observe that if, in addition to (10), we have $l \ge k$ then all 'large' singularities, i.e., those for which some of the corresponding differences from Lemma 3 are larger than $\max\left\{D, \|f^{(r)}\|_C/r!\right\}$, are in $\cup_{j=1}^{l^*}(a_j, b_j)$. Similar arguments to those for Q_n^* show that this is sufficient for the error to be $O(m^{-r})$.

Since $k = k(f)$ is not known, the condition $l \ge k$ can be guaranteed only by varying l with m. It turns out that it suffices to choose $l = l(m)$ such that $\lim_{m\to\infty} l(m) = \infty$ and

$$\limsup_{m\to\infty} \frac{l(m)\log m}{m} < \infty.$$

Indeed, then for sufficiently large m (dependent on f) we have (10) and $l(m) \geq k$, so that the total number of points used is

$$n \;=\; O(m + l(m)\log m) \;=\; O(m),$$

the total length of the intervals removed from the domain is $O(n^{-r}\log^{-1} n)$, and

$$|I(f) - Q_n^{**}(f)| \;=\; O(n^{-r}),$$

as claimed. □

We want to stress that the convergence n^{-r} cannot be improved. Indeed, the general results of Trojan [23], see [21, Sect. 2, Chapt. 10], imply that for any sequence of positive numbers δ_n converging to zero and for any ball in F_r^0 (with the norm defined as $\|f\| = \max\{\|f\|_C, \|f^{(r)}\|_C\}$) one can find functions f for which $|I(f) - Q_n(f)|$ converges no faster than $\delta_n n^{-r}$. Since δ_n can converge to zero arbitrarily slowly, faster convergence than n^{-r} cannot be guaranteed even for functions from F_r^0.

To complete the picture we now show that nonadaptive methods can converge at rate at most n^{-1} which means that the use of adaption is crucial in the asymptotic setting.

For any function $f : [0,T] \to \mathbb{R}$, $\varepsilon \in \mathbb{R}$, and $u \in [0,1)$, we let

$$f_{\varepsilon,u}(x) \;=\; \begin{cases} f(x) + \varepsilon & 0 \leq x < u, \\ f(x) & u \leq x \leq T, \end{cases}$$

and $f_{\varepsilon,1}(x) = f(x) + \varepsilon \;\forall x$. Obviously, if $f \in F_r^\infty$ then also $f_{\varepsilon,u} \in F_r^\infty$.

Lemma 5. *Let $\{Q_n\}_{n=1}^\infty$ be a sequence of nonadaptive methods, each using at most n function evaluations. Then for any $f \in F_r^\infty$ and any $\varepsilon \neq 0$ the set*

$$A \;=\; \left\{ u \in [0,1] : \quad \lim_{n\to\infty} n \cdot |I(f_{\varepsilon,u}) - Q_n(f_{\varepsilon,u})| \;=\; 0 \right\}$$

is of Lebesgue measure zero, i.e., $\lambda(A) = 0$.

Proof. Suppose Q_n uses the points $0 \leq x_1^n < x_2^n < \cdots < x_n^n \leq T$. Since for functions $f_{\varepsilon,u}$, $u \in [0,1]$, the information takes only $n+1$ different values

$$(\, f(x_1) + \varepsilon, \ldots, f(x_i) + \varepsilon, f(x_{i+1}), \ldots, f(x_n)\,)\,,$$

$0 \leq i \leq n$, the method $Q_n(f_{\varepsilon,u})$ takes at most $n+1$ different values. On the other hand, $I(f_{\varepsilon,u}) = I(f) + \varepsilon u$. Hence, for any $c > 0$ and any n the set

$$B_n^c \;=\; \left\{\, u \in [0,1] : \quad |I(f_{\varepsilon,u}) - Q_n(f_{\varepsilon,u})| \leq c(n+1)^{-1} \,\right\}$$

has measure at most $2c/|\varepsilon|$.

Letting

$$A_n^c \;=\; \left\{\, u \in [0,1] : \quad |I(f_{\varepsilon,u}) - Q_l(f_{\varepsilon,u})| \leq c(n+1)^{-1}\; \forall l \geq n \,\right\}$$

we have $A_n^c \subset B_n^c$, $A_1^c \subset A_2^c \subset \cdots$, and $A \subset \cup_{n=1}^\infty A_n^c$. Therefore

$$\lambda(A) \;\leq\; \lim_{n\to\infty} \lambda(A_n^c) \;\leq\; 2c/|\varepsilon|.$$

Since c can be arbitrarily small, $\lambda(A) = 0$. □

7. Numerical results

We implemented and tested the adaptive quadratures Q_n^* and Q_n^{**} using the Midpoint and Trapezoid rules ($r=2$), and Simpson's rule ($r=4$) as the basic quadrature S_r.

Note that both quadratures can be implemented (for any $r \ge 1$ and any S_r) in such a way that they run in $O(n)$ time. Indeed, this is obvious for Q_n^*. In case of Q_n^{**} the only difficulty lies in the selection of l^* indices corresponding to largest divided differences. However, this can be realized using a *heap*, see, e.g., [5, Sect. 7, Chapt. 2]. The heap of $O(n)$ divided differences can be built in $O(n)$ time, and then $O(l^*)$ removals of largest elements cost $O(l^* \log n)$ which is $O(n)$ since $l^* = O(n/\log n)$.

In all the tests the parameter b is set to one. The results are presented in the logarithmic scale, i.e., $-\log_{10}|I(f) - Q_n(f)|$ versus $\log_{10} n$. Each graph is the piecewise linear interpolation of data obtained by running the corresponding quadrature for $m = 10^{z/10}$ where z are successive integers from some range.

In **Fig. 1** to **Fig. 3** the adaptive quadratures (marked with 'a') are compared to the corresponding nonadaptive quadratures (marked with 'n'). We generally observe much smoother behavior of adaptive quadratures.

In **Fig. 1** the periodic function f_1 with one discontinuity is defined by $T = 6$, $g_{f_1}(x) = \sin(x)$, and $s_{f_1} = \pi$. In **Fig. 2** the comparison is done for the 'peak' function

$$f_2(x) \;=\; \sin(8x)/16 + \exp(-16|x-1|)$$

with $T = \pi$. Since f_2 is continuous, the nonadaptive quadratures converge in this case at rate n^{-2}; hence adaptive Midpoint and Trapezoid are not worse than their adaptive counterparts.

In **Fig. 3** we have $T = 3$ and

$$f_3(x) \;=\; \left(\exp\left(-(x-1)^2\right)\right) + \sum_{j=1}^{5} c_j \chi_{[0,s_j)}(x)$$

where $[c_1, c_2, c_3, c_4, c_5] = [0.8, -0.14, 0.06, -0.10, 0.2]$ and $s_j = j\pi/6$, $1 \le j \le 5$. This is an example of many well separated discontinuities, where the convergence n^{-r} of the adaptive quadratures occurs rather quickly.

Fig. 4 shows results for the three adaptive quadratures in case of a function with discontinuities very close one to another; f_4 is defined as f_3, however, we now have

$$[s_1, s_2, s_3, s_4, s_5] = [\pi/6, \pi/6 + 0.03, \pi/2, \pi/2 + 0.07, \pi/2 + 0.073].$$

The behavior of the errors is rather poor for a long time, and it suddenly becomes n^{-r} when all the singularities have been detected. It is clear that for such functions the error n^{-r} can occur arbitrarily late, which follows from the fact that the worst case error for F_r^∞ does not converge to zero.

In all the proceeding tests $l(m)$ was equal to the number $k(f)$ of singularities. In such cases the choice of D is not crucial since then the adaptive procedure is never performed more than k times. The situation is more subtle if $k(f)$ is not known. For example, **Fig. 5** shows the convergence of the adaptive Simpson for the function f_3, for $l(m) = \lfloor n/\log_2(m) \rfloor$, and for different values of D. Obviously, $D_1 = 0.5 = \|f_3^{(4)}\|/4!$ is the best choice.

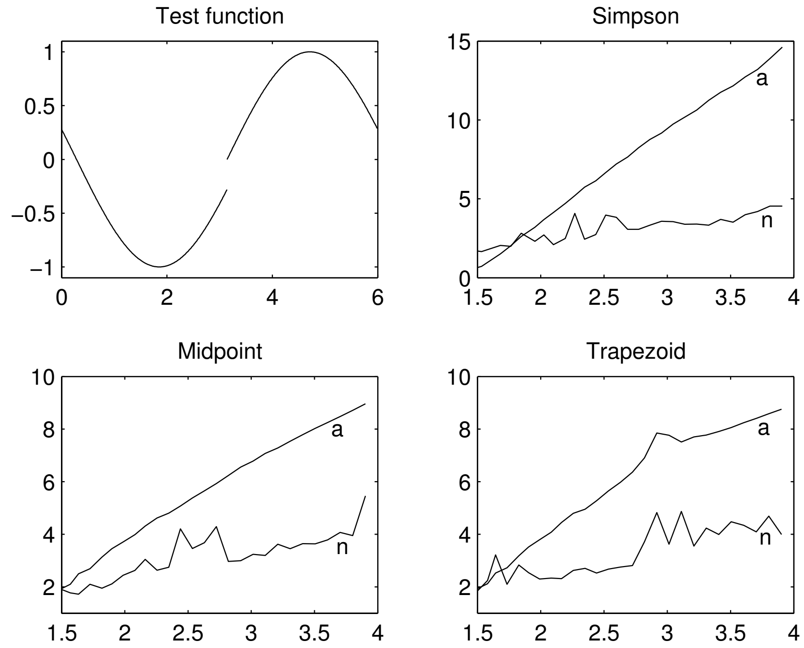

Fig. 1: Adaptive versus nonadaptive quadratures for the periodic function f_1 with one dicontinuity.

For the same reason, the choice of $l(m)$ is not crucial as long as $D = \|f^{(r)}\|_C/r!$ **Fig. 6** shows the convergence of the adaptive Simpson for f_3, for $D = 0$, and for different functions $l(m)$.

Acknowledgments. The authors would like to thank Yaxi Zhao for useful comments on this paper.

References

[1] K. E. Atkinson, *An Introduction to Numerical Analysis,* 2nd ed., Wiley, New York, 1989.

[2] N. S. Bakhvalov, On the optimality of linear methods for the operator approximation in convex classes of functions (in Russian), *Zh. Vychisl. Mat. Mat. Fiz.* **11** (1971), 1014-1018.

[3] W. Cheney and D. Kincaid, *Numerical Mathematics and Computing*, 4th ed., Brooks/Cole, New York, 1999.

[4] S. D. Conte and C. de Boor, *Elementary Numerical Analysis – An Algorithmic Approach*, 3rd ed., McGraw-Hill, New York, 1980.

[5] T. H. Corman, C. E. Leiserson and R. L. Rivest, *Introduction to Algorithms*, MIT Press, 1990.

[6] P. J. Davis and P. Rabinowitz, *Methods of Numerical Integration,* Academic Press, New York, 1975.

[7] C. de Boor, CADRE: An algorithm for numerical quadrature, *Mathematical Software*, 201-209, Academic Press, New York, 1971.

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

[9] I. Huerta, Adaption helps for some nonconvex classes, *J. Complexity* **2** (1986), 333-352.

[10] J. Kiefer, Optimum sequential search and approximation methods under minimum regularity assumption, *J. Soc. Indust. Appl. Math.* **5** (1957), 105-136.

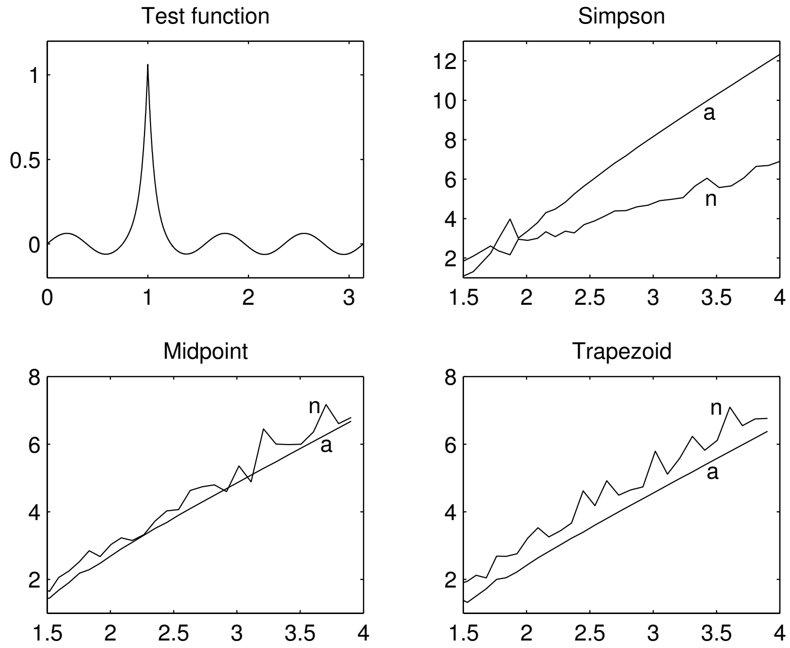

Fig. 2: Adaptive versus nonadaptive quadratures for the continuous ‘peak’ function f_2.

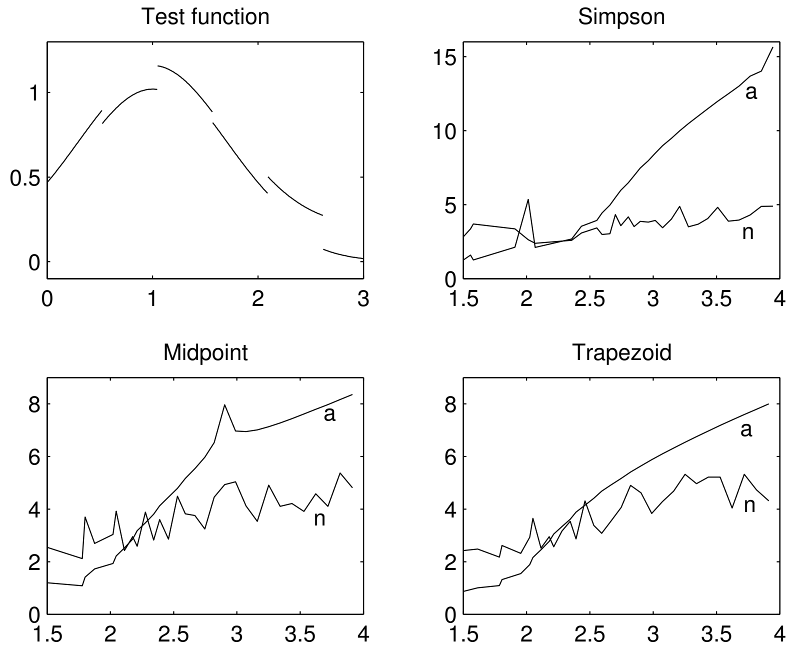

Fig. 3: Adaptive versus nonadaptive quadratures for the function f_3 with 5 well separated discontinuities.

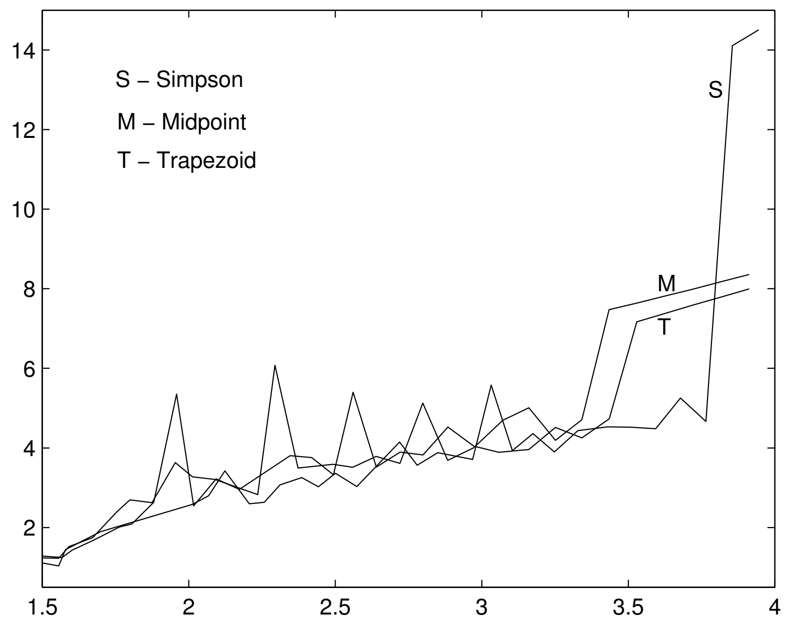

Fig. 4: Comparison of adaptive quadratures for the function f_4 with 5 very badly separated discontinuities.

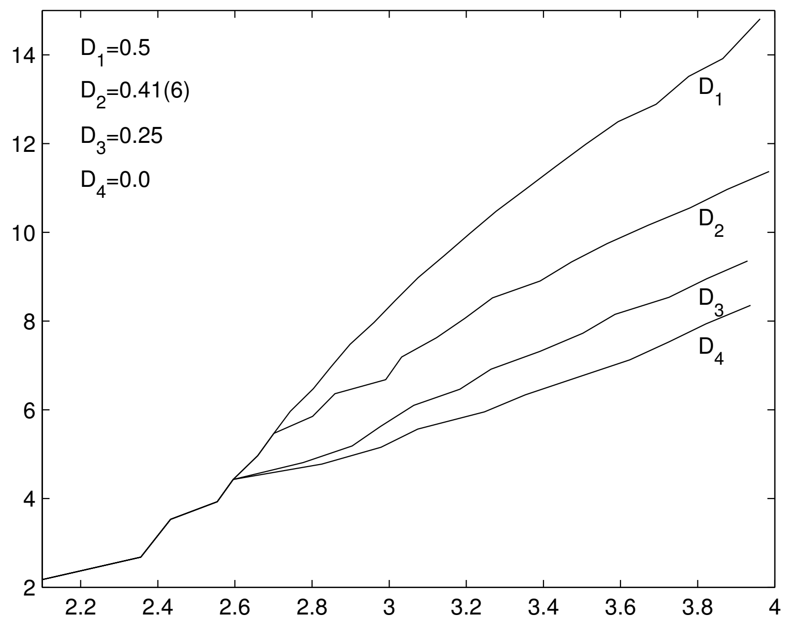

Fig. 5: Adaptive Simpson quadrature for f_3, for $l(m) = m/\log_2(m)$, and for different values of the threshold D.

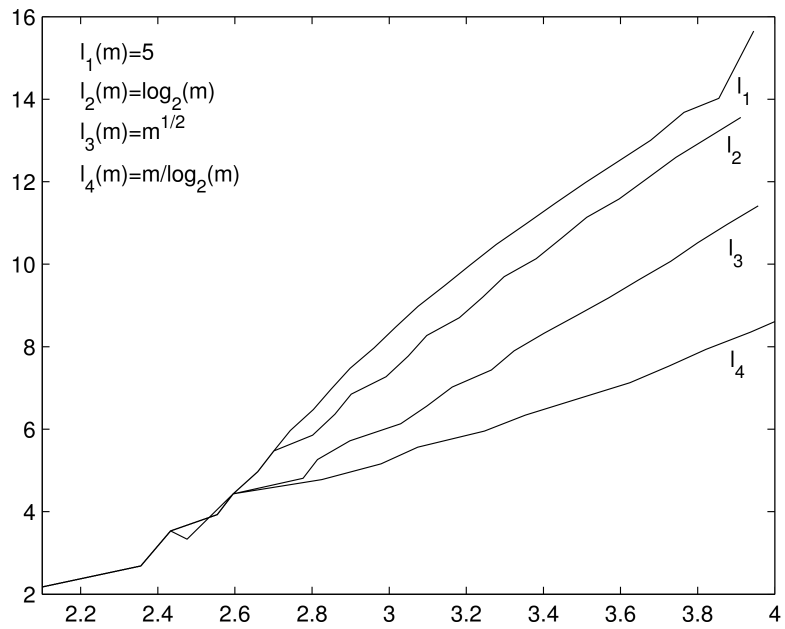

Fig. 6: Adaptive Simpson quadrature for f_3, for the threshold $D = 0$, and for different $l(m)$.

[11] A. R. Krommer and C. W. Ueberhuber, *Computational Integration*, SIAM, Philadelphia, 1998.

[12] J. M. Lyness, Notes on the adaptive Simpson quadrature routine, *J. Assoc. Comput. Mach.* **16** (1969), 483-495.

[13] W. M. McKeeman, Algorithm 145, adaptive numerical integration by Simpson's rule, *Comm. ACM* **5** (1962), 604.

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

[15] E. Novak, On the power of adaption, *J. Complexity* **12** (1996), 199-238.

[16] R. Piessens, E. deDoncker-Kapenga, C. Überhuber, and D. Kahaner, *QUADPACK: A Subroutine Package for Automatic Integration*, Springer Verlag, New York, 1983.

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

[18] A. Ralston and P. Rabinowitz, *A First Course in Numerical Analysis,* 2nd ed., McGraw-Hill, New York, 1978.

[19] J. R. Rice, A metalgorithm for adaptive quadrature, *J. Assoc. Comput. Mach.* **22** (1975), 61-82.

[20] J. F. Traub, G. W. Wasilkowski and H. Woźniakowski, When is nonadaptive information as powerful as adaptive information? *Proceedings of the 23rd IEEE Conference on Decision and Control*, 1536-1540, San Diego, 1984,

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

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

[23] G. M. Trojan, *Asymptotic Setting for Linear Problems,* unpublished Ph.D. dissertation, 1980; results presented in [21, Sect. 2, Chapt. 10].

[24] G. W. Wasilkowski, Information of varying cardinality, *J. Complexity* **1** (1986), 107-117.

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

Department of Mathematics, Informatics, and Mechanics, University of Warsaw, Banacha 2, 02-097 Warsaw, Poland
E-mail address: leszekp@mimuw.edu.pl

Department of Computer Science, University of Kentucky, 773 Anderson Hall, Lexington, KY 40506-0046, USA
E-mail address: greg@cs.uky.edu