SARAJEVO JOURNAL OF MATHEMATICS
Vol.4 (16) (2008), 31–38

NUMERICAL INTEGRATION OF FUNCTIONS GIVEN BY DATA POINTS

HAJRUDIN FEJZIĆ

Dedicated to Professor Fikret Vajzović on the occasion of his 80th birthday

Abstract. Let $a \le x_0 < x_1 < \cdots < x_n \le b$. We consider approximating integrals $\int_a^b f(x)\,dx$ of functions with bounded derivatives by sums of the form $\sum_{i=0}^n a_i f(x_i)$. We find sharp errors for these approximations and define the weights a_i, such that $\sum_{i=0}^n a_i f(x_i)$ is the best approximation of $\int_a^b f(x)\,dx$.

1. Introduction

Most numerical integration formulas are designed for equally spaced nodes and the error estimates are usually for functions with higher derivatives. In this paper we will discuss errors in approximating $\int_a^b f(x)\,dx$ by formulas of the form $\sum_{i=0}^n a_i f(x_i)$ where the nodes $a \le x_0 < x_1 < \cdots < x_n \le b$ are not necessarily equally spaced and for functions with just bounded first derivatives. We desire that the error be bounded by $c_n \|f'\|_\infty$ where c_n depends on the nodes $x_0, x_1, \ldots, x_n$ and the weights $a_0, a_1, \ldots, a_n$ only. We will show that for given nodes it is always possible to find the weights such that $c_n \to 0$ as $\max\{(x_0 - a), (b - x_n), (x_{i+1} - x_i) \mid 0 \le i \le n-1\} \to 0$. This condition can't be satisfied on the space of continuous functions as the following example shows.

Example 1. Let $x_i = a + \frac{i}{n}(b-a)$, $i = 0, 1, \ldots, n$. Define $f(x) = 1$ on intervals $\left(x_i + \frac{1}{3n}, x_{i+1} - \frac{1}{3n}\right)$, $f(x_i) = 0$ and extend $f(x)$ linearly on all of $[a, b]$. Then f is continuous on $[a, b]$, $\|f\|_\infty = 1$, $\int_a^b f(x)\,dx = \frac{2}{3}(b-a)$ but since for any choice of the weights $\sum_{i=0}^n a_i f(x_i) = 0$, we have $c_n \ge \frac{2}{3}(b-a)$ for all n.

2000 *Mathematics Subject Classification.* Primary 26A06, 26A24, 41A44; Secondary 26D07, 26D15, 26D20.

This paper was supported by Summer 2007 CSUSB TRC Grant.

Returning back to functions with bounded derivatives; since we require the error to be bounded by $c_n \|f'\|_\infty$, the formula $\sum_{i=0}^{n} a_i f(x_i)$ integrates $f(x) \equiv 1$ exactly, so the weights have to satisfy the condition $\sum_{i=0}^{n} a_i = (b-a)$. It turns out that this is also a sufficient condition for the error to be bounded by $c_n \|f'\|_\infty$. We will be in a position to compare different formulas that use the same nodes by comparing the corresponding sharp c_n's.

Definition 1. *We will say that c_n is sharp if there is a bounded derivative such that $\left|\int_a^b f(x)\,dx - \sum_{i=0}^{n} a_i f(x_i)\right| = c_n \|f'\|_\infty$. For given nodes $x_0 < x_1 < \cdots < x_n$ one formula is better than another if the corresponding sharp c_n has a smaller value.*

It may come as a surprise that for functions with bounded derivatives the composite trapezoidal rule is better than the composite Simpson's rule. First, notice that in order to compare the two rules we need $2n+1$ equally spaced nodes with $x_0 = a$ and $x_{2n} = b$. Namely in this case $c_{2n} = \frac{(b-a)^2}{8n}$ for trapezoidal and $c_{2n} = \frac{5(b-a)^2}{36n}$ for Simpson's rule respectively.

The trapezoidal and Simpson's rule for functions with bounded derivatives are studied in [1] and [2].

2. Numerical integration of functions with bounded first derivatives

For the rest of the paper; $[a,b]$ is a fixed interval and $a \le x_0 < x_1 < \cdots < x_n \le b$ are fixed nodes. In the introduction we gave an example why we need to impose smoother conditions on continuous functions, so we said that we will consider continuous functions with bounded derivatives. This condition can certainly be relaxed and we will do so by allowing that the derivative may not exist on a countable set.

Definition 2. *By Δ we denote the set of all continuous functions, that are differentiable at all but countabily many points in $[a,b]$ and such that their derivatives are bounded.*

We will consider approximating $\int_a^b f(x)\,dx$ with formulas of the form $\sum_{i=0}^{n} a_i f(x_i)$ where the weights satisfy the condition $\sum_{i=0}^{n} a_i = (b-a)$. In our first result we will use Heaviside function $H(x) = \begin{cases} 1 \text{ if } x > 0 \\ 0 \text{ if } x < 0 \end{cases}$ to express the error.

Theorem 1. *Let $f \in \Delta$. If the weights satisfy $\sum_{i=0}^{n} a_i = (b-a)$ then*

$$\left|\int_a^b f(x)\,dx - \sum_{i=0}^{n} a_i f(x_i)\right| \le \left(\int_a^b \left|(b-x) - \sum_{i=0}^{n} a_i H(x_i - x)\right| dx\right) \|f'\|_\infty .$$

Moreover $c_n := \int_a^b |(b-x) - \sum_{i=0}^{n} a_i H(x_i - x)|\,dx$ is a sharp bound.

Before we prove this theorem some remarks about the integrals used are in order. Recall that a function is Riemann integrable if and only if it is continuous almost everywhere. Thus the integrals in the statement of Theorem 1 are Riemann integrals. However, since bounded derivatives could be discontinuous on a set of positive measure we use Lebesgue integrals in our proof. Because the two integrals agree whenever the Riemann integral is defined the usual integral symbol is used for both integrals.

Proof. If $f \in \Delta$, then for every $x \in [a,b]$ by using the Lebesgue integral we can write $f(x) - f(a) = \int_a^x f'(t)\,dt$. Thus f is absolutely continuous and we can use integration by parts for Lebesgue integrals with $u = f(x)$ and $v = (x-b)$ to write

$$\int_a^b f(x)\,dx = (x-b)f(x)|_a^b + \int_a^b (b-x)f'(x)\,dx = (b-a)f(a) + \int_a^b (b-x)f'(x)\,dx.$$

On the other hand

$$\sum_{i=0}^{n} a_i \int_a^{x_i} f'(x)\,dx = \sum_{i=0}^{n} a_i f(x_i) - \sum_{i=0}^{n} a_i f(a) = \sum_{i=0}^{n} a_i f(x_i) - (b-a)f(a).$$

Combining these two equalities we obtain

$$\begin{aligned}\int_a^b f(x)\,dx - \sum_{i=0}^{n} a_i f(x_i) &= \int_a^b (b-x)f'(x)\,dx - \sum_{i=0}^{n} a_i \int_a^{x_i} f'(x)\,dx \\ &= \int_a^b \left[(b-x) - \sum_{i=0}^{n} a_i H(x_i - x)\right] f'(x)\,dx \qquad (1) \\ &\le \left(\int_a^b \left|(b-x) - \sum_{i=0}^{n} a_i H(x_i - x)\right| dx\right) \|f'\|_\infty = c_n \|f'\|_\infty .\end{aligned}$$

In order to better understand the error, c_n, it helps to graph $g_n(x) = (b - x) - \sum_{i=0}^{n} a_i H(x_i - x)$. A typical graph of $g_n(x)$ looks like function in the Figure 1. Notice that on $[x_i, x_{i+1}]$, $g_n(x) = -x + \left(b - \sum_{j=i+1}^{n} a_j\right)$ so the breaks in the -45° lines occur only at the nodes. Let $f(x) = \int_a^t f'(t)\,dt$ where $f' \equiv 1$ on intervals where $g_n(x)$ is positive and -1 where $g_n(x)$ is negative. Since f' may not be defined only at the nodes and where $g_n = 0$, we have that the continuous function f is differentiable at all but finitely many points and its derivative is equal to f'.

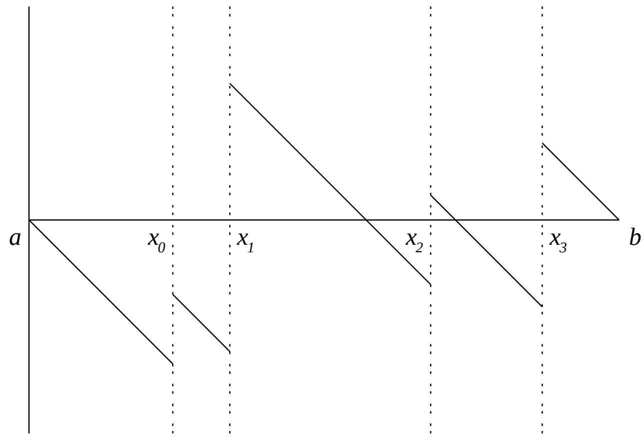

Figure 1: A typical graph of $g_3(x) = (b-x) - \sum_{i=0}^{3} a_i H(x_i - x)$

Thus $f \in \Delta$ and from (1) it follows that $\int_a^b f(x)\,dx - \sum_{i=0}^{n} a_i f(x_i) = \int_a^b |(b-x) - \sum_{i=0}^{n} a_i H(x_i - x)|\,dx = c_n \|f'\|_\infty$. Thus c_n is the sharp bound. □

Notice that $c_n = \int_a^b |g_n(x)|\,dx$ is the area bounded by $g_n(x)$ and the x-axis. Geometrically it is clear that this area is the smallest possible if the lines in the graph of $g_n(x)$ are crossing the midpoints of intervals $[x_i, x_{i+1}]$ as in the Figure 2.

In this case $g_n(\frac{x_i+x_{i+1}}{2}) = 0$ for $i = n-1, n-2, \ldots, 1, 0$. Hence the weights satisfy the system

$$\begin{cases} a_n = b - \frac{x_{n-1}+x_n}{2} \\ a_{n-1} + a_n = b - \frac{x_{n-2}+x_{n-1}}{2} \\ \vdots \\ a_1 + a_2 + \cdots + a_n = b - \frac{x_0+x_1}{2} \\ a_0 + a_1 + \cdots + a_n = b - a \end{cases}.$$

The unique solution is given by

$$a_n = b - \frac{x_{n-1} + x_n}{2},\ a_{n-1} = \frac{x_n - x_{n-2}}{2},\ a_{n-2} = \frac{x_{n-1} - x_{n-3}}{2}, \dots,$$
$$a_1 = \frac{x_2 - x_0}{2},\ a_0 = \frac{x_1 + x_0}{2} - a.$$

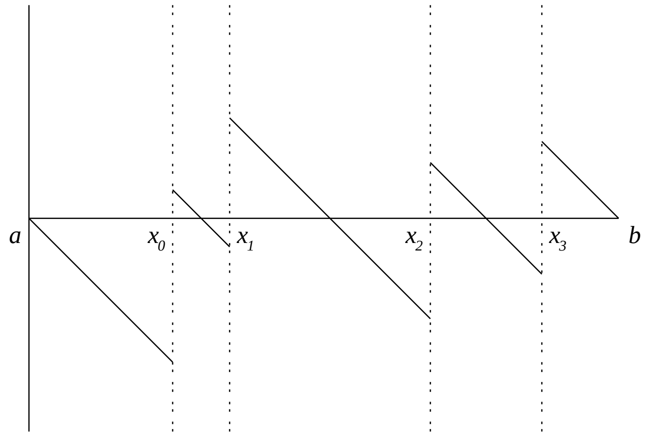

Figure 2: The graph of $g_3(x)$ that produces the smallest error

In this case $\int_a^b |g_n(x)|\, dx = \frac{(x_0-a)^2}{2} + \sum_{i=0}^{n-1} \frac{(x_{i+1}-x_i)^2}{4} + \frac{(b-x_n)^2}{2}$.

Corollary 1. *Let $[a, b]$ be a fixed interval with fixed nodes $a \le x_0 < x_1 < \cdots < x_n \le b$. The best estimate of $\int_a^b f(x)\, dx$ given by $\sum_{i=0}^n a_i f(x_i)$ for functions with bounded derivatives is achieved with weights $a_0 = \frac{x_1+x_0}{2} - a$, $a_1 = \frac{x_2-x_0}{2}, \dots, a_{n-2} = \frac{x_{n-1}-x_{n-3}}{2}$, $a_{n-1} = \frac{x_n-x_{n-2}}{2}$, $a_n = b - \frac{x_{n-1}+x_n}{2}$. In this case*

$$\left| \int_a^b f(x)dx - \sum_{i=0}^n a_i f(x_i) \right|$$
$$\le \left(\frac{(x_0-a)^2}{2} + \sum_{i=0}^{n-1} \frac{(x_{i+1}-x_i)^2}{4} + \frac{(b-x_n)^2}{2} \right) \|f'\|_\infty = c_n \|f'\|_\infty .$$

Moreover $c_n \to 0$ as $\max\{(x_0 - a), (b - x_n), (x_{i+1} - x_i) \mid 0 \le i \le n-1\} \to 0$.

Proof. Let $d_n = \max\{(x_0 - a), (b - x_n), (x_{i+1} - x_i) \mid 0 \le i \le n-1\}$. It only remains to show that $c_n \to 0$ as $d_n \to 0$. This is obvious since $c_n < d_n \left(\frac{x_0-a}{2} + \sum_{i=0}^{n-1} \frac{x_{i+1}-x_i}{2} + \frac{b-x_n}{2} \right) = d_n \frac{b-a}{2}$. □

The special case $n = 0$ of this Corollary is a well-known Ostrowski inequality (See [3].)

$$\left| \int_a^b f(x)\,dx - (b-a)f(x_0) \right| \le \left(\frac{(x_0-a)^2}{2} + \frac{(b-x_0)^2}{2} \right) \|f'\|_\infty .$$

Although usually for functions given by data points we do not have a control for selecting nodes, an interesting question is if we could select nodes, which formula is the best for integrating functions with bounded derivatives.

Corollary 2. *Let $[a,b]$ be a fixed interval. The best estimate, $\sum_{i=0}^{n} a_i f(x_i)$, of $\int_a^b f(x)\,dx$ for functions with bounded derivatives is*

$$\sum_{i=0}^{n} \frac{b-a}{n+1} f\left(a + \frac{b-a}{2(n+1)} + \frac{b-a}{(n+1)} i \right).$$

The error is bounded by $\frac{1}{4}\frac{(b-a)^2}{n+1} \|f'\|_\infty$.

Proof. By Corollary 1, if we introduce $h_0 = (x_0 - a)$, $h_{n+1} = (b - x_n)$, and for $1 \le i \le n$ we set $h_i = (x_i - x_{i-1})$ then the problem reduces to finding the minimum of the sum $\frac{1}{2}h_0^2 + \sum_{i=1}^{n} \frac{1}{4}h_i^2 + \frac{1}{2}h_{n+1}^2$ subject to $\sum_{i=0}^{n+1} h_i = b - a$. We can use Lagrange multipliers to find extrema. The solution to the system

$$h_0 = \lambda,\ h_{n+1} = \lambda,\ \frac{1}{2}h_i = \lambda \text{ for } 1 \le i \le n \text{ and } \sum_{i=0}^{n+1} h_i = b - a$$

is

$$h_0 = \frac{b-a}{2(n+1)},\ h_{n+1} = \frac{b-a}{2(n+1)} \quad \text{and } h_i = \frac{b-a}{(n+1)} \text{ for } 1 \le i \le n-1.$$

Since $\sum_{i=0}^{n+1} h_i = b - a$ it is routine to check that

$$\frac{1}{2}h_0^2 + \sum_{i=1}^{n} \frac{1}{4}h_i^2 + \frac{1}{2}h_{n+1}^2 = \frac{1}{4}\frac{(b-a)^2}{n+1} + \frac{1}{2}\left(h_0 - \frac{b-a}{2(n+1)} \right)^2 + \sum_{i=1}^{n} \frac{1}{4}\left(h_i - \frac{b-a}{(n+1)} \right)^2 + \frac{1}{2}\left(h_{n+1} - \frac{b-a}{2(n+1)} \right)^2.$$

Thus Lagrange multipliers produced the minimum sum of $\frac{1}{4}\frac{(b-a)^2}{n+1}$. Hence the nodes are $x_i = a + \frac{b-a}{2(n+1)} + \frac{b-a}{(n+1)} i$ for $0 \le i \le n$. From Corollary 1 it follows that the weights are $a_i = \frac{b-a}{n+1}$ for $0 \le i \le n$. □

In the introduction we gave the sharp error bounds for the trapezoidal and Simpson's rules. These errors can be obtained from Theorem 1, in the same way as the errors in the following examples.

Example 2. Boole's Rule:

$$\int_a^b f(x)\,dx = \frac{b-a}{90}\left[7f(a) + 32f(a+\frac{1}{4}(b-a)) + 12f(a+\frac{1}{2}(b-a)) \right.$$
$$\left. +32f(a+\frac{3}{4}(b-a)) + 7f(b)\right] - \frac{8}{945}\left(\frac{b-a}{4}\right)^7 f^{(6)}(\xi).$$

However if f is just differentiable by Theorem 1 the error is

$$c_4 := \int_a^b \left|(b-x) - \frac{b-a}{90}\left[32H(a+\frac{1}{4}(b-a)-x) + \right.\right.$$
$$\left.\left. 12H(a+\frac{1}{2}(b-a)-x) + 32H(a+\frac{3}{4}(b-a)-x) + 7\right]\right|\,dx.$$

Introducing the substitution $t = \frac{x-a}{b-a}$ we have

$$c_4 := (b-a)^2 \int_0^1 \left|(1-t) - \frac{1}{90}\left[32H(\frac{1}{4}-t) + \right.\right.$$
$$\left.\left. 12H(\frac{1}{2}-t) + 32H(\frac{3}{4}-t) + 7\right]\right|\,dt.$$

Using Maple we can evaluate the last integral to obtain $c_4 = \frac{239}{3240}(b-a)^2$.

Example 3. Composite Boole's Rule: First subdivide the interval $[a,b]$ into m equal subintervals and apply Boole's rule on each one of these m intervals. We can use Theorem 1 on the subintervals to obtain that the error over these intervals is $\frac{239}{3240}\left(\frac{b-a}{m}\right)^2$. Thus the total error is $\frac{239}{3240}\frac{(b-a)^2}{m}$. Since we used $(m+1)+3m = 4m+1$ points we have that $c_{4m} = \frac{239}{810}\frac{(b-a)^2}{4m}$. That c_{4m} is sharp follows easily from the fact that the errors over each subinterval are sharp.

References

[1] D. Cruz-Uribe and C. J. Neugebauer, *Sharp error bounds for trapezoidal rule and Simpson's rule,* JIPAM. J. Inequal. Pure Appl. Math. 3 (4)(2002), Article 49, 22 pp. (electronic).

[2] S. S. Dragomir, P. Cerone and J. Roumeliotis, *A new generalization of Ostrowski's integral inequality for mappings whose derivatives are bounded and applications in numerical integration and for special means,* Appl. Math. Lett., 13 (2000) 19–25.

[3] G. A. Anastassiou, *Ostrowski Type Inequalities,* Proc. Am. Math. Soc., 123 (12) (1995) 3775–3781.

(Received: July 31, 2007)
(Revised: October 7, 2007)

Department of Mathematics
California State University San Bernardino
92407 San Bernardino, CA
E–mail: hfejzic@csusb.edu