ON THE DEGREE OF APPROXIMATION OF CERTAIN QUADRATURE FORMULAS

By

A. L. O'Toole

National Research Fellow.

If $f(x)$ be a continuous function of period 2π, and if the interval under consideration, say the interval from 0 to 2π, be divided into m equal parts by the $m+1$ points $x_i = 2i\pi/m$, $i = 0, 1, 2, \ldots\ldots, m$, then the trigonometric sum of the nth order coinciding in value with $f(x)$ at the $m+1$ points x_i, or the trigonometric sum of the nth order lacking the term in $\sin nx$, is, according as $m = 2n+1$ or $m = 2n$,

$$\Phi_n(x) = \tfrac{1}{2}a_0 + a_1\cos x + a_2\cos 2x + \cdots\cdots + a_n\cos nx$$
$$+ b_1\sin x + b_2\sin 2x + \cdots\cdots + b_n\sin nx$$

or

$$u_n(x) = \tfrac{1}{2}a_0 + a_1\cos x + a_2\cos 2x + \cdots\cdots + \tfrac{1}{2}a_n\cos nx$$
$$+ b_1\sin x + b_2\sin 2x + \cdots + b_{n-1}\sin(n-1)x,$$

where

$$a_k = \frac{h}{\pi}\sum_{i=1}^{m} f(x_i)\cos kx_i, \qquad h = \frac{2\pi}{m},$$

$$b_k = \frac{h}{\pi}\sum_{i=1}^{m} f(x_i)\sin kx_i.$$

If the Fourier coefficients of $f(x)$ be denoted by

$$\alpha_k = \frac{1}{\pi}\int_0^{2\pi} f(x)\cos kx\,dx,$$

$$\beta_k = \frac{1}{\pi}\int_0^{2\pi} f(x)\sin kx\,dx,$$

then it has been shown[1] that the interpolating coefficients a_k and b_k are approximations to the Fourier coefficients α_k and β_k in the sense of the rectangle quadrature formula, in the sense of the trapezoid quadrature formula, in the sense of the average of the results of two applications of Simpson's formula, and in the sense of higher quadrature formulas. In other words, the simple rectangle formulas a_k and b_k are as good approximations to the areas α_k and β_k as the estimates given by the trapezoid rule, the average of two applications of Simpson's rule, or higher quadrature formulas.

It is the purpose of this note to discuss certain quadrature formulas and to observe some other conditions under which the rectangle formula will give as good an approximation as the more complicated formulas.

The most elementary and best known of the formulas are the rectangle formula, the trapezoid formula, and Simpson's formula. Many of the more complex rules are the results of attempts by different investigators[2] to improve by various devices the approximations given by these three simple rules.

Suppose the area under consideration is bounded by the curve $y = f(x)$, the x-axis and the ordinates at $x = a$ and $x = b$. If the interval from a to b be devided into n equal[3] parts, say of length h, by the $n+1$ points $x_0 = a, x_1, x_2, \cdots, x_{n-1}, x_n = b$, and if rectangles, each of width h and height y_i, $i = 0, 1, 2, \cdots, n-1$, be constructed, then the area as approximated by these n rectangles is

$$(1) \qquad A = h \sum_{\nu=0}^{n-1} y_\nu .$$

---

[1]D. Jackson, Some Notes on Trigonometric Interpolation, Amer. Math. Monthly, vol. xxxiii, no. 8, October 1927.

[2]See Runge and Willers, Encyklopädie Der Mathematischen Wissenschaften, Bd. II:3 (1915), pp. 45-176.

[3]Discussion from point of view of least squares, Otto Biermann, Monatshefte Fur Mathematik Und Physik, 14 (1903), pp. 226-242.

For unequal intervals see Jas. W. Glover, International Mathematical Congress, Toronto, 1924.

To find an expression for the error we assume the first derivative exists, so that for the first rectangle

$$f(x) = f(a) + (x-a)f'(u),$$

$$\int_a^{a+h} f(x)\,dx = hf(a) + \frac{h^2}{2}f'(z), \quad a < z < a+h.$$

Hence the error for the n rectangles is

$$E = \frac{h^2}{2}\sum_{v=1}^{n} f'(z_v) = \frac{(b-a)^2}{2n} f'(z), \quad a < z < b, \tag{1e}$$

i.e., an error of the order of $\frac{1}{n}$.

Let $n = mk$, $k = 1, 2, 3, \cdots$. If we approximate the area in the first k subintervals by a parabola of degree k coinciding in value with $f(x)$ at the first k values of x, then integrating Lagrange's interpolation formula an expression for the error is obtained. If k is odd then

$$E_1 = C_1\left(\frac{H}{2}\right)^{k+2}\frac{f^{(k+1)}(z)}{(n+1)!}, \quad H = kh,$$

where

$$C_1 = \frac{1}{(k+2)k^{k-1}}\int_{-1}^{1}(t^2-1)(k^2t^2-1^2)(k^2t^2-3^2)\cdots(k^2t^2-(k-2)^2)\,dt.$$

If k is even, then making use of Rolle's Theorem,

$$E_2 = C_2\left(\frac{H}{2}\right)^{k+3}\frac{f^{(k+2)}(z)}{k+2!}$$

where

$$C_2 = \frac{1}{k^{k-2}}\int_{-1}^{1}(t^2-1)(t^2)(k^2t^2-2^2)\cdots(k^2t^2-(k-2)^2)\,dt.$$

The error over the whole interval will be obtained by summing the m errors corresponding to each k subintervals.

If n trapezoids are formed by joining the ends of successive ordinates then the area as approximated by the sum of the areas of these trapezoids is

$$A = \frac{h}{2} \sum_{v=0}^{n-1} (y_v + y_{v+1}) \tag{2}$$

and the error is

$$E = -\frac{(b-a)^3}{12n^2} f''(z), \tag{2e}$$

i.e., an error of the order of $\frac{1}{n^2}$.

Simpson's formula may be obtained by passing second degree parabolas through the ends of three successive ordinates, that is $k=2$, and gives

$$A = \frac{h}{3}\left[2 \sum_{v=0}^{m} y_{2v} + 4\sum_{v=1}^{m} y_{2v-1} - (y_0 + y_{2m})\right], \quad n=2m. \tag{3}$$

The error is

$$E = -\frac{(b-a)^5}{180n^4} f^{iv}(z), \tag{3e}$$

i.e., an error of the order of $\frac{1}{n^4}$.

To illustrate the fact that sometimes the rectangle formula (1) gives a better approximation than the Simpson formula (3) these formulas will be applied to the problem of finding the area under the so-called normal curve of error. From a table[4] giving five places of decimals it is seen that the ordinates to the right of $x = 4.76$ and to the left of $x = -4.76$ are everywhere zero if the equation be written in the form $y = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ Divide the interval from $x = -4.80$ to $x = 4.80$ into eight partial intervals each of length 1.20. Formula (1) gives $A = .99998$ while

[4] Jas. W. Glover, Tables of Applied Mathematics in Finance, Insurance and Statistics.

(3) gives $A = .97834$, the same ordinates being used in each case.

There are three objections to the nature of Simpson's formula. They are the lack of smoothness at the points of intersection of the parabolas, the unequal weights attached to the odd and even numbered ordinates, and the requirement that the number of ordinates be odd.

Catalan[5] notices the lack of smoothness at the intersections of the parabolas used in setting up Simpson's rule and improves on it by passing parabolas through three successive ordinates and then retaining only the first half of each parabola except in the case of the last three ordinates where it is necessary to retain the whole parabola. To counterbalance the asymmetry introduced by these last three ordinates he repeats the process beginning with the last ordinat and then takes the arithmetic mean of the two results as his formula.
This gives

$$(4)\quad A = h\left[\sum_{v=0}^{n} y_v - \frac{5}{8}(y_0 + y_n) + \frac{1}{6}(y_1 + y_{n-1}) - \frac{1}{24}(y_2 + y_{n-2})\right].$$

And, of course, the error is still of the order of $\frac{1}{n^4}$. This formula has the additional advantage that it holds no matter whether n is even or odd.

Similarly Crotti[6] showed that the different weights attached to the odd and even numbered ordinates in Simpson's formula ic a disadvantage. And Parmentier[7] by subtracting Simpson's formula from twice Catalan's obtained a formula in which the weights are the reverse of those in Simpson's. Mansion[8] gave an alternative derivation of Catalan's formula, his derivation requiring, however, an even number of ordinates.

---

[5]E. Catalan, Nouvelles Annales, 1re series (1851), pp. 412-415.

[6]Crotti, Il Politechnio 33 (1885), pp. 193-207.

[7]Parmentier, Association française pour l'avancement des sciences, Session Grenoble, 1882.

[8]Mansion, Supplement zu Mathesis 1 (1881).

Catalan's formula may be thought of as the rectangle formula plus three correctional terms involving the first three and the last three ordinates. In the case of an even number of ordinates a formula[9] involving only two such correctional terms and giving an approximation of the order of the error in the single trapezoid, i.e., of the order of $\frac{1}{n^3}$, the error in a single trapezoid of width $\frac{(b-a)}{n}$ being $\frac{-f''(z)(b-a)^3}{12n^3}$, can be obtained by applying Simpson's formula to the first $2m-1$ ordinates and approximating the remaining area by the trepezoid rule. Repeat the process from the opposite end and take the arithmetic mean of the two results as the quadrature formula. This gives

$$(5) \qquad a = h\left[\sum_{v=0}^{n} y_v - \frac{7}{12}(y_0+y_n)+\frac{1}{12}(y_1+y_{n-1})\right], \; n=2m-1.$$

It is the only formula with just two correctional terms which will give even this order of approximation in general because any change in the coefficients of these end ordinates will introduce in general an error of the order of the error in the rectangle formula for a single subinterval, i.e., an error of the order of $\frac{1}{n^2}$.

Another important quadrature formula is called the three-eighths rule and is obtained by passing third order parabolas through four successive ordinates. It may be written

$$(6) \qquad A = \frac{3h}{8}\left[2\sum_{v=0}^{m} y_{3v} + 3\sum_{v=0}^{m-1} y_{3v+1} + 3\sum_{v=0}^{m-1} y_{3v+2} - (y_0+y_{3m})\right], \; n=3m.$$

The error is

$$(6e) \qquad E = -\frac{(b-a)^5}{400n^4} f^{iv}(z),$$

i.e., an error of the same order as the error corresponding to Simpson's formula. The error terms derived from the Lagrange

[9] Durand, Engineering News, Jan. 1894. J. Lipka, Graphical and Mechanical Computation, Part II, p. 226.

formula shows the advantage of using parabolas of even degree.

Besides the fact that the order of the error is the same as that in the case of Simpson's formula, this three-eighths formula has disadvantages similar to those mentioned in the case of Simpson's formula. There is still a lack of smoothness at the intersections of the parabolas; the weights attached to the ordinates are as undesirable as before; and the number of partial intervals must be a multiple of three.

It is possible however to do away with these disadvantages by proceeding as follows. Pass a third order parabola through the first four ordinates y_0, y_1, y_2, y_3. Retain only the area in the first two partial intervals. Pass a third order parabola through the four ordinates y_1, y_2, y_3, y_4 and retain only the area in the central interval. Proceed in this way retaining each time only the area in the central interval until the last four ordinates are reached where it will again be necessary to retain the area in two strips, viz., the last two partial intervals. The sum of these areas gives the required quadrature formula. It is

$$(7)\quad A = h\left[\sum_{v=0}^{n} y_v - \frac{2}{3}(y_0+y_n) + \frac{7}{24}(y_1+y_{n-1}) - \frac{1}{6}(y_2+y_{n-2}) + \frac{1}{24}(y_3+y_{n-3})\right].$$

This formula holds for any n greater than or equal to three. From the point of view of the order of the error this formula is, as one would expect, no better than Catalan's formula. As a matter of fact formula (7) can be obtained from formula (4) by subtracting from (4) $\frac{h}{24}(\Delta^3 y_0 - \Delta^3 y_{n-3})$ a quantity which, in general, is of the order of $\frac{1}{n^4}$.

If $n=4m$ and fourth order parabolas are used in approximating the area in four successive partial intervals then the formula is

$$(8)\quad A = \frac{4h}{45}\left[7\sum_{v=0}^{m} y_{4v} + 16\sum_{v=0}^{m-1} y_{4v+1} + 6\sum_{v=0}^{m-1} y_{4v+2} + 16\sum_{v=0}^{m-1} y_{4v+3} - \frac{7}{2}(y_0+y_{4m})\right].$$

The error is

$$E = -\frac{2(b-a)^7}{945n^6} f^{vi}(z), \tag{8e}$$

i.e., an error of the order of $\frac{1}{n^6}$.

Several modifications may be made to improve this formula. For instance if $n = 2m+1$ then apply the fourth degree parabola to the ordinates y_0, y_1, y_2, y_3, y_4 and retain only the area in the first three strips. Apply a fourth degree parabola to the ordinates y_2, y_3, y_4, y_5, y_6 and retain the area in the two central strips. And so on till in the final step it will be necessary to retain the area in the last three strips. Addition gives the formula

$$A = \frac{h}{720}\Big[896\sum_{v=0}^{m} y_{2v} + 544\sum_{v=1}^{m} y_{2v-1} - 653(y_0 + y_{2m}) + 374(y_1 + y_{2m-1}) \tag{9}$$

$$-256(y_2 + y_{2m-2}) + 106(y_3 + y_{2m-3}) - 19(y_4 + y_{2m-4})\Big]$$

A formula which holds for any n may be obtained by passing a fourth degree parabola through y_0, y_1, y_2, y_3, y_4 and retaining only the area between y_0 and y_2. Pass a fourth degree parabola through y_1, y_2, y_3, y_4, y_5 and retain only the area between y_2 and y_3. And so on, retaining only the area in one strip, until at the end it will be necessary to retain the area in the last three strips. Repeat the process beginning at the last ordinate and take the arithmetic mean. The result is

$$A = h\Big[\sum_{v=0}^{n} y_v - \frac{193}{288}(y_0 + y_n) + \frac{77}{240}(y_1 + y_{n-1}) - \frac{7}{30}(y_2 + y_{n-2}) \tag{10}$$

$$+ \frac{73}{720}(y_3 + y_{n-3}) - \frac{3}{160}(y_4 + y_{n-4})\Big].$$

This formula can be obtained in the case of an even number of ordinates by retaining three strips at the beginning, two from then on, reversing the process and taking the arithmetic mean.

Formulas (4), (5), (7) and (10) not only give, in general, at least as good approximations as Simpson's formula, the trapezoid formula, the three-eighths formula, and the fourth degree formula (8) respectively, but in addition have the important property that under certain conditions they show that the simple rectangle formula must give at least as good an approximation as the higher formulas. If $f(x)$ is a function such that the curve $y = f(x)$ actually, or at least for practical purposes, coincides with the x -axis to the left of $x=a$ and to the right of $x=b$, then in dividing the interval from a to b into h equal parts each of length h it will not affect the area required if two, one, three or four partial intervals of length h are marked off to the left of a and to the right of b, the number of such partial intervals corresponding to (4), (5), (7) and (10) respectively. Hence it is seen that under these conditions (4), (5), (7) and (10) reduce to the simple rectangle formula (1).

If the curve coincides with the x -axis at one end of the interval over which the area is required but does not at the other end then formulas (4), (5), (7) and (10) become respectively

$$(4a)\quad A=h\Big(\sum_{v=0}^{n} y_v - \frac{5}{8}y_n + \frac{1}{6}y_{n-1} - \frac{1}{24}y_{n-2}\Big),$$

$$(5a)\quad A=h\Big(\sum_{v=0}^{2m-1} y_v - \frac{7}{12}y_{2m-1} + \frac{1}{12}y_{2m-2}\Big),$$

$$(7a)\quad A=h\Big(\sum_{v=0}^{n} y_v - \frac{2}{3}y_n + \frac{7}{24}y_{n-1} - \frac{1}{6}y_{n-2} + \frac{1}{24}y_{n-3}\Big),$$

$$(10a)\quad A=h\Big(\sum_{v=0}^{n} y_v - \frac{193}{288}y_n + \frac{77}{240}y_{n-1} - \frac{7}{30}y_{n-2} + \frac{73}{720}y_{n-3} - \frac{3}{160}y_{n-4}\Big).$$

For example, consider again the normal curve of error and suppose that the area to the left of the ordinate at $x=o$ is required. Formulas (4a), (5a), (7a) and (10a) apply and for sixteen partial intervals give respectively $A=.49994$, $A=.49530$, $A=.50008$, and $A=.50002$, an extra partial interval to the left of $x=-4.80$ being used in the case of (5a) in order to have an odd number of intervals for that formula. Using thirty-two partial intervals the same formulas give $A=.49999$, $A=.49949$, $A=.50000$, and $A=.50000$ respectively.

If, as often happens, the values of ordinates outside the interval over which the area is required are known then even better quadrature formulas may be obtained. For example, suppose that in deriving formula (7) the ordinate y_{-1} at a distance of h to the left of y_0 and the ordinate y_{n+1} at a distance h to the right of y_n are known. Then it will not be necessary to retain the areas in double strips at the beginning and end of the interval, and the formula for the area over the interval from $x=a$ to $x=b$ is

$$(11)\quad A=h\left[\sum_{\nu=0}^{n} y_\nu-\frac{1}{24}(y_{-1}+y_{n+1})-\frac{1}{2}(y_0+y_n)+\frac{1}{24}(y_1+y_{n-1})\right].$$

It should be noted that in case y_{-1} and y_{n+1} are known Catalan's formula reduces to (11). And, similarly, in the case of the derivation of formula (10) it will be necessary to retain the area in a single strip each time except in the case of the last application of the fourth degree parabola when it will be necessary to retain the area in the two central strips. The formula arrived at is

$$(12)\quad A=h\left[\sum_{\nu=0}^{n} y_\nu-\frac{3}{160}(y_{-1}+y_{n+1})-\frac{83}{144}(y_0+y_n)+\frac{2}{15}(y_1+y_{n-1})\right.$$
$$\left.-\frac{11}{240}(y_2+y_{n-2})+\frac{11}{1440}(y_3+y_{n-3})\right].$$

Formulas (11) and (12) reduce to the rectangle formula (1) under the same conditions as in the cases of (4), (5), (7) and (10). Likewise when the curve coincides with the x-axis to the left of $x=a$ (11) and (12) become

(11a) $A = h\left(\sum_{v=0}^{n} y_v - \frac{1}{24} y_{n+1} - \frac{1}{2} y_n + \frac{1}{24} y_{n-1}\right)$, and

(12a) $A = h\left(\sum_{v=0}^{n} y_v - \frac{3}{160} y_{n+1} - \frac{83}{144} y_n + \frac{2}{15} y_{n-1} - \frac{11}{240} y_{n-2} + \frac{11}{1440} y_{n-3}\right)$

If we apply formula (11a) to finding the area under the normal curve to the left of the ordinate at $x = o$ and take $n = 4$, $h = 1.20$, $a = -4.80$ then we find $A = .49999$. In other words, in this case (11a) gives as good a result with six ordinates as (4a) or (7a) give with thirty-three ordinates or (6a) with thirty-four ordinates.

Quadrature formulas involving parabolas of degree higher than four have been obtained but they are to be used with caution on account of the great freedom they allow the approximating curves. However, modifications similar to those in this paper could also be made for these higher formulas. And the effect of any number of ordinates outside the ends of the interval could be noted.

This note will be concluded with a remark on the effect of errors in the data giving the values of the ordinates. Suppose the quadrature formula is $A = h(a_0 y_0 + a_1 y_1 + a_2 y_2 + \cdots\cdots + a_n y_n)$ and suppose further that each y_i is subject to an error e_i, $i = 0, 1, 2, 3, \cdots, n$. If e is the greatest of the absolute values of the e_i then the error in A cannot be greater than $he(a_0 + a_1 + a_2 + \cdots + a_n)$ if $a_0, a_1, a_2, \cdots\cdots, a_n$ are all positive, as will be true if parabolas of the fourth degree or lower are used. But $h(a_0 + a_1 + a_2 + \cdots + a_n) = (b-a)$ if the area is to be four from $x = a$ to $x = b$. Hence the error in A due to errors in the data is not greater than $e(b-a)$. When parabolas of degree higher than four are used the coefficients in the quadrature formula are not always positive.

A. L. O'Toole