Analysis of Numerical Integration Methods

By Carl De Vries

Calculus II – Honors – Fall 2013

Introduction:

There several numerical integration techniques that can be used to determine the definite integral of a function. Numerical methods are useful when analytical integration is very difficult. However, they only provide an approximation of the definite integral and they are subject to truncation error and round off error. The approximation is determined by finding the area of the function under the curve. The approximation does not necessarily represent an area though. It may represent a variety of different values. For example, say the function f(x) describes the forces in a system with respect to time. The definite integral of the function between [a,b] is equal to the work done in the system as well as the area underneath the curve.

Each technique has its own pros and cons. The best method for each situation can be determined by examining the constraints given by the problem or given by the method of calculation. This paper explores two fundamental Newton-Cotes methods. The first is the Trapezoidal Rule and the second method is Simpson's 1/3 Rule.

Theory:

Newton-Cotes Methods:

The Newton-Cotes methods are the most frequently used numerical integration techniques. This is because they provide accurate answers and they are easier to implement than other numerical methods. The Newton-Cotes methods approximate the definite integral of a function by approximating the original curve with a number of simple curves. The Trapezoidal Rule and Simpson's 1/3 Rule are two closed Newton-Cotes methods. The closed methods are used to find the definite integral of a function between two concrete data points. These two data points are the upper limit of integration [x=a] and the lower limit of integration [x=b]. While not described here, open Newton-Cotes methods are typically used to approximate improper integrals or used to find a solution to ordinary differential equations.

The Trapezoidal Rule and Simpson's Rule approximate the definite integral of a function by finding the area under the curve between two limits [a,b]. To reach better approximations, each method can divide the range between the upper limit of integration and the lower limit of integration into multiple subintervals of equal width. The number of intervals must be specified at the beginning of the problem and is represented by the variable [n]. Increasing the number of intervals will increase the accuracy of the answer to a certain point. If too many intervals are used the time to calculate the problem and the amount of round off error may become an issue. The area under the curve is calculated using the summation of the area for all the subintervals.

Figure 1 is a plot of the curve $f(x)=x^2-x-6$. The shaded area represents the area under the curve. There are three unique sub-areas which make up the total area under the curve. They can be identified by the direction of the shading under the curve. Both Trapezoidal Rule and Simpson's 1/3 Rule will be used to approximate either the total area or a sub area of this function.

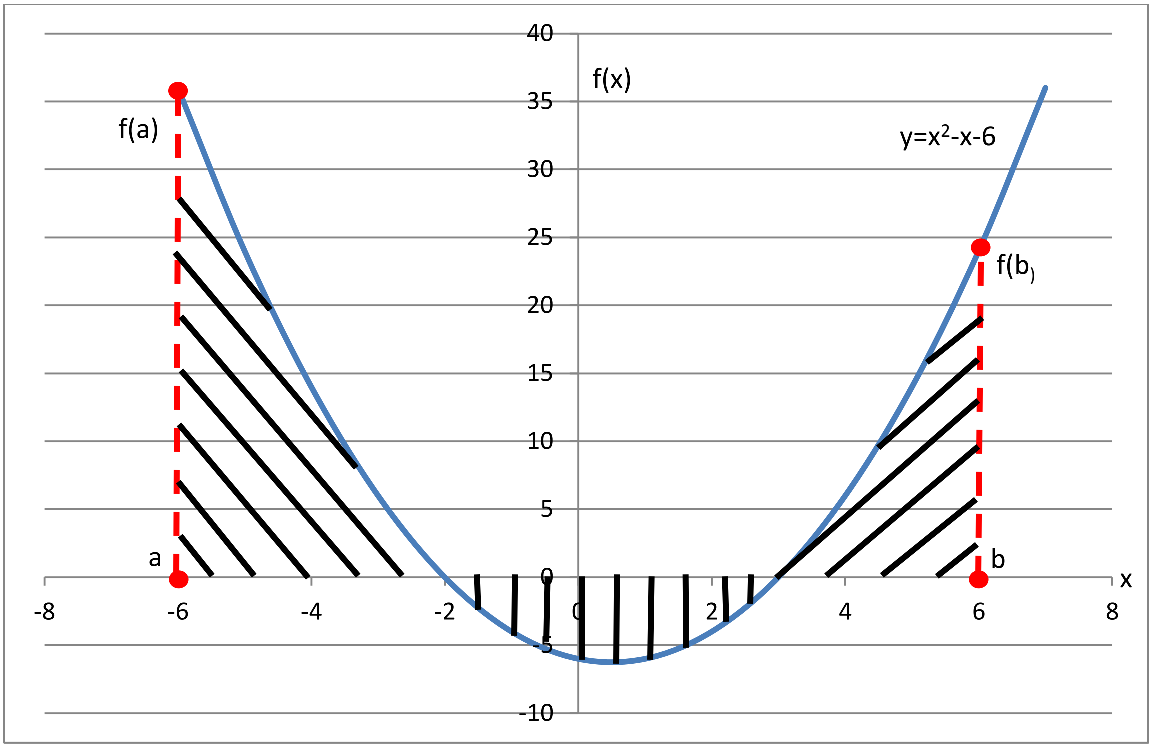

Figure 1 depicts the area under the curve between the values x= 6 and x = -6.

Trapezoidal Rule:

Trapezoidal Rule approximates the curve of the function using a first order approximation. This means a straight line is used to approximate the function between two points. It is named the Trapezoidal Rule because connecting the bounds between $f(x_0)$ and $f(x_n)$ forms a trapezoid. Solving for the area of the trapezoid yields an approximation of the area under the curve for the given subinterval. The areas for each subinterval are then added up to determine the total area under the curve. If only one interval is used between [a,b], the approximation of the definite integral will be less accurate due to a poor curve approximation. As the range between [a,b] is divided into more subintervals the approximation will become more accurate.

A comparison of Figure 2 and Figure 3 shows the difference in the curve approximation when a different number of subintervals is used. Figure 2 only has a single interval between [a,b] and Figure 3 uses two subintervals between [a,b]. When using more than one interval, each subinterval bound is assigned a variable $[x_i]$. The lower limit of integration[a] is equal to x_0 and the upper limit of integration [b] is equal to x_n. Figure 2 and Figure 3 show a sub-section of the graph from Figure 1. The limits of integration for Figure 2 and Figure 3 are a = -6 and b = -2.

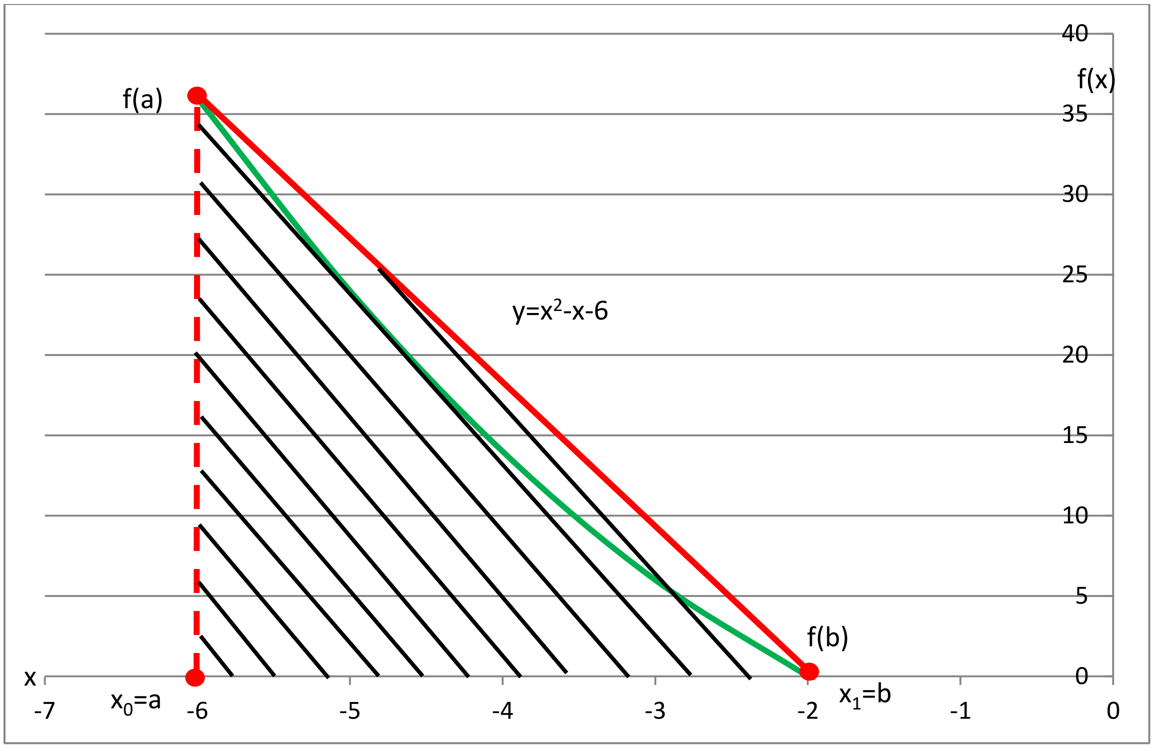

Figure 2 shows a curve approximation (solid red) for the function f(x) = x2-x-6 (solid green). The upper limit is at a = -6 and the lower limit is at b = -2. This approximation uses a single interval (n = 1).

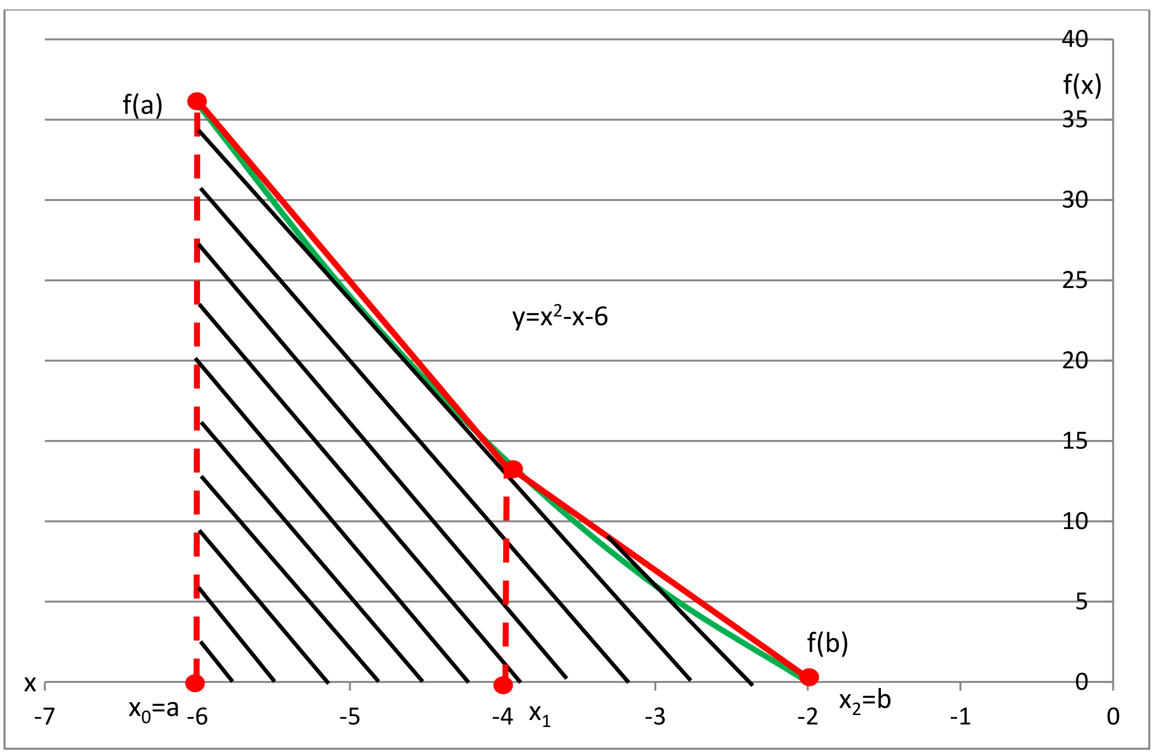

Figure 3 shows a curve approximation for the function f(x) = x^2-x-6. The upper limit is at a = -6 and the lower limit is at b= -2. This approximation uses two intervals (n=2).

The approximation in Figure 2 overestimates the area under the curve more than the approximation in Figure 3. Using more intervals will make the curve approximation for a subinterval more accurate. Thus, the approximation of the total area below the curve will be more accurate. In Figure 2 and Figure 3 the curve approximation is an overestimate. This is because the function is concave up between the interval [a,b]. Using a concave down function would produce an approximation which underestimates the area below the curve. If an approximation is being made for a third order function (cubic) or greater, the approximation may contain a combination of overestimates and underestimates due to changes in concavity. This is something to take into consideration depending on the application of the problem.

To approximate the area of the curve, the Trapezoid Rule starts with the equation for the area of a trapezoid. The height of the trapezoid is represented by [h]. When using the Trapezoidal rule, the height of the trapezoid equals the width of each subinterval. This width is the same for all intervals in the approximation. The height (interval width) is found by determining the range between the limits of integration and dividing by the number of intervals.

$$h = \frac{b-a}{n}$$

When the function is graphed, the height runs horizontally along the x-axis. The trapezoid may appear to be on its side. The bases of the trapezoid are given by s_1 and s_2, respectively. The bases are equal to the value f(x_i). Using more than one interval requires that each internal base (f(x_1) through f(x_{n-1})) be added twice. This is because the value f(x_i) will represent b_1 for one trapezoid and b_2 for the preceding trapezoid. To accomplish this, all of the values from f(x_1) to f(x_{n-1}) are summed and then multiplied by two. The outer bounds x_0 and x_n are sides on only a single trapezoid. Thus, they are not multiplied by two. The following steps show the how to derive the equation for the Trapezoidal Rule.

$$A = \frac{1}{2} * h * \ (s_1 + s_2)$$

Start with the equation for the area of a trapezoid.

$$A = \frac{1}{2} * h * \ (f(x_0) + 2 * \sum_{i=1}^{n-1} f(x_i) + \ f(x_n))$$

Substitute (b1 + b2) with the summation for all the internal f(x_n) values.

.

$$A = \frac{1}{2} * \frac{b-a}{n} * [f(x_0) + 2 * \sum_{i=1}^{n-1} f(x_i) + f(x_n)]$$

Substitute the width of each interval for the height [h].

.

$$A = \frac{b-a}{2n} * [f(x_0) + 2 * \sum_{i=1}^{n-1} f(x_i) + f(x_n)]$$

Combine terms to produce the final equation.

Simpson's 1/3 Rule:

Simpson's 1/3 Rule is unique because it requires an even number of intervals to work. This is because subinterval approximations are calculated two at a time. Thus, the smallest number of intervals that can be used is two. The intervals from [x_0, x_1] and [x_1, x_2] are used as the first two subintervals. A second degree polynomial that goes from x_0 to x_2 and passing through x_1 is used to approximate the curve in the given subinterval. An area approximation for the two subintervals can be found using the following equation.

$$A = \frac{h}{3} * [f(x_0) + 4 * f(x_1) + f(x_2)]$$

This equation is comparable to using the Trapezoidal Rule with a single interval. Using more intervals with Simpson's Rule will yield more accurate approximations. The number of subintervals must always be even. Using two more subintervals for our approximation yields the following equation.

$$A = \frac{h}{3} * [f(x_0) + 4 * f(x_1) + f(x_2)] + \frac{h}{3} * [f(x_2) + 4 * f(x_3) + f(x_4)]$$

Combine the [h/3] terms from both subinterval approximations.

$$A = \frac{h}{3} * ([f(x_0) + 4 * f(x_1) + f(x_2)] + [f(x_2) + 4 * f(x_3) + f(x_4)])$$

Next, combine all the terms with a coefficient of four.

$$A = \frac{h}{3} * (\left[f(x_0) + 4\left(f(x_1) + f(x_3)\right) + f(x_2) + f(x_2) + f(x_4)\right]$$

Then, combine all the remaining like terms.

$$A = \frac{h}{3} * ([f(x_0) + 4(f(x_1) + f(x_3)) + 2 * f(x_2) + f(x_4)]$$

From this notation it appears that all values of $f(x_i)$ when [i] is an odd number will be multiplied by four. Also, with the exception of the final side $f(x_4)$, all the values of $f(x_i)$ when [i] is an even number are multiplied by two. Note that the values $f(x_0)$ and $f(x_4)$ are only added once. This is because they are the outside edges of the approximation area. The next step is to put all the function values multiplied by four into summation notation. The same will be done for the function values multiplied by two. This is an easier notation to read when the number of intervals begins to get large.

$$A = \frac{h}{3} * \left[f(x_0) + 4 \sum_{i=1,3}^{3} f(x_i) + 2 \sum_{j=2}^{2} f(x_j) + f(x_4) \right]$$

The next step is to substitute for the height [h]. As with the Trapezoidal Rule, the height is equal to the difference between the limits of integration [x = a and x = b], and is then divided by the number of intervals.

$$A = \frac{b-a}{3n} * \left[f(x_0) + 4 \sum_{i=1,3}^{3} f(x_i) + 2 \sum_{j=2}^{2} f(x_j) + f(x_4) \right]$$

The equation can be put into a general equation that can be used with any number of intervals as long as it's an even number.

$$A = \frac{b-a}{3n} * \left[f(x_0) + 4 \sum_{i=1,3,5}^{n-1} f(x_i) + 2 \sum_{j=2,4,6}^{n-2} f(x_j) + f(x_n) \right]$$

Simpson's 1/3 Rule tends to produce more accurate approximations because the second degree polynomials are a better fit to the original curve than first degree approximations. The curve approximations with Simpson's Rule create less error due to overestimations and underestimations. However, Simpson's 1/3 Rule doesn't work well for certain functions and it only works with an even number of intervals.

Coding:

The code for implementing Trapezoidal Rule and Simpson's Rule depend heavily on a loop structure. The loop structure is used to sum all of the $f(x_i)$ values for each method. The Trapezoidal Rule only has one summation, and thus, it only requires a single loop. Simpson's 1/3 Rule has two separate summations. One is for accumulating all of the function values at odd intervals. The other is for accumulating the function values at all the even intervals. For both methods, the summations are the most calculation intensive component. This is because the number of iterations through the loop is dependent on the number of intervals that will be used.

Both the Trapezoidal Rule and Simpson's 1/3 Rule require the same four inputs to run. The first input is the function. Next the program needs the upper limit of integration and the lower limit of integration. The last input needed is the number of intervals to use. The number of intervals needed depends on the method being used. For Simpson's Rule, the number of intervals must be positive. Entering this number correctly is important because my implementation does not sanitize input to ensure a valid parameter is entered.

Code Validation:

I validated my code using two different problems. The first problem is function with a known answer for the definite integral. For this problem, the results tables include approximations with different numbers for intervals.

The known function I used was $f(x) = x^2-x-6$. The first step was to solve analytically for the definite integral. I used the value a = -6 for the upper limit and I used the value b = 6 for the lower limit. The answer to the analytical solution is 72. The code validation for each method returned the following results.

Trapezoidal Rule:

Parameters:
$f(x) = x^2-x-6$
a = -6
b = 6

Known solution:
A = 72

Intervals, n	Interval width, h	Area
12	1	74
24	0.5	72.5
48	0.25	72.125
96	0.125	72.03125
192	0.0625	72.0078125

Simpson's 1/3 Rule:

Parameters:
$f(x) = x^2 - x - 6$
$a = -6$
$b = 6$

Known solution:

Intervals, n	Interval width, h	Area
12	1	72
24	0.5	72
48	0.25	72
96	0.125	72
192	0.0625	72

For Trapezoidal Rule, the results clearly show an increase in accuracy as the number of intervals increases. The original approximation, 74, is greater than the known analytical answer of 72. This should be expected with the given parameters. Figure 1 shows that the curve is concave up from [-6,0.5]. This is a larger interval than the concave down portion from [0.5,6]. Approximations from the Trapezoidal Rule always overestimate the true area when a curve is concave up. This can be seen in Figure 2. Because there are a larger number of subintervals with an overestimated area, the total area is also slightly over estimated.

The approximations from the Trapezoidal Rule do converge on the true area and the approximations even become fairly accurate. However, this only occurs when the number of intervals increases. As the number of intervals increases so does the amount of time required to find an approximation.

The results from Simpson's 1/3 Rule validation are harder to interpret than the Trapezoidal Rule. With only twelve intervals Simpson's 1/3 Rule approximated the area under the curve at exactly 72. It should be noted that the approximation of exactly 72 does not mean Simpson's Rule gives exact answers. In fact, the result is still an approximation. The calculator simply ran out of spaces for additional digits and it rounded the answer to exactly 72. While it's nice that Simpson's Rule was very accurate in this case, the results don't tell us much.

The second problem is an application problem. The problem was selected from the book *Digital Computations for Numerical Methods.* The problem asks for a solution which describes the deflection of an aircraft wing. The deflection in the wing, δ, is given by the following equation.

$$\delta = \int_0^L \frac{Mx}{EI}\, dx$$

The variable, M, is the internal moment of the wing. The variable E is the modulus of elasticity of the material for the wing. The variable I represents the moment of inertia for the wing. The final variable, L, is the upper limit and represents the length of the wing. Each variable is assigned the following values.

$M = \frac{wx^2}{2}$ where $w = 2\ kips/ft$

$L = 40\ ft$

$E = 4.32\ \times 10^6$

$I(x) = 30 \times 10^{-9}x^4 + 5.0 \times 10^{-6}x^3 + 0.3 \times 10^{-3}x^2 + 8.0 \times 10^{-3}x + 0.1$

After plugging in all the variables I was left with the following integral.

$$\delta = \frac{1}{4.32\ \times 10^6} \times \int_0^{40} \frac{x^3}{30 \times 10^{-9}x^4 + 5.0 \times 10^{-6}x^3 + 0.3 \times 10^{-3}x^2 + 8.0 \times 10^{-3}x + 0.1} dx$$

The book calculated the approximations using Simpson's Rule. They found two different approximations. One approximation used four intervals and the second used eight. I used my Simpson's Rule program to find an approximation for the integral, and then I multiplied the answer by the coefficient using the basic calculator functions. The following table shows the approximations from my Simpson's Rule program.

Method	Number of intervals, n	Given approximation, δ_t (ft)	My approximation, δ_A (ft)
Simpson's Rule	4	0.1900	0.1900
Simpson's Rule	8	0.1901	0.1900

My first approximation, using four intervals, came out to be just under 0.1900. After rounding my answer to the appropriate digit, I matched the approximation from the book. The second approximation came is close to the book's approximation, but it's slightly off. The digit following the second zero was a four. I think it's possible my calculator truncated an answer earlier on in the program, and thus my final approximation was just a bit to0 low to round up to 0.1901.

Summary:

I came to several conclusions during my analysis of the Trapezoidal Rule and Simpson's 1/3 Rule. The first conclusion is that approximating the functions using higher order polynomials tends to yield more accurate answers with fewer intervals. I came to this conclusion by looking at the data from my first code validation. The Simpson's Rule program made much closer approximations than the Trapezoidal Rule. The downfall to Simpson's Rule is that it doesn't always work well with functions that oscillate. It's also apparent that as the number of intervals increases the time it takes to reach an approximation also increases. This isn't very surprising considering more intervals require additional iterations through the loop structure. I didn't notice a significant difference between the run-times for Simpson's Rule and the Trapezoid rule. Simpson's Rule requires two loop structures; however, one loop only processes odd intervals and the other loop only processes even intervals. The total number of iterations for each method should still be the same. The only difference I found is in the number of operation performed outside the loop. The time these operations take won't vary with the number of intervals, so I don't think there's a large difference in the run time even as the number of intervals gets very large. For the problem I validated, Simpson's 1/3 Rule was the clear winner. In the future, I would lean towards using the Simpson's Rule program rather than the Trapezoid Rule.