Integration: Part 1

September 24, 2015

Integration is introduced on a stand-alone basis. Integrals are defined as the limit of a sum, leading to numerical methods for their evaluation. The fundamental theorem of calculus allows elementary functions to be integrated. Some advanced techniques are in the next module.

Contents

1 Definition and properties

Integration has been introduced through the example of work done by a force. This module repeats some of the material, but on a stand-alone and general basis.

1.1 Definition

Let $F(x)$ be a smooth function with x defined on the real line (or a real interval). The integral of F over an interval (a, b), is defined as

$$\int_a^b F(x)\,dx \equiv \lim_{\Delta x\to 0} \sum F(x)\,\Delta x \qquad (1)$$

The RHS is a shorthand for the following procedure:

- Chop the interval (a, b) into short segments of length Δx.
- In each interval evaluate $F(x)$ at some point and calculate the product $F(x)\,\Delta x$.
- Add them all up.
- Repeat the procedure for smaller and smaller Δx until the value converges.

The notation is meant to be suggestive:

$$\sum \mapsto \int \quad , \quad \Delta x \mapsto dx \qquad (2)$$

In physics we often omit the limit symbol and regard Δx as small or infinitesimal:

$$\int_a^b F(x)\,dx = \sum F(x)\,\Delta x \qquad (3)$$

1.2 Graphical representation

Before taking the limit
The algorithm is illustrated in **Figure 1**, and the dot shows where the function is evaluated in a typical interval — the point is deliberately chosen randomly. Each term in the sum (3) is the area of a strip. Provided the function is smooth, evaluating

the function at different points leads to only a tiny difference which vanishes when $\Delta x \to 0$.

To show this more precisely, **Figure 2a** and **Figure 2b** illustrate two ways of evaluating the function. The difference is shown in **Figure 2c**, where

$$\begin{aligned} \text{width} &= \Delta x \\ \text{height} &\approx \text{slope} \times \Delta x \propto \Delta x \\ \text{area} &\propto (\Delta x)^2 \end{aligned} \tag{4}$$

The number of intervals $\propto (\Delta x)^{-1}$. So the total error is $\propto \Delta x \propto N^{-1}$ if the entire interval is chopped into N segments.

Upon taking the limit
If we chop into many many tiny intervals, then the area becomes equal to *the area under the curve*, as illustrated in **Figure 3**. The segments Δx can have different lengths, so long as all the lengths approach zero in the limit.

Sign convention
Since the integral involves $F\,\Delta x$, we need to state the sign convention for each factor.

If $F(x)$ is negative, then the integral is negative. This is also true for the contribution of a sub-interval (**Figure 4**): if the graph of $F(x)$ lies below the axis, then the "area under the curve" is negative.

For the sign of Δx, the convention is as follows. Given an integral

$$\int_a^b (\ldots)\,dx \tag{5}$$

Define a series of points $x_0 = a$, $x_1, x_2, \ldots, x_{N-1}$, $x_N = b$. The lengths of the intervals are

$$(\Delta x)_i = x_i - x_{i-1} \tag{6}$$

for $i = 1, \ldots, N$. Thus

$$\sum_{i=1}^{N} (\Delta x)_i = b - a \tag{7}$$

This definition is valid whether $b > a$ (the "normal" case) or $b < a$, or even if the sequence $\{x_n\}$ is not monotonic (i.e., there is some "doubling back"), so long as the lengths of all the segments approach zero. Some "unusual" possibilities are illustrated in **Figure 5**.

But in the most common case where the sequence $\{x_n\}$ goes monotonically (and typically in equal steps) from a to b, it means that Δx has the same sign as $b - a$. If $b < a$, then there is a minus sign coming from Δx.

There are two complementary points of view: We need an intuitive understanding about the signs; but at the same time, the rules of integration will *automatically* take care of the signs.

Problem 1
Determine the signs of the following integrals

$$\int_\pi^0 \sin x\,dx \quad , \quad \int_\infty^1 \frac{2}{x^2}\,dx \tag{8}$$

The second integral is related to the PE of an attractive inverse-square force. §

1.3 Additivity

Two additivity (or linearity) properties follow trivially from the definition.

Additivity in the integrand

$$\begin{aligned} &\int_a^b [\alpha_1 F_1(x) + \alpha_2 F_2(x)]\,dx \\ &= \alpha_1 \int_a^b F(x)\,dx + \alpha_2 \int_a^b F_2(x)\,dx \end{aligned} \tag{9}$$

Additivity in the interval

$$\int_a^c F(x)\,dx = \int_a^b F(x)\,dx + \int_b^c F(x)\,dx \tag{10}$$

which is valid for any a, b, c *without* requiring that $a < b < c$.

1.4 Indefinite integral

We are often interested in how the integral depends on the upper limit, holding the lower limit fixed, e.g.,

$$\int_r^x F(x')\,dx' \tag{11}$$

where r is any convenient reference point. The notation is often simplified.

First, if the reference point is changed from r to s, the difference is just the constant

$$C = \int_r^s F(x')\,dx' \tag{12}$$

So we can omit the lower limit, with the understanding that there is an undetermined additive constant C.

Second, when there is no danger of confusion, the dummy variable in the integrand is also written as x. Thus we have (see later if you do not know how to integrate yet)

$$\int^x x^2\,dx = (1/3)x^3 \tag{13}$$

and "plus constant" is understood. Such expressions are said to be *indefinite integrals*.

For a definite integral, we evaluate at the two limits and take the difference, e.g.,

$$\begin{aligned} & \int_a^b x^2\,dx \\ = & (1/3)x^3\Big|_a^b = (1/3)b^3 - (1/3)a^3 \end{aligned} \tag{14}$$

2 Numerical methods

The only general method to evaluate an integral is "chop and add", based on the definition. This section introduces several numerical methods with increasing sophistication; beginning students should focus on the first two. The accuracy is analyzed in terms of N, the number of intervals adopted.

We illustrate with the integral

$$I = \int_0^1 \frac{1}{1+x^2}\,dx \tag{15}$$

which (see next module) has the exact value

$$\frac{\pi}{4} = 0.785\,398\,138\,397\,448\ldots \tag{16}$$

All numerical algorithms and results are in the spreadsheet num.xlsx. Each method is shown on one sheet, with two versions (a, b) in which N differs by a factor of 2.

2.1 General formalism

Chop the range of integration (a, b) of length $L = b - a$ into N intervals. A typical interval is $(x-c, x+c)$, where x is the midpoint and $2c = \Delta x = L/N$ is the length, assumed small.

The integral over the interval is $\bar{F} \cdot (2c)$, where $\bar{F}$ is the true average of F:

$$\bar{F} = \frac{1}{2c}\int_{-c}^{c} F(x+\xi)\,d\xi \tag{17}$$

All methods are based on approximating $\bar{F}$ by a weighted average of F values in the interval

$$\tilde{F} = \sum_j w_j F(x + \alpha_j c) \tag{18}$$

where the points at α_j $(-1 \le \alpha_j \le 1)$ are to be sampled with weights w_j $(\sum_j w_j = 1)$.

2.2 Naive method

The naive method (based on the definition) is to evaluate F at one arbitrary point in each interval, say the leftmost (L) or rightmost (R) point (i.e., $\alpha = -1$ or $+1$); see Table 1 and sheets 1, 2 in num.xlsx. The error scales as N^{-1}, and it is easy to obtain accuracy at the few percent level for $N \sim 10$.

		N	value	FE
1a	L	10	0.8100	3.0×10^{-2}
1b	L	20	0.7978	1.6×10^{-2}
2a	R	10	0.7600	-3.3×10^{-2}
2b	R	20	0.7728	-1.6×10^{-2}

Table 1. Four different ways of using the naive method. FE is the fractional error.

2.3 Taking midpoint

It is better to evaluate F at the midpoint (M), in other words using $\alpha = 0$; results are in Table 2 and sheet 3 in num.xlsx. The error scales as N^{-2}. It is easy to obtain 4 figure accuracy.

		N	value	FE
3a	M	10	0.7856	2.6×10^{-4}
3b	M	20	0.7855	6.6×10^{-5}

Table 2. Two different evaluations using the midpoint. FE is the fractional error.

2.4 Two adjustable points per interval*

** This subsection is more advanced and can be skipped.*

We can do better with two adjustable points per interval (2AP). Let

$$\tilde{F} \quad = \quad \frac{1}{2}\left[F(x{-}\alpha c) + F(x{+}\alpha c)\right] \qquad (19)$$

in other words $w_1 = w_2 = 1/2$, $-\alpha_1 = \alpha_2 = \alpha$. It turns out that the best choice is $\alpha = 1/\sqrt{3}$, and the error decreases rapidly as N^{-4}. Results are in Table 3 and sheet 4 in num.xlsx. There is extremely high accuracy with only small N.[1]

		N	value	FE
4a	2AP	5	0.7854	9.0×10^{-9}
4b	2AP	10	0.7854	1.4×10^{-10}

Table 3. Two different evaluations using two adjustable points per interval. FE is the fractional error.

2.5 Three fixed points per interval*

** This subsection is more advanced and can be skipped.*

We can also use three fixed points (3FP). For an interval $(x-c, x+c)$, sample at $x-c, x, x+c$. The adjustable parameter is the weight. For symmetry and requiring the total weight to be unity, this means using

$$\tilde{F} \quad = \quad w[F(x{-}c) + F(x{+}c)] + (1{-}2w)F(x) \qquad (20)$$

It turns out that if $w = 1/6$, then the leading error will go as N^{-4}. This means the relative weights of the three terms are $1, 4, 1$.

Simpson's rule
Denote the three points of the first interval as 1L, 1M, 1R and those of the second interval as 2L, 2M, 2R etc. Then notice that 1R = 2L, so the number of function evaluations is $\sim 2N$ and not $\sim 3N$, i.e., not more than for two adjustable points, but with the advantage that the points are now regular.

In short, if we organize all the points (1L, 1M, 1R=2L, 2M, 2R=3L, ...) in a sequence with function values F_n, then the relative weights are $w_n = 1, 4, 2, 4, 2, 4, \ldots, 2, 4, 1$. (Consider the first entry of weight 2. This consists of 1 from 1R and 1 from 2L.) So *Simpson's rule* is

$$\int_a^b F(x)\,dx \quad = \quad \frac{\sum w_n F_n}{\sum w_n}\,(b-a) \qquad (21)$$

Results are in Table 4 and sheet 5 in num.xlsx. As before, there is extremely high accuracy even with small N.[2]

		N	value	FE
5a	3FP	5	0.7854	-1.3×10^{-8}
5b	3FP	10	0.7854	-2.0×10^{-10}

Table 4. Two different evaluations using three fixed points per interval. FE is the fractional error.

2.6 Summary

Table 5 summarizes the convergence properties of various methods.

p	where	accuracy
1	any	N^{-1}
1	midpoint	N^{-2}
2	adjustable	N^{-4}
3	fixed	N^{-4}

Table 5. Summary of different methods using p points per interval and their accuracy in terms of the number of intervals N.

Problem 1
How would you formulate an algorithm with 3 adjustable points per interval and how would the error depend on N? (Do this after you study the Appendix.) §

Improving efficiency
The spreadsheet shows the logic but has not been optimized for efficiency.

[1] In this example, there is an improvement by a factor of 64 when N is doubled, whereas we expect a factor of only $2^4 = 16$. This phenomenon, as well as the fantastic accuracy, is an accident in the present example, where the N^{-4} error term very nearly vanishes, leaving the next term N^{-6}. See the Problem in the Appendix.

[2] Again, an accident in this example eliminates the N^{-4} term, making the error go as N^{-6}.

- Each Δx is evaluated by subtracting two neighbouring values of x. But for uniform intervals, we can simply refer to a constant value, and save many subtractions.
- If the function values encountered are F_i, then the algorithm in the spreadsheet is, say, $\sum_i (F_i \,\Delta x)$, but it is better to change the algorithm to $(\sum_i F_i)\,\Delta x$, which saves many multiplications.

These improvements become important when N is very large.

Using the numerical method
Until you learn other software packages (e.g., Mathematica, MATLAB, FORTRAN, C++), you may like to use the spreadsheet provided as a template for evaluating integrals numerically. All you need to do is to replace (and then copy) the formula for evaluating F.

Problem 2
Using any of the numerical schemes, evaluate

$$\int_0^{1/2} \frac{dx}{\sqrt{1-x^2}} \tag{22}$$

to 4 digits. §

Problem 3
Likewise evaluate

$$G(\xi) \quad = \quad \int_0^{2\pi} \frac{d\theta}{(1+2\xi\cos\theta+\xi^2)^{1/2}} \tag{23}$$

for $\xi = 0.3, 0.5, 0.7$. This integral arises in determining the rate of precession of a planetary orbit (e.g., that of Mercury) due to other planets.[3]

3 Fundamental theorem of calculus

The fundamental theorem of calculus states that *integration is the inverse of differentiation.*

3.1 Proof

Let

$$\Phi(x) \quad \equiv \quad \int_0^x F(x)\,dx \tag{24}$$

Then

$$\begin{aligned} & \Phi(x+\Delta x) - \Phi(x) \\ = \; & \int_a^{x+\Delta x} F(x)\,dx - \int_a^x F(x)\,dx \\ = \; & \int_x^{x+\Delta x} F(x)\,dx = F(x)\,\Delta x \end{aligned} \tag{25}$$

Re-arranging terms

$$\begin{aligned} F(x) &= \frac{\Phi(x+\Delta x) - \Phi(x)}{\Delta x} \\ &= \frac{d\Phi(x)}{dx} \end{aligned} \tag{26}$$

In all the above, Δx is supposed to be infinitesimal.

3.2 Two statements

The above statement can be written as

$$\frac{d}{dx}\int_a^x F(x)\,dx \quad = \quad F(x) \tag{27}$$

or schematically

$$\frac{d}{dx}\int^x (\ldots)\,dx \quad = \quad (\ldots) \tag{28}$$

Or we can write it as

$$\int^x F(x)\,dx \quad = \quad \Phi(x) \tag{29}$$

and inserting (26) into this we get

$$\int^x \frac{d}{dx}\Phi(x)\,dx \quad = \quad \Phi(x) \tag{30}$$

or schematically

$$\int^x \frac{d}{dx}(\ldots)\,dx \quad = \quad (\ldots) \tag{31}$$

In short

$$\frac{d}{dx}\int = \int \frac{d}{dx} = I \tag{32}$$

where I is the identity operation.

In fact, the notation of calculus makes all this intuitive. Start from (30):

$$\begin{aligned} & \int \frac{d}{dx}\Phi(x)\,dx \qquad \text{[Cancel } dx\text{]} \\ = \; & \int d\Phi \qquad \text{[Cancel } \textstyle\int \text{ against } d\text{]} \\ = \; & \Phi \end{aligned} \tag{33}$$

[3]See KH Lo, K Young and BYP Lee, "Advance of Perihelion", Am. J. Phys., 81, 695 (2013). doi 10.1119/1.4813067.

3.3 Applications

Using this inverse property, we immediately get the following results:

$$\begin{aligned}\int^x x^n\,dx &= \frac{1}{n+1}x^{n+1} \quad , \quad n \neq -1\\ \int^x \cos x\,dx &= \sin x\\ \int^x \sin x\,dx &= -\cos x\\ \int^x e^x\,dx &= e^x \end{aligned} \tag{34}$$

To prove these, just differentiate the RHS. In the first example, n does not have to be an integer.

A special case
In the integration of a power, the case of $1/x$ is excluded. We shall deal with this in the next module through change of variables. But it is interesting to present here another (and less conventional) derivation which emphasizes continuity with the other "normal" cases.

Consider, for $n \neq 0$ but small

$$\begin{aligned}&\int_u^v x^{n-1}\,dx = \frac{1}{n}\,x^n\big|_u^v\\ &= \frac{1}{n}\left(v^n - u^n\right) = \frac{1}{n}\left(e^{n\ln v} - e^{n\ln u}\right)\\ &= \frac{1}{n}\left[(1 + n\ln v + \ldots) - (1 + n\ln u + \ldots)\right]\\ &= \ln v - \ln u + O(n)\end{aligned} \tag{35}$$

in which the exponentials have been expanded in powers of n. Taking $n \to 0$ gives

$$\int_u^v x^{-1}\,dx = \ln v - \ln u \tag{36}$$

or for the indefinite integral

$$\int^x x^{-1}\,dx = \ln x \tag{37}$$

In fact, we can say that,

$$\int^x x^{n-1}\,dx = C(n) + \ln x + O(n^2) \tag{38}$$

where, as $n \to 0$, the only singular part is in $C(n)$ — which does not matter. This derivation shows that $\ln x$ is "like" $(1/0)x^0$. This technique of isolating an infinity in one place where it does not matter is important in the theory of renormalization in quantum field theory.

Turning (37) around, we have

$$\frac{d}{dx}\ln x = \frac{1}{x} \tag{39}$$

We then also have the result

$$\int^x \ln x\,dx = x\ln x - x \tag{40}$$

easily verified by differentiating the RHS.

Appendix

A Derivation of numerical algorithms

The true average of F is

$$\begin{aligned}\bar{F} &= \frac{1}{2c}\int_{-c}^{c} F(x+\xi)\,d\xi\\ &= \frac{1}{2c}\int_{-c}^{c}\left[\sum_{n=0}^{\infty}\frac{1}{n!}F^{(n)}\xi^n\right]d\xi\\ &= \sum_{n=\text{even}}^{\infty}\frac{1}{(n+1)!}F^{(n)}c^n\end{aligned} \tag{41}$$

where F is expanded in a Taylor series, $F^{(n)}$ is the nth derivative evaluated at x, and the extra factor of $1/(n+1)$ comes from the integral $\int \xi^n\,d\xi$.

On the other hand

$$\begin{aligned}\tilde{F} &= \sum_j w_j F(x+\alpha_j c)\\ &= \sum_j w_j\left[\sum_n \frac{1}{n!}F^{(n)}(\alpha_j c)^n\right]\end{aligned} \tag{42}$$

which, upon reversing the order of summation, gives

$$\begin{aligned}\tilde{F} &= \sum_n^{\infty}\frac{D_n}{n!}F^{(n)}c^n\\ D_n &= \sum_j w_j\alpha_j^n\end{aligned} \tag{43}$$

The terms with odd n will be zero if the points and weights are chosen symmetrically.

It is obvious that (a) D_0 gives the correct value of 1 if the weights add up to unity; (b) $D_1 = 0$ if the points and weights are symmetrical. The two advanced methods (2AP, 3FP) both use one adjustable parameter to ensure $D_2 = 1/3$, namely

$$\sum_j w_j \alpha_j^2 = \frac{1}{3} \tag{44}$$

If $\tilde{F}$ gives correctly the c^2 term and the c^3 term vanishes, the leading error will go as $c^4 \propto N^{-4}$.

Two adjustable points

The choice in (19) is $w_1 = w_2 = 1/2$, $-\alpha_1 = \alpha_2 = \alpha$. When placed into (44), this gives the condition

$$\alpha = \frac{1}{\sqrt{3}} \tag{45}$$

Three fixed points

The choice in (20) is $w_1 = w_2 = w$, $w_3 = 1 - 2w$; $\alpha_1 = -1$, $\alpha_2 = 1$, $\alpha_3 = 0$. Putting this into (44) gives

$$w + w + 0 = \frac{1}{3} \tag{46}$$

or $w = 1/6$. In other words, the weights to be attached to the points $(x-c, 0, x+c)$ (or L, M, R) are 1/6, 4/6, 1/6, or in relative terms $1, 4, 1$.

Problem 4

Both methods above match the $F^{(2)}c^2$ term, so the leading error should be $\propto F^{(4)}c^4 \propto N^{-4}$. Yet the errors in Table 3 and Table 4 go as N^{-6}. Why?

Let the integral be divided into N intervals centered at x_j, each of length $2c$. The leading term in the error is proportional to

$$\begin{aligned} & \sum_j F^{(4)}(x_j) \approx \frac{1}{2c} \int_a^b F^{(4)}(x)\, dx \\ = & \frac{1}{2c} \left[F^{(3)}(b) - F^{(3)}(a) \right] \end{aligned} \tag{47}$$

Show that the square bracket on the RHS in (47) is zero in the particular example. Therefore the leading error will be the *next* term, which goes as N^{-6}. §

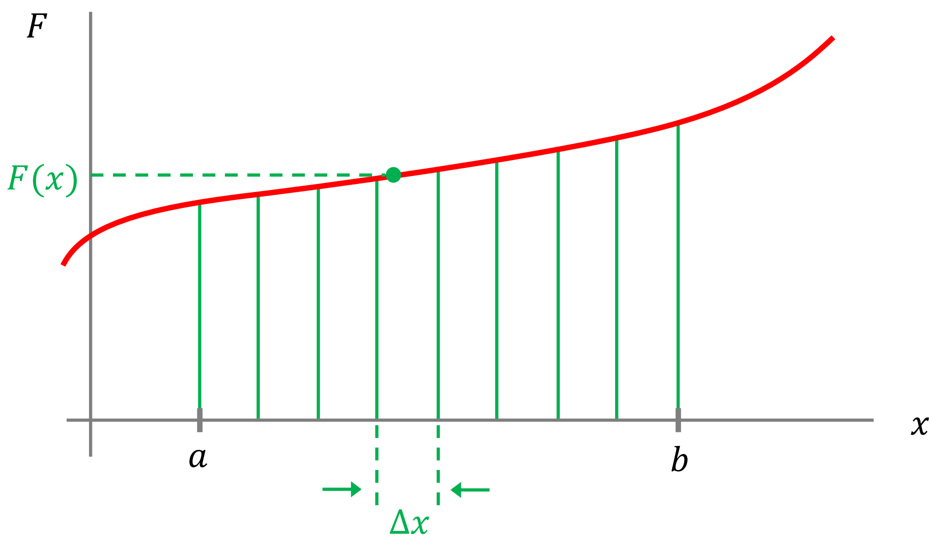

Figure 1

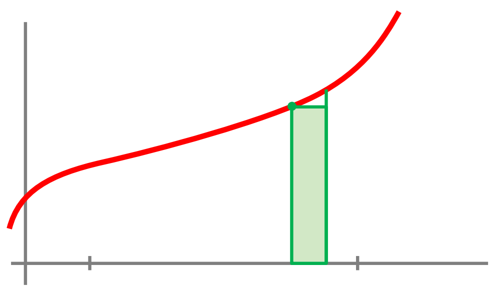

Figure 2a

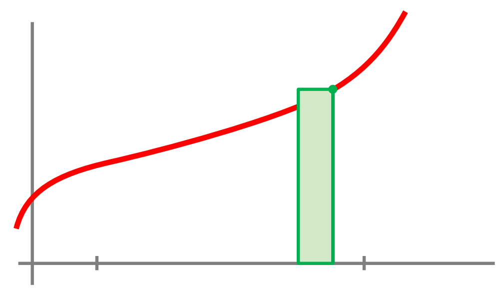

Figure 2b

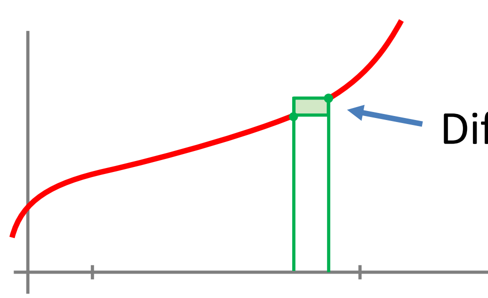

Figure 2c

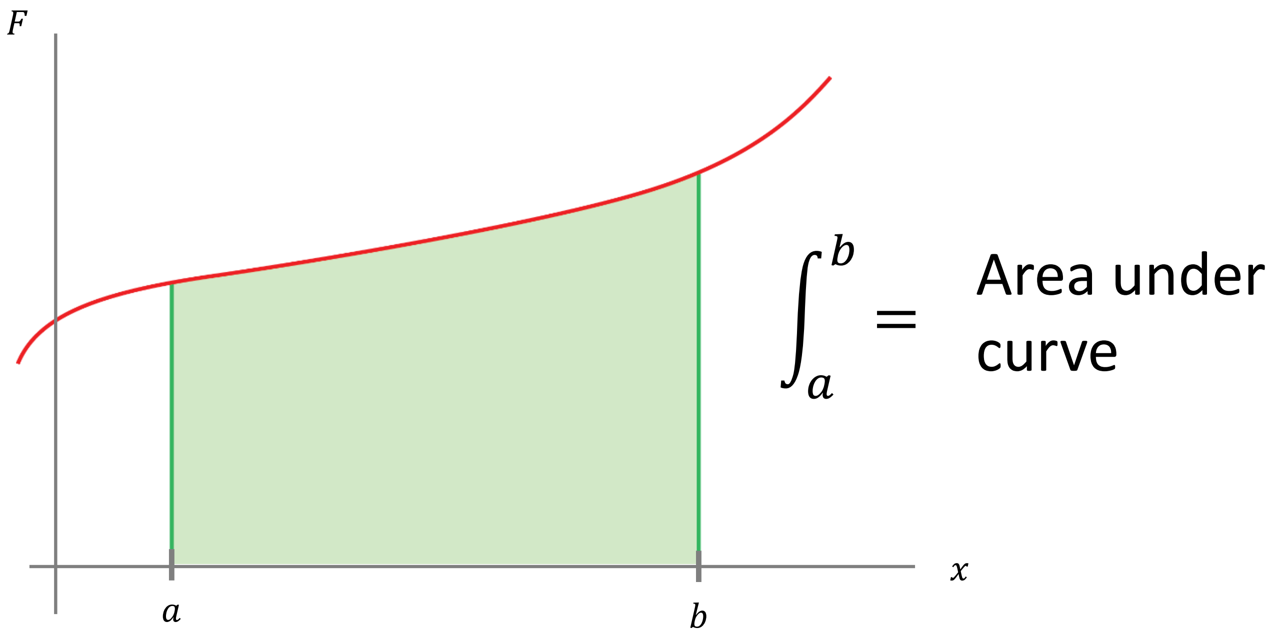

Figure 3

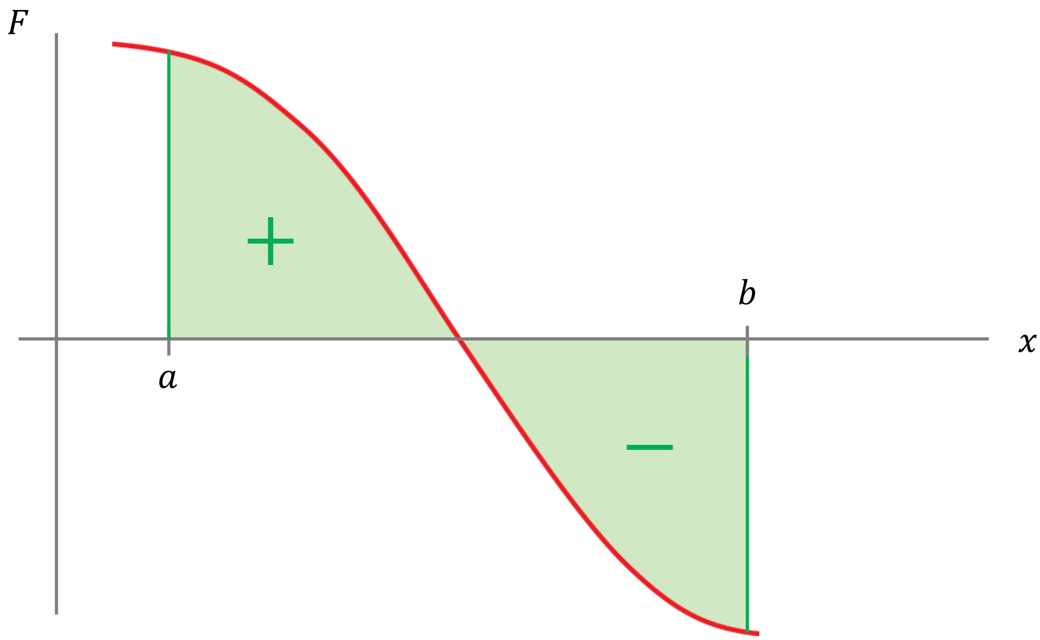

Figure 4

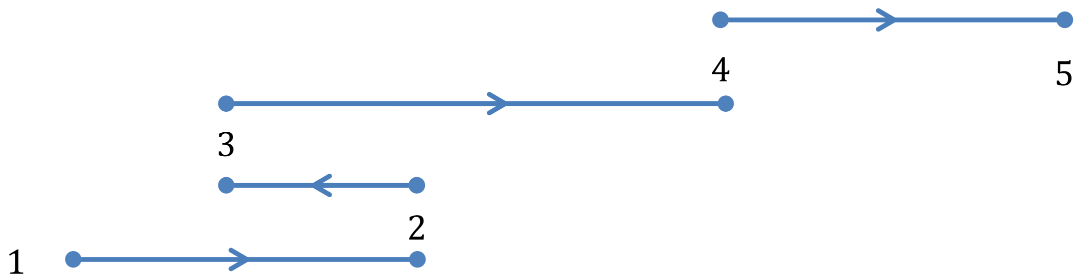

Figure 5