LIMITS AND INTEGRABILITY: BASIC CONCEPTS AND GENERAL RULES

by David Levermore

20 October 1999

This is a survey of some basic facts about integration that will be covered on our third exam. It supplements the material covered in the book (Chapter 3 and Sections 7.5 and 7.6) and the class lectures. It covers the definite integral as a limit of Riemann sums, its interpretations in terms of area and averages, the Fundamental Theorem of Calculus, its interpretation in terms of total change, general properties of the definite integral, and numerical integration.

1. THE DEFINITE INTEGRAL

1.1: Integrability. Let f be a function defined over an interval $[a, b]$. The definite integral of f over $[a, b]$ is defined to be the limit of Riemann sums. These sums are constructed by first dividing $[a, b]$ into n subintervals with points x_k that are ordered so that

$$a = x_0 < x_1 < \cdots < x_{n-1} < x_n = b\,.$$

The k^{th} subinterval is then $[x_{k-1}, x_k]$ and its length is given by

$$\Delta x_k \equiv x_k - x_{k-1}\,. \tag{1.1}$$

A so-called general Riemann sum associated with these subintervals has the form

$$\sum_{k=1}^{n} f(p_k)\,\Delta x_k\,, \quad \text{where } p_k \text{ is some point in } [x_{k-1}, x_k]. \tag{1.2}$$

If f is not too badly behaved, it is reasonable to expect that if one lets n get larger and larger while choosing the points x_k in such a way so as to make each Δx_k get smaller and smaller, then no matter how the points p_k are chosen, the Riemann sums will always converge to the same number. When this is the case, one says f is **integrable** over the interval $[a, b]$ and the limiting number is called the **definite integral** of f over $[a, b]$ and is denoted by

$$\int_a^b f(x)\,dx\,. \tag{1.3}$$

An important fact that you should know, which was given without proof, is the following.

The Integrability Theorem. If either f is monotonic over $[a, b]$, or f is bounded over $[a, b]$ and is continuous at all but a finite number of points in $[a, b]$ then it is integrable over $[a, b]$. In particular, if f is continuous over $[a, b]$ then it is integrable over $[a, b]$.

The notations '$\int$' and 'dx' that appear in (1.3) parallel the notations '$\sum$' and 'Δx' that appear in the approximating Riemann sums (1.2). The symbol '$\int$' is an old fashion letter 's' and reflects the fact the definite integral arises as the limit of sums. The symbol 'dx' is not just an ornament. As we will see later, it can be thought of as an infinitesimally small Δx. In applications where f and x each have units of measure, the units of each Riemann sum will be the units of $f(p_k)\,\Delta x_k$ — namely, the product of the units of f times the units of x. The limiting definite integral will therefore have these same units. By assigning units of x to the dx in (1.3), the units of the definite integral are made clear.

You should remember that the value of the definite integral (1.3) is a *number* that depends only on f, which is called the **integrand**, and a and b, which are respectively called the lower and upper **endpoints** or **limits of integration** . It *does not* depend x, which is called the **variable of integration**. Thereby, one has that

$$\int_a^b f(x)\,dx = \int_a^b f(z)\,dz = \int_a^b f(t)\,dt = \cdots .$$

The variable of integration is therefore sometimes referred to as a "dummy" variable.

1.2: Definite Integrals of Monotonic Functions. If $[a, b]$ is divided into n uniform subintervals, the length of each subinterval is then given by

$$\Delta x \equiv \frac{b-a}{n} . \tag{1.4}$$

The k^{th} subinterval is then $[x_{k-1}, x_k]$ where the points x_k are given by the formula

$$x_k \equiv a + k\Delta x\,, \quad \text{for } k = 0, 1, \cdots, n. \tag{1.5}$$

The so-called **right-hand sum** corresponds to the choice $p_k = x_k$ in (1.2) because x_k is the right-hand endpoint of $[x_{k-1}, x_k]$. It is denoted by RIGHT_n. The so-called **left-hand sum** corresponds to the choice $p_k = x_{k-1}$ in (1.2) because x_{k-1} is the left-hand endpoint of $[x_{k-1}, x_k]$. It is denoted by LEFT_n.

It should be clear from the graph of f that if f is increasing over $[a, b]$ then

$$\mathrm{LEFT}_n \le \int_a^b f(x)\,dx \le \mathrm{RIGHT}_n\,,$$

while if f is decreasing over $[a, b]$ then

$$\mathrm{RIGHT}_n \le \int_a^b f(x)\,dx \le \mathrm{LEFT}_n\,.$$

However, one can show (can you?) that

$$\mathrm{RIGHT}_n - \mathrm{LEFT}_n = \big(f(b) - f(a)\big)\,\Delta x\,.$$

From (1.4) one sees that Δx goes to zero as n goes to infinity. Hence, both the right-hand and left-hand sums must converge to the definite integral of f.

1.3: The Definite Integral and Area. When a function f is positive over an interval $[a, b]$ then the definite integral of f over $[a, b]$ may be interpreted as giving the area of the region enclosed by the vertical lines $x = a$ and $x = b$, the x-axis $y = 0$, and the curve $y = f(x)$. Roughly speaking, one can say that

$$\int_a^b f(x)\,dx = \text{the area of the region below } f \text{ and over } [a, b]. \tag{1.6}$$

This interpretation can be used to evaluate the definite integral when the integrand describes a geometric region for which you know how to compute the area.

Example; One sees that

$$\int_0^r \sqrt{r^2 - x^2}\,dx = \frac{1}{4}\pi r^2\,,$$

because the region under the curve $y = \sqrt{r^2 - x^2}$ over $[0, r]$ is one quarter of the disk centered at the origin of radius r.

Example; One sees that

$$\int_a^b (mx + k)\,dx = \frac{(ma + k) + (mb + k)}{2}(b - a) = m\,\frac{b^2 - a^2}{2} + k(b - a)\,,$$

because the region under the curve $y = mx + k$ over $[a, b]$ is a trapezoid with base $(b - a)$ and heights $(ma + k)$ and $(mb + k)$.

In the above examples the regions described by the integrals were simple geometric shapes for which you know formulas for the area. The same approach can be used when the region described by the integral can be decomposed into several such simple geometric shapes.

Example: You can use this approach to show that

$$\int_0^{\sqrt{3}} \sqrt{4 - x^2}\,dx = \frac{2\pi}{3} + \frac{\sqrt{3}}{2}\,,$$

by decomposing the region into a pie slice and a triangle. More generally, you can use a similar decomposition to show for $|b| \leq r$ that

$$\int_0^b \sqrt{r^2 - x^2}\,dx = \frac{r^2}{2}\sin^{-1}(b/r) + \frac{b}{2}\sqrt{r^2 - b^2}\,.$$

It is clear that only very few definite integrals can be evaluated by this approach.

When a function f takes both positive and negative values over an interval $[a,b]$ then the definite integral of f over $[a,b]$ may be interpreted as giving the so-called signed area of f over $[a,b]$. Specifically, if one considers the region enclosed by the vertical lines $x=a$ and $x=b$, the x-axis $y=0$, and the curve $y=f(x)$, then

$$\begin{aligned}\int_a^b f(x)\,dx = &\text{ the area of the part of the region that lies above the } x\text{-axis} \\ &- \text{the area of the part of the region that lies below the } x\text{-axis.}\end{aligned} \tag{1.7}$$

This interpretation can also be used to evaluate the definite integral when the integrand describes a geometric region that can be decomposed into simple geometric shapes for which you know how to compute the area.

1.4: The Definite Integral and Average Value. The definite integral of a function f over an interval $[a,b]$ may also be interpreted in terms of the average value of f over that interval. Specifically, one has that

$$\text{the average value of } f \text{ over } [a,b] = \frac{1}{b-a}\int_a^b f(x)\,dx\,. \tag{1.8}$$

The motivation for this interpretation is as follows. Suppose that $[a,b]$ is divided into n uniform subintervals. The length of each subinterval is therefore $\Delta x = (b-a)/n$. Suppose that we select a point p_k from the k^{th} subinterval for each of the n subintervals. The n points $\{p_1, p_2, \cdots, p_n\}$ are thereby uniformly distributed throughout $[a,b]$. Now it is clear from our usual understanding of averages that

$$\text{the average of the values } \{f(p_1), f(p_2), \cdots, f(p_n)\} = \frac{1}{n}\sum_{k=1}^{n} f(p_k)\,.$$

But because $\Delta x = (b-a)/n$, this expression for the average value of f over the points p_k can be recast in terms of a Riemann sum as

$$\frac{1}{n}\sum_{k=1}^{n} f(p_k) = \frac{1}{b-a}\sum_{k=1}^{n} f(p_k)\,\Delta x\,.$$

As n get larger and larger, then no matter how the points p_k are selected, these expressions for average values of f will therefore converge to the right-hand side of (1.8).

Geometrically, (1.8) states that the average value of f over $[a,b]$ is the height of the rectangle whose base is $[a,b]$ (which has a width of $b-a$) and whose signed area equals the signed area of the region enclosed by the vertical lines $x=a$ and $x=b$, the x-axis $y=0$, and the curve $y=f(x)$. You should understand this interpretation well enough to recognize when to apply it in a word problem.

2. THE FUNDAMENTAL THEOREM OF CALCULUS

The most important theorem in calculus relates the concept of the derivative to that of the definite integral. It is the following.

The Fundamental Theorem of Calculus. If F is any differentiable function whose derivative F' is integrable over $[a,b]$ then

$$F(b) - F(a) \equiv F(x)\Big|_a^b = \int_a^b F'(x)\,dx\,. \tag{2.1}$$

In particular, (2.1) holds whenever F is continuously differentiable over $[a,b]$ (which means that F is differentiable and F' is continuous (hence, integrable) over $[a,b]$.)

Remark: The requirement that F' be integrable is necessary. Consider F defined by

$$F(x) = \begin{cases} x^2 \cos\left(\dfrac{1}{x^2}\right) & \text{for } x \neq 0\,, \\ 0 & \text{for } x = 0\,. \end{cases}$$

This function is differentiable everywhere with F' given by

$$F'(x) = \begin{cases} \dfrac{2}{x} \sin\left(\dfrac{1}{x^2}\right) + 2x \cos\left(\dfrac{1}{x^2}\right) & \text{for } x \neq 0\,, \\ 0 & \text{for } x = 0\,. \end{cases}$$

Because F' is unbounded near $x = 0$, its definite integral on the right-hand side of (2.1) does not exist (in the sense defined in Section 1.1) over any interval containing $x = 0$.

2.1: The Definite Integral Gives Total Change. The Fundamental Theorem of Calculus may be interpreted as stating that the **total change** of a quantity $F(x)$ between $x = a$ and $x = b$ is given by the definite integral from a to b of $F'(x)$, its rate of change with respect to x. For example, if $s(t)$ gives the position of an object as a continuously differentiable function of time t over $[a,b]$ then its velocity is given by $v(t) = s'(t)$, and the total distance traveled between $t = a$ and $t = b$ is given by

$$s(b) - s(a) = \int_a^b v(t)\,dt\,.$$

More generally, if F is any continuously differentiable function of time t over $[a,b]$ then the instantaneous rate of change of $F(t)$ at time t is given by $F'(t)$, and the total change of $F(t)$ between $t = a$ and $t = b$ is given by

$$F(b) - F(a) = \int_a^b F'(t)\,dt\,.$$

You should understand this interpretation of the Fundamental Theorem well enough to recognize when to apply it in a word problem.

2.2: The Evaluation of Definite Integrals. The Fundamental Theorem of Calculus yields the most powerful tool with which to evaluate definite integrals. It implies that if f is integrable over $[a, b]$ and **IF** you know a function F such that $F' = f$ then

$$\int_a^b f(x)\,dx = F(x)\Big|_a^b = F(b) - F(a)\,. \tag{2.2}$$

The 'big **IF**' in the application of this theorem for a given integrand f is in finding a function F such that $F' = f$. Such a function F is called a **primitive** or **antiderivative** of f. Correspondingly, finding such a function for a given f is called **integration** or **antidifferentiation** of f.

Given a simple enough analytic expression defining a function, you should be able to find a primitive of the function by inspection using your knowledge of derivatives. For example, the functions

$$g(t) = t^3 + 2t\,, \qquad w(z) = (z-1)^2\,, \qquad f(x) = \sin(3x)\,,$$

respectively have primitives

$$G(t) = \frac{1}{4}t^4 + t^2\,, \qquad W(z) = \frac{1}{3}(z-1)^3\,, \qquad F(x) = -\frac{1}{3}\cos(3x)\,.$$

Therefore any definite integral with the above g, w, or f as integrands may be easily evaluated using the Fundamental Theorem. Examples that make use of the above primitives are

$$\int_0^2 t^3 + 2t\,dt = \left(\frac{1}{4}t^4 + t^2\right)\Big|_0^2 = (4+4) - (0+0) = 8\,,$$

$$\int_{-2}^2 (z-1)^2\,dz = \frac{1}{3}(z-1)^3\Big|_{-2}^2 = \frac{1}{3} - \left(-\frac{27}{3}\right) = \frac{28}{3}\,,$$

$$\int_0^{\frac{\pi}{6}} \sin(3x)\,dx = -\frac{1}{3}\cos(3x)\Big|_0^{\frac{\pi}{6}} = -\frac{1}{3}\cos(\tfrac{\pi}{2}) + \frac{1}{3}\cos(0) = 0 + \frac{1}{3} = \frac{1}{3}\,.$$

Notice that there are plenty of chances to make sign errors in such calculations. However, if you take a moment to visualize the integrand, the correct sign of the definite integral might be obvious. For example, do you see that in each of the examples above the sign of the integrand is positive over the interval of integration?

3. OTHER PROPERTIES OF THE DEFINITE INTEGRAL

The definite integral has many general properties with which you should become familiar, some of which we will review in this section. You should try to understand each of these properties at least three ways: graphically through areas, analytically through the Fundamental Theorem of Calculus, and numerically through approximating sums.

3.1: Endpoints of Integration. Let f be any function that is integrable over the intervals indicated. Then

$$\begin{aligned} \int_a^b f(x)\,dx &= -\int_b^a f(x)\,dx\,, \\ \int_a^c f(x)\,dx &= \int_a^b f(x)\,dx + \int_b^c f(x)\,dx\,. \end{aligned} \tag{3.1}$$

These are expressed in words as "exchanging the endpoints of integration changes the sign of the integral" and "the integral over an interval that is divided into subintervals is the sum of the integrals over those subintervals" respectively. They hold no matter how a, b and c are ordered.

3.2: Linear Combinations of Integrands. Given any two functions f and g that are integrable over an interval $[a, b]$, and a constant k, the functions kf and $f + g$ are also integrable over an interval $[a, b]$, and their definite integrals are given by

$$\begin{aligned} \int_a^b k\,f(x)\,dx &= k\int_a^b f(x)\,dx\,, \\ \int_a^b f(x) + g(x)\,dx &= \int_a^b f(x)\,dx + \int_a^b g(x)\,dx\,. \end{aligned} \tag{3.2}$$

These are expressed in words as "the integral of a multiple is the multiple of the integral" and "the integral of a sum is the sum of the integrals" respectively.

Recall that the linear combinations of n given functions $\{f_1, f_2, \cdots, f_n\}$ are all those functions of the form $k_1 f_1 + k_2 f_2 + \cdots + k_n f_n$ for some choice of n constants $\{k_1, k_2, \cdots, k_n\}$. In other words, the linear combinations are all those function that can be built up from the given functions $\{f_1, f_2, \cdots, f_n\}$ by repeated multiplication by constants and addition. If each of the given functions $\{f_1, f_2, \cdots, f_n\}$ is integrable over an interval $[a, b]$ then repeated applications of the multiplication and sum rules (3.2) show that each such linear combination is also integrable over an interval $[a, b]$ and its definite integral over $[a, b]$ is given by

$$\begin{aligned}&\int_a^b \big(k_1 f_1(x) + k_2 f_2(x) + \cdots + k_n f_n(x)\big)\,dx \\ &= k_1 \int_a^b f_1(x)\,dx + k_2 \int_a^b f_2(x)\,dx + \cdots + k_n \int_a^b f_n(x)\,dx\,.\end{aligned} \tag{3.3}$$

This is expressed in words as "the integral of a linear combination is the linear combination of the integrals". It is important to understand that this rule need not be memorized because all it does is embody repeated applications of (3.2). That is to say, if you have truly mastered (3.2) then (3.3) will seem obvious to you and will not need to be memorized.

Given any two functions f and g that are integrable over an interval $[a, b]$, a particular instance of the linear combination rule (3.3) is

$$\int_a^b f(x) - g(x)\,dx = \int_a^b f(x)\,dx - \int_a^b g(x)\,dx\,. \tag{3.4}$$

This is expressed in words as "the integral of a difference is the difference of the integrals". As an instance of (3.3), this rule also should seem obvious based on a mastery of (3.2).

3.3: Even and Odd Symmetries. If f is an even function over a symmetric interval $[-b, b]$, and if f is integrable over $[0, b]$, then f is integrable over $[-b, 0]$ with

$$\int_{-b}^0 f(x)\,dx = \int_0^b f(x)\,dx\,, \tag{3.5}$$

and f is integrable over $[-b, b]$ with

$$\int_{-b}^b f(x)\,dx = 2\int_0^b f(x)\,dx\,. \tag{3.6}$$

Relations (3.5) and (3.6) should be evident to you from the symmetry of the graph of f about the vertical axis.

If f is an odd function over a symmetric interval $[-b, b]$, and if f is integrable over $[0, b]$, then f is integrable over $[-b, 0]$ with

$$\int_{-b}^0 f(x)\,dx = -\int_0^b f(x)\,dx\,, \tag{3.7}$$

and f is integrable over $[-b, b]$ with

$$\int_{-b}^b f(x)\,dx = 0\,. \tag{3.8}$$

Relations (3.7) and (3.8) should be evident to you from the symmetry of the graph of f about the origin.

Relation (3.8) is particularly useful because it allows you to evaluate some definite integrals that can not be evaluated any other way. For example, one sees that

$$\int_{-1}^{1} \tan(x^5)\, dx = 0\,,$$

because the integrand is odd and the interval of integration is symmetric about the origin. Even in cases that can be evaluated by another method, this approach is quicker if it applies. For example,

$$\int_{-\frac{\pi}{6}}^{\frac{\pi}{6}} \tan^3(x)\, dx = 0\,,$$

because the integrand is odd and the interval of integration is symmetric about the origin.

3.4: Periodic and Antiperiodic Symmetries. If f has period p, and if f is integrable over $[a, b]$, then f is integrable over $[a+p, b+p]$ with

$$\int_{a+p}^{b+p} f(x)\, dx = \int_{a}^{b} f(x)\, dx\,. \tag{3.9}$$

If moreover f is integrable over any interval of length p, say $[a, a+p]$, then it is integrable over every interval of length p with

$$\int_{b}^{b+p} f(x)\, dx = \int_{a}^{a+p} f(x)\, dx\,, \tag{3.10}$$

for every b. Relations (3.9) and (3.10) should be evident to you from the periodic symmetry of the graph of f.

If f has antiperiod p, and if f is integrable over $[a, b]$, then f is integrable over $[a+p, b+p]$ with

$$\int_{a+p}^{b+p} f(x)\, dx = -\int_{a}^{b} f(x)\, dx\,. \tag{3.11}$$

If moreover f is integrable over $[a, b+p]$ where $a < b$ then

$$\int_{a}^{b+p} f(x)\, dx = \int_{b}^{a+p} f(x)\, dx\,. \tag{3.12}$$

By setting $b = a + p$ in (3.12), one finds that if f is integrable over any interval of length p then it is integrable over every interval of length $2p$ with

$$\int_{a}^{a+2p} f(x)\, dx = 0\,. \tag{3.13}$$

In other words, if f has antiperiod p and it is integrated over any interval of length twice p, the integral is zero. Relations (3.11), (3.12) and (3.13) should be evident to you from the antiperiodic symmetry of the graph of f.

Relation (3.13) is particularly useful because it allows you to easily evaluate some definite integrals that are hard to evaluate any other way. For example, one sees that

$$\int_0^{4\pi} \sin(\cos(x))\, dx = 0\,,$$

because the integrand has antiperiod π and the integration is over an interval whose length is an even multiple of π. Even in cases that can be evaluated by another method, this approach is quicker if it applies. For example,

$$\int_{-\pi}^{\pi} \cos^5(x)\, dx = 0\,,$$

because the integrand has antiperiod π and the integration is over an interval whose length is an even multiple of π.

3.5: Comparison. Let f and g be functions that are integrable over an interval $[a, b]$. If $f \leq g$ over $[a, b]$ then

$$\int_a^b f(x)\, dx \leq \int_a^b g(x)\, dx\,.$$

If $f < g$ over (a, b) then

$$\int_a^b f(x)\, dx < \int_a^b g(x)\, dx\,.$$

The way that these are most often used is in instances where one can evaluate the integral for one of the functions, where one obtains a bound on the other. For example, because one sees that $e^{x^2} < e^4$ over $(0, 2)$, one sees that

$$\int_0^2 e^{x^2}\, dx < \int_0^2 e^4\, dx = 2e^4\,.$$

This upper bound is very crude. You will see that if you evaluate the integral on the left numerically. Similarly, because one sees that $1 < e^{x^2}$ over $(0, 2)$, one sees that

$$2 = \int_0^2 1\, dx < \int_0^2 e^{x^2}\, dx\,.$$

This lower bound is also very crude.

4. NUMERICAL INTEGRATION

There are definite integrals for which no exact value is known. In such a case one must resort to approximating the value of the integral by so-called **numerical integration** or **quadrature** methods. Most calculators now have a routine that approximately evaluates

$$\int_a^b f(x)\,dx$$

for any given an integrand f and endpoints of integration a and b. The methods they use are advanced versions of the ones we study.

All the methods we will study divide the interval $[a, b]$ into n uniform subintervals. The length Δx of each subinterval is given by

$$\Delta x \equiv \frac{b-a}{n}. \tag{4.1}$$

The k^{th} subinterval is then $[x_{k-1}, x_k]$ where

$$a = x_0 < x_1 < \cdots < x_{n-1} < x_n = b,$$

with x_k given by the formula

$$x_k \equiv a + k\Delta x. \tag{4.2}$$

Most basic numerical integration methods associated with these subintervals are built-up from one or more Riemann sums of the form

$$\left(\sum_{k=1}^{n} f(p_k)\right)\Delta x, \quad \text{where } p_k \text{ is some point in } [x_{k-1}, x_k]. \tag{4.3}$$

When f is positive such a sum approximates the area under the curve $y = f(x)$ over the k^{th} subinterval by the area of a rectangle of height $f(p_k)$. Given an interval $[a, b]$ and a number of subintervals n, you should be able to compute Δx and the points x_k using (4.1) and (4.2). Given moreover an integrand f and a rule for choosing the points p_k, you should be able to set up a Riemann sum of the form (4.3).

1: The Left-Hand and Right-Hand Rules. The so-called **left-hand rule** is denoted by LEFT_n and corresponds to the Riemann sum (4.3) with the choice $p_k = x_{k-1}$, which is the left-hand endpoint of the k^{th} subinterval. The so-called **right-hand rule** is denoted by RIGHT_n and corresponds to the Riemann sum (4.3) with the choice $p_k = x_k$, which is the right-hand endpoint of the k^{th} subinterval. These rules are related by the identity

$$\text{RIGHT}_n - \text{LEFT}_n = \big(f(b) - f(a)\big)\,\Delta x. \tag{4.4}$$

Can you see this both graphically and analytically?

The left-hand and right-hand rules are clearly both exact for constant functions. It is also clear that if f is increasing over $[a, b]$ then the left-hand rule gives an underestimate while the right-hand rule gives an overestimate:

$$\mathrm{LEFT}_n \leq \int_a^b f(x)\,dx \leq \mathrm{RIGHT}_n\,.$$

Similarly, if f is decreasing over $[a, b]$ then the right-hand rule gives an underestimate while the left-hand rule gives an overestimate:

$$\mathrm{RIGHT}_n \leq \int_a^b f(x)\,dx \leq \mathrm{LEFT}_n\,.$$

Hence, if f is either increasing over $[a, b]$ or decreasing over $[a, b]$ then the right-hand and left-hand rules are each accurate to within $|\mathrm{RIGHT}_n - \mathrm{LEFT}_n|$. But by (4.4), we know that

$$|\mathrm{RIGHT}_n - \mathrm{LEFT}_n| = |f(b) - f(a)|\,\Delta x\,.$$

Hence, if f is either increasing over $[a, b]$ or decreasing over $[a, b]$ then the error ERROR_n made by either the left-hand or right-hand rule satisfies

$$|\mathrm{ERROR}_n| \leq |f(b) - f(a)|\,\Delta x\,. \tag{4.5}$$

This upper bound for the size of the error can be made as small as you wish by picking n large enough. It decreases like $1/n$ as n increases.

Advanced methods show that if f is any differentiable function with $f(a) \neq f(b)$ then as n gets larger and larger the leading order errors of LEFT_n and RIGHT_n are given by

$$\begin{aligned} \mathrm{LEFT}_n - \int_a^b f(x)\,dx &\sim -\frac{1}{2}\big(f(b) - f(a)\big)\,\Delta x\,, \\ \mathrm{RIGHT}_n - \int_a^b f(x)\,dx &\sim \frac{1}{2}\big(f(b) - f(a)\big)\,\Delta x\,. \end{aligned} \tag{4.6}$$

Notice that these errors have opposite signs, about equal magnitude, and decrease like $1/n$ as n increases. Moreover, they are consistent with (4.4) and (4.5).

2: The Midpoint and Trapezoidal Rules. The so-called **midpoint rule** is denoted by MID_n and corresponds to the Riemann sum (4.3) with the choice $p_k = \frac{1}{2}(x_{k-1} + x_k)$, which is the midpoint of the k^{th} subinterval. The **trapezoidal rule** is denoted by TRAP_n and is the average of the left-hand and right-hand Riemann sums given by

$$\text{TRAP}_n = \tfrac{1}{2}\left(\text{LEFT}_n + \text{RIGHT}_n\right). \tag{4.7}$$

For this combination the leading order errors of LEFT_n and RIGHT_n given by (4.6) cancel.

Both the midpoint and trapezoidal rules can be thought of as approximating the area under the curve over each subinterval by that of a trapezoid. In the case of the midpoint rule the top of the trapezoid is given by the tangent line at the midpoint of the subinterval, while in the case of the trapezoidal rule the top of the trapezoid is given by the secant line associated with the endpoints of the subinterval. The midpoint and trapezoidal rules are therefore both exact for linear functions. This way of looking at them also shows that if f is concave up over $[a,b]$ then the midpoint rule gives an underestimate while the trapezoidal rule gives an overestimate:

$$\text{MID}_n \leq \int_a^b f(x)\,dx \leq \text{TRAP}_n\,.$$

Similarly, if f is concave down over $[a,b]$ then the trapezoidal rule gives an underestimate while the midpoint rule gives an overestimate:

$$\text{TRAP}_n \leq \int_a^b f(x)\,dx \leq \text{MID}_n\,.$$

Hence, if f is either concave up over $[a,b]$ or concave down over $[a,b]$ then the midpoint and trapezoidal rules are each accurate to within $|\text{TRAP}_n - \text{MID}_n|$.

In practice you will find that the midpoint rule is always better than the trapezoidal rule when f is either concave up over $[a,b]$ or concave down over $[a,b]$. This can be understood analytically by noticing that the midpoint and trapezoidal rules are related by

$$\text{TRAP}_{2n} = \tfrac{1}{2}\text{TRAP}_n + \tfrac{1}{2}\text{MID}_n\,. \tag{4.8}$$

Hence, if f is concave up over $[a,b]$ then

$$\text{MID}_n \leq \int_a^b f(x)\,dx \leq \text{TRAP}_{2n} = \tfrac{1}{2}\text{TRAP}_n + \tfrac{1}{2}\text{MID}_n\,,$$

whereby

$$0 \leq \int_a^b f(x)\,dx - \text{MID}_n \leq \frac{1}{2}\left(\text{TRAP}_n - \text{MID}_n\right) \leq \text{TRAP}_n - \int_a^b f(x)\,dx\,.$$

Similarly, if f is concave down over $[a,b]$ then

$$0 \leq \text{MID}_n - \int_a^b f(x)\,dx \leq \frac{1}{2}\big(\text{MID}_n - \text{TRAP}_n\big) \leq \int_a^b f(x)\,dx - \text{TRAP}_n\,.$$

Hence, if f is either concave up over $[a,b]$ or concave down over $[a,b]$ then the midpoint rule is better than the trapezoidal rule. Can you see this graphically?

Recall that when is either increasing over $[a,b]$ or dereasing over $[a,b]$ the size of the error made by either the left-hand or right-hand rule satisfies the upper bound given by (4.5). Similarly, there is an upper bound for the size of the error made by the midpoint and trapezoidal rules when f is either concave up over $[a,b]$ or concave down over $[a,b]$. Because we have already showed that the midpoint rule is better than the trapezoidal rule in those cases, all that remains is to find an upper bound for the size of the error made by the trapezoidal rule. If f is concave up over $[a,b]$ then the trapezoidal rule gives an overestimate for the integral. The integral can be underestimated over $[a,b]$ by replacing f with its tangent line approximation at x_k over each subinterval $[x_k - \frac{1}{2}\Delta x, x_k + \frac{1}{2}\Delta x]$. (A picture should make this clear.) When this approximation is integrated over $[a,b]$ one finds that

$$\text{TRAP}_n - \frac{1}{8}\big(f'(b) - f'(a)\big)\,(\Delta x)^2 \leq \int_a^b f(x)\,dx\,,$$

whereby

$$\text{TRAP}_n - \int_a^b f(x)\,dx \leq \frac{1}{8}\big(f'(b) - f'(a)\big)\,(\Delta x)^2\,.$$

Similarly, if f is concave down over $[a,b]$ then

$$\int_a^b f(x)\,dx - \text{TRAP}_n \leq -\frac{1}{8}\big(f'(b) - f'(a)\big)\,(\Delta x)^2\,.$$

Hence, if f is either concave up over $[a,b]$ or concave down over $[a,b]$ then the errors MID-ERROR_n and TRAP-ERROR_n made by the midpoint and trapezoidal rules satisfy

$$|\text{MID-ERROR}_n| \leq |\text{TRAP-ERROR}_n| \leq \frac{1}{8}\,|f'(b) - f'(a)|\,(\Delta x)^2\,. \tag{4.9}$$

This upper bound for the size of the error can be made as small as you wish by picking n large enough. It decreases like $1/n^2$ as n increases. This is a much better rate of convergence than that given by (4.5) for the left-hand and right-hand rules.

Advanced methods show that if f is any twice differentiable function with $f'(a) \neq f'(b)$ then as n gets larger and larger the leading order errors of MID_n and TRAP_n are given by

$$\begin{aligned} \mathrm{MID}_n - \int_a^b f(x)\,dx &\sim -\frac{1}{24}\big(f'(b) - f'(a)\big)\,(\Delta x)^2\,, \\ \mathrm{TRAP}_n - \int_a^b f(x)\,dx &\sim \frac{1}{12}\big(f'(b) - f'(a)\big)\,(\Delta x)^2\,. \end{aligned} \tag{4.10}$$

Notice that the error for the midpoint rule is about half the size of that for the trapezoidal rule. Moreover, they have opposite signs, decrease like $1/n^2$ as n increases, and are consistent with (4.9).

3: Simpson's Rule. The best numerical integration method we study is **Simpson's rule**. It is denoted by SIMP_n and is given by

$$\mathrm{SIMP}_n = \tfrac{2}{3}\mathrm{MID}_n + \tfrac{1}{3}\mathrm{TRAP}_n\,. \tag{4.11}$$

For this combination the leading order errors of MID_n and TRAP_n given by (4.10) cancel. If you recall that TRAP_n was defined by (4.7) as the average of the left-hand and right-hand Riemann sums, Simpson's rule can be expressed in terms of Riemann sums as

$$\mathrm{SIMP}_n = \tfrac{1}{6}\mathrm{LEFT}_n + \tfrac{2}{3}\mathrm{MID}_n + \tfrac{1}{6}\mathrm{RIGHT}_n\,.$$

It can be checked that Simpson's rule is exact for cubic functions.

Advanced methods show that if f is any four times differentiable function with $f'''(a) \neq f'''(b)$ then as n gets larger and larger the leading order error of SIMP_n is given by

$$\mathrm{SIMP}_n - \int_a^b f(x)\,dx \sim \frac{1}{2880}\big(f'''(b) - f'''(a)\big)\,(\Delta x)^4\,.$$

Notice that the error of Simpson's rule decreases like $1/n^4$ as n increases.

Because it follows from (4.8) that

$$\mathrm{MID}_n = 2\mathrm{TRAP}_{2n} - \mathrm{TRAP}_n\,,$$

another way to think of Simpson's rule (4.11) is

$$\mathrm{SIMP}_n = \tfrac{4}{3}\mathrm{TRAP}_{2n} - \tfrac{1}{3}\mathrm{TRAP}_n\,.$$

This way to think of Simpson's rule is related to the way one thinks of more advanced numerical integration methods.