(*) **Summation formulas**

(i) $\sum_{k=1}^{n} f(k) \pm g(k) = \sum_{k=1}^{n} f(k) \pm \sum_{k=1}^{n} g(k)$

(ii) $\sum_{k=1}^{n} cf(k) = c \sum_{k=1}^{n} f(k)$

(iii) $\sum_{k=1}^{n} c = nc$

(iv) $\sum_{k=1}^{n} k = \frac{n^2}{2} + \frac{n}{2} \quad \left(= \frac{n(n+1)}{2} \right)$

(v) $\sum_{k=1}^{n} k^2 = \frac{n^3}{3} + \frac{n^2}{2} + \frac{n}{6} \quad \left(= \frac{n(n+1)(2n+1)}{6} \right)$

(vi) $\sum_{k=1}^{n} k^3 = \frac{n^4}{4} + \frac{n^3}{2} + \frac{n^2}{4} \quad \left(= \frac{n^2(n+1)^2}{4} \right)$

(*) **Area calculation 1 — the easy way**

In class, we considered the problem of calculating the area of the region R bounded by the lines $y = x + 1$, $y = 0$ (the x-axis), $x = 0$ (the y-axis) and $x = 3$. This region is illustrated below in Figure 1

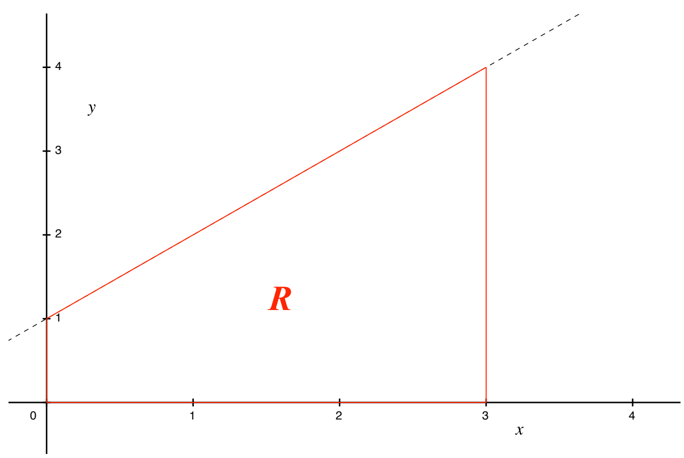

Figure 1: Region R

The quick and easy way to do this is to use simple formulas from geometry. For example, we see that the region R is a triangle (with corners at $(0,1), (3,1)$ and $(3,4)$) on top of a rectangle (with corners $(0,0), (0,1), (3,1)$ and $(3,0)$), as illustrated in Figure 2 below.

The rectangle has width 3 and height 1, so its area is $3 \times 1 = 3$, and the triangle has base 3 and height 3, so its area is $\frac{1}{2}(3 \times 1) = 1.5$. This means that the area of the region is

$$\text{area}(R) = 3 + 1.5 = 4.5.$$

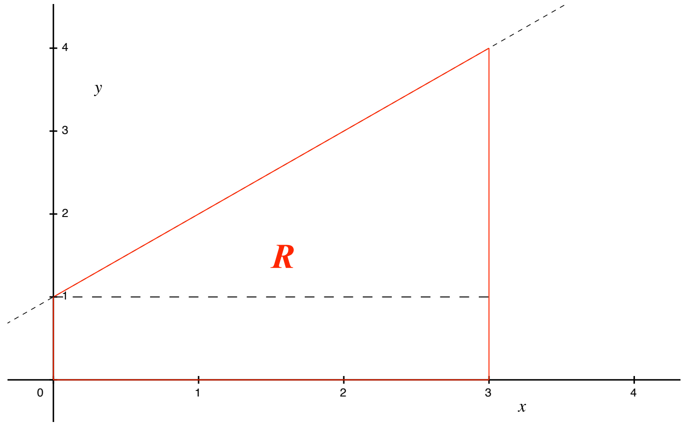

Figure 2: Triangle on top of rectangle — the dotted line is their common side.

(*) **Area calculation 2 — the interesting way**

Area is defined in terms of squares and by extension, rectangles. The area of a square with side a is *by definition* a^2. The area of a rectangle with sides a and b is, by a simple generalization (or by definition, if you prefer) ab.

The formula for the area of a triangle is obtained by observing that a triangle is half of a rectangle that has the same height and width (base) as the triangle. The areas of other geometric shapes *with straight edges* are obtained in similar ways. Specifically, any 2-dimensional region bounded by (finitely many) straight edges can be partitioned into (finitely many) rectangles and/or triangles. The area of such a region is then equal to the sum of the areas of its rectangular/triangular parts.

The question is: *how do we find the area of regions that are bounded by curves that aren't straight?* For example, where does the formula, $A = \pi r^2$, for the area A of a circle of radius r come from?

The answer is — also using rectangles (or triangles, in some cases, like the case of a circle). The problem is that we can't partition such a region into finitely many rectangles/triangles exactly. The solution to this problem is to approximately cover the region in question with finitely many rectangles (or triangles) to obtain an approximation to the area of the region (equal to the sum of the areas of the rectangles/triangles). Then, we repeat this process with more and more rectangles — which we take to be thinner and thinner — so that they cover the region more and more accurately (with smaller and smaller errors), so the approximation to the area of the region becomes more and more accurate.

Then we take a limit.

I will illustrate with the region R from before.

The first step is to divide the interval $[0, 3]$ on the x-axis into n subintervals,

$$[x_0, x_1],\ [x_1, x_2],\ [x_2, x_3],\ \ldots, [x_{k-1}, x_k],\ \ldots, [x_{n-1}, x_n],$$

where $x_0 = 0$ and $x_n = 3$. We call this collection of intervals a *partition* of $[0, 3]$. Furthermore, to keep things simple, we can choose the intervals to all have the same width, so

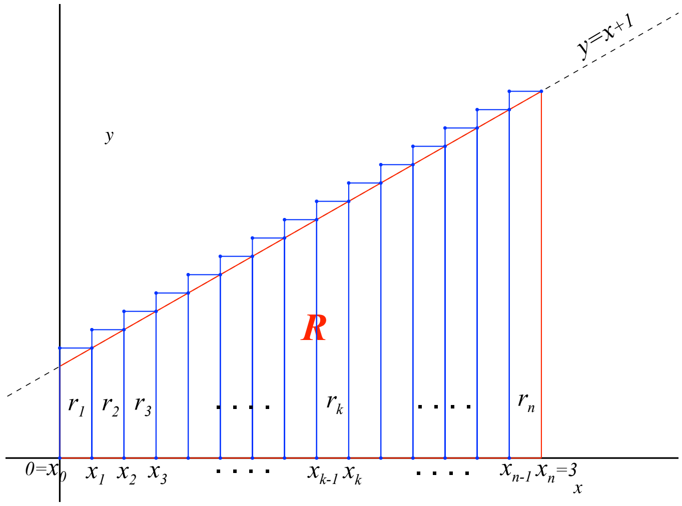

Figure 3: Covering the region R with rectangles.

that

$$x_1 - x_0 = x_2 - x_1 = x_3 - x_2 = \cdots = x_k - x_{k-1} = \cdots = x_n - x_{n-1} = \frac{3}{n}$$

because we have divided and interval of length 3 into n equal parts. With this choice, we see that $x_1 = x_0 + \frac{3}{n} = 0 + \frac{3}{n} = \frac{3}{n}$, $x_2 = x_1 + \frac{3}{n} = 2 \cdot \frac{3}{n}$, $x_3 = x_2 + \frac{3}{n} = 3 \cdot \frac{3}{n}$ and so on. In general, for $k = 1, \ldots, n$, we have

$$x_k = k \cdot \frac{3}{n} = \frac{3k}{n}.$$

Next, we cover the region R with n rectangles, $r_1, r_2, \ldots, r_n$, where for each k, r_k covers the vertical strip from $[x_{k-1}, x_k]$ to the line $y = x + 1$, as depicted in Figure 3, above. The height of r_k is $x_k + 1 = \frac{3k}{n} + 1$, so the area of r_k is

$$\text{area}(r_k) = \overbrace{\frac{3}{n}}^{\text{base}} \cdot \overbrace{\left(\frac{3k}{n} + 1\right)}^{\text{height}} = \frac{9k}{n^2} + \frac{3}{n}.$$

From Figure 3, we see that the sum of the areas of the rectangles r_k is approximately equal to the area of the region R (it's a little bit of an overestimate), so (using the summation formulas above)

$$\begin{aligned}
\text{area}(R) \approx \sum_{k=1}^{n} \text{area}(r_k) &= \sum_{k=1}^{n} \left(\frac{9k}{n^2} + \frac{3}{n}\right) \\
&= \sum_{k=1}^{n} \frac{9k}{n^2} + \sum_{k=1}^{n} \frac{3}{n} = \frac{9}{n^2}\left(\sum_{k=1}^{n} k\right) + \cancel{n} \cdot \frac{3}{\cancel{n}} \\
&= \frac{9}{n^2}\left(\frac{n^2}{2} + \frac{n}{2}\right) + 3 = \frac{9}{\cancel{n^2}} \cdot \frac{\cancel{n^2}}{2} + \frac{9}{n^{\cancel{2}}} \cdot \frac{\cancel{n}}{2} + 3 \\
&= 7.5 + \frac{9}{2n}
\end{aligned}$$

We already know that area$(R) = 7.5$, so we can see that we are on the right track. Moreover, since this works for any number n of rectangles, we also see that the error of approximation, $9/2n$, grows smaller and smaller as n grows larger and larger. We can express this more formally by writing

$$\text{area}(R) = \lim_{n\to\infty} \left(\sum_{k=1}^{n} \text{area}(r_k) \right) = \lim_{n\to\infty} \left(\sum_{k=1}^{n} \frac{9k}{n^2} + \frac{3}{n} \right) = \lim_{n\to\infty} 7.5 + \frac{9}{2n} = 7.5.$$

Observation: The advantage of the 'more interesting' method is that it generalizes, as we shall see. The advantage of the easy method is that it was easy and allowed us to test the new method in a situation where we knew what the answer was.

(*) The definite integral

The ***definite integral*** of the function $y = f(x)$ on the interval $[a, b]$ is denoted by

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

and is defined by the limit

$$\int_a^b f(x)\,dx = \lim_{n\to\infty} \left(\sum_{j=1}^{n} f(x_j^*) \cdot \Delta x_j \right),$$

assuming that the limit exists. In this definition we assume that for each n:

(i) $a = x_0 < x_1 < x_2 < \cdots < x_{n-1} < x_n = b$.
The collection of subintervals $\{[x_0, x_1], [x_1, x_2], \ldots, [x_{n-1}, x_n]\}$ is called a ***partition*** of the interval $[a, b]$.

(ii) x_j^* is *a point* (chosen as we please) in the interval $[x_{j-1}, x_j]$, i.e., $x_{j-1} \leq x_j^* \leq x_j$.

(iii) $\Delta x_j = x_j - x_{j-1}$, for $j = 1, 2, \ldots, n$. This is the *length* of the j^{th} subinterval, $[x_{j-1}, x_j]$.

For the definition above to be useful, we require that widths of ***all*** of the subintervals in the n^{th} partition become smaller and smaller, as the number n of subintervals grows larger and larger. To make this requirement more precise, we define δ_n to be the length of the longest subinterval in the partition,†

$$\delta_n = \max_{1\leq j\leq n} \Delta_j.$$

With this notation, the requirement described above can be expressed by writing

(iv) $\lim_{n\to\infty} \delta_n = 0$.

Comments:

- The definite integral $\int_a^b f(x)\,dx$ returns a *numerical value*.‡

†This number δ_n is called the *diameter* of the partition.
‡In contrast to the indefinite integral $\int f(x)\,dx$ which returns a collection of functions.

- If the function $f(x)$ is *continuous*, then the limit defining the integral $\int_a^b f(x)\,dx$ always exists independently of how the partitions are chosen and of how the points x_j^* are chosen. This means that we can choose the partitions and x_j^*s to make calculating the integral as easy as possible (more on that later).
- We were led to the definition above by considering the problem of computing the area. However, the definite integral is *not defined in terms of area.* Computing area is just one of many applications of definite integration.
- The sum, $\sum_{j=1}^n f(x_j^*) \cdot \Delta x_j$, appearing in the definition of the definite integral is called a *Riemann sum.*[§]

(*) **Right-hand and left-hand sums**

Since the value of the limit defining the definite integral does not depend on the way that the partitions are chosen (when the integrand is continuous), we can choose things to be convenient.

With this in mind, the simplest way to form a partition of the interval $[a, b]$ is to divide it into n ***equal*** subintervals. This choice means that all of the subintervals in the partition have (the same) length:

$$\Delta x = \Delta x_j = (x_j - x_{j-1}) = \frac{b-a}{n}.$$

Since all of the intervals have the same length, $\delta_n = \frac{b-a}{n}$ and condition (iv) is seen to be satisfied in this case, because

$$\lim_{n\to\infty} \delta_n = \lim_{n\to\infty} \frac{b-a}{n} = 0.$$

Second, this choice also makes the values of the endpoints $x_0, x_1, x_2, \dots, x_n$ easy to determine. Specifically,

$$x_0 = a,\ x_1 = a + \frac{b-a}{n},\ x_2 = a + 2\cdot\frac{b-a}{n},\ \dots,\ x_j = a + j\cdot\frac{b-a}{n},\ \dots,\ x_n = a + n\cdot\frac{b-a}{n} = b,$$

because x_j is exactly j steps of size Δx to the right of a.

Next, we have to choose x_j^* (the point where $f(x)$ is evaluated in $[x_{j-1}, x_j]$). The two most common choices here are $x_j^* = x_j$ (the right-hand endpoint) or $x_j^* = x_{j-1}$ (the left-hand endpoint).

With the choices $\Delta x_j = \frac{b-a}{n}$ and $x_j^* = x_j$, the resulting Riemann sum,

$$\sum_{j=1}^n f(x_j)\Delta x_j = \sum_{j=1}^n f\left(a + j\cdot\frac{b-a}{n}\right)\cdot\left(\frac{b-a}{n}\right),$$

is called a *right-hand sum* for the integral $\int_a^b f(x)\,dx$.

Likewise, with $\Delta x_j = \frac{b-a}{n}$ and $x_j^* = x_{j-1}$, the resulting Riemann sum,

$$\sum_{j=1}^n f(x_{j-1})\Delta x_j = \sum_{j=1}^n f\left(a + (j-1)\cdot\frac{b-a}{n}\right)\cdot\left(\frac{b-a}{n}\right),$$

is called a *left-hand sum* for the integral $\int_a^b f(x)\,dx$.

[§]Named for Bernhard Riemann. The integral, so defined, is called the Riemann integral.

(*) **Example**

Find the area of the region bounded by the curve $y = x^2$ and the lines $y = 0$, $x = 1$ and $x = 3$. To do this, we observe that if $f(x) \geq 0$ for x in the interval $[a, b]$, then the area of the region bounded by $y = f(x)$, $y = 0$ (the x-axis), $x = a$ and $x = b$ is given by the definite integral $\int_a^b f(x)\,dx$.

I will use *right-hand sums* to calculate

$$\text{area}(\mathcal{R}) = \int_1^3 x^2\,dx,$$

where $\mathcal{R}$ is the region in question, as illustrated in Figure 4, below.

To compute the integral, we use the following steps: (i) Determine the partition corresponding to n, which means finding Δx_j and the endpoints of the subintervals $x_0, x_1, x_2, \ldots, x_n$. (ii) Write down the right-hand sum corresponding to this partition and then simplify it. (iii) Compute the limit of the right-hand sums as $n \to \infty$.

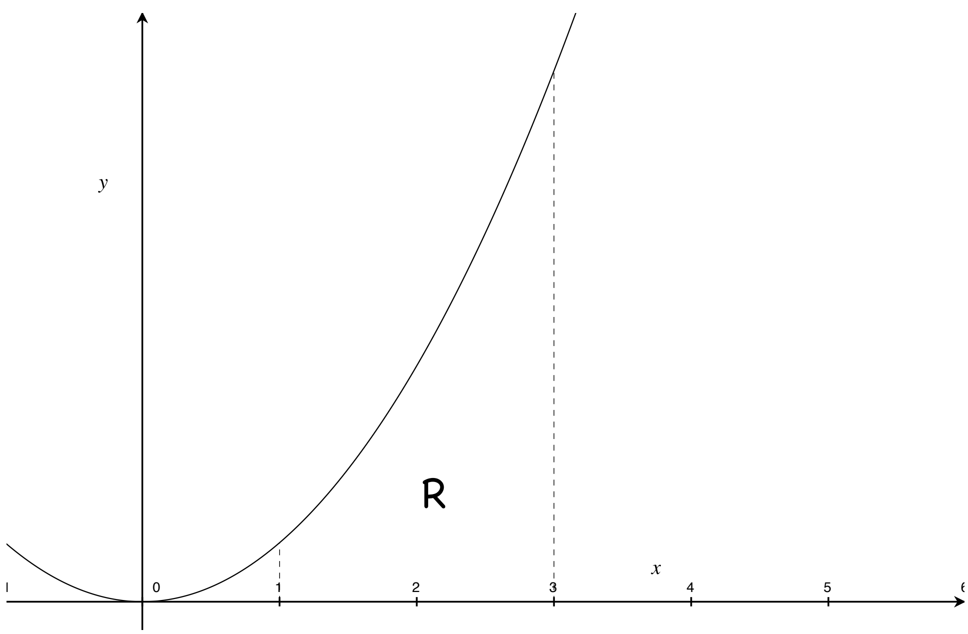

Figure 4: The region in Example 1.

Step (i) Divide the interval $[1, 3]$ into n equal pieces, which means that

- $\Delta x_j = \dfrac{3-1}{n} = \dfrac{2}{n}$, and therefore
- $x_j = 1 + j \cdot \dfrac{2}{n} = 1 + \dfrac{2j}{n}$.

Step (ii) Write down the righthand sum and simplify

$$
\begin{aligned}
RHS(n) = \sum_{j=1}^{n} f(x_j)\Delta x_j = \sum_{j=1}^{n}\left(1+\frac{2j}{n}\right)^2\cdot\frac{2}{n} &= \frac{2}{n}\sum_{j=1}^{n}\left(1+\frac{4j}{n}+\frac{4j^2}{n^2}\right)\\
&= \frac{2}{n}\left(\sum_{j=1}^{n} 1 + \frac{4}{n}\sum_{j=1}^{n} j + \frac{4}{n^2}\sum_{j=1}^{n} j^2\right)\\
&= \frac{2}{n}\cdot n + \frac{8}{n^2}\cdot\frac{n(n+1)}{2} + \frac{8}{n^3}\cdot\frac{n(n+1)(2n+1)}{6}\\
&= 2 + \frac{4n^2+4n}{n^2} + \frac{8n^3+12n^2+4n}{3n^3}\\
&= 2+4+\frac{8}{3}+\frac{4}{n}+\frac{4}{n}+\frac{4}{3n^2} = \frac{26}{3}+\frac{8}{n}+\frac{4}{3n^2}
\end{aligned}
$$

Step (iii) Compute the integral (and therefore the area of $\mathcal{R}$) by finding the limit of the righthand sums as $n \to \infty$

$$
\begin{aligned}
\text{area}(\mathcal{R}) = \int_1^3 x^2\,dx &= \lim_{n\to\infty}\left(\sum_{j=1}^{n} f(x_j)\Delta x_j\right)\\
&= \lim_{n\to\infty}\left(\sum_{j=1}^{n}\left(1+\frac{2j}{n}\right)^2\cdot\frac{2}{n}\right)\\
&= \lim_{n\to\infty}\left(\frac{26}{3}+\frac{8}{n}+\frac{4}{3n^2}\right) = \frac{26}{3}.
\end{aligned}
$$

Observation: The indefinite integral of $f(x) = x^2$ is

$$\int x^2\,dx = \frac{x^3}{3} + C,$$

in other words, if $F(x)$ is an antiderivative of $f(x) = x^2$, then $F(x) = \dfrac{x^3}{3} + C$. This is interesting, because if $F(x)$ is any one of these antiderivatives, then

$$F(3) - F(1) = \left(\frac{3^3}{3} + \cancel{C}\right) - \left(\frac{1^3}{3} + \cancel{C}\right) = \frac{27-1}{3} = \frac{26}{3} = \int_1^3 x^2\,dx.$$

This is ***not*** a coincidence.