This week's notes are about the theory of integration; the notation and details will be complicated, but we will NOT be using most of it for computation (we will compute integrals in week 4 notes). The important thing to understand is this overall "story":

- Informally, the definite integral is the area under a graph (p5-11, §5.2 in textbook).
- The definite integral is defined to be a limit of something called a Riemann sum, and is painfully hard to compute by hand (p12, §5.3-5.4 in textbook).
- The Fundamental Theorem of Calculus (FTC) says that a definite integral of f can be calculated using its antiderivative (i.e. by finding a function F with $f = \frac{dF}{dx}$). This is much easier than using the definition (p , §5.5 in textbook).
- Many interesting geometric quantities are limits of Riemann sums. By rewriting these as multiple integrals and using FTC, we can evaluate some of them using antiderivatives (week 5 notes, §14 in textbook).

This story is extremely important because only a tiny proportion of elementary functions have elementary antiderivates. (An elementary function is a function that is "built out of" $x^n, e^x, \ln x, \sin x, \cos x$.) In other words, the integral of most familiar functions is something that we do not have a name for. So, in almost all applications, functions are integrated numerically using Riemann sums.

Sigma notation for sums (§5.1)

Integration is about adding many things together, so it's useful to have some notation for sums.

Definition: If m and n are integers with $m \leq n$, and f is a function defined at $m, m+1, \ldots, n$, then

$$\sum_{i=m}^{n} f(i) = f(m) + f(m+1) + \cdots + f(n).$$

In this formula, i is the *index of summation*, m is the *lower limit* and n is the *upper limit*. Note that the index of summation i is a "dummy variable" and can be changed without changing the value of the sum, i.e. $\displaystyle\sum_{i=m}^{n} f(i) = \sum_{j=m}^{n} f(j)$.

Examples:

$$\sum_{i=2}^{5} i^2 = \underbrace{2^2}_{i=2} + \underbrace{3^2}_{i=3} + \underbrace{4^2}_{i=4} + \underbrace{5^2}_{i=5}.$$

$$\sum_{j=5}^{n} jx^j = 5x^5 + 6x^6 + \cdots + (n-1)x^{n-1} + nx^n.$$

Definition: If m and n are integers with $m \leq n$, and f is a function defined at $m, m+1, \ldots, n$, then

$$\sum_{i=m}^{n} f(i) = f(m) + f(m+1) + \cdots + f(n).$$

The function $f(i)$ can itself be a sum (with a different index of summation) - in the example below, $f(i) = \displaystyle\sum_{j=2}^{4} \frac{x^i}{i+j}$.

Example:

$$\sum_{i=3}^{4} \sum_{j=2}^{4} \frac{x^i}{i+j} = \sum_{i=3}^{4} \frac{x^i}{i+2} + \frac{x^i}{i+3} + \frac{x^i}{i+4}$$

$$= \underset{\substack{i=3\\j=2}}{\frac{x^3}{3+2}} + \underset{\substack{i=3\\j=3}}{\frac{x^3}{3+3}} + \underset{\substack{i=3\\j=4}}{\frac{x^3}{3+4}} + \underset{\substack{i=34\\j=2}}{\frac{x^4}{4+2}} + \underset{\substack{i=4\\j=3}}{\frac{x^4}{4+3}} + \underset{\substack{i=4\\j=4}}{\frac{x^4}{4+4}}.$$

Some properties of sums:

- If A and B are constants, then $\displaystyle\sum_{i=m}^{n}(Af(i)+Bg(i)) = A\sum_{i=m}^{n} f(i) + B\sum_{i=m}^{n} g(i)$;

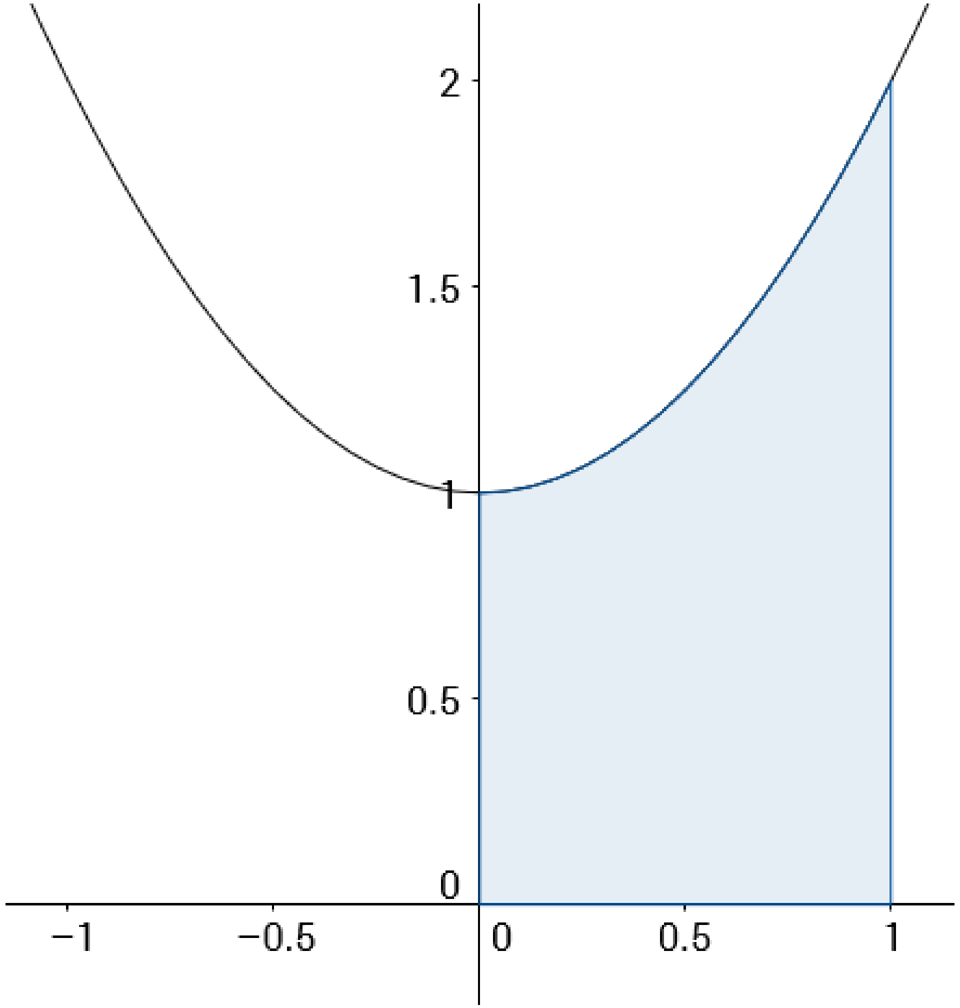
  Example: $\displaystyle\sum_{i=1}^{n}\frac{i^2+i}{3} = \frac{1}{3}\sum_{i=1}^{n} i^2 + \frac{1}{3}\sum_{i=1}^{n} i$ and $\displaystyle\sum_{i=1}^{n}\frac{i^2+i}{n} = \frac{1}{n}\sum_{i=1}^{n} i^2 + \frac{1}{n}\sum_{i=1}^{n} i$

- $\displaystyle\sum_{i=1}^{n} 1 = \underbrace{\overset{i=1}{1} + \overset{i=2}{1} + \ldots + \overset{i=n}{1}}_{n \text{ times}} = n.$

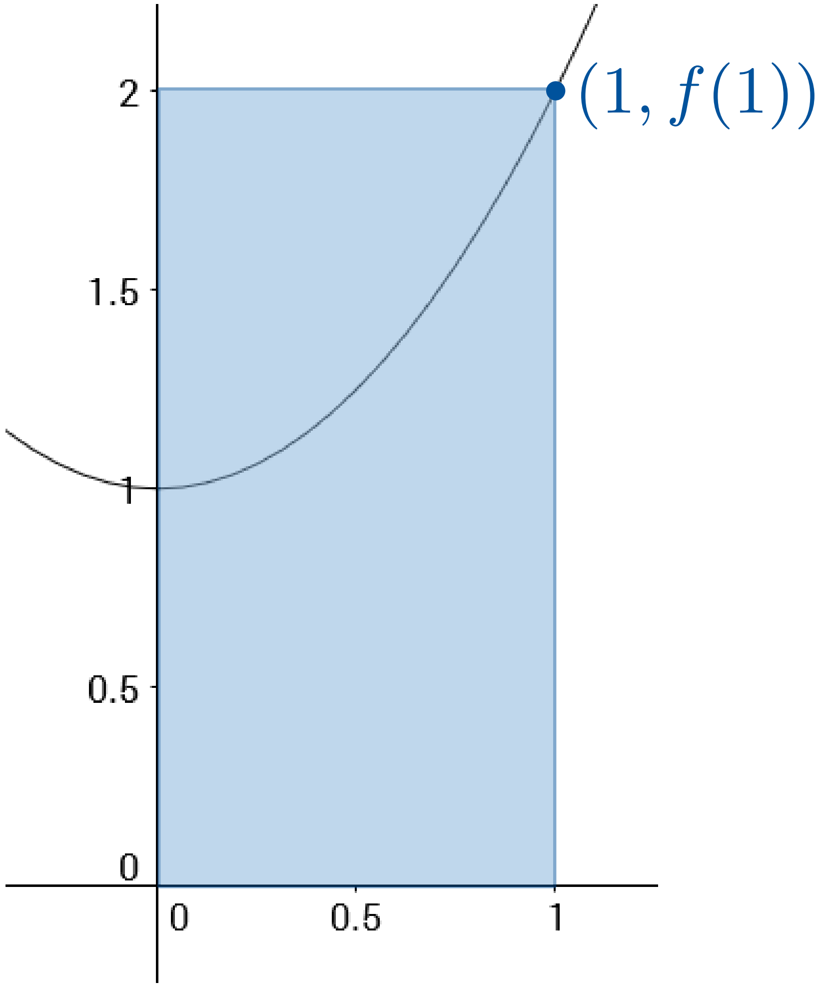
  Example: Combining the two properties, $\displaystyle\sum_{i=1}^{n}\frac{1}{n} = \frac{1}{n}\sum_{i=1}^{n} 1 = \frac{1}{n}n = 1.$

§5.2: Area under a graph

Suppose we want to find the area of the region bounded by the lines $x = 0$, $x = 1$, $y = 0$ and the graph of $f(x) = x^2 + 1$.



A first step might be to approximate the region by this rectangle:

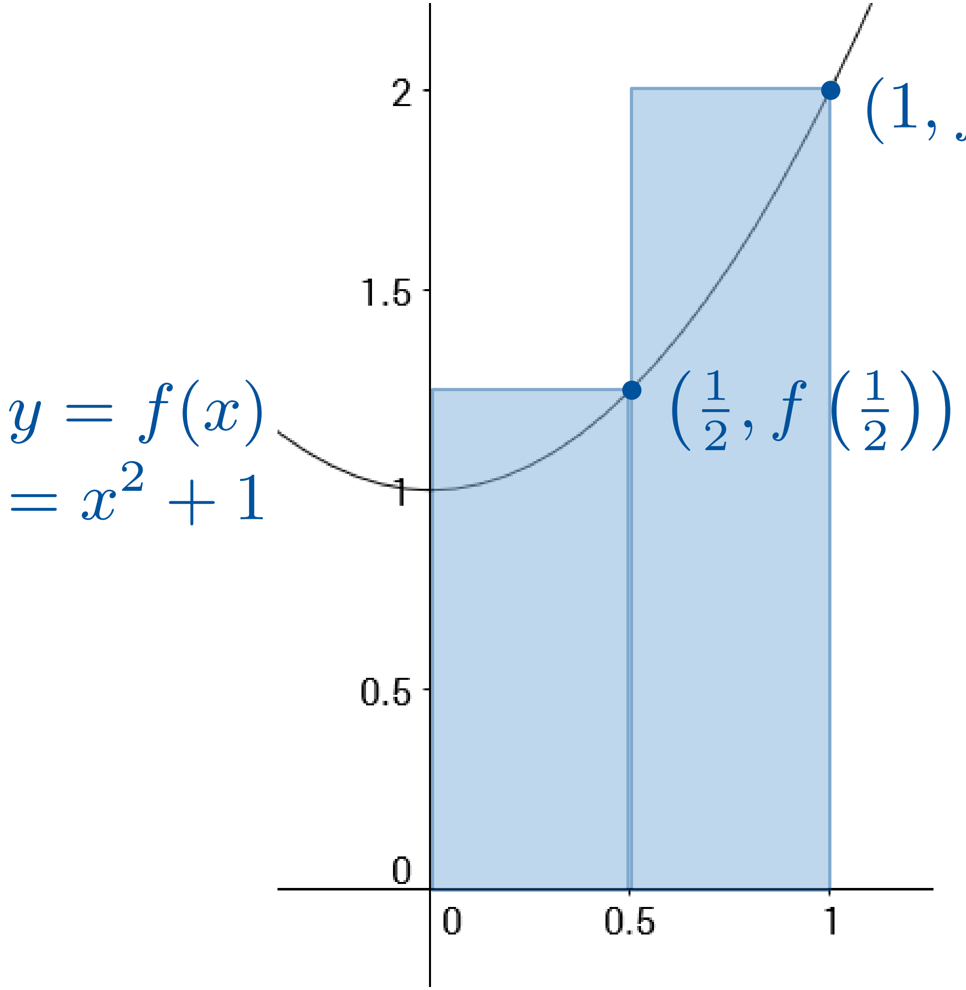


Approximate area
$= \text{width} \times \text{height} = 1f(1) = 2.$

We obtain a better approximation by using two rectangles:

Approximate area

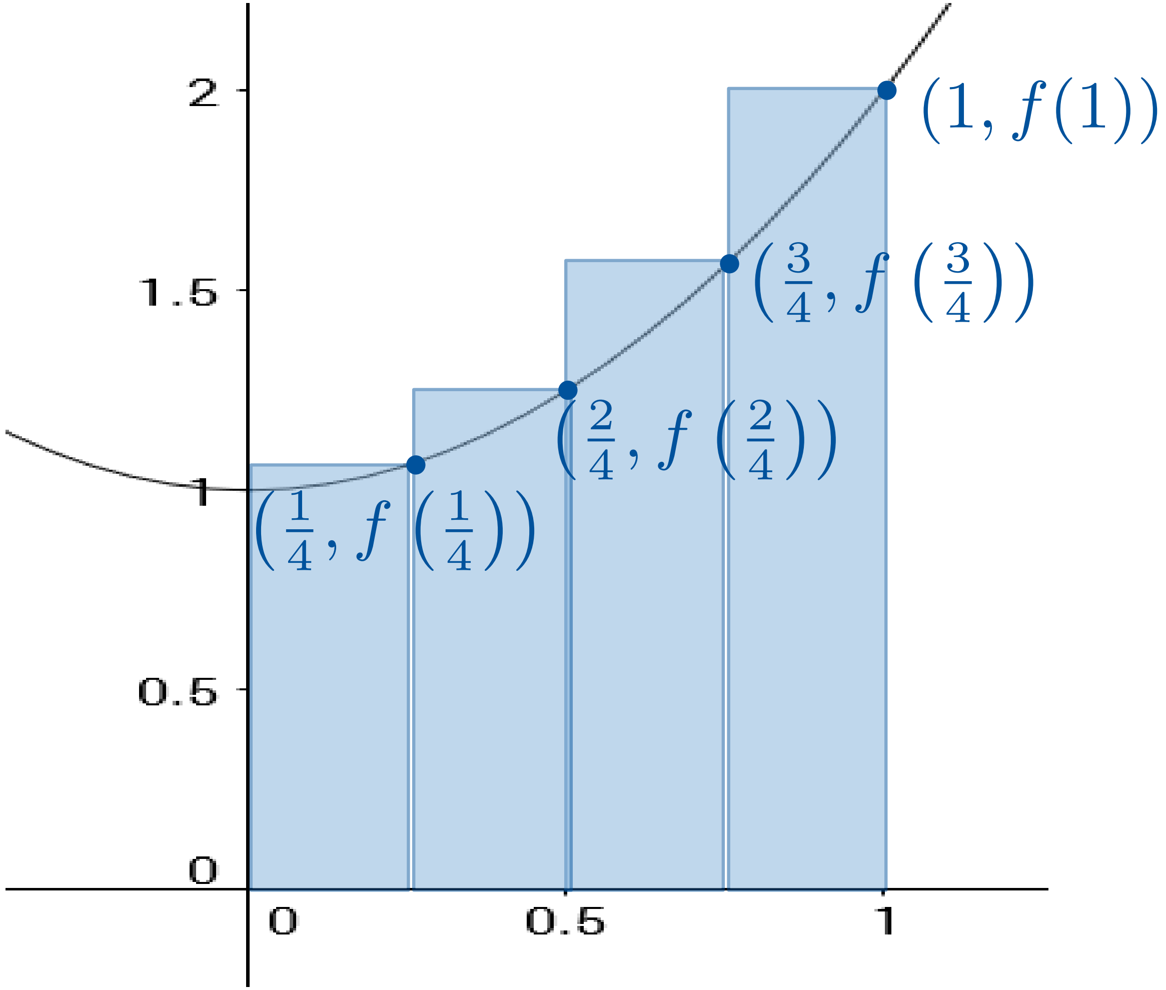
$$= \frac{1}{2} f\left(\frac{1}{2}\right) + \frac{1}{2} f(1) = \frac{1}{2}\frac{5}{4} + \frac{1}{2} 2 = 1.625.$$



We have an even better approximation using four rectangles:

$$\frac{1}{4} f\left(\frac{1}{4}\right) + \frac{1}{4} f\left(\frac{2}{4}\right) + \frac{1}{4} f\left(\frac{3}{4}\right) + \frac{1}{4} f(1)$$

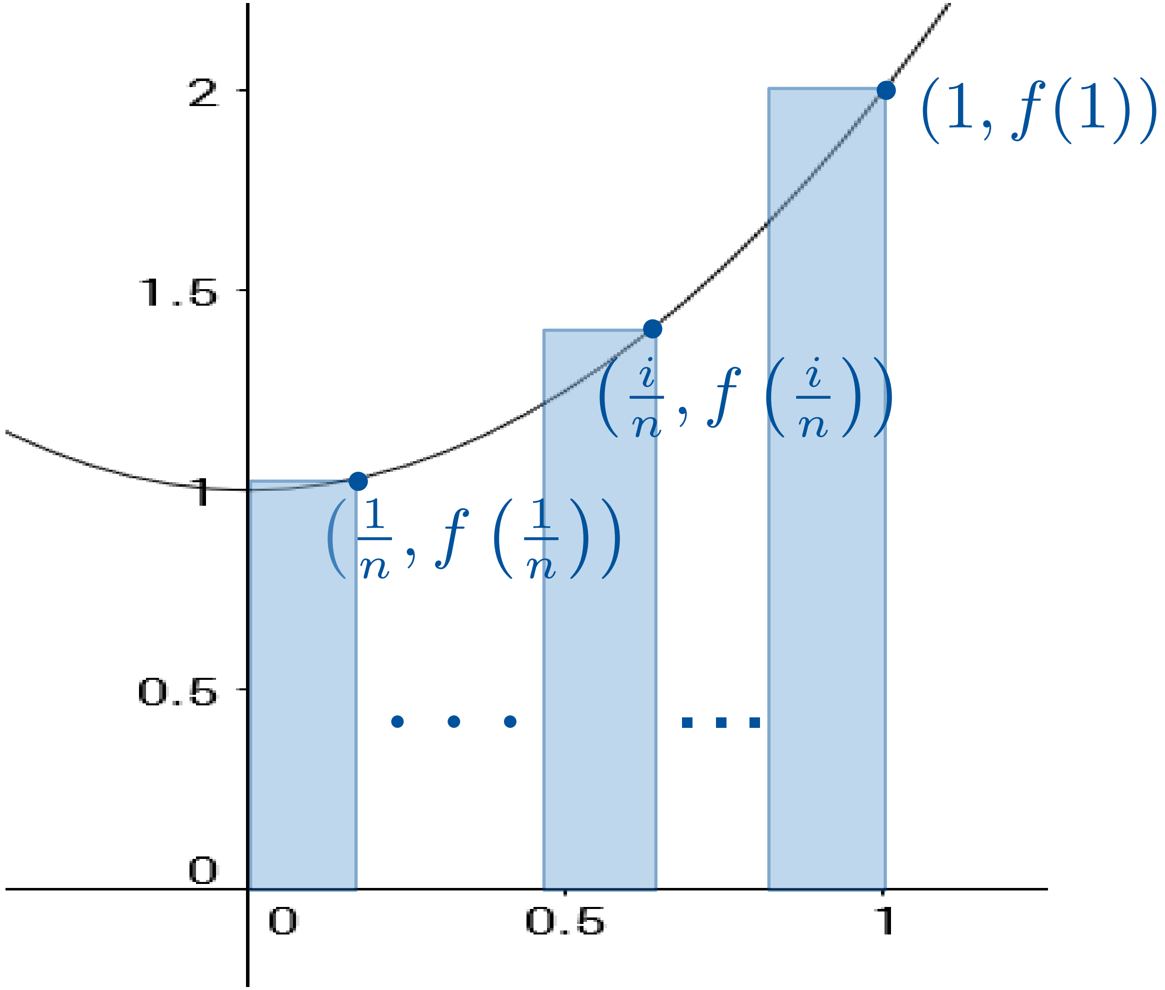
$$= 1.46875.$$



The approximate area using n rectangles is

$$\frac{1}{n}f\left(\frac{1}{n}\right)+\frac{1}{n}f\left(\frac{2}{n}\right)+\dots\frac{1}{n}f\left(\frac{i}{n}\right)+\cdots+\frac{1}{n}f(1)=\sum_{i=1}^{n}\frac{1}{n}f\left(\frac{i}{n}\right),$$

because the ith rectangle has width $\dfrac{1}{n}$ and height $f\left(\dfrac{i}{n}\right)$.



Remembering $f(x)=x^2+1$, this approximate area is:

$$\begin{aligned}\sum_{i=1}^{n}\frac{1}{n}\left(\left(\frac{i}{n}\right)^2+1\right)&=\sum_{i=1}^{n}\left(\frac{i^2}{n^3}+\frac{1}{n}\right)\\&=\sum_{i=1}^{n}\frac{i^2}{n^3}+\sum_{i=1}^{n}\frac{1}{n}\\&=\frac{1}{n^3}\left(\sum_{i=1}^{n}i^2\right)+1.\end{aligned}$$

because of the properties of sums (p4)

From the last page: the approximate area using n rectangles is $\left(\frac{1}{n^3}\sum_{i=1}^{n} i^2\right) + 1.$

Fact: $\sum_{i=1}^{n} i^2 = \frac{n(n+1)(2n+1)}{6}.$

(This formula is unimportant for the rest of the class so we will not prove it, see §5.1 Theorem 1c in textbook.)

So the approximate area using n rectangles is

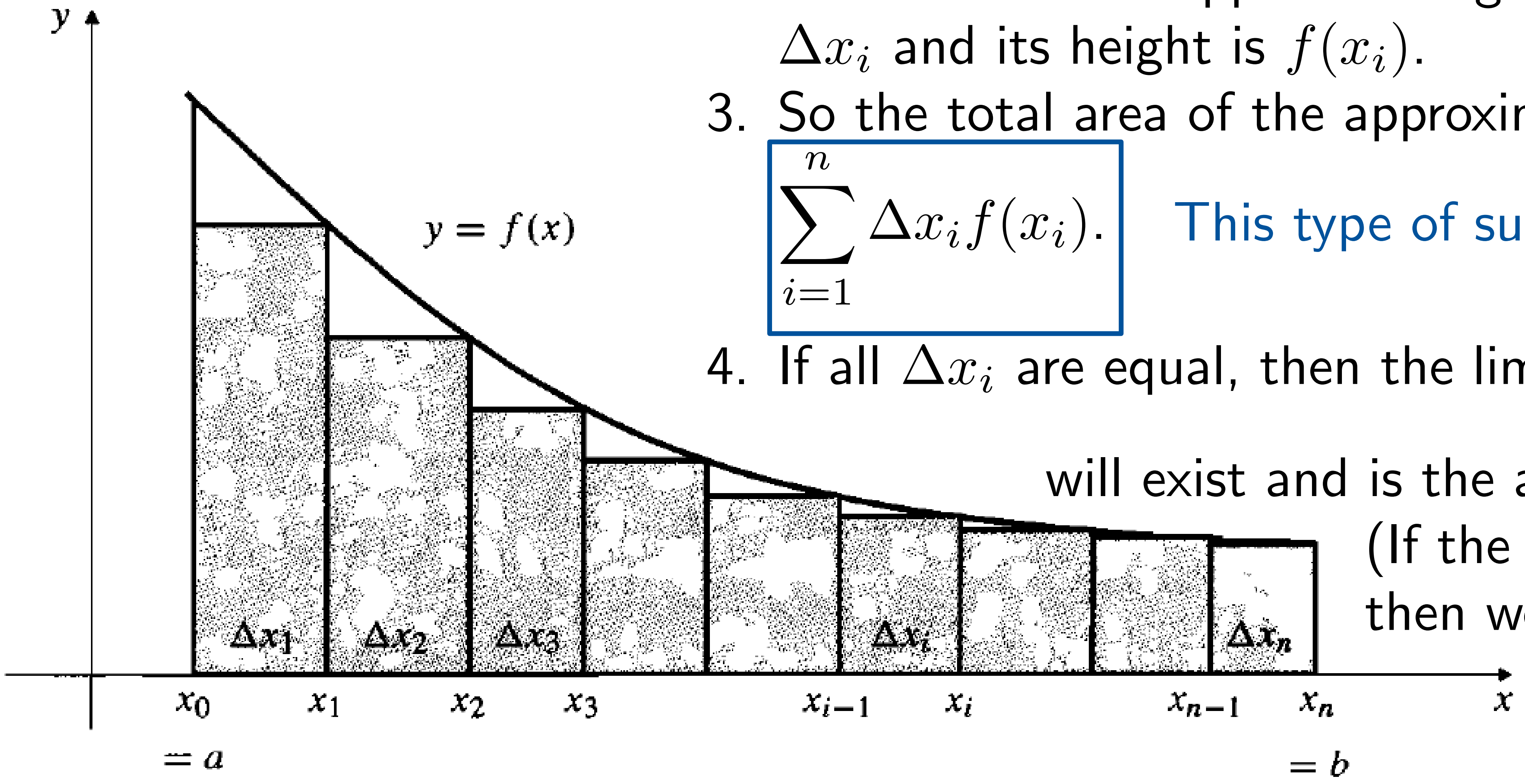
$$\frac{1}{n^3}\frac{n(n+1)(2n+1)}{6} + 1 = \frac{4}{3} + \frac{1}{2n} + \frac{1}{6n^2}.$$

Because our approximation becomes more and more accurate as we use more and more rectangles, the true area must be the limit

$$\lim_{n\to\infty} \frac{4}{3} + \frac{1}{2n} + \frac{1}{6n^2} = \frac{4}{3}$$

(This type of computation is important theoretically, but we will rarely compute like this.)

In general, to find the area under the graph of a continuous, positive function $f : [a,b] \to \mathbb{R}$:

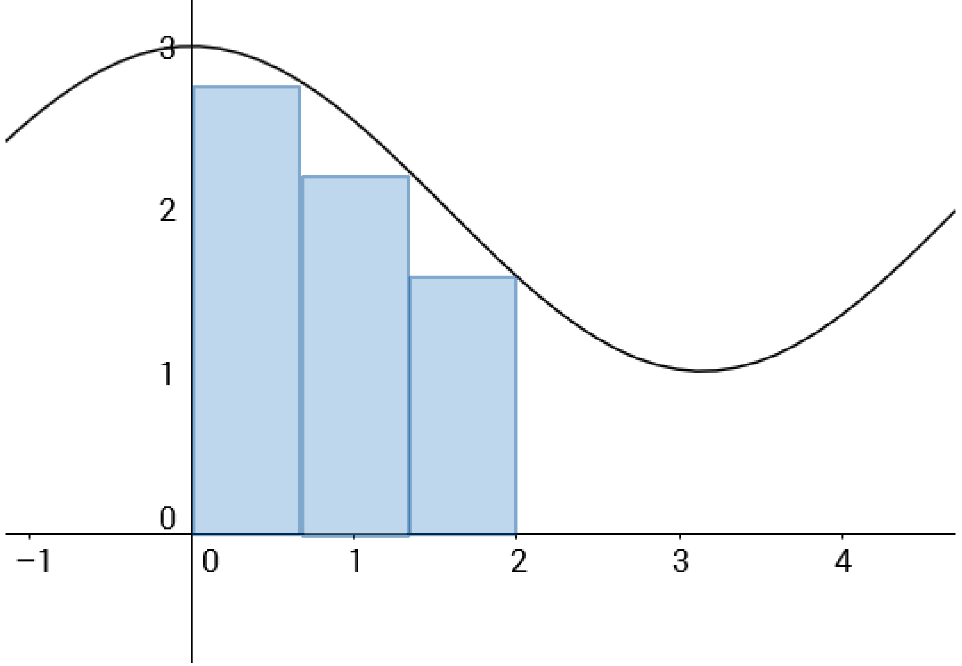


1. Divide $[a,b]$ into n subintervals by choosing x_i satisfying $a = x_0 < x_1 < \cdots < x_n = b$. Let $\Delta x_i = x_i - x_{i-1}$.
2. Consider the ith approximating rectangle: its width is Δx_i and its height is $f(x_i)$.
3. So the total area of the approximating rectangles is $\boxed{\sum_{i=1}^{n} \Delta x_i f(x_i).}$ This type of sum is a *Riemann sum*
4. If all Δx_i are equal, then the limit $\lim_{n\to\infty} \sum_{i=1}^{n} \Delta x_i f(x_i)$ will exist and is the area under the graph. (If the Δx_i are not all equal, then we have to choose x_i carefully.)

Example: Consider the function $f : [0, 2] \to \mathbb{R}$ given by $f(x) = 2 + \cos x$.

a. Use a Riemann sum with 3 subintervals of equal width to approximate the area under the graph of f.

b. Express the exact area under the graph of f as a limit of a Riemann sum.



Answer:

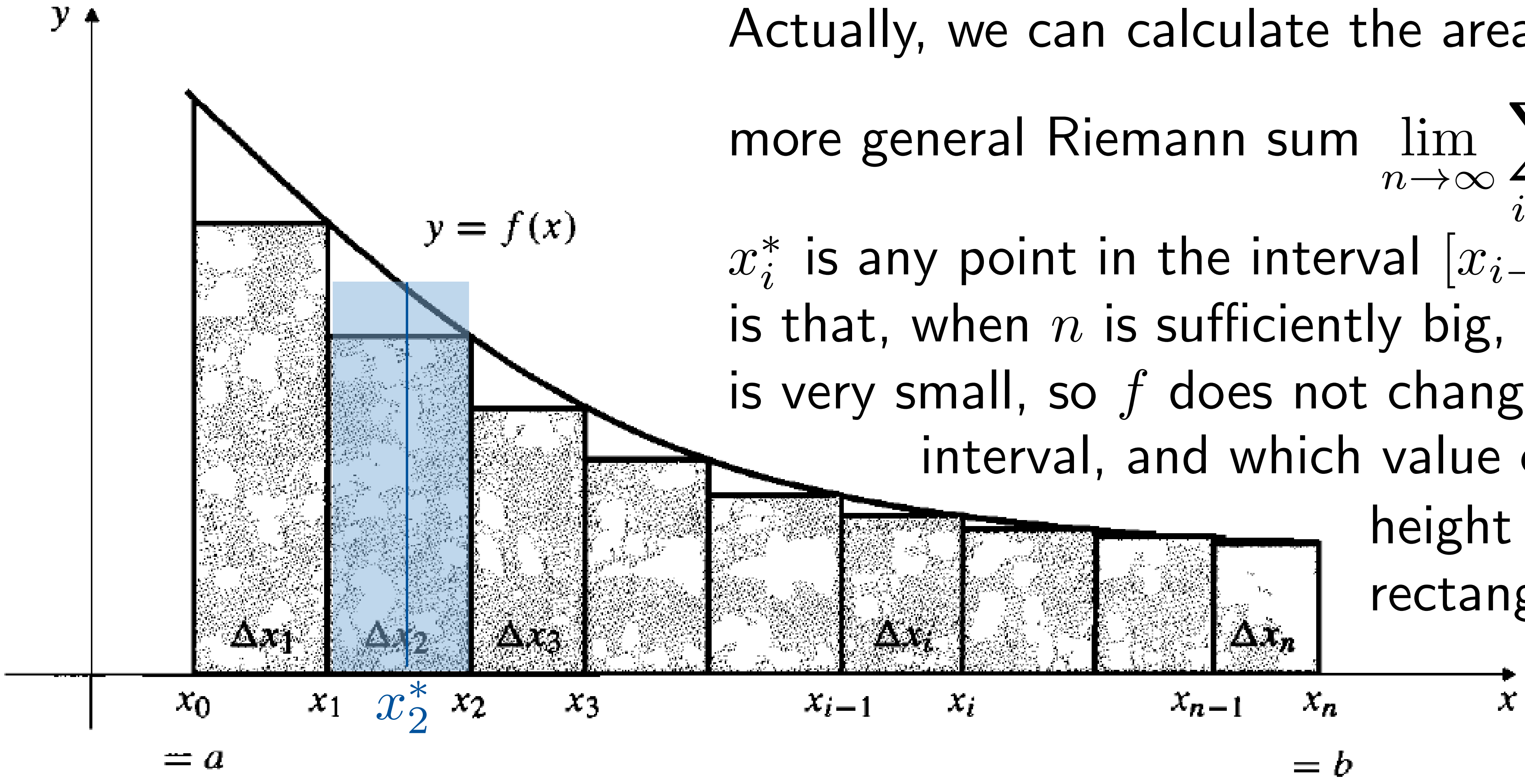
a. To divide $[0, 2]$ into 3 subintervals of equal width, take $\Delta x_i = \dfrac{2}{3}$, so $x_0 = a = 0$, $x_1 = \dfrac{2}{3}$, $x_2 = \dfrac{4}{3}$, $x_3 = b = 2$. So the Riemann sum is

$$\sum_{i=1}^{3} \Delta x_i f(x_i) = \frac{2}{3}\left(2 + \cos\frac{2}{3}\right) + \frac{2}{3}\left(2 + \cos\frac{4}{3}\right) + \frac{2}{3}\left(2 + \cos 2\right).$$

b. To divide $[0, 2]$ into n subintervals of equal width, take $\Delta x_i = \dfrac{2}{n}$, so $x_i = \dfrac{2}{n} i$.

So the area under the graph is $\displaystyle \lim_{n\to\infty} \sum_{i=1}^{n} \Delta x_i f(x_i) = \lim_{n\to\infty} \sum_{i=1}^{n} \frac{2}{n}\left(2 + \cos\frac{2i}{n}\right)$.

Let $f : [a, b] \to \mathbb{R}$ be a continuous, positive function, and $a = x_0 < x_1 < \cdots < x_n = b$ a division of $[a, b]$ into n subintervals of equal width Δx_i. We saw (p9) that the area under the graph of f is $\lim_{n\to\infty} \sum_{i=1}^{n} \Delta x_i f(x_i)$.

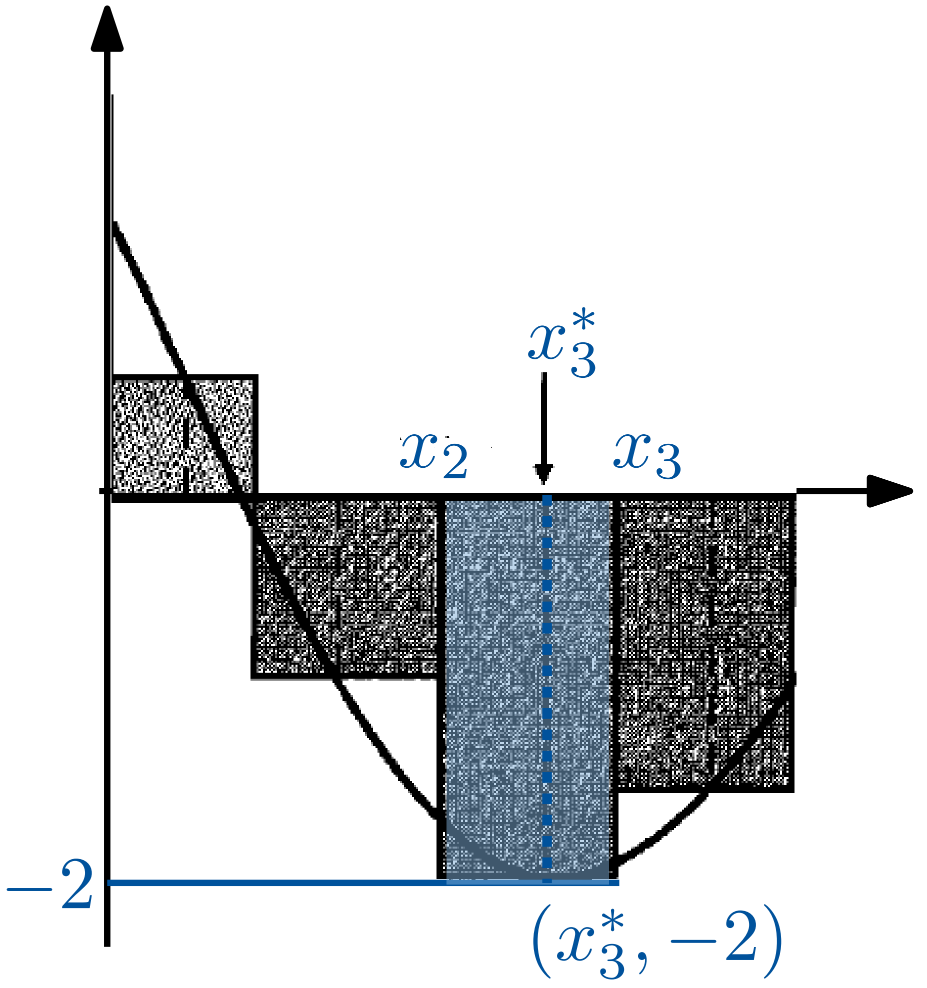


Actually, we can calculate the area as the limit of the more general Riemann sum $\lim_{n\to\infty} \sum_{i=1}^{n} \Delta x_i f(x_i^*)$, where x_i^* is any point in the interval $[x_{i-1}, x_i]$. The intuition is that, when n is sufficiently big, the interval $[x_{i-1}, x_i]$ is very small, so f does not change much within the interval, and which value of f we use as the height of the approximating rectangles will not make much difference.

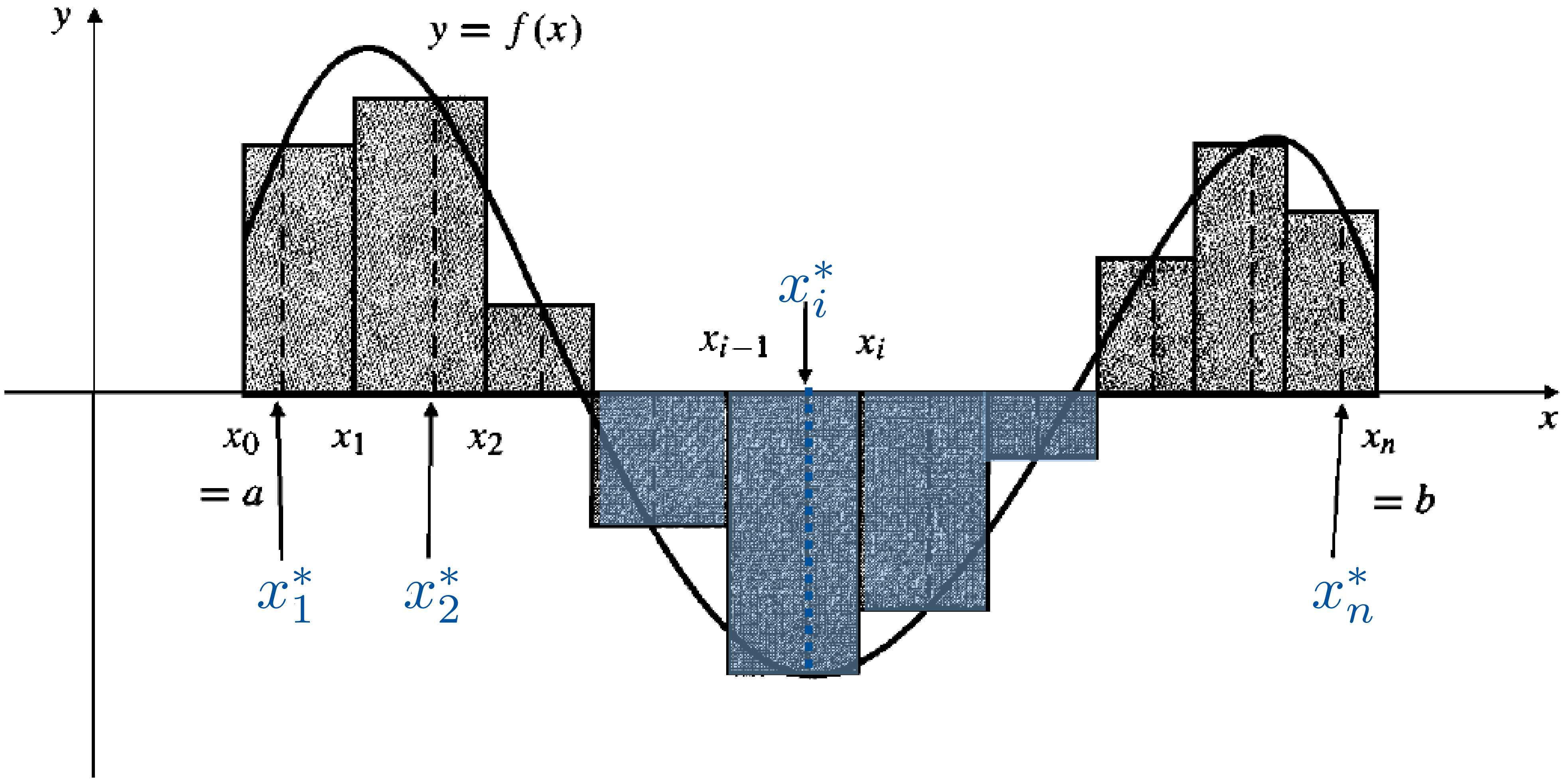
§5.3-5.4: The Definite Integral

For functions $f : [a, b] \to \mathbb{R}$ taking both positive and negative values, the Riemann sum $\displaystyle\sum_{i=1}^{n} \Delta x_i f(x_i^*)$ is still defined. But what does this mean when f is negative?

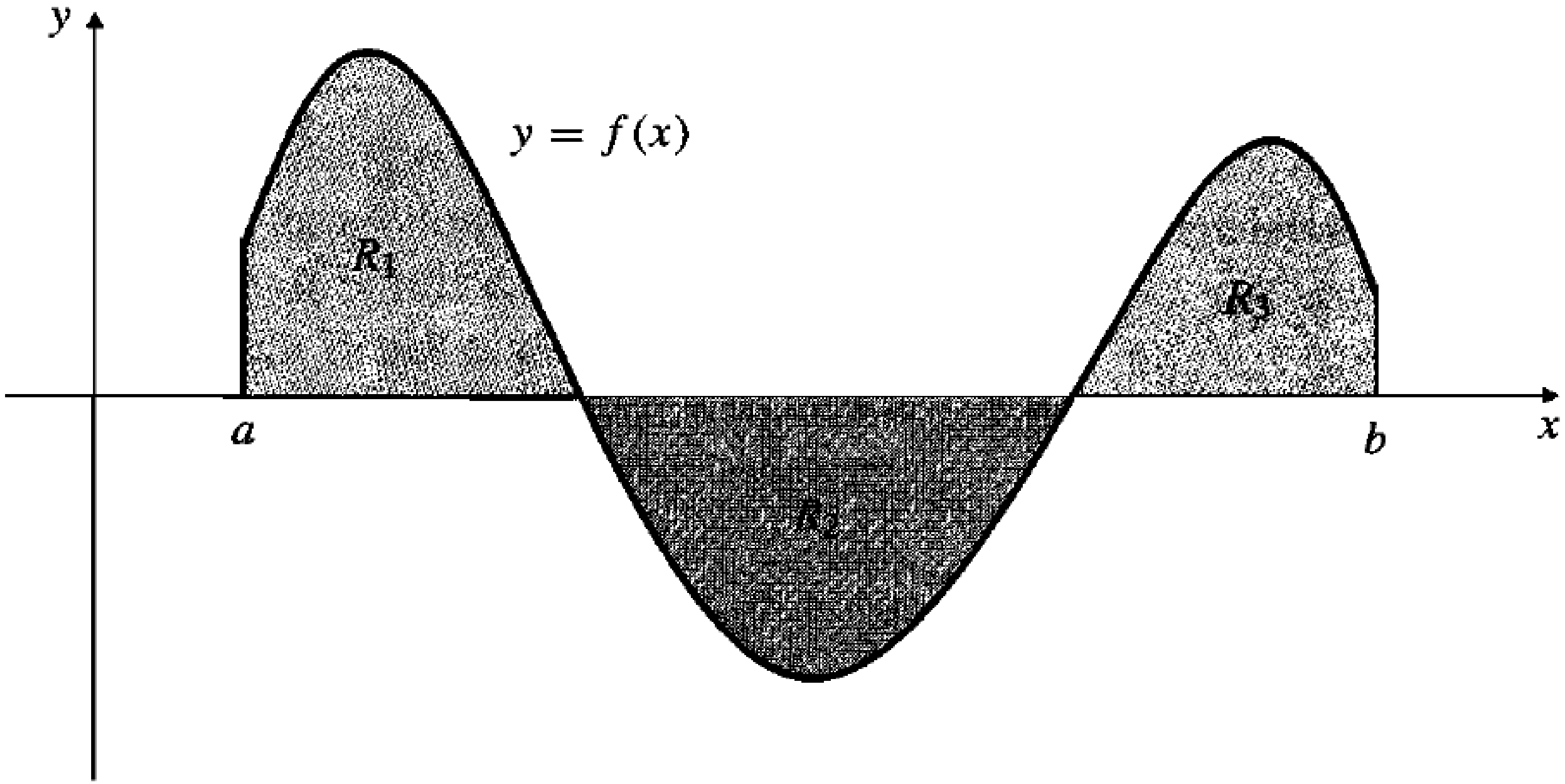
To answer this, suppose $f(x_3^*) = -2$ in the diagrammed example.
Then the 3rd term in the Riemann sum is $\Delta x_3(-2)$.
The height of the 3rd (blue) rectangle in the diagram is 2. So its area is $\Delta x_3 2$, the negative of the 3rd term in the Riemann sum.



So the Riemann sum $\sum_{i=1}^{n} \Delta x_i f(x_i^*)$ is the area of the grey rectangles, which are above the x-axis and below the graph, minus the area of the blue rectangles, which are below the x-axis and above the graph.



So the limit $\lim_{n\to\infty}\sum_{i=1}^{n}\Delta x_i f(x_i^*)$ is the signed area: the total area below the graph and above the x-axis, minus the total area above the graph and below the x-axis.



The signed area is an interesting quantity: for example, if f is velocity, then the signed area is the change in position. So let's define this to be the integral.

Definition: Let $a = x_0 < x_1 < \cdots < x_n = b$ be a division of $[a, b]$ into n subintervals of equal width Δx_i, and let x_i^* be a point in $[x_{i-1}, x_i]$. A function $f : [a, b] \to \mathbb{R}$ is *integrable* if $\displaystyle\lim_{n\to\infty} \sum_{i=1}^{n} \Delta x_i f(x_i^*)$ exists and is independent of the choice of x_i^* in $[x_{i-1}, x_i]$. The value of this limit is the *integral of f on $[a, b]$* (or the integral of f from a to b):

$$\int_a^b f(x)\,dx = \lim_{n\to\infty} \sum_{i=1}^{n} \Delta x_i f(x_i^*).$$

It is hard to use this definition to prove that a function is integrable. Luckily, we have the following theorem:

Theorem 2: Continuous functions are integrable: If f is (piecewise) continuous on $[a, b]$, then f is integrable on $[a, b]$.

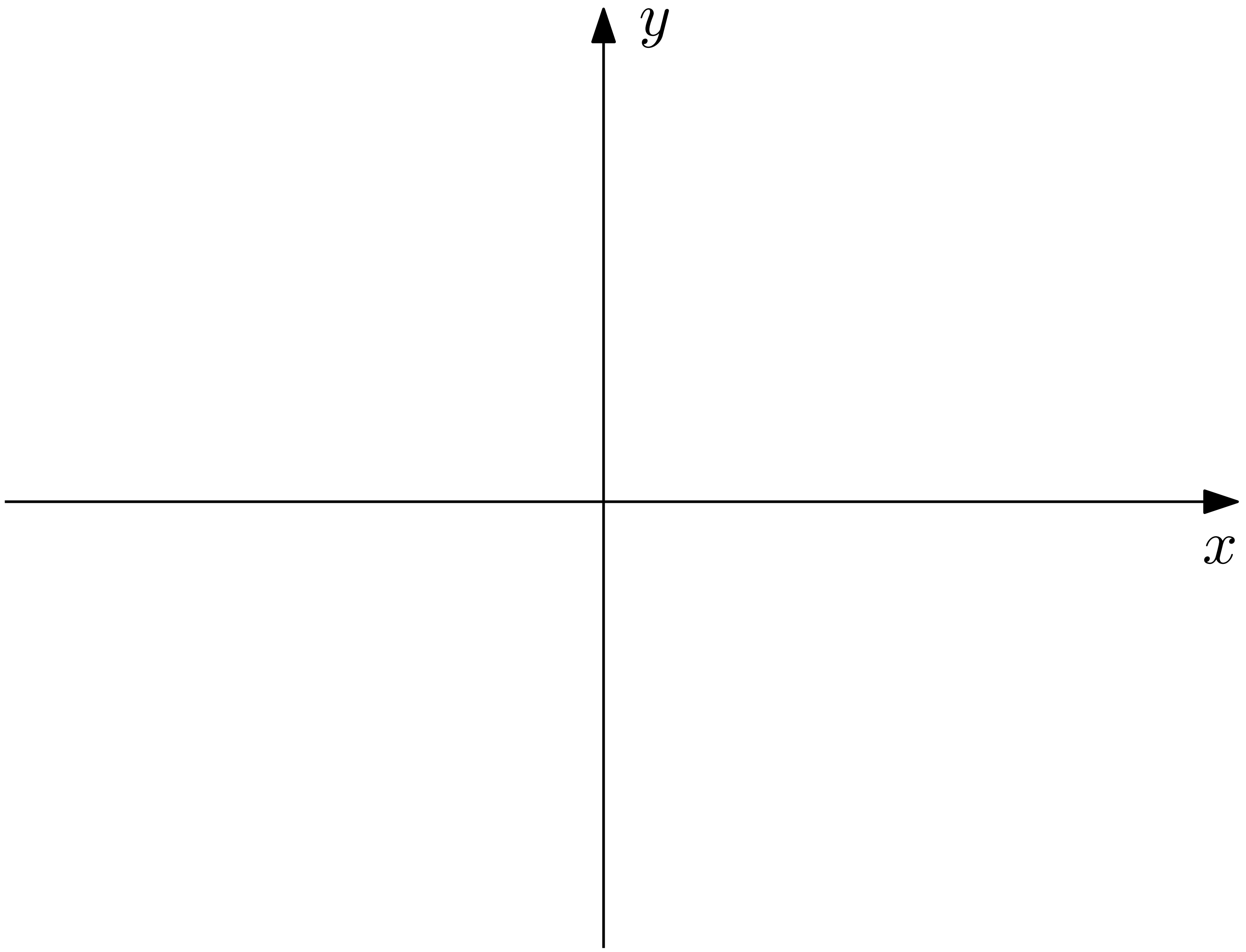
Terminology of the various parts of the integral symbol $\int_a^b f(x)\,dx$:

- $\int$ is the *integral sign* - it is a long S for "sum".
- a is the *lower limit of integration* and b is the *upper limit of integration*.
- f is the *integrand*, the function that is being integrated.
- dx tells us that the *variable of integration* is x. The variable of integration is a dummy variable like the index of summation (p2), we can change it without changing the value of the definite integral, e.g. $\int_a^b f(x)\,dx = \int_a^b f(t)\,dt$.

Important:

- The definite integral is a number, not a function.
- The symbol $\int f(x)\,dx$, without any limits of integration, is the *indefinite integral* or antiderivative. It is a function of x, whose derivative is f. At the moment we do not know that it is related to the definite integral.

Example: By drawing a graph and using geometry, determine $\int_1^2 2 - x\,dx$.



It will be useful to define $\int_a^b f(x)\,dx$ when $a > b$, so we can put variables in the limits of the integral without worrying about which limit is bigger (e.g. p21). The convention which makes all our later theorems work is

$$\int_a^b f(x)\,dx = -\int_b^a f(x)\,dx,$$

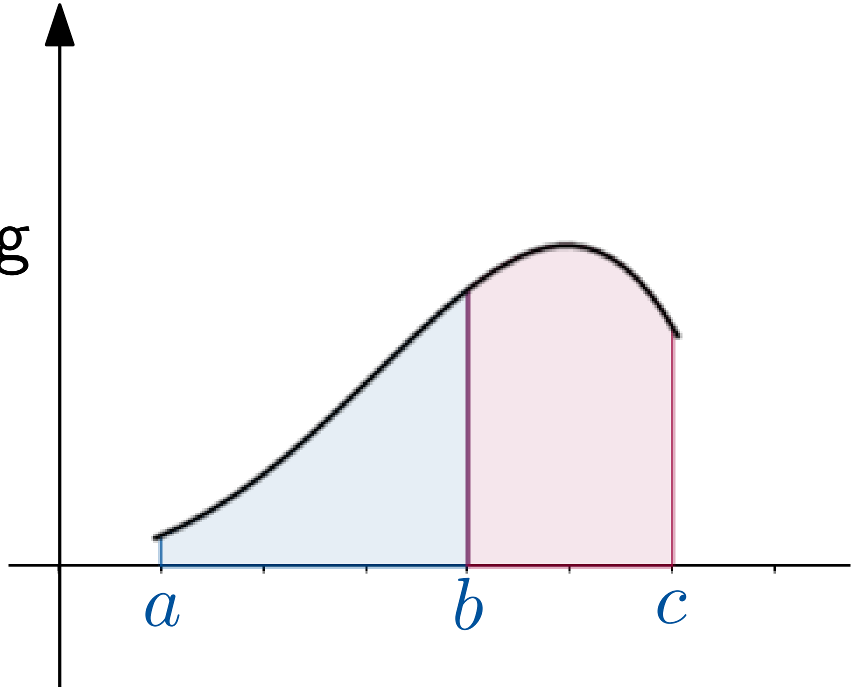
i.e. reversing the limits of integration changes the sign of the integral.

Important properties of the definite integral (the labelling follows §5.4 Theorem 3 in textbook):

c. An integral depends linearly on the integrand: if A and B are constants, then $\int_a^b Af(x) + Bg(x)\,dx = A\int_a^b f(x)\,dx + B\int_a^b g(x)\,dx$. This comes from the corresponding property of Riemann sums (p4).

d. An integral depends additively on the interval of integration:

$$\int_a^b f(x)\,dx + \int_b^c f(x)\,dx = \int_a^c f(x)\,dx.$$

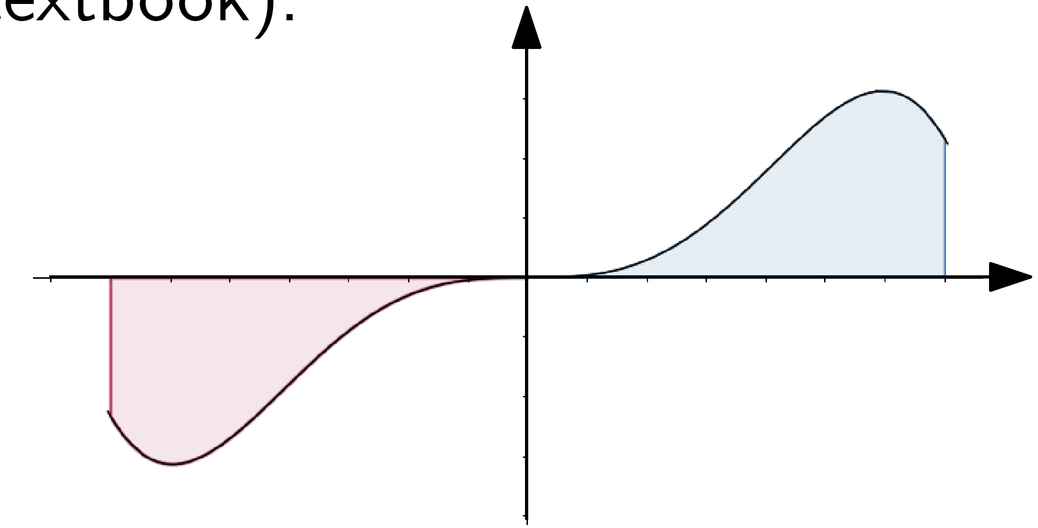


For the case $a < b < c$, this is believable from thinking about integrals as signed areas. When a, b, c are in another order, we need to use identity/definition from the previous page.

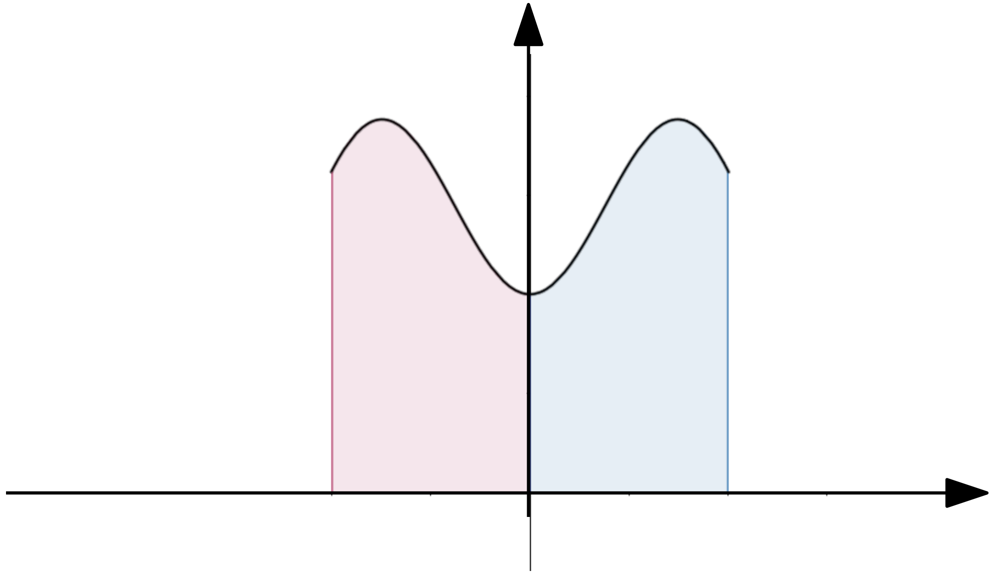
We can deduce from d. that

a. $\int_a^a f(x)\,dx = 0$.

The following two properties shows how to use symmetry to simplify some integrals (the labelling follows §5.4 Theorem 3 in textbook):



g. If f is an odd function ($f(-x) = -f(x)$), then $\displaystyle\int_{-a}^{a} f(x)\,dx = 0$.



h. If f is an even function ($f(-x) = f(x)$), then $\displaystyle\int_{-a}^{a} f(x)\,dx = 2\int_{0}^{a} f(x)\,dx$.

§5.5: The Fundamental Theorem of Calculus

This important theorem is in two parts:

Theorem 5: Fundamental Theorem of Calculus (FTC): Let $f : [a, b] \to \mathbb{R}$ be a continuous function.

FTC1. The cumulative area function $F : [a, b] \to \mathbb{R}$ defined by $F(x) = \int_a^x f(t)\, dt$ is differentiable, and is an antiderivative of f, i.e. $F'(x) = f(x)$.

FTC2. If $G : [a, b] \to \mathbb{R}$ is any antiderivative of f (i.e. $G'(x) = f(x)$), then

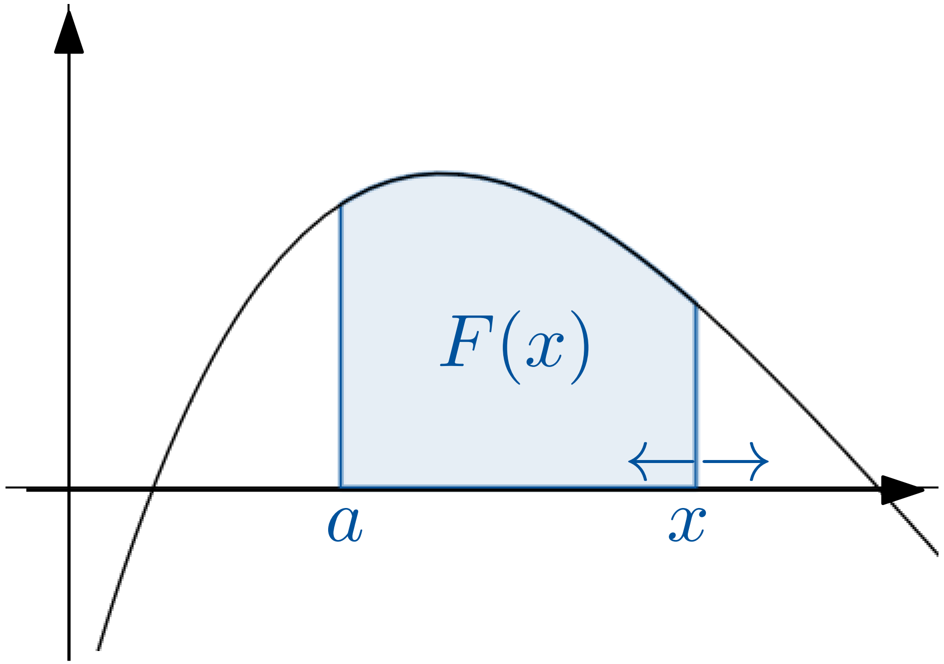
$$\int_a^b f(x)\, dx = G(b) - G(a).$$

FTC1 explains how to differentiate a cumulative area function, and is mainly for theoretical use.

FTC2 explains how to compute a definite integral if you can find the antiderivative of the integrand - this will be very useful to us.

FTC1 will be "obvious" if we understand the cumulative area function

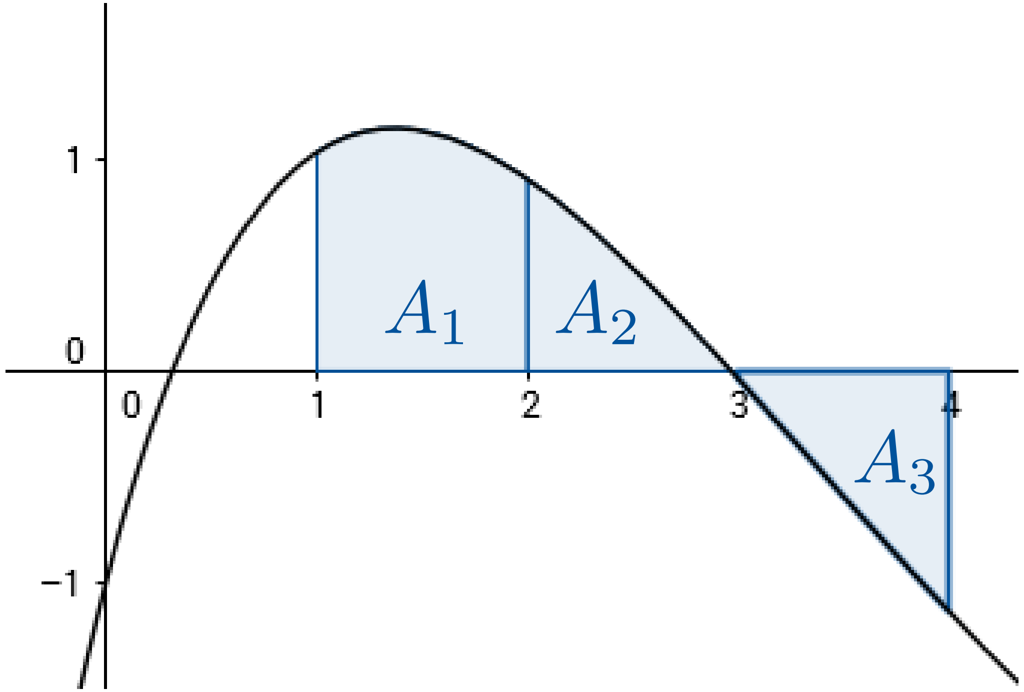
$$F(x) = \int_a^x f(t)\,dt.$$



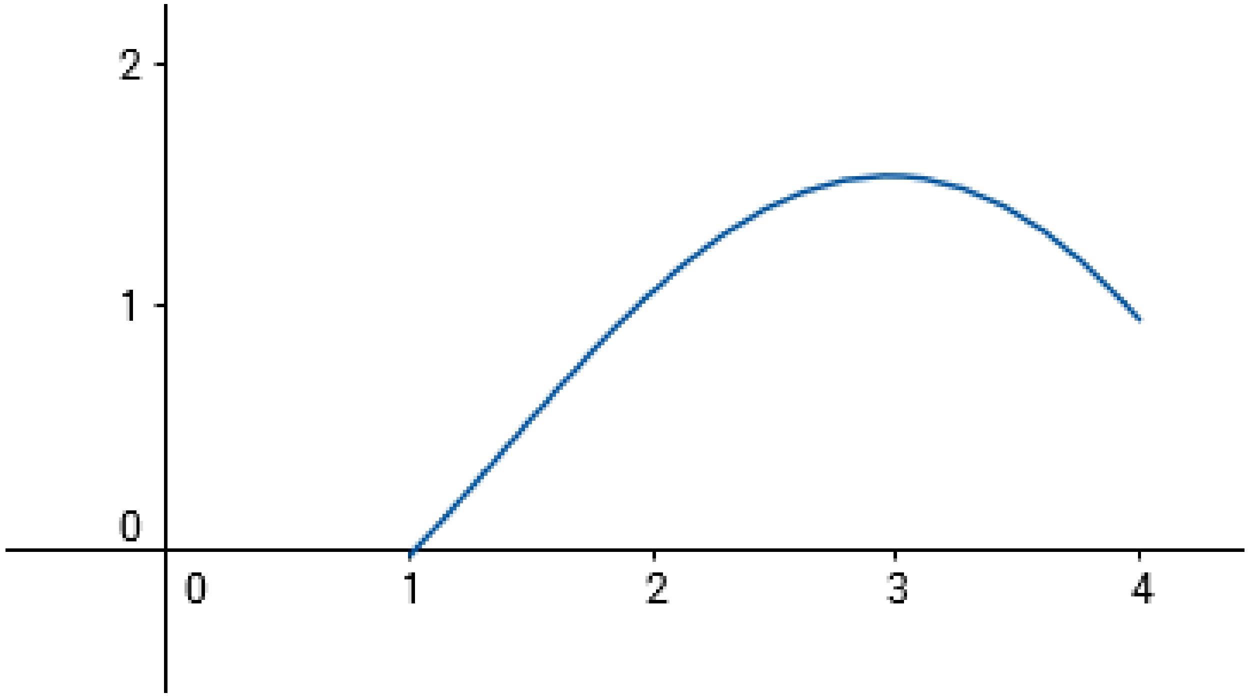
First note that such a function is defined whether $x \geq a$ or $x < a$, because of our definition / identity (p18) that reversing the limits of an integral changes its sign.

Despite the slightly scary formula, cumulative area functions are very natural: for example, if $f(t)$ is the rate that a company is earning money at time t, then $F(x)$ is the total money earned from time a to time x. (Cumulative area functions are also very important in probability.)

Suppose this is the graph of $f : [1, 4] \to \mathbb{R}$:

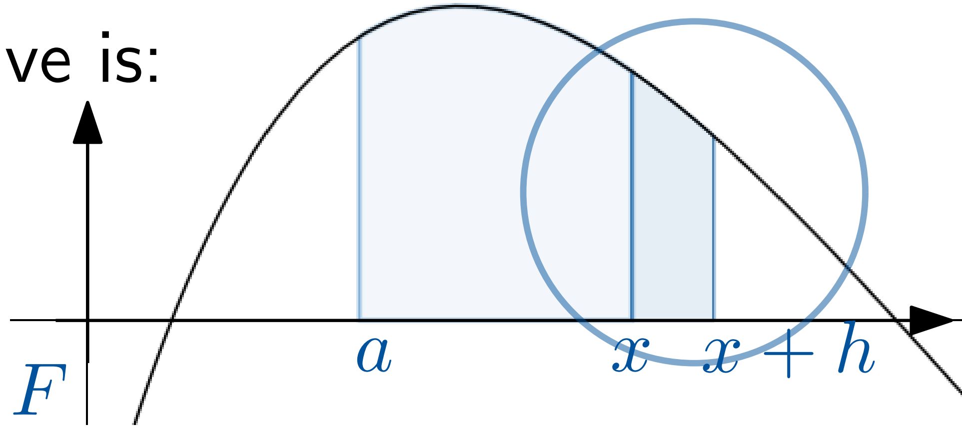


Let's sketch its cumulative area function $F(x) = \int_1^x f(t)\,dt$.



- $F(1) = \int_1^1 f(t)\,dt = 0$ by the properties of definite integrals.
- $F(2) = \int_1^2 f(t)\,dt = A_1$, which is a positive number.
- $F(3) = \int_1^3 f(t)\,dt = A_1 + A_2$. Since $A_2 > 0$, we must have $F(3) > F(2)$, but $A_2 < A_1$ so the increase in F between 2 and 3 is less than it was between 1 and 2
- $F(4) = \int_1^4 f(t)\,dt = A_1 + A_2 - A_3$, so $F(4) < F(3)$.

Observe that we were sketching $F(x)$ by considering the increase or decrease of F, i.e. the derivative of F. This derivative is:



$$F'(x) = \lim_{h\to 0} \frac{F(x+h) - F(x)}{h} \quad \text{definition of derivative}$$

$$= \lim_{h\to 0} \frac{1}{h} \left[\int_a^{x+h} f(t)\,dt - \int_a^{x} f(t)\,dt \right] \quad \text{definition of } F$$

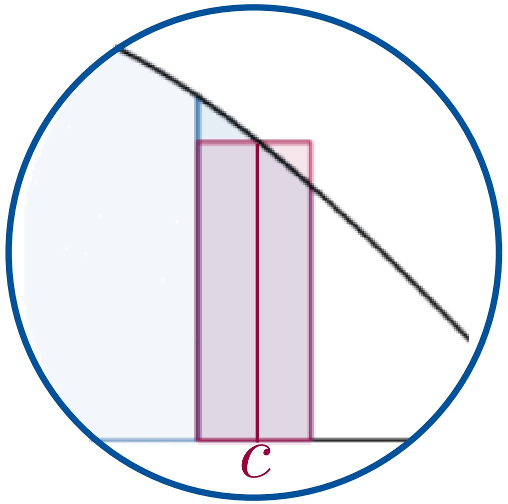
$$= \lim_{h\to 0} \frac{1}{h} \left[\left(\int_a^{x} f(t)\,dt + \int_x^{x+h} f(t)\,dt \right) - \int_a^{x} f(t)\,dt \right]$$

additive dependence on the domain of integration (d, p19)

$$= \lim_{h\to 0} \frac{1}{h} \int_x^{x+h} f(t)\,dt.$$

By the Mean Value Theorem for Integrals (later, §5.4), there is a number $c \in [x, x+h]$ such that $\int_x^{x+h} f(t)\,dt = hf(c)$. So

$$F'(x) = \lim_{h\to 0} \frac{1}{h} hf(c) = \lim_{h\to 0} f(c) = f(x).$$



The previous page proved FTC1: $F(x)=\int_a^x f(t)\,dt$ is an antiderivative of f.

Now we use FTC1 to prove FTC2: $\int_a^b f(t)\,dt = G(b)-G(a)$ for any antiderivative G of f.

Because G and F are both antiderivatives of f, we must have $F(x)=G(x)+C$ for some constant C.

So

$$\begin{aligned}
\int_a^b f(t)\,dt &= F(b) && \text{definition of } F \\
&= F(b)-F(a) && \text{because } F(a)=\int_a^a f(t)\,dt = 0 \\
&= (G(b)+C)-(G(a)+C) && \text{using } F(x)=G(x)+C \\
&= G(b)-G(a).
\end{aligned}$$

To simplify the notation when using FTC2, we write $F(x)\big|_a^b$ to mean $F(b) - F(a)$.
(The alternative notation $[F(x)]_a^b$ will also be accepted.)
Recall that the symbol $\int f(x)\,dx$ means the general antiderivative of f. So FTC2 says $\displaystyle\int_a^b f(x)\,dx = \left(\int f(x)\,dx\right)\bigg|_a^b$.

Redo Example: (Q1 ex. sheet #5) Compute $\displaystyle\int_{-3}^{1} 2x\,dx$ using FTC2.

Redo Example: (p5-8) Compute $\int_0^1 x^2 + 1\,dx$ using FTC2.

Redo Example: (p10) Compute $\int_0^2 2 + \cos x \, dx$ using FTC2.

As the previous examples showed, it's useful to know some common, simple antiderivatives:

$$\int x^r \, dx = \frac{x^{r+1}}{r+1} + C \text{ if } r \neq -1.$$

$$\int \sin x \, dx = -\cos x + C.$$

$$\int \cos x \, dx = \sin x + C.$$

$$\int e^x \, dx = e^x + C.$$

$$\int \frac{1}{x} \, dx = \ln |x| + C.$$

These can be proved by differentiating the right hand side, e.g. for the last line:

if $x > 0$, then $\ln |x| = \ln x$, and $\frac{d}{dx} \ln x = \frac{1}{x}$.

if $x < 0$, then $\ln |x| = \ln(-x)$, and $\frac{d}{dx} \ln(-x) = \frac{1}{-x}(-1) = \frac{1}{x}$.

Some other useful antiderivatives that will be provided to you in exams:

$$\int \frac{1}{\sqrt{1-x^2}}\,dx = \sin^{-1} x + C.$$
$$\int \frac{1}{1+x^2}\,dx = \tan^{-1} x + C.$$

These can be proved by differentiating the right hand sides, using implicit differentiation (see §3.5 of textbook).

Warning: FTC2 only works for continuous integrands. For example, it cannot be applied to $\frac{1}{x^2}$ on an interval containing 0, where the function is not defined.

$\int_{-1}^{1} \frac{1}{x^2}\, dx \neq \left(\frac{-1}{x}\right)\Big|_{-1}^{1} = -2$ - we will see (§6.5) that the associated area is in fact infinite.

(Integrals like these, on an interval containing points where the integrand is not defined, are called improper integrals. These regions do sometimes have finite area - we will explore this later.)