Chapter 2

The Fundamental Theorem of Calculus

We now proceed to outline the major themes of Calculus, a discussion which culminates in the Fundamental Theorem of Calculus, which is a big part of our motivation for the rest of the course.

2.1 Lecture 3: The Concepts of the Derivative and of the Integral

The two major problems about functions that we seek to solve using Calculus are:

- the tangent problem
- the area problem

2.1.1 The tangent problem

Consider the graph of a function \( y = f(x) \) and a point \( x = a \) in the domain of \( f \). The tangent line to \( f \) at \( a \) is the unique line that passes through the point \((a, f(a))\) and has the same slope at the curve at that point.

We think of the tangent line to a curve in many equivalent ways:

- it is the best possible straight-line approximation to \( f(x) \) near \( x = a \), in the sense that it gives more accurate estimates of \( f(x') \) for \( x' \) near \( a \) than any other straight line passing through \((a, f(a))\)
- if an object were constrained to follow the path \( y = f(x) \) at a fixed speed, and released from constraint at \((a, f(a))\), then it would follow the tangent line from that point onwards (in either direction)
- the slope of the tangent line at \( x = a \) is a measure of the instantaneous rate of change of \( y \) with respect to \( x \), at that moment.

The tangent line does not always exist; for example, the point \((0, 0)\) on the curve \( y = |x| \) doesn’t have a unique, well-defined tangent line according to any of the above “heuristic definitions” (and as we’ll agree later: the function \(|x|\) is not differentiable at \( x = 0 \)).

These are good ways of thinking about the tangent line, but this is mathematics! So we want to

- find ways to approximate the tangent line
- determine how to compute it exactly.

First off: the tangent line to \( y = f(x) \) at \( x = a \) has to go through \((a, f(a))\); therefore, to specify the tangent line it is enough to give its slope. This is why the slope of the tangent line is the number we want to determine.

Approximating the slope of the tangent line

Consider \( y = f(x) \) at the point \( x = a \). Let \( h \) be a small nonzero value.

The line through \((a, f(a))\) and \((a + h, f(a + h))\) is called a secant line; for very small values of \( h \), the slope of the secant line approximates the slope of the tangent line, as we can see through examples.
A formula for the slope of the secant line is just the rise over the run. If \( h > 0 \) then this is
\[
\frac{\Delta y}{\Delta x} = \frac{f(a + h) - f(a)}{(a + h) - a} = \frac{f(a + h) - f(a)}{h}
\]
and if \( h < 0 \) then this is
\[
\frac{\Delta y}{\Delta x} = \frac{f(a) - f(a + h)}{a - (a + h)} = \frac{f(a) - f(a + h)}{-h} = \frac{f(a + h) - f(a)}{h}.
\]
That is, we get the same formula for the slope of the secant line whether \( h \) is positive or negative. Good.

This formula is, however, invalid for \( h = 0 \). Unfortunately (or fortunately) that is exactly the value of \( h \) which would give us the slope of the tangent line!

**Determining the slope of the tangent line**

We define the slope of the tangent line to \( y = f(x) \) at \( x = a \) to be
\[
f'(a) = \lim_{h \to 0} \frac{f(a + h) - f(a)}{h}
\]
(if the limit exists) and we call this value the *derivative of* \( f \) *at* \( a \).

**Example 2.1.** Consider the curve \( y = x^2 \) at the point \( x = 1 \). Then
\[
f'(1) = \lim_{h \to 0} \frac{f(a + h) - f(a)}{h} = \lim_{h \to 0} \frac{f(1 + h) - f(1)}{h} = \lim_{h \to 0} \frac{(1 + h)^2 - 1^2}{h} = \lim_{h \to 0} \frac{1 + 2h + h^2 - 1}{h} = \lim_{h \to 0} \frac{2h + h^2}{h} = \lim_{h \to 0} (2 + h)
\]
That is, if we plot the slope of the secant line for every \( h \) except zero, we get points on a straight line \( y = 2 + x \) — and so we deduce that the limit is \( f'(1) = 2 \).

**A concrete interpretation: the velocity problem**

Given an object moving along a linear path such that its position at time \( t \) is given by \( f(t) \), we can interpret:

- The slope of the secant line from \((a, f(a))\) to \((a + h, f(a + h))\), given by
  \[
  \frac{f(a + h) - f(a)}{h}
  \]
is just the displacement divided by the time it took to move it; hence the average velocity on the time interval \([a, a + h]\).

- The slope of the tangent line at \((a, f(a))\) is therefore the *instantaneous velocity* of the particle at that instant of time \( t = a \).
2.1.2 The area problem

We can find the area of almost any region which is entirely bounded by straight lines, using geometry. The Greeks tackled many other classical shapes (notably Archimedes). But what about irregular shapes? How can we determine such an area?

We’ll first simplify the problem to a region with 3 straight sides and one interesting one; if we can solve the area problem for such a figure, then the general case isn’t hard to figure out, by cutting the region appropriately.

So suppose \( f \) is a nonnegative function and \([a, b]\) is an interval contained in the domain of \( f \). Consider the region which is cut out by the \( x \)-axis below, the line \( x = a \) on the left, the line \( x = b \) on the right, and the curve \( y = f(x) \) on the top.

We denote this area by the symbol

\[
\int_{a}^{b} f(x) \, dx.
\]

Here, \( a \) and \( b \) are called the limits of integration, \( f(x) \) is the integrand, and the symbol \( dx \) at the end is just part of the notation; think of it like a right parenthesis and the symbol \( f_{a}^{b} \) as being the left parenthesis.

We want to:

- find ways to approximate the area of this region
- determine how to compute it exactly.

The area can fail to exist if the function is just too bizarre (like \( f(x) = 0 \) on rationals and 1 on irrationals) but in fact integration is quite robust (far more so than differentiation); \( f \) doesn’t have to be continuous, and if it missing a point or two we can just ignore that.

**Approximating the area under the curve** \( y = f(x) \) between \( x = a \) and \( x = b \)

In elementary school, we estimated the area of a region by tracing it onto graph paper and counting the squares. That’s not a bad idea.

- We begin by dividing the interval \([a, b]\) into \( n \) slices (called subintervals), for some positive integer \( n \).
- Over each subinterval, draw a rectangle whose height is \( f(x) \), where \( x \) is (say) the left endpoint of the subinterval.
- We see that the area covered by the rectangles is close to the area under the curve; and as \( n \) gets larger, we get a better and better fit.
- We add up the areas of all the rectangles; this is an approximation to the area under the curve.

To state this precisely, let us simplify things a bit by supposing that all our rectangles have the same width (from now on!). We introduce some notation.

- \( a = x_0 < x_1 < \cdots < x_{n-1} < x_n = b \) are points along the interval (the endpoints of the subintervals).
- Say each subinterval has the same length, which we traditionally call \( \Delta x \): so \( \Delta x = x_1 - x_0 = x_2 - x_1 = \cdots = \Delta x = x_n - x_{n-1} \). We also have that
  \[
  \Delta x = \frac{b-a}{n}
  \]
- On the interval \([x_i, x_{i+1}]\), the left endpoint is \( x_i \) so the height of the rectangle is \( f(x_i) \).

So our estimate (using left endpoints) is

\[
L_n = f(x_0) \Delta x + f(x_1) \Delta x + \cdots + f(x_{n-1}) \Delta x
\]

and we call this a left Riemann sum with \( n \) subintervals.

We could also do this using right endpoints, and we’d get

\[
R_n = f(x_1) \Delta x + f(x_2) \Delta x + \cdots + f(x_n) \Delta x
\]

the right Riemann sum with \( n \) subintervals.
Example 2.2. Let’s estimate the area under the curve \( y = x^2 \) between \( x = 0 \) and \( x = 1 \), using \( L_4 \) and \( R_4 \).

Our intervals have endpoints

\[
0 < 0.25 < 0.5 < 0.75 < 1
\]

and \( \Delta x = \frac{1-0}{4} = 0.25 \), yes. Then

\[
L_4 = f(0) \Delta x + f(0.25) \Delta x + f(0.5) \Delta x + f(0.75) \Delta x
\]

\[
= 0(0.25) + (0.25)^2(0.25) + (0.5)^2(0.25) + (0.75)^2(0.25)
\]

\[
= \frac{14}{64} = 0.2185
\]

whereas

\[
R_4 = f(\frac{1}{4}) \frac{1}{4} + f(\frac{1}{2}) \frac{1}{4} + f(\frac{3}{4}) \frac{1}{4} + f(1) \frac{1}{4}
\]

\[
= \frac{1}{4} \left( \frac{1}{16} + \frac{1}{4} + \frac{9}{16} + 1 \right)
\]

\[
= \frac{30}{64} = 0.46875
\]

and moreover, we see that \( L_4 \) underestimates the area and \( R_4 \) overestimates the area, so

\[
0.2185 \leq \text{area} \leq 0.46875.
\]

This isn’t very impressive; but once we start taking more and more subintervals (using the applet in class, for example, or by computer program) we see that the values get extremely close together.

**Determining the area under the curve**

This process of taking left and right Riemann sums with more and more subintervals converges to the actual area. If only we could take infinitely small subintervals (meaning, infinitely many rectangles!), we could get the area exactly. So we define

\[
\int_a^b f(x) \, dx = \lim_{n \to \infty} L_n = \lim_{n \to \infty} R_n
\]

when these two limits exist and coincide. This value is called the integral of \( f \) between \( a \) and \( b \).

(We’ll be a bit more general and a bit more precise next time; we’ll also do an example to see how this limit can be made completely accurate and give a real answer (just like the derivative does).)

**A concrete interpretation: the distance problem**

Suppose now that a particle moves along a linear path and \( f(t) \) is its velocity at time \( t \).

Recall that if an object moves with constant velocity \( v \) then its displacement from time \( t = a \) to \( t = b \) is the product \( v \times (b - a) \).

The number \( f(t_i)(t_{i+1} - t_i) = f(t_i) \Delta t \) is the product of the velocity of the particle at time \( t_i \) and time \( t_{i+1} - t_i \). If the velocity didn’t change over that interval, then this number would be exactly the displacement of the particle over that time interval.

In other words, if the intervals are short enough, then the area of each rectangle approximates the displacement of the particle over that time interval. Adding up the areas of all the rectangles means finding the total displacement of the particle.

In this case, \( \int_a^b f(t) \, dt \) is the total displacement of the particle between time \( t = a \) and time \( t = b \).

Our amazing realization: So the derivative lets you figure out instantaneous velocity from the displacement function; and the integral lets you figure out the displacement from the velocity function.
2.2 Lecture 4: The integral

Example 2.3. Let’s compute $\int_0^1 x^2 \, dx$.

So let $n$ be the number of intervals. Then $\Delta x = \frac{b-a}{n} = \frac{1}{n}$ in this case, and so

$$R_n = \frac{1}{n} \left( \frac{1}{n} \right)^2 + \frac{1}{n} \left( \frac{2}{n} \right)^2 + \cdots + \frac{1}{n} \left( \frac{n-1}{n} \right)^2 + \frac{1}{n} \left( \frac{n}{n} \right)^2$$

$$= \frac{1}{n} \left( \frac{1}{n^2} + \frac{2^2}{n^2} + \cdots + \frac{n^2}{n^2} \right)$$

$$= \frac{1}{n^3} (1^2 + 2^2 + 3^2 + \cdots + n^2)$$

Now we need to use the formula for the sum of the first $n$ squares (which you can prove by induction, for example):

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

Plugging this into our expression for $R_n$, we get

$$R_n = \frac{1}{n^3} \left( \frac{n(n+1)(2n+1)}{6} \right) = \frac{2n^2 + 3n + 1}{6n^2} = \frac{1}{3} + \frac{1}{2n} + \frac{1}{6n^2}$$

As $n \to \infty$, these latter terms go to zero, so we say that

$$\lim_{n \to \infty} R_n = \frac{1}{3}.$$ 

We can do the same for the left sums as well:

$$L_n = \frac{1}{n} \left( 0^2 + \frac{1}{n} \left( \frac{1}{n} \right)^2 + \frac{1}{n} \left( \frac{2}{n} \right)^2 + \cdots + \frac{1}{n} \left( \frac{n-1}{n} \right)^2 \right)$$

$$= \frac{1}{n} \left( \frac{1}{n^2} + \frac{2^2}{n^2} + \cdots + \frac{(n-1)^2}{n^2} \right)$$

$$= \frac{1}{n^3} (1^2 + 2^2 + 3^2 + \cdots + (n-1)^2)$$

Now put $n - 1$ into the formula to see that the sum of the first $n - 1$ squares is

$$1^2 + 2^2 + 3^2 + \cdots + (n-1)^2 = \frac{(n-1)n(2n-1)}{6}$$

so that

$$L_n = \frac{1}{n^3} \left( \frac{(n-1)n(2n-1)}{6} \right) = \frac{2n^2 - 3n + 1}{6n^2} = \frac{1}{3} - \frac{1}{2n} + \frac{1}{6n^2}$$

which again yields

$$\lim_{n \to \infty} L_n = \frac{1}{3}.$$ 

Therefore we conclude that

$$\int_0^1 x^2 \, dx = \frac{1}{3}.$$