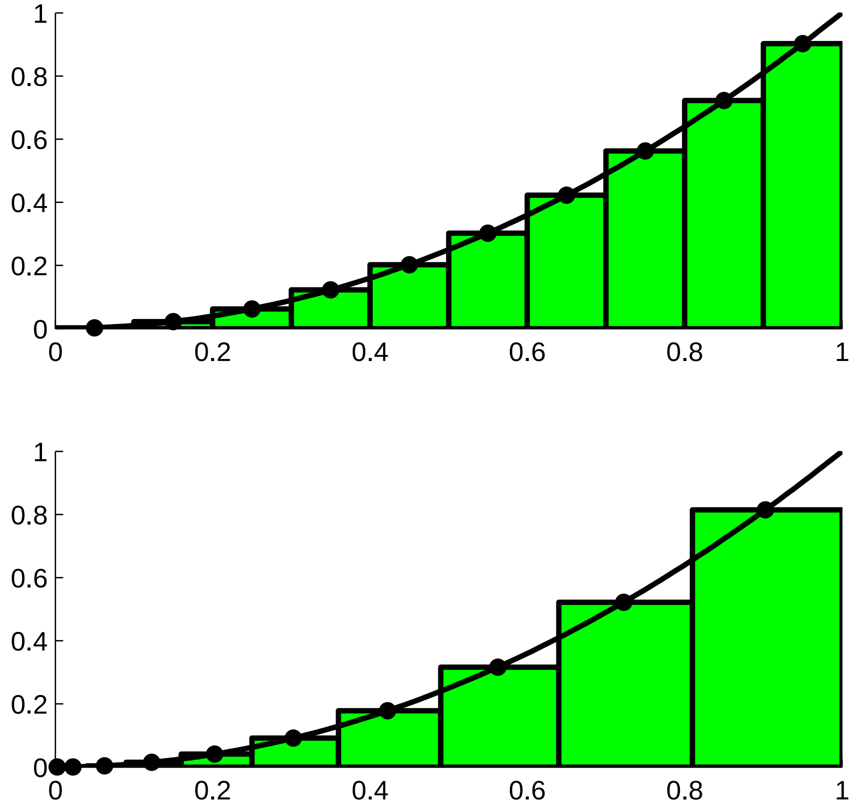

Figure 2: Computing area in two different ways. The pictures show two approximations of the same area under parabola resulting from the use of the Midpoint Rule with ten subdivisions. The difference is in the description of the curve. In the top figure the parabola is described as $y = x^2$, $0 \leq x \leq 1$, while in the bottom figure the same parabola is described with parametric equations $x = t^2$, $y = t^4$, $0 \leq t \leq 1$. The top and bottom Riemann sums are .3325 and .3292, respectively. As an exercise, compute these sums for one hundred subdivisions—this will help you follow the rest of the handout.

Substitution in integrals

As you learned in Calculus I, one way to find the area under $y = f(x)$, $a \le x \le b$ is to compute the integral:

$$A = \int_a^b f(x)\, dx.$$

So far this has been the only approach to finding areas because, until recently, we could only describe curves with equations of the form $y = f(x)$. However, we now have the idea of parametrization at our disposal. This opens new possibilities some of which we will now explore.

Imagine a point particle moving along the graph of $y = f(x)$ without reversing direction. Let $x(t)$ be the x-coordinate of that particle, expressed as a function of time t. The y-coordinate is then a composite function $y(t) = f(x(t))$. For simplicity, let us assume that the particle starts its motion at time $t = 0$ at the point $(a, f(a))$ and ends its motion at time $t = 1$ at $(b, f(b))$; in order to avoid pedantry, we will introduce further assumptions as the need for them arises.

As our imaginary particle moves from $(a, f(a))$ to $(b, f(b))$, it traces out the graph of $y = f(x)$. It stands to reason that we should be able to compute the area under that graph using parametric equations of motion. To this end, let us subdivide the time interval $[0, 1]$ into N equal subintervals of size Δt with equispaced points $t_0 = 1 < t_1 < t_2 \dots t_{N-1} < t_N = 1$. Next sample the x-coordinate of the particle on this subdivision: that is, compute $x_n = x(t_n)$ for $n = 0, \dots, N$. Notice that x_n's are not, in general, equispaced. Nevertheless, we can use the resulting nonuniform partition of $[a, b]$ to set up an appropriate Riemann sum. It will be instructive to do that in two takes.

Take one

It may seem simple and natural to sample the y-coordinate of the particle at t_n's—the same way the x-coordinate was sampled—and use the Right Endpoint Rule, say. This leads to the sum

$$A_N = \sum_{n=1}^{N} y_n\,(x_n - x_{n-1}) = \sum_{n=1}^{N} f(x(t_n))\,(x(t_n) - x(t_{n-1})),$$

whose limit is, indeed, the requisite area A. The problem now is to recognize the limit

$$\lim_{N\to\infty}\sum_{n=1}^{N} f(x(t_n))\,(x(t_n)-x(t_{n-1})) \tag{2}$$

as a definite integral. This is where things get tricky.

One big clue is that the integral representing (2) should have t as its variable of integration. This suggests that we must somehow introduce Δt into the Riemann sum. A standard way of accomplishing that is to multiply and divide Equation (2) by Δt:

$$\lim_{N\to\infty}\sum_{n=1}^{N} f(x(t_n))\,\frac{x(t_n)-x(t_{n-1})}{\Delta t}\,\Delta t.$$

Notice that we now have the precursor of the differential dt placed in the correct spot. Yet, there is more. Notice the fraction—it looks conspicuously like a difference quotient.

Recall that we have encountered this situation before in the discussion of arc length. There our approach was to use the Mean Value Theorem. *If you forgot MVT, review it before reading on.* Assuming that $x(t)$ is differentiable and applying MVT (in the same way as in the discussion of arc length), we can proceed to write the limit of the Riemann sum as:

$$\lim_{N\to\infty}\sum_{n=1}^{N} f(x(t_n))\,x'(c_n)\,\Delta t. \tag{3}$$

Here x' is the (time) derivative of x and c_n is the point lying somewhere in $[t_{n-1}, t_n]$ such that

$$\frac{x(t_n)-x(t_{n-1})}{\Delta t} = x'(c_n).$$

Equation (3) strongly hints that the limit is the integral

$$A = \int_0^1 f(x(t))\,x'(t)\,dt. \tag{4}$$

Yet the transition is not obvious because the functions $f(x(t))$ and $x'(t)$ are sampled at different points! The former is sampled at t_n's—the even subdivision of the time interval $[0, 1]$; meanwhile, the latter is sampled at c_n's–the points that arose from the use of MVT. Try to think of a way around this difficulty before reading the answer in the next section.

Take 2

Zen teaches its practitioners that solving a problem requires the removal of the cause of the problem. Our problem is that $f(x(t))$ and $x'(t)$ are sampled at different points. The cause of the problem is the use of the Right Endpoint rule. To remove the cause, let us use a different rule for forming the Riemann sum. In fact, let us sample the y-coordinates at the points not at t_n but at c_n's!

In other words, once we sample the x-coordinate of the particle at $t_n's$, we invoke MVT to generate c_n's. We then sample the y-coordinate there, *at c_n's.* The result is the limit

$$\lim_{N\to\infty} \sum_{n=1}^{N} f(x(c_n))\, x'(c_n)\, \Delta t.$$

which, by definition, equals the integral in Equation (4).

Substitution rule

Now, in addition to the "Calculus I" way of finding area under a graph, we can find the same area using Equation (4). Therefore:

$$\int_a^b f(x)\, dx = \int_0^1 f(x(t))\, x'(t)\, dt.$$

More generally, we can have

$$\int_a^b f(x)\, dx = \int_{t_1}^{t_2} f(x(t))\, x'(t)\, dt, \tag{5}$$

where $x(t)$ is any monotone, differentiable function with $x(t_1) = a$ and $x(t_2) = b$.

Take a close look at Equation (5)—you may recognize it as the familiar Substitution Rule. It may be that you learned the latter as "u-substitution":

$$\int_a^b f(u(x))\, u'(x)\, dx = \int_{u(a)}^{u(b)} f(u)\, du. \tag{6}$$

If that is the case, compare Equations (5) and (6): you should be able to see that they express one and the same idea using slightly different notation.

Why, you may ask, have we not started with the Substitution Rule? Indeed, substituting x in

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

with $x(t)$ gives the area under the curve in terms of its parametric equations right away, without longwinded arguments involving MVT.

The short answer is that you need to *understand* the Substitution Rule. That entails, among other things, knowing where the rule comes from. Besides there are a few other reasons why I wanted you to go through the above derivation:

1. One can never have enough practice setting up Riemann sums. Especially in Calculus II.

2. If you want to understand mathematics, you must study mathematical *reasoning.* Otherwise the subject will be forever obscured by rules.

3. The derivation in this handout has a lot of parallels with the derivation of the formula for *arc length.* Chances are, this will help you understand arc length a little better.

4. Some of the Riemann sums in this handout foreshadow profound generalizations of the Riemann integral. Although we will not study those generalizations, it is good to have seen them.

And speaking of longwinded arguments, I could have made this handout much shorter, without extra takes. Yet, contrary to what you may think, brevity would not have improved readability.

Most people, with few exceptions, find mathematical textbooks hard or impossible to read. Have you ever wondered why? One reason, paradoxically, is the relentless pursuit of logical order. To a mathematician, it makes perfect sense to discuss limits first, then continuity, then differentiability, followed by optimization, related rates, and so on. Indeed, Calculus concepts can be and usually are presented in perfectly organized linear progression. Likewise, every Calculus derivation can be arranged as a perfectly ordered chain of arguments and deductions; indeed, this is how standard Calculus textbooks are written. However the logical order is different from the *natural* order that governs human thought. The human animal relies much more on intuition and experience than cold hard logic.

Any professional mathematician can, after some deliberation, present a perfect derivation of Equation (5). However, most mathematicians would not be able to do that right away, without preparation because... they are human, despite the popular opinion. Even the best of us first need to try a few things, jot down some ideas, notice that, say, MVT needs to be invoked and so on. Only then can we organize the argument in logical order and present it to the astonishment of the general public. I hope that, having read this handout, you have a better idea of how mathematics is actually *done*. Do not be fooled by the "director's cut".

Exercises

1. Perform the indicated substitution and evaluate the resulting integral (you may need to simplify the integrand using algebra or trigonometry). Don't forget to validate your work using `Matlab`:

   (a) $\displaystyle\int_0^{\frac{1}{2}} \frac{1}{\sqrt{1-x^2}}\,dx, \quad x = \sin(t)$

   (b) $\displaystyle\int_0^1 x\,e^{-x^2}\,dx, \quad x^2 = t$

   (c) $\displaystyle\int_1^2 \frac{\ln(x)}{x}\,dx, \quad \ln(x) = t$

2. Try to work out the integrals below using substitution. If you cannot find the right substitution, present an attempt and explain why it failed.

   (a) $\displaystyle\int_0^1 x\,(1-x)^5\,dx.$

   (b) $\displaystyle\int_0^a \frac{dx}{1+x^2}.$

3. The length of the parabola $y = x^2$, $0 \le x \le 1$ is given by the following integral:

   $$L = \int_0^1 \sqrt{1+4\,x^2}\,dx.$$

   Work it out using the substitution $2\,x = \dfrac{e^t - e^{-t}}{2}$. Hint: after the substitution, write the expression under the radical as a square.

4. Use `Matlab` to plot the curve $x = 2\cos(t)$, $y = 3\sin(t)$, $0 \leq t \leq 2\pi$. Do you recognize it? Present the code and its output.

5. Find the area enclosed by the curve in the previous exercise. Hint: use the double angle formula to simplify the integrand.