Some Commentary About Friday's Class

I think that all students in Math 103 have at least a basic understanding of integrals. But there are several different ways to think about single variable integrals; most have at least some use, but some of them have shortcomings. Here I will describe a few of them.

1. **Area**

   Perhaps the most geometrically satisfying perspective on single variable integrals involves area. In particular, when f is a non-negative continuous function, we can think of

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

   as representing the area between the graph of f and the x-axis, between $x = a$ and $x = b$.

   But what about in the cases when the function f is not non-negative? Well, we all know that there is a way to deal with this. Namely, we end up conceding that the integral doesn't actually represent simply area – instead, we count the area as being negative if the curve is below the x-axis...

   What??? How did this happen? If we claim to think of an integral as being area, then how do we find ourselves in a situation where sometimes area counts as being negative? Negative area is certainly not a natural geometric concept.

   The fact is, even though there is a somewhat satisfying geometric picture relating integrals to area, they are NOT the same! And the relationship between the two is arguably one of the less natural interpretations of an integral.

2. **The Fundamental Theorem**

   Some students that are reliant on computational tools are under the mistaken impression that an integral of a continuous function f is defined as

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

   where F is an antiderivative of f, sometimes written as

   $$\int f(x)\,dx$$

   This is an important fact about integrals, no doubt. But this also is NOT the definition of an integral! Of course the equation above represents a true theorem, called the Fundamental Theorem of Calculus. But it is important to realize that there is a difference between a theorem and a definition! The above is a theorem, that can be proved – definitions can NOT be proved; all we can do is motivate a definition.

3. **Indefinite Integrals**

This unfortunate term means the same as the term "antiderivative". My best guess is that the term was actually motivated by the notation typically used to represent antiderivatives, namely

$$\int f(x)\,dx$$

Since the same symbol "$\int$" that is used to represent integrals is used to represent antiderivatives, it might not seem unreasonable to use the word "integral" to represent antiderivatives as well... and to avoid confusion, they throw in the word "indefinite", since an antiderivative has no bounds.

In my mind, this sloppiness in semantics is very unfortunate. Using the symbol "$\int$" to represent antiderivatives is bad enough (surely they could have dreamed up some other symbol!)... but there simply is no reason to then create a new term, "indefinite integral" when we already had a perfectly satisfactory and descriptive term, "antiderivative".

For another thing, it is NOT an integral!

For yet another, this abuse can make the fundamental theorem confusing... While it seems worthwhile and not obvious to have a theorem claiming one can compute integrals by using antiderivatives, it seems much less impressive and "fundamental" to claim that one can use integrals to compute, well, integrals.

4. **Riemann Sums**

While Riemann sums may have originally been used for the purpose of computing areas under curves, the fact is that Riemann sums themselves are really much more interesting than area is. Not only can they be used to define and compute the area under a curve, but they can also be used to represent and then compute an enormous variety of other things.

Riemann sums give us the most fundamental and meaningful definition that we can give for an integral of a continuous function in this class. Namely,

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

This should be familiar to students in this class.

The symbol "$\int$" can be thought of as representing the "$\lim \sum$" on the right side of the equation; and similarly the "dx" can be thought of as representing the "Δx" on the right side. So, an integral most fundamentally is a sum – a special kind of sum involving a limit.

Of course this relates to the area under the curve in a familiar way. For a given value of n, the sum is in fact the sum of the areas of rectangles that closely approximate slices of the area between the curve and the function; and as n gets larger, the differences between the rectangles and the slices of area become more negligibly small.

From this perspective we can see why it is that we were forced to think of the integral as being negative when f is negative... because for those terms in the Riemann sum, we have the product of a negative f and a positive Δx.

5. **A Different Interpretation of Riemann Sums and Integrals**

The Riemann sum above is written in such a way that it can be clearly interpreted as having to do directly with area; note that each of the terms in the sum is a product of a height (a value of the function) with a width (the width of a subinterval of $[a, b]$).

We might write it thus as

$$\lim \sum \Delta A$$

where $\Delta A = f(x_i)\Delta x$ (I've suppressed the indices of the limit and the sum, since they are usually the same.)

Of course since we think of "$\int$" as representing "$\lim \sum$", we might rewrite this as

$$\int dA$$

This is a shorthand, of course – the "dA" actually represents $\Delta A = f(x)\Delta x$. But it is a shorthand for a Riemann sum, and a Riemann sum is the definition of an integral... so, this is a reasonable shorthand to use.

From this perspective, the complete computation of area, from the beginning, might go something like this: We have the interval $[a, b]$, which we subdivide into intervals of width dx. To an arbitrary such subinterval $[x, x + dx]$, we associate the slice of area that is directly above it, whose area we will call dA. This area dA has height $f(x)$ and width dx, so we have $dA = f(x)\,dx$. Thus we obtain the total area A as

$$A = \int dA = \int f(x)\,dx$$

The interval of x with which we have associated these pieces of the area we are computing is $[a, b]$, so we conclude the final integral that this shorthand has helped us to determine is

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

6. **Using Integrals and Riemann Sums to Represent Other Things**

But we don't have to interpret a Riemann sum as a sum of little pieces of area. For example, we could think of it as a sum of little pieces of volume, or a sum of little pieces of length, or little pieces of work, or little pieces of force,...

In fact, we can use a Riemann sum to represent the sum of any collection of things that we individually associate to little pieces $[x, x+dx]$ of the interval $[a, b]$. And as above, we can use integral notation as a shorthand for Riemann sum notation.

For example, let's compute the volume of the cone we obtain by rotating the line $y = 3x$ around the x-axis, between $x = 0$ and $x = 2$:

We think of the interval $[0, 2]$ as being broken up into tiny subintervals $[x, x + dx]$. To each of these subintervals we will associate the slice of volume that we obtain by slicing perpendicularly to the x-axis.

The volume V is the sum of these little pieces of volume dV. Each of these slices is a cylinder with area A and width equal to the width of the corresponding subinterval, which has width dx. The area A of course is a circle, with radius y.

So our computation becomes simply

$$V = \int dV = \int A\,dx = \int \pi y^2\,dx = \int_0^2 \pi(3x)^2\,dx$$

All of the intermediate steps are shorthands for Riemann sums, and once we have it simplified into a form with only a single variable, we include the bounds, which are just the bounds of the interval to which we associated the slices of volume in the first place.

As another less geometric example, let's compute the mass of a straight rod of length 6 (which we will think of as being the interval $[0, 6]$), with mass density given as $\rho(x)$ at a point x on the rod.

To a subinterval $[x, x + dx]$, we associate the mass contained within that subinterval, which we will write as dm. That mass is the density times the length, so we have

$$m = \int dm = \int_0^6 \rho(x)\,dx$$

7. **The Area Under A Parametric Curve**

In class we talked about how to compute the area under a parametric curve given by $x = f(t)$, $y = g(t)$, $t \in [\alpha, \beta]$.

The first approach we took was to assume that the curve passed the vertical line test, and thus that we could interpret y as being a function of x, even if we could not realize that algebraically. Given this assumption, we could then write

$$A = \int dA = \int_{f(\alpha)}^{f(\beta)} y\,dx$$

Of course we could not evaluate this directly since we don't know any direct algebraic relationship between x and y. However, if we change variables by using the substitution $x = f(t)$, the y simply becomes $g(t)$ (because $y = y(x)$ is the value of y when x is $f(t)$, namely, the value of y at time t, which is $g(t)$).

So the integral becomes

$$A = \int dA = \int_{f(\alpha)}^{f(\beta)} y\,dx = \int g(t)d(f(t)) = \int_{\alpha}^{\beta} g(t)f'(t)dt$$

Of course this makes the assumption that the curve passes the vertical line test. And of course, a parametric curve is under no obligation to do so! So, what do we do if the curve fails the vertical line test?

Well, before we answer that question, let's talk about a different way to approach the problem we just solved above. So for the moment, let's again assume that the curve passes the vertical line test.

Instead of writing down the area first as an x-integral, let's just directly write the area as a t-integral. We know we are interested in the interval $t \in [\alpha, \beta]$, so all we have to do is associate to every subinterval $[t, t + dt]$ a piece of the area that we are interested in.

Here's how we will do it: The image of $[t, t + dt]$ is the interval $[f(t), f(t + dt)]$ on the x-axis; and above that interval, there is a slice of area. That is the slice of area we will associate to $[t, t + dt]$, and that we will call dA.

In order to write down the integral, we are going to need to be able to write down the height and width of that slice of area. The height is no problem of course, it is just the height of the curve, namely the value of y, when $x = f(t)$... which of course is $y = g(t)$.

The width can be thought of two ways. Algebraically, one can say

$$f(t + dt) - f(t) = \frac{f(t + dt) - f(t)}{dt}\,dt = f'(t)\,dt$$

Or one can interpret $f(t + dt) - f(t)$, the change in f, as just being df, which is $df(t) = f'(t)\, dt$.

Knowing the height and width of the slice in question, we can then just write the area directly as an integral in terms of t:

$$A = \int dA = \int g(t)\, df = \int_\alpha^\beta g(t) f'(t)\, dt$$

Note that we have the same answer we got before. But in this case, the approach we took does not rely on the assumption that the curve satisfies the vertical line test. So, while we don't know what exactly this expression represents in the case that the curve does something more interesting, we can at least motivate some sense in writing it down.

So far in the computations we have been implicitly assuming that the parametric curve is moving to the right... in other words, that $f(t)$ is an increasing function. Based on this assumption, we conclude that the width (a distance, and thus a positive number) is actually $f'(t)\, dt$, because both of those factors are positive.

But what happens if the curve is moving to the left? Namely, what happens if f' is negative? In that case, the width is the negative of $f'(t)\, dt$, because the width has to be positive and the expression above is negative. So, $dA = -f'(t)\, dt$, and thus $-dA = f'(t)\, dt$.

So, whenever the curve is moving to the left, the integral we have written down is not counting pieces of area, but instead is counting negative pieces of area.

Let's now think about the curve we drew in class that failed the vertical line test. The curve started moving to the right, then it turned down and around back to the left, and then again down and around back to the right. Let's say the curve is moving to the right for $t \in [\alpha, t_1]$, to the left for $t \in [t_1, t_2]$, and then to the right for $t \in [t_2, \beta]$.

Let's consider what we get when we compute the integral

$$\int_\alpha^\beta g(t) f'(t)\, dt$$

On the first piece while the curve is moving to the right, we pick up all of the area under that part of the curve. As we proceed through the second part of the curve where it is moving to the left, the integral picks up the negative of the area under that part of the curve... this effectively "subtracts off" the area underneath that part of the curve that had been counted in the first part, and effectively leaves only the part between the first two pieces of the curve. Proceeding with the third part of the curve, moving to the right, the integral adds in the area under that part.

What we are left with is, surprisingly, exactly the area between the curve and the x-axis, even though the curve can not be represented as the graph of a function with y as a function of x!

Similar reasoning will allow one to conclude what happens in the case that a curve loops around on itself, or does other interesting things. Students should draw a few interesting curves, and try to conclude what the integral above would compute in those instances.