Lecture for Week 13 (Secs. 6.1–2)

Sums and Areas

The big sigma ($\sum$) is nothing to panic over. It's just a notation, and a very convenient one. For example,

$$\sum_{j=1}^{n} j^2$$

means what we would otherwise write as

$$1 + 2^2 + 3^2 + \cdots + n^2.$$

One advantage of this notation is that it facilitates letting the number of terms, n, be a variable instead of a definite number. (Later, n will sometimes be replaced by ∞.) Another advantage is that the *summand*, j^2, provides a "pattern" or "template" for the terms, whereas the expression with the dots is rather vague. (Somebody else might have a different idea from you of how to fill in the missing terms.)

Next semester will pay a lot of attention to sums, especially those with infinitely many terms (*series*). For the moment, however, our concern is just with defining integrals as limits of sums, and thence using sums to approximate integrals and vice versa. (This is analogous to the use of derivatives to approximate finite differences ($\Delta y = y(x + \Delta x) - y(x)$) and vice versa.)

For the sake of having exactly computable examples of integrals, textbooks make use of formulas for the sums of powers of the integers. The only one of these worth memorizing is

$$\sum_{j=1}^{n} j = \tfrac{1}{2}n(n+1).$$

It has a famous, easy proof, reportedly discovered by Gauss when he was in grade school (see p. 366).

Integrals are closely related to areas, so we start by talking about area. The basic principle is that the area of a region is the sum of the areas of (disjoint) regions that make it up (see the third part of Fig. 2, p. 370, for instance). That allows us to define the areas of arbitrary polygons in terms of the known formulas for the areas of triangles and (most fundamentally) *rectangles*.

For a shape with curved boundaries, we cut the inside up into rectangles, but we can't quite do that for the region right next to the boundary. The best we can do is to get better and better approximations by taking smaller and smaller rectangles. (When Archimedes started this subject, he used other kinds of polygons (remember p. 41), but the end result is the same.)

We are mostly concerned with the "area under the graph" of a function, as in all the Figures in Sec. 6.2 *except* Fig. 2. That is, we assume (for this week only) that the values of $f(x)$ are nonnegative, and consider the region bounded above by the graph of f, below by the horizontal axis, and on the sides by two vertical lines, $x = a$ and $x = b$. Then it is natural to approximate the region by a bunch of narrow vertical rectangles, side by side.

For a first try, let's assume that all the rectangular strips have the same width,

$$\Delta x = \frac{b-a}{n} \quad \text{if there are } n \text{ strips.}$$

And, let's say that the height of each strip is the value of the function at the right end of the strip (so the graph passes through the upper right corner of each rectangle, as in Fig. 3). Then the total area of the strips is

$$\sum_{j=1}^{n} f(a + j\Delta x)\Delta x.$$

So we intuitively expect the area of the curved region to be

$$\lim_{\Delta x \to 0} \sum_{j=1}^{n} f(a + j\Delta x)\Delta x.$$

In fact, we would like to take this formula as the *definition* of the area.

What's wrong with that? Well, mathematicians have a responsibility to check that results of definitions don't depend on arbitrary elements in the definition in some silly way. For example, what if we evaluated the function at the *left* ends of the strips instead of the right?

More seriously, suppose we consider the area under the graph between a and c and also the area between c and b ($a < c < b$). Then the area between a and b should be the sum of those, right? But that is not obvious from our definition, if $b - a$ is a rational number and $c - a$ is irrational, say.

To sweep all such worries under the rug right at the start, the standard definition allows the widths of the strips to be different, provided that they all go to 0 in the end. Also, the height of a strip is allowed to be the value of $f(x)$ at *any* x in that strip's base. (That is, the graph must cut through the top of each rectangle.)

The price of this improved approach is a need for a more cumbersome notation. We must assume that the interval $[a, b]$ is divided into small intervals by points x_j where

$$a \equiv x_0 < x_1 < x_2 < \cdots < x_{n-1} < x_n \equiv b.$$

The width of the jth strip is $\Delta x_j = x_j - x_{j-1}$. The height of the jth strip is $f(x_j^*)$ for some x_j^* in the interval $[x_{j-1}, x_j]$. Then the total area of the strips is

$$\sum_{j=1}^{n} f(x_j^*)\Delta x_j \, .$$

Finally, therefore, we can define the area under the graph to be

$$A = \lim_{\|P\|\to 0} \sum_{j=1}^{n} f(x_j^*)\Delta x_j \, ,$$

where $\|P\| = \max \Delta x_j$. That is, we consider *infinitely many* ways of cutting up the area into strips, and we demand that the limit must be the same for *all possible* choices of the numbers x_j and x_j^*, subject only to the condition that the width of all the strips goes to zero in the limit.

For the function a few slides back that behaved differently at rational and irrational numbers, the limit will not exist — the sum is wildly different depending on whether the x_j^* are rational or not. So we get rid of that pathological function by saying it just doesn't have an area.

For "sensible" functions (in particular, continuous ones) all the different choices of partitions (x_j, x_j^*) will give the same limit for A. In particular, numerically it doesn't matter whether we choose $x_j^* = x_j$ (the *right sum* I discussed earlier), or $x_j^* = x_{j-1}$ (the *left sum*), or $x_j^* = \frac{1}{2}(x_{j-1} + x_j)$ (the *midpoint sum)*, or choose x_j^* so that $f(x_j^*)$ is the smallest value of f on the j interval (*inscribed* rectangles, yielding the *lower sum*), or the largest value (*circumscribed* rectangles, yielding the *upper sum*).

Note that for many standard, simply calculated examples, f is an increasing function, and therefore the upper sum is the same thing as the right sum, the lower sum same as the left sum (see Figs. 11 and 12). In general, though, the left sum will be neither an upper nor a lower sum, etc. (see Fig. 10).

Having said all these things about areas, I won't need to say them again about definite integrals.