MATH-115 (DUPRÉ) SPRING 2010 LECTURES

1. **LECTURE** MONDAY 11 JANUARY 2010

Today we discussed the basic course policy and where to find the syllabus online when it is posted. My office is Gibson Hall Room 309 and my office hours will be Monday, Wednesday, and Friday, 9 AM to 10 AM or 1 PM to 2 PM. I will often be in my office at other times, so if you need to see me outside of class just come up to my office. If I am not there, look for me in the Math Department Library or elsewhere on the fourth floor.

We discussed some basic problems that calculus deals with, the problem of tangency and velocity as well as the problem of finding areas and volume when the boundaries are curved. We discussed the fact that the two problems are related and their relation is the Fundamental Theorem of Calculus. We noted that if a two dimensional region has a curved boundary but is "long and skinny", then its area is very accurately estimated as length multiplied by width. For instance, we thought of the problem of computing the area of the center stripe of a highway passing through mountains. For a straight highway the area is simply length multiplied by width. But what is the effect of curvature on the area of the median stripe? We can think of the effect of curvature on area and volume for long skinny things as related to how flexible they are. A thick rope cannot bend as easily as thread. Thin copper wire bends more easily than thick copper wire. This is because to bend something it must contract on one side and stretch on the other. How much contracting and stretching for a given curve depends on the thickness. Thus as the thickness goes to zero, the we find complete flexibility. This is due to the fact that the area is very accurately approximated as length multiplied by width for a region which is very thin.

We discussed the idea that the rate of change of area due to a moving boundary is simply the length of the moving boundary multiplied by the "outward" velocity of the moving boundary. Thus, we figured that when a beach is eroding at a certain rate, the length of the beach front multiplied by the erosion rate gives the rate of land loss. The same considerations would apply for instance to the rate at which the area of an oil spill on the ocean increases due to the boundary of the spill moving out from the spill. If the spill is contained except for a ten mile length of boundary and that boundary is moving out at two miles per hour at a given instant, then at that instant, the area of the spill is increasing at a rate of twenty square miles per hour.

2. **LECTURE** WEDNESDAY 13 JANUARY 2010

Today we began by discussing the general idea of a **function** or **map**. If S and T are sets, then a **FUNCTION** f from S to T, denoted $f : S \longrightarrow T$, is a **RULE** which assigns to each member x of S, a member $f(x)$ of the set T. Functions are also called maps or mappings. This is a very general concept, since we have absolutely no restrictions on what sets can be used for making functions and no restrictions on the rules that can be used to define functions. For instance to start, we let S be the set of points on a smooth surface and let T be the set of all planes in three dimensional space. We defined the function $f : S \longrightarrow T$ in this case to be the rule which assigns to the point x in the surface S the plane $f(x)$ which is tangent to S at the point x. As another example, we defined L to be the set of lines in three dimensional space and defined $g : S \longrightarrow L$ by requiring $g(x)$ be the line through x perpendicular to the tangent plane $f(x)$. It is customary to call $g(x)$ the *normal line* to S at x.

We noted that it is useful to think of a function as an input-output device so when $f : S \longrightarrow T$, then S is the set of allowable inputs for the device f and T is a set which contains the outputs. If $g : T \longrightarrow U$, then we can take the output of f and use it for the input of g and the result is a rule which takes inputs in S and gives outputs in U. This combined rule is called the **COMPOSITE** of f and g, denoted $g \circ f$. Thus,

$$(g \circ f)(x) = g(f(x)).$$

If S is any set, there is always the **IDENTITY MAP** of S denoted id_S which is simply the rule

$$id_S(x) = x,$$

for every x in S.

We will usually be concerned only with the situation of functions $S \longrightarrow T$ where S and T are sets of numbers. In this case we can often picture the function with its graph. If f, g are functions on any set S with real number values, then we can form $f + g$ and fg as well as their quotient f/g. The rules are simply

$$(f + g)(x) = f(x) + g(x),$$

$$(fg)(x) = f(x)g(x),$$

and

$$(f/g)(x) = \frac{f(x)}{g(x)},$$

and where we note that we have to replace S by the smaller set consisting of those members x of S for which $g(x) \neq 0$, in the last case for the quotient.

3. **LECTURE** FRIDAY 15 JANUARY 2010

Today we began by reviewing the general definition of a function or mapping given in the previous lecture. We then defined union and intersection for arbitrary sets. We also defined the **CARTESIAN PRODUCT** of arbitrary sets. Given any sets A and B, which includes the possibility they are the same or maybe have some members in common. Their union is denoted $A \cup B$, their intersection is denoted $A \cap B$, and their cartesian product is denoted $A \times B$. As for the definitions,

$A \cup B = \{x \text{ such that } x \text{ is in } A \text{ or } x \text{ is in } B\}$,

$A \cap B = \{x \text{ such that } x \text{ is in both } A \text{ and } B\}$,

$A \times B = \{(x, y) \text{ such that } x \text{ is in } A \text{ and } y \text{ is in } B\}$.

Keep in mind that for ordered pairs (a, b) and (h, k), by definition, equality of ordered pairs

$$(a, b) = (h, k)$$

means that both $a = h$ and $b = k$.

If $f : S \longrightarrow T$ is any function, then we define the graph of f denoted $Graph(f)$ as the subset of $S \times T$ given by

$Graph(f) = \{(x, y) \text{ in } S \times T \text{ such that } y = f(x)\}$.

We noted that this means every function can be simply viewed as a subset of a cartesian product, but not every subset of a cartesian product is the graph of a function. For the subset G of $S \times T$ to be the graph of a function, it is necessary that every member of S is the first entry of an ordered pair in G and no two different ordered pairs in G can have the same first entry. When these conditions are satisfied, then G defines the function $g : S \longrightarrow T$ where for x in S to find $g(x)$ we search through all the order pairs in G until we find the one ordered pair whose first entry is x, so this ordered pair is (x, b) where b is in T. Then $g(x) = b$.

We discussed the fact that the general idea of a function is just too general and allows too many pathological examples. At first, for calculus, we only consider functions $f : S \longrightarrow T$, where S and T are sets of numbers. In fact, we want to restrict our attention to functions whose graphs are curves in the plane which are "continuous" and "piecewise smooth", pictorially. For the tangency problem then we want to look at the simplest functions whose graphs are straight lines, since the tangent line to a curve in particular is a straight line. We call a function **LINEAR** if its graph is a straight line. Since a line is determined by two points, we used that fact and similar triangles to arrive at the **POINT-SLOPE** form for the equation of a line:

$$y = m(x - a) + b.$$

This line obviously has slope m and passes through the point (a, b). In more detail, if (x, y) and (a, b) are two points (given in coordinates) on the line L in the plane, then using similar triangles, we notice that the ratio

$$m = \frac{\Delta y}{\Delta x} = \frac{y - b}{x - a},$$

is the same no matter what the two points are, as long as they are both points on L. Thus, the point-slope form of the equation of the line is just the result of multiplying out the denominator $\Delta x = x - a$.

We discussed examples beginning with unit conversion from Celsius to Fahrenheit for temperature and unit conversion in general. We then observed that in looking at the picture of the tangent line to a graph, that in many applied situations where we know the relation between two quantities is reasonably smooth, that for small changes in input, the output is approximately on the tangent line which is linear. Thus, two nearby values can determine the tangent line approximately giving a linear function which can be useful over a limited range of inputs.

We also observed that in many situations where we have linear functions to deal with, we have $f(0) = 0$, so to find the slope m it suffices to know $f(a)$ for a single non-zero number a. We worked examples showing that the area $A(s)$ of a slice of pizza of radius R whose edge arc length is s must be given by the simple formula

$$A(s) = \frac{1}{2}Rs.$$

Likewise, we showed that if we have a region on the surface of a solid ball of area A, then it subtends a "conical" shaped solid at the center of the ball whose volume $V(A)$ is a linear function of A, and the result was

$$V(A) = \frac{1}{3}RA.$$

You should notice the similarity in the two formulas, and suspect that the number in the denominator has to do with the dimension we are dealing with in each case.

Finally, we noted that most numerical quantities do not have any natural units-units are somewhat arbitrary, but in the case of angles, the most undeniably natural unit is the revolution, since we can all agree on what one revolution is, without any need to measure. The degree is then defined as $(1/360)$ revolution and better for our purposes will be what is called the **RADIAN** measure for angle. One radian is by definition $(1/2\pi)$ revolution, which is a little less than 60 degrees, around 57 degrees.

Next time we will discuss how to find the arc length of the segment of circular arc given the angle and radius.

4. **LECTURE** MONDAY 18 JANUARY 2010

NO CLASS MEETING TODAY BECAUSE OF HOLIDAY (MARTIN L. KING, JR)

5. **LECTURE** WEDNESDAY 20 JANUARY 2010

Today we discussed rate of change and instantaneous rate of change and its relation to velocity. We observed that we are always free to think of the graph of any function as illustrating a motion of a bead on a vertical wire, no matter what the graph is originally trying to express. We noted that when two curves meet tangentially, the corresponding motions have the same instantaneous velocity at the meeting point whereas when two curves cross each other it represents a motion where one object passes right through the other in ghost-like fashion. This means the tangent line velocity must be the same as the instantaneous velocity at the point of tangency. We can think of the bead on the wire as having a speedometer which then tells its instantaneous speed at each instant, and its velocity is positive when the bead moves in the upward direction and negative when moving downward. In particular, we noted that straight lines represent motion at constant velocity, and that the slope of the line gives that constant velocity. In particular, since a horizontal line represents a motion consisting of a bead sitting at a fixed point without moving, this means that any constant function has velocity zero. Since our graphs are graphs of number valued functions of input numbers, we can make sense of $f + g$, fg, and f/g whenever we have to functions f and g. By thinking of f as giving the vertical motion of a spaceship and g as giving the vertical motion relative to the spaceship of a bead on a vertical wire fixed in the spaceship, we reasoned that $h = f + g$ gives the motion of the bead relative to the fixed vertical axis used to locate the spaceship itself. Likewise, if $v_f(t)$ gives the velocity for the motion f at time t, then clearly for the bead on the wire in the spaceship it is the case that $v_h(t) = v_f(t) + v_g(t)$. For instance, if an astronaut in the spaceship thinks the bead is moving up at 2 units per second and if an outside observer sees the spaceship moving up at 7 units per second at that same instant, then the outside observer would think the bead is moving up at the rate of 9 units per second. This means it must be the case that at any time t we have

$$v_{(f+g)}(t) = v_f(t) + v_g(t)$$

which is called the simple **ADDITION RULE** for velocities.

For multiplication the situation is a little more complicated. Here, we begin by imagining an ink spill which is contained on part of its boundary by weighted rubber hoses, but where there are three separate lengths of boundary, B_1, B_2, B_3 which are not contained and are thus moving out as the spill expands, so they are functions of time. Assume that B_1 has length l_1, that B_2 has length l_2, and that B_3 has length l_3. Also assume that the boundary B_k is moving out in the direction normal (which means perpendicular) to the boundary at each point with velocity v_k. Let A_k be the area increase due to the spreading spill through the boundary B_k. Consequently, $l_1, l_2, l_3, v_1, v_2, v_3$, and A_1, A_2, A_3 all depend on time. Let A_0 be the area to start at time $t = 0$, so the area at time t is simply

$$A(t) = A_0 + A_1(t) + A_2(t) + A_3(t).$$

Notice that in a very small duration of time dt, from t to $t + dt$, the area A_k increases by an amount roughly $l_k(t)w$, where w is the width of a narrow strip along the boundary. The w is of course due to the boundary moving during the small amount of time, and since its velocity in the outward direction is v_k, it must move out by an amount $v_k dt$, as distance moved is velocity multiplied by elapsed time, and we can assume that dt is so small that the velocity stays the same during this small amount of time. The area increase due to increasing A_k or leakage along B_k is therefore just the tiny amount dA_k given by

$$dA_k = l_k w = l_k v_k dt.$$

This means that the rate of change of area or the velocity of the area function A_k is

$$v_{A_k} = \frac{dA}{dt} = l_k v_k,$$

so by our addition rule for velocities, the sum total rate of increase of area of the ink spill is

$$v_A = v_{A_0} + v_{A_1} + v_{A_2} + v_{A_3} = v_{A_0} + l_1 v_1 + l_2 v_2 + l_3 v_3.$$

Of course, as A_0 is just a constant, we know that $v_{A_0} = 0$, so at each instant, in more detail,

$$v_A(t) = l_1(t)v_1(t) + l_2(t)v_2(t) + l_3(t)v_3(t).$$

Notice that there could be many pieces of moving boundary, and at each instant to compute the rate that area is increasing we just need to know at that particular instant for each piece what the length of that piece is and its normal velocity, again all at that particular instant. We then just multiply each length by the velocity at which it moves and then add up all the results. In fact, if the velocity is varying along the moving boundary, we could imagine that boundary as consisting of very many very tiny bits of boundary, each so small that along each little bit the boundary is moving all at the same velocity, and then apply the same procedure, but with many bits of boundary instead of only three. Of course this would be very laborious, so we will not do it here. However, this result for ink spills can apply to any plane region which is varying because some of its boundary pieces are moving outward (or inward, which would be negative boundary velocity). In particular, we can apply this to find the velocity of a product fg. Imagine a rectangle whose sides are along the x and y axes with one corner at the origin of coordinates and fixed there. The opposite corner at time t is at the point with coordinates $(f(t), g(t))$. Thus, as time passes, the rectangle changes its size and shape. Notice that at time t, the edge along the vertical line through $(f(t), 0)$ has length $g(t)$ at time t and thus this vertical edge moves at velocity $v_f(t)$. On the other hand, the horizontal edge of the rectangle, in the horizontal line through $(0, g(t))$, has length $f(t)$ at time t and moves with velocity $v_g(t)$. The area of this rectangle at time t is just

$$A(t) = f(t)g(t),$$

so we must have

$$v_{fg}(t) = v_A(t) = v_f(t)g(t) + f(t)v_g(t),$$

which we call the **PRODUCT RULE** for velocities. Notice this is saying that to find the velocity of the product we do not simply multiply the velocities. We will however find a much more useful rule for velocities which does involve multiplication when we deal with composite functions.

6. **LECTURE** FRIDAY 22 JANUARY 2010

Today we discussed functions of several variables and the first section of chapter 9 in the text as well as the rules we have so far for working out velocity and rate of change.

In general, if S and T are sets, remember a function $f : S \longrightarrow T$ is a rule which assigns a member denoted $f(s)$ in the set T to each member s of the set S. In this situation, we call S the **DOMAIN** of the function f, and we call T the **CODOMAIN** of the function f. We will primarily deal with the case where S and T are sets of real numbers, but more generally than that is the situation where T is a set of real numbers, but S is some set whose members can be described by numerical systems of tags. For instance, if S is a set of points on the surface of the Earth, then each point can be assigned a longitude and a latitude giving two numbers (x, y) which specify the point. More precisely, if p is a point on the surface of the Earth, then we can denote by $x(p)$ its longitude expressed in radians and denote by $y(p)$ its latitude. You will notice here that actually x and y are real number valued functions whose domains are the set of all points on the surface of the Earth. Another useful real number valued function here could be z where we define $z(p)$ to be the height above sea level, so when $z(p)$ is negative, you are below sea level if you are located at the point p. As well, we could consider the function g where $g(p)$ is the temperature in degrees Celsius at the point p. Now, to picture the function g, over a limited region of the Earth's surface such as over the surface of the United States, we can use the functions x and y to make a map of the United States on the flat plane with rectangular coordinates (x, y). Then in three dimensional space, we imagine plotting over the point p with coordinates (x, y) a point at height $z = g(p)$, The result of doing this is to create a surface in three dimensional space which represents the temperature function over the whole United States. For instance, we expect the points on this surface to be relatively high over points of Texas and Florida and much lower over points up North. Notice that over the United States, we might as well regard the temperature as depending on the two real number variables x and y simultaneously. That is, here we could write $z = g(x, y)$. If we want to study how temperature varies as we move about, to describe the rate of change of temperature near a specific point with coordinates (x_0, y_0), we could just hold the $y-$coordinate fixed at value y_0, and see how temperature varies as we allow x alone to vary. We have in effect then formed a function of one real number variable $f(x) = g(x, y_0)$. The rate of change of this function f is the rate of change of temperature in the East-West direction as we move towards the West (increasing longitude moves you West), certainly says something about the temperature variation. If we hold x fixed at the value x_0 and allow only y to vary, we find a new function $h(y) = g(x_0, y)$ and its rate of change tells us the rate of change of temperature in the North-South direction as we move North. For instance the coordinates of New Orleans are $(\pi/2, \pi/6)$, and from here as we move West we might expect the temperature to rise so the East-West rate is a positive number whereas if we move North we expect the temperature to fall, so we would expect the rate of change in the North-South direction to be negative. If we move a little bit from New Orleans which involves both some Westward and Northward movement, then it turns out that for small moves we can add the temperature change resulting from a pure Westward move to the temperature change from a pure Northward move to get the total resulting temperature change. For instance if moving Δx to the West results in a two degree increase in temperature and moving Δy to the North results in a 5 degree drop in temperature, then at the point with coordinates $(\pi/2 + \Delta x, \pi/6 + \Delta y)$, we would expect the temperature to drop by three degrees.

In general, we could have a function of many variables and to study its rate of change at a specific point in its domain, we can fix all but one of the variables and let only that variable change to work out the rate of change with respect to that variable alone. Doing that for each of the variables gives us the overall rate of change of the function. In applications, most functions we need to deal with have many variables, but can be analyzed in this way by reducing to many different functions of a single variable. This means that even though almost all the problems we will deal with only involve a single variable, the results can then be used to deal with practical problems involving many variables.

Last time we worked out some simple rules for figuring out velocities of motions given as functions. Thus, we observed that if f is any function, then we can think of its graph as giving a one dimensional motion and the slope of its tangent line at the point $(t, f(t))$ is the velocity of that motion which we denoted by $v_f(t)$. As this notation is somewhat cumbersome, we will henceforth use the notation

$$f'(x) = v_f(t).$$

Then we view $f'(t) = v_f(t)$ as giving us the **RATE OF CHANGE** of f at the input t. Notice that we in effect have a new function denoted f' whose value $f'(t)$ at t tells us the slope of the tangent line to the graph of f at the point $(t, f(t))$. The function f' is also called the **DERIVATIVE** of f and the process of passing from f to f', that is the process of finding f' from f is called **DIFFERENTIATION**.

If f and g are any two functions, we observed that $f + g$ can be viewed as giving the motion of a bead moving inside a spaceship as observed outside the spaceship when f gives the motion of the spaceship and g gives the motion as it would be seen by an astronaut inside the spaceship. Remember the result is that

$$v_{(f+g)}(t) = v_f(t) + v_g(t)$$

which in our new notation becomes

$$(f + g)'(t) = f'(t) + g'(t)$$

and therefore as functions we have simply

$$(f + g)' = f' + g'.$$

Notice this says that to differentiate a sum of terms you just differentiate each term and sum the results. Obviously then, the same applies no matter how many terms there are. We call this the **ADDITION RULE FOR DIFFERENTIATION**.

For dealing with the product fg of the functions f and g, we noticed that we can view the rate of change as the rate of change of the area of a rectangle. Since we knew how to find rates of change of areas of ink spills in terms of velocities of moving boundaries and their lengths, we found that

$$v_{(fg)}(t) = v_f(t)g(t) + f(t)v_g(t).$$

In our simplified notation, this becomes

$$(fg)'(t) = f'(t)g(t) + f(t)g'(t)$$

for each input t, and therefore

$$(fg)' = f'g + fg'.$$

This is then called the **PRODUCT RULE FOR DIFFERENTIATION**. It is tricky at first, but to remember it, notice when you see a product of two things that needs to be differentiated, you just differentiate the first and copy the second beside it, write down a plus sign, then copy the first factor and beside that finally put the derivative of the second factor:

$$DIFF\ A\ PRODUCT = (DIFF)(COPY) + (COPY)(DIFF).$$

Remember for any linear function the rate of change is just the slope, so if $f(x) = m(x-a)+b$, then $f'(x) = m$ for every x and this means that the derivative function f' is just the constant function with value m. Any constant function has for its graph a horizontal straight line which therefore has slope zero. Thus

$$c' = 0$$

for any constant c, or

$$(CONSTANT)' = 0.$$

Notice this means that if f is linear, as f' is constant, when we differentiate f' we must get zero:

$$f'' = 0, \ any\ linear\ function\ f.$$

In particular, the simplest linear function is the identity function on the set of real numbers, $f(x) = x$ or $y = x$ which we can often simply denote by x. Thus, as its slope is one, we have

$$x' = 1.$$

To calculate the derivative of x^2 whose graph is a parabola, we just use the product rule:

$$(x^2)' = (xx)' = x'x + xx' = 1x + x1 = x + x = 2x.$$

This means we have the rather simple result

$$(x^2)' = 2x.$$

To calculate $(x^3)'$ we just use the product rule again:

$$(x^3)' = (x^2x)' = (x^2)'x + x^2x' = (2x)x + x^2 = 2x^2 + x^2 = 3x^2,$$

so

$$(x^3)' = 3x^2.$$

There seems to be a simple rule here that we could simply follow in terms of powers. In fact if we use the product rule again, we find

$$(x^4)' = 4x^3.$$

This would lead us to guess that if we keep applying the product rule we find that for any positive integer power n we have

$$(x^n)' = nx^{n-1}.$$

This is called the **POWER RULE FOR DIFFERENTIATION** and it says simply take the exponent and move down as a coefficient and reduce the exponent by one. We observed that if we assume this rule is true for $n = 1000$, then it is also true for $n = 1001$, because of the product rule. In fact for any n, if we assume the power rule works for that n, then it also works for $n + 1$. What about $n = 0$? Notice that $x^0 = 1$ is therefore constant so its derivative is zero, $(x^0)' = 0$, and using the power rule for the case $n = 0$ would give

$$(x^0)' = 0x^{-1} = 0,$$

which is the correct answer, so the power rule also works for $n = 0$. On the other hand, we used the fact that x, the identity function has graph a straight line of slope one to get $x' = 1$ which is $(x^1)' = 1 = (1)(1) = 1x^0$, and this is the power rule for $n = 1$. If we had tried to use the power rule for $n = 0$ with the product rule to find x', we would have

$$x' = (x^1x^0)' = x'x^0 + x(x^0)' = x'1 + x0 = x'$$

which means we do not find out $x' = 1$, the only thing the product rule tells us here is that $x' = x'$ which we already knew. But the product rule and the power rule for $n = 1$ does imply the power rule for $n = 2$ as we showed above, and in fact for any n, if we assume the power rule works for that n, then the product rule tells that the power rule also works for $n+1$. This means that the power rule must work for all non-negative integers.

What we have here is a case of what is called the **PRINCIPLE OF MATHEMATICAL INDUCTION**. In general, suppose that you have an infinite sequence of statements

$$S_1, S_2, S_3, ..., S_n, S_{n+1}, ...$$

and suppose that you can show that for each n it is the case that S_n implies S_{n+1}. If you can show that S_1 is true, then it must be the case that S_n is true for every positive integer n.

The product rule also tells us how to easily differentiate a constant multiple of any function f which we already know how to differentiate. If we know how to differentiate f and we want the derivative of cf where c is any constant, then as $c' = 0$, the product rule here gives

$$(cf)' = c'f + cf' = 0f + cf' = cf'$$

which means constants simply come out front of the differentiation:

$$(cf)' = cf'.$$

Thus the derivative of $7x^4$ is just

$$(7x^4)' = 7(x^4)' = (7)(4x^3) = 28x^3.$$

Notice that in general to differentiate any term of the form cx^n we just get

$$(cx^n)' = (cn)x^{n-1}.$$

This is a slight generalization of the power rule and it says simply take the original exponent and multiply by the original coefficient to get the new coefficient and to get the new exponent just lower the original exponent by one. You should practice this rule so that you can do it quickly and effortlessly. Combined with the addition rule, you can differentiate any polynomial, so for example

$$(8x^5 + 4x^7 - 9x^{12})' = 40x^4 + 28x^6 - 108x^{11}.$$

Thus if f is the function

$$f(x) = 8x^5 + 4x^7 - 9x^{12},$$

and you want to know the equation of the tangent line to the graph of f at the point $(3, f(3))$, then to start you calculate $f(3) = -4772277$, so you know this tangent line passes through the point

$$(3, -4772277).$$

Next you need the slope of this tangent line which is $f'(3)$ which you can calculate using the expression just found above, namely

$$f'(x) = 40x^4 + 28x^6 - 108x^{11}.$$

We need to put $x = 3$ in this expression to get the slope of the tangent line. Therefore

$$f'(3) = -19108224,$$

We now know that the tangent line we are looking for has slope $m = -19108224$ and passes through the point $(3, -4772277)$. Using the point-slope form of the equation of a line we have now

$$y = [-19108224](x - 3) - 4772277$$

is the equation of the tangent line to the graph of f at the point of interest here.

7. **LECTURE** MONDAY 25 JANUARY 2010

Before beginning the lecture I reminded you all that we will have quizzes on Wednesdays, so this Wednesday will be our first quiz covering the first three sections of the book and what we have covered in the lectures on tangent lines and differentiation.

We began by reviewing our rules for differentiation from the preceding lecture and observing that the product rule for differentiation was a special case of the idea that when a plane region changes shape by having a moving boundary, to calculate the rate of change of area, for each piece of moving boundary we multiply its length by the velocity with which it moves out (in direction perpendicular to the boundary), and then add up all of these products from all the moving boundary pieces. To prove the product rule for differentiation, we used a rectangle with one side of length $f(t)$ and the other side of length $g(t)$ and observed that for the two straight edges which move, one has length $g(t)$ and moves out with velocity $f'(t)$ and the other has length $f(t)$ and moves out with velocity $g'(t)$, so as the area at time t must be $A(t) = f(t)g(t)$, we have

$$(fg)' = A' = f'g + fg'.$$

Last time we also used the product rule for differentiation to prove the power rule for differentiation. We also used the product rule to prove the general power rule for differentiating any monimial,

$$(cx^n)' = (nc)x^{n-1}.$$

Notice this includes the case of a constant multiple of a power, but the product rule tells us that

$$(cf)' = cf'$$

for any function f whenever c is a constant. Combined with the addition rule for differentiation,

$$(f+g)' = f' + g',$$

our rules so far allow us to differentiate any polynomial.

Remember the power rule says that for any non-negative exponent n, the power function $f(x) = x^n$ has derivative

$$f'(x) = (x^n)' = nx^{n-1}.$$

We remarked that even though we have not demonstrated the fact, that the power rule actually works for all power functions with any *real number* exponents. Thus, for instance, as

$$x^{1/2} = \sqrt{x},$$

we can use the power rule to differentiate the square root function. Thus

$$(\sqrt{x})' = (x^{1/2})' = \frac{1}{2}x^{-1/2} = \frac{1}{2\sqrt{x}}.$$

Next, we observed that for the general case of rate of change of area with moving curved boundaries, we needed to give a better argument that for a curved path of constant width W and length L the area is $A = LW$. To see this, we began by reviewing our formulas for area and arc length for sectors of discs. First, using linearity, we recalled that the area of a sector or radius R whose circular arc length is S, that the area is simply

$$A = \frac{1}{2}RS.$$

We also reviewed radian measure for angles and used linearity to work out conversion from degrees to radians. Thus, by definition, one radian is $1/(2\pi)$ revolutions or what is the same thing, one revolution is 2π radians. On the other hand, one degree is $1/360$ revolutions or one revolution is 360 degrees. We used linearity to work out the method of conversion back and forth from degrees to radians. Clearly doubling the angle doubles either measure of the angle, so they must be related by a simple linear relationship. In particular, for converting from radians to degrees, we must have a simple linear equation of the form

$$\alpha = k\theta,$$

where θ is the angle expressed in radians, and α is the same angle expressed in degrees. This means we only need to determine the constant k, and to do this we only need a single non zero angle for which we know both angle measures. But,

2π radians = one revolution = 360 degrees.

This means that

$$360 = k2\pi,$$

and therefore

$$k = \frac{180}{\pi}.$$

Notice also, that the equation gives us

$$\theta = \frac{1}{k}\alpha = \frac{\pi}{180}\alpha,$$

as the equation for converting from degrees to radians. Our two equations for converting back and forth are thus

$$\alpha = \frac{180}{\pi}\theta$$

and

$$\theta = \frac{\pi}{180}\alpha.$$

Notice that in either equation, if you imagine the units inserted, so next to the π is the word radian, next to the 180 is the word degree, and next to the θ is the word radian and next to α is the word degree, then in either conversion equation, the units cancel giving the same units on either side.

The reason for the radian measure of angle becomes clear when we work out the formula for arc length of a circular sector in terms of the angle of the sector and its radius. Keeping the radius fixed, we easily see that the arc length of the circular sector must be a linear function of the angle measure, so we must have a constant k with

$$S = k\theta,$$

where θ is the angle in radians. In case the angle is 2π, we know this is one full revolution, and therefore the sector is a full circle whose circumference is $2\pi R$. We therefore have for this case,

$$2\pi R = k2\pi,$$

so canceling the common factors we find simply $k = R$, and therefore

$$S = R\theta$$

is the formula for arc length of a circular sector of radius R making an angle of θ radians. The simplicity of this formula when the angle is expressed in radians is the main reason for the radian measure of angles, and is why this measure is pervasive in physics and engineering.

Recall, that we had used linearity to express the area of a sector of a disc in terms of the arc length of its circular sector boundary, and the simple formula was

$$A = \frac{1}{2}SR,$$

where S is the arc length and R is the radius. Combining this formula with the formula for the arc length of a circular sector we have

$$A = \frac{1}{2}SR = \frac{1}{2}(R\theta R) = \frac{1}{2}R^2\theta,$$

where R is the radius and θ is the angle of the sector expressed in radians.

We can imagine that if we have a pathway of constant width W in the plane, that we would measure its length as the length of the curve going right down the middle of the path. If such a path has for boundaries two circular sectors of radius r and R with common center and common angle θ, assuming $r < R$, then we must have the center line of the path is a circular sector whose radius is midway between r and R, that is it must be a circular sector of radius B which is the average of r and R,

$$B = \frac{1}{2}[r + R],$$

whereas the width of the pathway is simply

$$W = R - r.$$

Now, the arc length of the path center line is simply

$$L = B\theta = \frac{1}{2}\theta[r + R],,$$

whereas the area of the pathway is the difference in area of the two disc sectors,

$$A = \frac{1}{2}\theta R^2 - \frac{1}{2}\theta r^2 = \frac{1}{2}\theta(R^2 - r^2).$$

Now, we can factor $R^2 - r^2$ as

$$R^2 - r^2 = [R + r][R + r],$$

and substituting this into the area expression gives

$$A = \frac{1}{2}\theta[R + r][R - r] = \theta B[R - r] = LW,$$

and we see that the formula for area as length multiplied by width works even for pathways of constant width which bend along circular centerlines. Since we can imagine any path of constant width as having a centerline consisting of many many very very tiny circular sectors, this means that any pathway of constant width has area LW, where L is the length of the centerline and W is the constant width.

For the argument that the rate of change of area of an ink spill due to a piece of moving boundary of length L moving with velocity v normal to the boundary must be given by

$$\frac{dA}{dt} = Lv,$$

we need merely now observe, that the added area dA during the elapsed time dt is a very thin path whose centerline length is approximately L with the approximation ever better as the elapsed time goes to zero, so that as the width is vdt, its area is very very close to $Lvdt$, so as dt goes to zero in an appropriate sense,

$$dA = Lvdt,$$

and

$$\frac{dA}{dt} = Lv.$$

We also worked examples of finding tangent lines, using the fact that once we know the derivative function f' of the function f, we can then calculate the slope of the tangent line to the graph of f at any point $(c, f(c))$ with c in the domain of f', since the domain of f' is certainly contained in the domain of f. Using the slope of the tangent line at the specific point of interest allows us to write down the equation of the tangent line in point-slope form. Thus, to find the equation of the tangent line to the curve $y = x^2$ at the point $(3, 9)$, we are dealing with the function $f(x) = x^2$. So $f'(x) = 2x$, by the power rule and therefore $f'(3) = (2)(3) = 6$ is the slope of the tangent line to the graph of f. Consequently, in point-slope form, the equation of the tangent line is

$$y = 6(x - 3) + 9.$$

We observed that using the tangent line equation, we can approximate values of the function for inputs near the point of tangency. For instance, using the tangent line equation just found for $f(x) = x^2$, we know that $(2.05)^2$ is approximately

$$y = 6(2.05 - 2) + 9 = 9 + (6)(1/20) = 9.3.$$

We worked several examples using the functions x^2, x^3, and $x^{1/2}$. For instance, the tangent line approximation for the square of 2.05 is not very significant, as it can be worked out easily by hand. However, if we need an approximate value of the square root of 83, we can use the tangent to the graph of $f(x) = \sqrt{x}$ at the point $(81, 9)$. Since the power rule gave us

$$f'(x) = \frac{1}{2\sqrt{x}},$$

we have using $x = 81$, that the tangent line has slope

$$f'(81) = \frac{1}{18},$$

and therefore the tangent line to the graph of f at the point $(81, 9)$ is simply

$$y = \frac{1}{18}(x - 81) + 9.$$

Since when we look at the graph of f we can clearly see the tangent line stays very close to the graph of f, it appears that putting $x = 83$ in the tangent line equation should give a value for y which is very close to the square root of 83. This is obviously nine and one ninth. Likewise, nine and one sixth is approximately the square root of 84.

8. **LECTURE** WEDNESDAY 27 JANUARY 2010

Today we reviewed our work so far and had **QUIZ 1**. We will continue with differentiation and its applications on Friday and there may also be a quiz on Friday.

9. **LECTURE** FRIDAY 29 JANUARY 2010

Today we reviewed the rules for differentiation and discussed a new rule for differentiating composite functions called the **CHAIN RULE**. The Chain Rule says that to differentiate the composite $g \circ f$ we have

$$(g \circ f)' = (g' \circ f) \cdot f,$$

which means that for each x in the domain of this derivative we have

$$(g \circ f)'(x) = [g(f(x)]' = [g'(f(x))][f'(x)].$$

For instance, if we want to differentiate the function

$$h(x) = (x^7 + 3x^4 - 9)^{39},$$

we know we could with lots of work expand the thirty ninth power out and use our previous rules, as they can be used to differentiate any polynomial. However, it is much simpler to apply the chain rule here. We should recall that in general, whenever we see grouping symbols such as parenthesis, brackets, radicals, fractions, we might be dealing with a composite function. In our example, we can view h as the result of substituting

$$f(x) = x^7 + 3x^4 - 9$$

for y in the function g where

$$g(y) = y^{39}.$$

The chain rule says begin by differentiating g which we do with the power rule here,

$$g'(y) = 39y^{38}.$$

For the next step, the chain rule says we must substitute $y = f(x)$ in the derivative g' of g. This gives

$$g'(f(x)) = (39)(x^7 + 3x^4 - 9)^{38},$$

but this is not all. Next the chain rule says we need $f'(x)$, and using our rules we have easily

$$f'(x) = 7x^6 + 12x^3.$$

Finally, the chain rule says that to get the derivative of

$$h(x) = (g \circ f)(x) = (x^7 + 3x^4 - 9)^{39},$$

we need to multiply the result we have for $g'(f(x))$ by the derivative $f'(x)$ which was just computed for f. The result is

$$h'(x) = g'(f(x))f'(x) = (39)(x^7 + 3x^4 - 9)^{38}(7x^6 + 12x^3).$$

Now we have explained every step here, but for the actual practical calculation of derivatives, we can look at the final result and see a simpler way. Begin by noticing that when we differentiated g, we were just differentiating in a sense with respect to the parenthesis. That is, think of the first step as looking at

$$(x^7 + 3x^4 - 9)^{39},$$

and ignoring what is inside the parenthesis and concentrate on seeing in your mind's eye the symbols

$$(\qquad\quad)^{39}$$

and "differentiate with respect to the parenthesis" getting

$$(39)(\qquad\quad)^{38},$$

and leaving enough space to copy inside the parenthesis what is in the original function getting

$$(39)(7x^6 + 12x^3)^{38},$$

and now concentrating on what is inside the parenthesis, simply differentiate it an put the result as a factor to the side getting the final result

$$(39)(7x^6 + 12x^3)^{38}(7x^6 + 12x^3).$$

You must practice using this rule so that you can do it smoothly whenever you see a composite function, especially whenever you see parenthesis.

We worked several examples using the chain rule. For instance, if we want to differentiate

$$h(x) = \sqrt{7x^6 + 12x^3},$$

then we see that we can begin by using the rules for exponents to express the radical as the one half power, so

$$h(x) = (7x^6 + 12x^3)^{1/2},$$

and then we proceed just as in the previous example.

We next observed that to see that the chain rule is true, we can think of x as being time t, and think of $y = f(t)$ as expressing the height of a lady bug on a vertical wire at time t. If temperature depends on height as $T = g(y)$, then to say $g'(y_0) = 3$ means that at location y_0 on the wire, temperature is increasing at 3 units of temperature per unit increase in height. To be more definite, suppose we measure position in centimeters (cm) and temperature in degrees Fahrenheit (deg), and time in minutes (min). If $g'(y_0) = 3$ at the specific location y_0, it does not mean that if you go up one centimeter your temperature will go up by 3 degrees. It means that if temperature *were* to increase steadily with height at the rate at which it increases exactly when $y = y_0$, then the temperature *would* increase by 3 degrees if you were to increase your height by one centimeter. Thus at y_0 we say that temperature is increasing at the rate of 3 deg/cm. If you actually do move up one centimeter, your temperature might even be lower. This is the same situation as looking at your speedometer and knowing that if it is 2pm and your speedometer reads 60 mph, then you will not necessarily be 60 miles away at 3pm, since at 2:30pm, you might just turn around and go back the other way. In any case, the meaning of instantaneous rate of change can be usefully thought of as a hypothetical statement about what would happen if the instantaneous rate were to stay exactly the same and never change. Now, suppose the lady bug arrives at position y_0 on the vertical wire, and at the instant of arrival happens to be moving at 2 cm/min. Likewise, this is an instantaneous velocity, so if that rate were to never change, then one minute later the lady bug would be two centimeters higher, and if the temperature rate of increase were also to be 3 deg/cm everywhere on the wire, then the lady bug would feel a 6 degree increase in temperature one minute later, which is 6 deg/min. If the lady bug's velocity is 5 cm/min, then one minute later the lady bug would feel the temperature to have increased 15 degrees which is 15 deg/min. Thus, realizing that this is only hypothetical, just like the speedometer reading, we must realize that the lady bug's sensation of temperature increase must be that at the instant the lady bug is at y_0, the temperature is increasing at the rate of $3v_0$ if her upward velocity is v_0 at that instant. Now, to say t_0 is the time the lady bug arrives at y_0, means that $y_0 = f(t_0)$, and to say her velocity is v_0 means $v_0 = f'(t_0)$. So we are seeing here that the instantaneous time rate of increase of temperature experienced by the lady bug is the product of $g'(y_0)$ and $f'(t_0)$. On the other hand, the temperature on the lady bug at time t is

$$T(t) = g(f(t)) = (g \circ f)(t),$$

so we see here

$$T'(t) = g'(f(t))f'(t),$$

and this is exactly what the chain rule gives us. In general, if something changes at a certain rate with respect to change of position, then multiplying our speed by the rate of change at our position in the direction we are going gives the rate of change in time we experience for that something which could be temperature or pressure or any of an infinity of possibilities.

We next noticed that if we have a circular ink spill, then the radius of the spill is some function of time $r = r(t)$, and therefore the boundary of the spill moves with velocity

$$v(t) = r'(t).$$

On the other hand, the moving boundary at time t is a circle of radius $r(t)$ which therefore has length $C(t)$ given by

$$C(t) = 2\pi r(t),$$

using the formula for the circumference of a circle. But, the area at time t is simply

$$A(t) = \pi[r(t)]^2,$$

so now the chain rule tells us the rate of change of area is

$$A'(t) = \pi(2)r(t)r'(t) = 2\pi r(t)v(t) = C(t)v(t),$$

so at each instant of time t, the rate of increase of area of the ink spill is simply the velocity of the moving boundary multiplied by the velocity at which it moves in the direction perpendicular to the boundary, at each instant of time. Notice, this is a special case of the general rule which we previously reasoned holds in complete generality, but here we have both the velocity of the moving boundary and the area in terms of the boundary so we can see that the chain rule verifies the general rule.

As a second geometric example, we can see that if we have an exploding ball of stuff and the radius at time t is $r(t)$, then the velocity of the boundary at time t is $r'(t)$, and the surface area of the ball at time t is

$$A(t) = 4\pi[r(t)]^2.$$

On the other hand, the volume of the ball at time t is simply

$$V(t) = \frac{4}{3}\pi[r(t)]^3,$$

so by the chain rule,

$$V'(t) = \frac{4}{3}\pi(3)[r(t)]^2[r'(t)] = 4\pi[r(t)]^2 v(t) = A(t)v(t),$$

and again we see that the rate of increase of volume is at each instant simply the product of the surface area of the moving boundary multiplied by the velocity at which it moves out in a direction perpendicular to the boundary.

We observed that the chain rule allows us to study rates of change for functions of several variables by simply thinking of all the variables as changing in time, so all the variables become functions of time t. For instance, if

$$g(x, y) = x^3y + xy^2,$$

then we can imagine that $x = f(t)$ is some function of time and $y = h(t)$ is also some function of time, then we can form the new function of time

$$k(t) = g(f(t), h(t)),$$

which is a function of time only. Thus, by using various different choices for f and h, we can get some handle on the idea of rate of change for g. Using the chain rule, we see that as

$$k(t) = g(f(t), h(t)) = [f(t)]^3[h(t)] + [f(t)][h(t)]^2,$$

we can differentiate using the chain rule together with the sum and product rules to find that

$$k'(t) = (3)[f(t)]^2[f'(t)][h(t)] + [f(t)]^3[h'(t)] + [f'(t)][h(t)]^2 + [f(t)][(2)[h(t)]^1[h'(t)].$$

If we replace $f(t)$ by x everywhere and $h(t)$ by y everywhere, we have simply

$$k'(t) = (3x^2)yf'(t) + x^3h'(t) + [f'(t)]y^2 + x(2y)h'(t).$$

Notice that the $f'(t)$ and $h'(t)$ always appear to the first degree-no higher powers, because of the chain rule. Each time you apply the chain rule to a term, you have a series of terms each having either $f'(t)$ as a factor or $h'(t)$ as a factor. Therefore, in the end, we can factor these expressions out. In our example, for instance, we have

$$k'(t) = [3x^2y + y^2]f'(t) + [x^3 + 2xy]g'(t).$$

It is interesting that the coefficient expressions in x and y for each term can be easily found in the following simple way. To find the coefficient of $f'(t)$ simply differentiate g as a function of x thinking of y as a constant. To find the coefficient of $h'(t)$ simply differentiate g with respect to y thinking of x a a constant. In fact, this is so useful, we have a special symbol for it. If g is any function of any number of variables, and if w is one of those variables, we denote by $\partial_w g$ the result of differentiating g with respect to w when thinking of all the other variables as constant. We then see we have shown

$$k'(t) = [\partial_x g(x, y)]f'(t) + [\partial_y g(x, y)]h'(t),$$

and this is actually a far reaching generalization of all of our rules so far. That is, this last equation is true for any function g of any variables x and y and any functions f and h, provided only that all are differentiable and that $k(t) = g(f(t), h(t))$ is defined, meaning for the values of t we use that $(f(t), g(t))$ is a point in the domain of the function g.

10. **LECTURE** MONDAY 1 FEBRUARY 2010

Today we reviewed the rules for differentiation and worked some examples. We noted that if $G(x, y)$ is a function of the two variables x and y, given by some algebraic expression, then whenever we assume that x and y are both functions of the single variable t, then we have a new function, say h which is a function of t given by

$$h(t) = G(x(t), y(t)),$$

so we can use the rules for differentiation to find $h'(t)$. We noted that whenever we do, because of the chain rule, we will get a sum of terms each having a factor of $x'(t)$ or a factor of $y'(t)$, but never any powers of x' or y' nor any products $x'y'$. This means that we can always factor out the x' from all the terms in which it appears and do the same for y' in which case we end up with expressions or function $M(x, y)$ and $N(x, y)$ so that we have

$$h' = M(x, y)x' + N(x, y)y',$$

where we have left out the t's for simplicity, so actually, we have

$$h'(t) = M(x(t), y(t))x'(t) + N(x(t), y(t))y'(t).$$

Now, leaving out the t then for simplicity, and looking at

$$h' = Mx' + Ny',$$

where M and N are expressions in x and y, we notice that we did not have any restriction on how we make x and y depend on t. In particular, if we require that $x(t) = t$ and $y = \mathit{constant}$, then $x'(t) = 1$, and $y'(t) = 0$, so for this choice we find

$$h' = M.$$

But notice that for this choice we have $x = t$, and $y = \mathit{constant}$, so differentiating with respect to t in this case is just differentiating G with respect to x treating y as a constant. This is called **PARTIAL DIFFERENTIATION**, whenever there is more than one variable, and we differentiate with respect to one of the variables treating all other variables as constant. When dealing with situations where there are several variables, it is sometimes difficult to keep in mind which variable we are differentiating with respect to, so instead of simply writing f' for the derivative with respect to a variable, we actually include the variable we are differentiating with respect to in the notation. For instance, if $y = f(x)$, we usually simply write f' for the derivative of f, and this would mean $f'(x)$ is the derivative of f evaluated at x. But it also means $f'(t)$ is the value of f' at t. However, if we have understood we are dealing with two variables x and y and that y is depending on x through the equation $y = f'(x)$, then it is often convenient to write one of the expressions

$$f' = D_x f = D_x y = \frac{dy}{dx} = \frac{df}{dx}$$

for the derivative f'. All are saying the same thing, but depending on circumstances, we may need to be reminded of different things with our notation. Thus, when we write

$$\frac{dy}{dx},$$

we are understanding that y is a function of x, so there is some rule in effect, say f, for which any specific value of x gives the value $y = f(x)$, and dy/dx is expressing the rate of change of y with respect to x, so

$$f' = \frac{dy}{dx}.$$

Thus, if $y = x^2$, then

$$\frac{dy}{dx} = 2x.$$

In case there is more than one variable in an expression and we choose to differentiate with respect to only one of the variables treating all others as constants, then we call the process partial differentiation, and to differentiate with respect to say u, we denote the operation with

$$\partial_u = \frac{\partial}{\partial x}.$$

Thus, if G is a function of the variables u, v, w, x, y, z, for instance, then if we treat v, w, x, y, z all as constants, then G becomes a function of only the single variable u, so ordinary differentiation with respect to u but treating all the other variables constant would be denoted by

$$\partial_u G$$

or

$$\frac{\partial G}{\partial x}$$

or

$$\partial_u G(u, v, w, x, y, z).$$

Thus,

$$\partial_u G = \frac{\partial G}{\partial x} = \partial_u G(u, v, w, x, y, z),$$

as all are denoting exactly the same thing.

Likewise,

$$\partial_v G = \frac{\partial G}{\partial v}$$

denotes the result of differentiating with respect to v with all the variables u, w, x, y, z treated as constants. Thus, we can say that the equation which we began with for the derivative of $h = G(x, y)$ with respect to t, which to repeat is

$$h' = Mx' + Ny',$$

could be better expressed in our notation as

$$\frac{dh}{dt} = [\partial_x G(x, y)]\frac{dx}{dt} + [\partial_y (x, y)]\frac{dy}{dt},$$

or leaving out the variables, we can get the same idea across with

$$\frac{dh}{dt} = [\partial_x]\frac{dx}{dt} + [\partial_y]\frac{dy}{dt}.$$

Thus,

$$M(x, y) = \partial_x G(x, y)$$

and

$$N(x, y) = \partial_y G(x, y),$$

or for short,

$$M = \partial_x G$$

and

$$N = \partial_y G.$$

But, when we look at the equation

$$h' = Mx' + Ny',$$

we have to know what variable we are differentiating with respect to and what variables h depends on, whereas when we look at

$$\frac{dh}{dt} = [\partial_x G]\frac{dx}{dt} + [\partial_y G]\frac{dy}{dt},$$

we can see clearly that G is a function of the variables x and y and that both these variables are depending on t and that h is the result of having $G(x, y)$ depend on t as a result of the way x and y are depending on t.

We can notice for convenience, that our differentiation rules apply just as well to partial differentiation, but with partial differentiation, we have when dealing with functions of the variables say u, v, w, x, that

$$\begin{array}{cccc} \partial_u u = 1 & \partial_u v = 0 & \partial_u w = 0 & \partial_u x = 0 \\ \partial_v u = 0 & \partial_v v = 1 & \partial_v w = 0 & \partial_v x = 0 \\ \partial_w u = 0 & \partial_w v = 0 & \partial_w w = 1 & \partial_w x = 0 \\ \partial_x u = 0 & \partial_x v = 0 & \partial_w x = 0 & \partial_x x = 1 \end{array}$$

so we notice that we have a simple pattern here, all the diagonal values are 1's and all the off diagonal values are 0's. Thus in case of only two variables x and y, we have

$$\begin{array}{cc} \partial_x x = 1 & \partial_x y = 0 \\ \partial_y x = 0 & \partial_y y = 1 \end{array}$$

so any expression involving x and y alone can be differentiated with our rules and remembering this simple array of values for partial derivatives of the variables themselves. Thus, the fact that $\partial_x y = 0$ is obvious, if we are treating y as constant when differentiating with respect to x.

One utility of the equation

$$\frac{dh}{dt} = [\partial_x G]\frac{dx}{dt} + [\partial_y G]\frac{dy}{dt},$$

is that it contains all the rules for differentiation in a single equation. For instance, to get the Addition or Sum Rule for differentiation which says that

$$(f + g)' = f' + g',$$

we can take G simply to be the function

$$G(x, y) = x + y.$$

We choose x and y to depend on t by setting

$$x = f(t)$$

and

$$y = g(t).$$

Then it is easy to see that

$$\partial_x G = \partial_x(x + y) = 1,$$

because when we differentiate with respect to x we treat y as constant, so the expression $G(x, y) = x + y$ when y is constant is simply a linear function of x with slope one. Likewise, for the same reason we have

$$\partial_y G = \partial_y(x + y) = 1.$$

But we also have here

$$\frac{dx}{dt} = f'$$

$$\frac{dy}{dt} = g',$$

and

$$(f+g)' = h' = \frac{dh}{dt} = [\partial_x G]\frac{dx}{dt} + [\partial_y G]\frac{dy}{dt} = 1 \cdot f' + 1 \cdot g' = f' + g'.$$

To prove the Product Rule we take

$$G(x,y) = xy,$$

then as y is constant when applying ∂_x, we see that G is simply a line through the origin with slope y when differentiating G with respect to x, so

$$\partial_x G = \partial_x(xy) = y,$$

and likewise, reversing the roles of x and y in this process,

$$\partial_y G = \partial_y(xy) = x.$$

Applying the same process now as we did for the sum rule, as now $h = fg$, we find

$$(fg)' = h' = [\partial_x G]\frac{dx}{dt} + [\partial_y G]\frac{dy}{dt} = yf' + xg' = f'g + fg'.$$

For the Chain Rule, we simply take

$$G(x,y) = f(y),$$

instead of $x = f$, so now

$$h = f(y) = f \circ y = f \circ g$$

and

$$\partial_x G = 0$$

and

$$\partial_y G = f'$$

resulting in

$$(f \circ g)' = [\partial_x G]\frac{dx}{dt} + [\partial_y G]\frac{dy}{dt} = 0 \cdot \frac{dx}{dt} + f' \cdot g' = (f' \circ g) \cdot g'.$$

11. **LECTURE** WEDNESDAY 3 FEBRUARY 2010

Today we reviewed for QUIZ 2 and then gave QUIZ 2 in class.

12. **LECTURE** FRIDAY 5 FEBRUARY 2010

Today we began reviewing the problems on PRACTICE TEST 1 for computing derivatives from tabulated information when there are several variables involved.

Specifically, if H depends on the variables u, v, w and each of these variables depends on x, then we can form the composite function h with the rule

$$h(x) = H(u(x), v(x), w(x)).$$

We recalled that our differentiation rules tell us that

$$D_x h = [\partial_u H] \cdot D_x u + [\partial_v] \cdot D_x v + [\partial_w] \cdot D_x w,$$

and this means for each specific x in the domain of h we have

$$\begin{aligned} h'(x) = D_x h(x) = [\partial_u H(u(x), v(x), w(x))] \cdot D_x u(x) \\ +[\partial_v (u(x), v(x), w(x))] \cdot D_x v(x) \\ +[\partial_w (u(x), v(x), w(x))] \cdot D_x w(x). \end{aligned}$$

When we are given actual rules for calculating H, u, v, w for which our differentiation rules apply, we can then calculate the various derivatives to find $h'(x)$ for any x. However, in many applications, we never know the actual rules, but have theoretical reason to suspect they exist. Often, we must rely on experimental measurement to build enough tabulated values to be able to apply the formula at a specific value of x. As an example here, imagine that you find yourself at the controls of a Boeing 747 in flight with no one to help you learn how to fly the plane (say everyone on board got sick and passed out except you). You would find yourself flying along at first without having control of the plane, and it would seem that everything is fine. However, that cannot keep on indefinitely, as the plane will run out of fuel and crash. You must figure out how to control the plane. You will see throttle controls for the engines and a wheel in front of you that may look strange at first, since it is not round, and pedals for your feet. What you would definitely not do is grab the controls and start changing them all radically. Since the plane is flying as is, you could assume that the positions of the controls will maintain the plane in level flight. You might then experiment a little by making very very tiny adjustments to the controls to see what happens. If you try to turn the wheel very slightly to the left as you would when turning left while driving a car, you will find that it takes some pressure, but applying a very tiny pressure you will find that the plane starts to rotate over sideways to your left, as if it wants to start flying on its left side if more pressure is applied. You would probably be inclined at this point to put opposite pressure as if turning right, and this causes the plane to rotate back the other way to its original position. After a little fiddling here, you would probably get the plane to fly straight and level again. We could say that if H is the angle of tilt for the plane about its front to back axis, you have begun to see how H depends on pressure x applied to the wheel in the left right direction for various values near $x = 0$. In the process, you may have discovered that the wheel can be pushed forward and back. When you put slight forward pressure on the wheel, the nose of the plane starts to go down, and your natural reaction at this point would probably be that down is not where you want to go, so you would pull back on the wheel and the nose of the plane would come back up, and again, a little fiddling around you would get familiar with how to make the plane fly straight and level again, and you would as well see how to control both variables x and y, where y is the pressure applied to the wheel in the forward direction. For instance, you would have noticed that a small change in y does not change H, and this would tell you that $\partial_y H = 0$, a very useful piece of information here.

At this point what you really know is $\partial_x H$ and $\partial_y H$, and if K denotes the up angle of the nose, you have discovered $\partial_y K$. You will have also noticed that slight pressure left or right on the wheel does not change K, and this means that $\partial_x K = 0$. If you apply a little pressure to the left pedal, the plane will start to turn to the left ever so slightly, but you will find that in fact you are mostly simply flying a little sideways without actually turning, and you will feel a slight upward pressure of the right pedal on your foot simultaneously-the two pedals are not independent of one another, they work together. You might also start studying the dials and notice that you have an altimeter which tells you your altitude, and a rate of climb indicator which tells you your rate of climb up or down. If you apply pressure to one of the pedals and the plane starts to move slightly sideways, then careful monitoring of the dials will show that you are slightly losing altitude. You will find the same if you apply forward pressure to the wheel or if you apply left turning or right turning pressure to the wheel. Thus, if A is altitude, we know that $\partial_x A$ is negative when H is not zero and $\partial_y A$ is negative when y is positive. When y is negative so we are pulling back on the wheel, then A increases, so $\partial_y A$ is positive at $y = 0$. Thus, your altitude depends in a rather complex way on all of the control settings. If you can see the ground below and you pay close attention, when you increase pressure on the left pedal causing the plane to go slightly sideways, you will notice that you do start to move in a direction more to the left, but what may be counter-intuitive here is that to turn more definitely to the left, you need to rotate the wheel to the left at the same time. This will cause the plane to begin to go left and lose altitude at the same time. To compensate for the altitude loss here in a left turn, you can simultaneously pull back on the wheel pulling up on the nose. Once you have mastered keeping the altitude constant while turning left, you know how to control the plane effectively as far as being able to control where you are going. If you slightly pull back on the engine throttle controls, you will hear the engines slightly change pitch and you will feel as if you are slightly falling and losing altitude. More throttle and the plane wants to resume level flight. You are now flying the airplane. You are in a very dangerous situation, but by applying very small changes to the controls and noticing carefully what happens, you have a chance to survive.

In business situations, the business man is usually interested in profit, P which can be an unknown function of many variables such as demand, supply, availability of various raw materials, variables which we might call u, v, w, x, y, z. In fact all these depend on time t in some way we do not have much control over or know much about. By noticing the result of small changes each of the variables while keeping all the others fixed, we can work out, for given values $u_0, v_0, w_0, x_0, y_0, z_0$ not only the value $P(u_0, v_0, w_0, x_0, y_0, z_0)$, but in addition, we can get approximate values for the partial derivatives

$$\begin{aligned}
&\partial_u P(u_0, v_0, w_0, x_0, y_0, z_0),\\
&\partial_v P(u_0, v_0, w_0, x_0, y_0, z_0),\\
&\partial_w P(u_0, v_0, w_0, x_0, y_0, z_0),\\
&\partial_x P(u_0, v_0, w_0, x_0, y_0, z_0),\\
&\partial_y P(u_0, v_0, w_0, x_0, y_0, z_0),\\
&\partial_z P(u_0, v_0, w_0, x_0, y_0, z_0).
\end{aligned}$$

Likewise, for another given set of circumstances we would have other fixed values

$$u_1, v_1, w_1, x_1, y_1, z_1$$

for the variables and making slight changes in the variables and studying the effects on P we could determine the values of P and its partial derivatives at

$$(u_1, v_1, w_1, x_1, y_1, z_1).$$

All the information from doing these studies as various specific input values of these variables would be summarized in a table of values. Now as far as the time variation is concerned, that could be dealt with by studying at a specific time t_0, how all the variables are changing as t increases slightly forward in time from t_0. Thus, we would find out not only the values of the variables at t_0, but also there rates of change at t_0. Likewise, this might be done at later times $t_1, t_2, t_3, ...t_m$. All this information could be tabulated, giving the values of the variables and their derivatives at each of these values of t. Our equation above then allows us to calculate $D_x h(t)$ if t appears in the table including the variable t and if the resulting values of the variables u, v, w, x, y, z appears in the table for P and its partial derivatives.

13. **LECTURE** MONDAY 8 FEBRUARY 2010

Today we continued reviewing for TEST 1.

14. **LECTURE** WEDNESDAY 10 FEBRUARY 2010

Today TEST 1.

15. **LECTURE** FRIDAY 12 FEBRUARY 2010

CLASS DID NOT MEET

16. **LECTURE** MONDAY 15 FEBRUARY 2010

CLASS DID NOT MEET BECAUSE OF MARDI GRAS

17. **LECTURE** WEDNESDAY 17 FEBRUARY 2010

CLASS DID NOT MEET

18. **LECTURE** FRIDAY 19 FEBRUARY 2010

CLASS DID NOT MEET

19. **LECTURE** MONDAY 22 FEBRUARY 2010

Today we discussed exponential functions and their derivatives. Keep in mind that up until now, all our examples have never had variables in the exponent position, rather only in the base position. For instance, the functions

$$g(x) = x^5,$$
$$h(x) = x^\pi,$$
$$r(x) = x^{1/3}$$

are examples of POWER FUNCTIONS. The variable in each of these examples is in the base position, and the exponent is FIXED or CONSTANT. In an exponential function, the base is a fixed constant, and the exponent is allowed to be variable. For instance, if

$$f(x) = 2^x,$$

we noted that for each unit increase in x, the height of the graph doubles. This function is a simple example of an exponential function, and we see that it increases EXPLOSIVELY. In fact, where exponential functions appear in models of chemical reactions, they often indicate an explosive reaction. We noted that the domain of the exponential function is all real numbers whereas the range is all positive real numbers. Thus, any positive number can be expressed as a power of 2. If b is any number bigger than one, then for some number c we have $2^c = b$ and therefore

$$b^x = (2^c)^x = 2^{cx}.$$

Remember, when you take any function g and form $g(cx)$ it merely squeezes or expands the graph in the horizontal direction. Increasing the base increases the value of c and squeezes the graph making it go up faster, whereas decreasing the base decreases the value of c causing the graph to expand horizontally and appear to increase less rapidly. Thus, all exponential function graphs have the same basic form, they are the result of horizontally squeezing or expanding the graph of $y = 2^x$. To find the base of an exponential function by looking at the graph, you have to actually measure. Let f_b denote the exponential function with base b, so

$$f_b(x) = b^x.$$

We can see that if $b \geq 2$, then the graph of f_b does not cross the diagonal line. As $f_b > 0$ always, the inverse function has all positive real numbers for its domain and all real numbers for its range. the inverse function to f_b is denoted $\log_b$ and is called the logarithm function to base b. Thus we always have

$$b^{\log_b x} = x, \ x > 0,$$

and

$$\log_b(b^x) = x, \ \mathit{all}\ x.$$

Next we must try to differentiate the exponential function. Since

$$f_b'(x) = \lim_{h \to 0} \frac{f_b(x+h) - f_b(x)}{h},$$

we know that for very tiny h we have

$$f_b'(x) =^{app}= \frac{b^{x+h} - b^x}{h} = \frac{b^h \cdot b^x - 1 \cdot b^x}{h} = \frac{b^h - 1}{h} \cdot b^x.$$

But if we take $x = 0$ in this last expression, as $b^0 = 1$, we find that

$$f_b'(0) = \lim_{h \to 0} \frac{b^h - 1}{h},$$

and therefore we have

$$f_b'(x) = f_b'(0) \cdot b^x = f_b'(0) \cdot f_b(x).$$

This equation is somewhat amazing since it says that as soon as we know the slope of the tangent line where the graph crosses the $y-$axis, we just multiply that slope by $f_b(x)$ to get the slope at the point (x, b^x) on the graph of f_b.

Next, we noticed that by adjusting b properly, we can make $f_b'(0) = 1$. In fact, for $b = 1$, the exponential function is obviously constant with value one, so the slope of every tangent line is zero. On the other hand, by making b very big we see that the slope of the tangent line where the graph crosses the $y-$ axis can be made very large. to see this in more detail, notice that if $c > 1$, then

$$f_b(cx) = b^{cx} = (b^c)^x = f_{b^c}(x),$$

and by the chain rule,

$$f_{b^c}'(x) = [f_b(cx)]' = f_b'(cx) \cdot c,$$

so taking $x = 0$ gives

$$f_{b^c}'(0) = c \cdot f_b'(0).$$

This means that as soon as we realize that for b bigger than one $f_b'(0)$ cannot be zero and in fact must be a positive number, we can choose c to make an exponential function whose slope at the y intercept is any number we want. Clearly, the simplest choice for a base is the base for which the tangent slope at the y intercept is simply one. This choice is denoted by the symbol e in every mathematics book. Thus,

$$f_e'(0) = 1$$

and therefore

$$f_e'(x) = f_e(x).$$

That is, to differentiate e^x you just copy!!

To repeat

$$(e^x)' = e^x.$$

Thus by the chain rule, for any differentiable function g we have

$$[e^{g(x)}]' = e^{g(x)} \cdot g'(x).$$

To differentiate any exponential function you just use the chain rule and logarithm to base e, since

$$b^x = e^{\log_e(b^x)} = e^{x \cdot \log_e(b)},$$

therefore

$$[b^x]' = e^{x \cdot \log_e(b)} \cdot \log_e(b) = \log_e(b) \cdot b^x.$$

Next, we need to find how to compute the number e itself as best we can. It turns out that e is irrational like π and its decimal expansion therefore has no pattern at all. No matter how many decimals of e you see, from them alone you cannot predict what comes next, there is no pattern. We can see without to much trouble that e is between two and three, but we want to get a means of getting very accurate approximations. to do this, we note that by definition we must have

$$\lim_{h\to 0} \frac{e^h - 1}{h} = 1,$$

which means that for very very tiny h we have

$$\frac{e^h - 1}{h} =^{app}= 1,$$

so

$$e^h =^{app}=^{app}= 1 + h,$$

and therefore

$$e =^{app}= (1+h)^{1/h}$$

for very very tiny h or in other words,

$$e = \lim_{h\to 0}(1+h)^{1/h}.$$

Since $1/h$ gets very big as h becomes very small, we can likewise say that

$$e = \lim_{x\to\infty}(1+\frac{1}{x})^x.$$

For instance with large integer values of x say for n a very large integer, we have

$$(1+\frac{1}{n})^n$$

should be very close to e. You can check this easily with your calculator.

One of the common every day uses of the exponential function is to deal with compound interest. To get a preliminary idea of this connection, imagine that you are very lucky and you place a certain bet on the roulette wheel at the casino which pays off at the rate r for every dollar you put down, so if you initially put down P_0 dollars, then you win rP_0 dollars in addition to your original bet P_0 which you get to keep. Thus, after one spin of the wheel you have

$$P_0 + rP_0 = P_0(1+r)$$

dollars. Now suppose that you "let it ride" and are lucky enough to keep on winning. After two spins you have

$$P_0(1+r)(1+r) = P_0(1+r)^2,$$

and you see that after n spins of the wheel you would have

$$P_0(1+r)^n$$

dollars on the table in bets provided of course that you kept on winning. Notice this is an exponential function of n. We will see how this relates to compound interest in the next lecture.

20. **LECTURE** WEDNESDAY 24 FEBRUARY 2010

Today we discussed compound interest and continuously compounded interest. As our first example, we imagined a bet distributed over the roulette table which paid off giving a 1/11 increase rate for each spin of the wheel. We observed that if our initial investment was P_0, then after n spins of the wheel, letting our bets "ride", we would have $P(n)$ where

$$P(n) = P_0(1 + [1/11])^n,$$

assuming that we won on every spin. More generally, if we earn r dollars for each dollar invested, then after n spins we have

$$P(n) = P_0(1 + r)^n.$$

We see that this is an exponential function with base $1 + r$. If T is the time from one spin to the next, we could say that in time t there are $n = t/T$ spins, so that if this value of n is a whole number, then again, if $P(t)$ is the value at time t, then

$$P(t) = P_0(1 + r)^{t/T} = P_0[(1 + r)^{1/T}]^t,$$

and this is an exponential function with base

$$b = (1 + r)^{1/T}.$$

In financial situations, we usually like to compare investments by their interest rates in a standard way. To begin with, assume that a bank offers annual interest rate r so that if you had initially deposited P_0 dollars, then after a year the bank would deposit $P_0 r$ dollars into your account. After t years, if you left all the money in the account and never withdrew any, then the account balance would be

$$P(t) = P_0(1 + r)^t,$$

which is an exponential function with base $b_1 = 1 + r$. Now you could think that if you came into the bank after 6 months or one half of a year, that the bank should reasonably give you half of your interest, and if they did, you would have interest amount

$$\frac{1}{2}P_0 r = P_0\frac{r}{2},$$

which means you could think of the semi-annual interest rate as simply $r/2$. Given that the bank would allow you to draw interest every 6 months, you could put this interest back into the account and let it stay in until the end of the year. For the second 6 months of the year you would be paid interest on the amount

$$P(1/2) = P_0(1 + [r/2])$$

so at the end of the year your bank balance would be

$$P(1) = P_0(1 + [r/2])(1 + [r/2]) = P_0(1 + [r/2])^2,$$

and therefore after t years your bank balance would be

$$P(t) = P_0(1 + [r/2])^{2t} = P_0[(1 + [r/2])^2]^t,$$

which is an exponential function with base b_2, where

$$b_2 = (1 + [r/2])^2.$$

What is happening here is that for the second half of the year you are earning interest on the interest from the first half of the year as well as interest on the original deposit, and that is what is called **COMPOUND INTEREST**. Notice that the argument that you deserve to be paid interest after only 6 months is that your money should be considered to be continuously earning money, after all that is why the bank wants your deposit-it is using the money to earn money continually. But, if this is the case, you could just as well argue that you should be paid interest at the rate $r/3$ if you came in every third of a year to collect interest and re-deposit it. This obviously by the same reasoning leads to the conclusion that your balance after t years is

$$P(t) = P_0[(1 + [r/3])^3]^t$$

which is an exponential function with base b_3 where

$$b_3 = (1 + [r/3])^3.$$

Notice that you could argue that you should not have to bother coming into the bank at all if you are just going to redeposit the earned interest, so the bank should just automatically increase your bank balance. But if this is the case, why only do it every third of a year. Why not every month or better still every day or even every second. This leads to the concept of **CONTINUOUSLY COMPOUNDED INTEREST**. Obviously, if you compound n times during the year, then

$$P(t) = P_0 b_n^t,$$

where the base b_n for the exponential function is

$$b_n = (1 + [r/n])^n.$$

For continuously compounded interest then the base should be b_∞ where

$$b_\infty = \lim_{n\to\infty}(1 + [r/n])^n.$$

We observed in the last lecture that

$$e = \lim_{x\to\infty}(1 + [1/x])^x.$$

We can use this fact to calculate b_∞. By taking n to be sufficiently large, we can make n/r enormous, so we can think of

$$e =^{app}= (1 + \frac{1}{n/r})^{n/r} = (1 + [r/n])^{n/r},$$

and therefore raising both sides to the power r gives

$$e^r =^{app}= (1 + [r/n])^n = b_n.$$

Making n larger and larger just leads to better and better approximation here, so obviously

$$b_\infty = e^r.$$

Thus we can say that at annual interest rate r continuously compounded, your bank balance at time t should be simply

$$P(t) = P_0(e^r)^t = P_0 e^{rt}.$$

Now we need to keep in mind that exponential functions are sort of like lines. If you know any two points on an exponential curve, then you can determine the curve's equation, just like if you know two points on a line you can determine the equation for the whole line. As an example, suppose we have an investment which is worth one thousand dollars after 2 years and worth four thousand dollars after 7 years. To evaluate such an investment you should compare it to a continuously compounded interest account which has the same payoffs at the times you know. For the bank account you have

$$P(t) = P_0 b^t,$$

and thus you want this bank account to have

$$P(2) = 1000$$

and

$$P(7) = 4000.$$

This means we have

$$4000 = P(7) = P_0 b^7$$

and

$$1000 = P(2) = P_0 b^2.$$

dividing the last equation into the one above we see that P_0 cancels out and we find

$$4 = \frac{P(7)}{P(2)} = \frac{P_0 b^7}{P_0 b^2} = \frac{b^7}{b^2} = b^5.$$

This means that

$$b = 4^{1/5},$$

so now we know the base of our exponential function. We therefore can now write

$$P(t) = P_0(4^{1/5})^t = P_0 \cdot 4^{t/5}.$$

Notice this is very similar to finding the slope of a line when you know two points on the line. To find the value of P_0 is analogous to using the point-slope form of the equation of a line and getting the equivalent slope-intercept form of the equation. Once you have the slope of the line, since you started with two points on the line, you can use either point to write down the point-slope equation of the line, and then rearrange it to get the slope-intercept form of the line's equation. Here with the exponential function, we do the similar thing. For instance, using $P(2) = 1000$, we have

$$1000 = P(2) = P_0 b^2 = P_0(4^{1/5})^2 = P_0 4^{2/5}.$$

Therefore,

$$P_0 = \frac{1000}{4^{2/5}} = (1000)(4^{-2/5}).$$

This means that the exponential function describing a continuously compounded interest account which would yield these same results after 2 years and after 7 years would have to balance $P(t)$ at time t given by

$$P(t) = (1000)(4^{-2/5})(4^{1/5})^t = (1000) \cdot 4^{(t-2)/5}.$$

If instead we had used the other point on the curve, we would have

$$4000 = P(7) = P_0 \cdot 4^{7/5},$$

so

$$P_0 = \frac{P(7)}{4^{7/5}} = (4000)(4^{-7/5}).$$

But,

$$(4000)(4^{-7/5}) = (4000)(4^{-1-(2/5)}) = (4000)(\frac{1}{4})(4^{-2/5}) = (1000)(4^{-2/5}),$$

which is just what we had before using $P(2) = 1000$. Thus, you can use either point to find P_0 which is the $y-$intercept of the exponential curve, and is clearly $P(0) = P_0$. Now we are in a position to find the annual interest rate. If the rate is r, then we know the base is

$$b = e^r$$

for continuously compounded interest, and here $b = 4^{1/5}$, so therefore using the natural logarithm which is the logarithm to base e, we have

$$r = \log_e(b) = \ln(b) = \ln(4^{1/5}) = \frac{1}{5}\ln(4) = (.2)(1.386294361) = .2772588722.$$

Now lets consider the rate of growth for such accounts or investments from the standpoint of continuously compounded interest. Remember that

$$(e^x)' = e^x,$$

which means that

$$\frac{d}{dx}e^x = e^x.$$

By the chain rule, then for the account with annual interest rate r continuously compounded we have, setting $x = rt$,

$$\frac{d}{dt}P(t) = \frac{d}{dt}P_0 \cdot e^{rt} = P_0\frac{d}{dx}e^x \cdot \frac{dx}{dt} = P_0e^x \cdot r = P_0e^{rt} \cdot r = P(t) \cdot r.$$

This is a very simple result:

$$\frac{d}{dt}P(t) = P(t) \cdot r.$$

For instance, in the small time interval from t to $t + \Delta t$, the amount earned, ΔP is given to very good approximation by using the approximate equation

$$\frac{\Delta P}{\Delta t} =^{app}= P(t) \cdot r,$$

and therefore

$$\Delta P =^{app}= P(t) \cdot r \cdot \Delta t.$$

For instance, if you have an investment currently worth one billion dollars invested at an annual interest rate of ten percent continuously compounded, then over the next 24 hours or one day you earn about

$$(1000000000)(.1)(1/365) = 273972.6027 \ dollars.$$

In fact over the next hour, even more accurately you will earn

$$11415.52511 \ dollars.$$

The thing to notice here is that always, at each instant, the growth rate is proportional to the amount that is there at that instant,

$$\frac{dP}{dt} = P \cdot r,$$

where for short we write P for $P(t)$. This simple fact is what characterizes exponential curves, and we will see that there are many more applications of this simple fact.

21. **LECTURE** FRIDAY 26 FEBRUARY 2010

Today we discussed applications of exponential and logarithmic functions to problems of growth and decay. We began be recalling the fact that exponential functions are similar to lines in that an exponential curve is determined by knowing two points on its graph. This becomes clearer when we apply the laws of logarithms. Remember that the logarithm to base b is the inverse function to the exponential function with base b. Thus

$$\log_b(b^x) = x, \ \textit{all } x,$$

and

$$b^{\log_b(x)} = x, \ \textit{all } x > 0.$$

As a consequence, each law of exponents determines a law of logarithms. For instance, the fact that $b^0 = 1$ means that

$$\log_b(1) = 0.$$

The fact that always

$$b^{x+y} = b^x b^y$$

gives the law

$$\log_b(xy) = \log_b x + \log_b y, \ \textit{all } x, y > 0.$$

The fact that always

$$(b^x)^y = b^{xy}$$

gives the law

$$\log_b(x^p) = p \cdot \log_b x, \ \textit{all } p \textit{ and all } x > 0.$$

We also know from the inverse function relations above that $A = B$ if and only if $\log_b A = \log_b B$, for any positive numbers A and B. Thus the exponential equation

$$Y = A \cdot b^x$$

is equivalent to the equation

$$\log_b(Y) = \log_b(A) + tx,$$

which is a simple linear equation. In the other direction, if

$$y = a + mx$$

is the equation of a line, then exponentiating both sides using base c gives

$$c^y = c^{a+mx} = c^a \cdot (c^m)^x,$$

so with $Y = c^y, \ A = c^a$ and $b = c^m$ we find the usual form of exponential function

$$Y = A \cdot b^x.$$

In problems of growth, a convenient measure of the growth rate is the **DOUBLING TIME**, which we can denote by D. Thus for the exponential function

$$Y(t) = Ab^t,$$

we see that $Y(0) = A$, and therefore

$$2A = Y(D) = A \cdot b^D,$$

so cancelling the A's gives

$$2 = b^D$$

and therefore

$$b = 2^{1/D}.$$

This means that the exponential growth function Y is simply

$$Y(t) = A \cdot 2^{t/D}.$$

You can see that it is easy to calculate $Y(t)/A$ whenever t is an integer multiple of D.

In problems of decay such as with radioactive substances, the convenient characterization is the **HALF LIFE**, H which is the time it takes for something to decay by half. Thus, with the exponential decay function Y, where

$$Y(t) = A \cdot b^t,$$

we see that again $A = Y(0)$, so

$$\frac{1}{2}A = Y(H) = A \cdot b^H,$$

so cancelling this time gives

$$\frac{1}{2} = b^H,$$

and therefore

$$b = (1/2)^{1/H} = 2^{-1/H}.$$

When we put this value of b in for the expression for $Y(t)$, we find

$$Y(t) = A \cdot 2^{-t/H} = A \cdot (\frac{1}{2})^{t/H}.$$

Again, it is now easy to calculate $Y(t)/A$ whenever t is an integer number of half-lives.

To get a feel for how the doubling time relates to growth, we calculate that an investment compounded continuously at a ten percent annual rate has a doubling time of about 6.9 years or just under 7 years. To get a feel for half-life, we estimated that 100 grams of a substance with atomic weight 100 for which each atomic decay produces two energetic neutrons would initially conservatively produce around a trillion neutrons per second as it decays.

22. **LECTURE** MONDAY 1 MARCH 2010

Today we reviewed for the quiz on Wednesday. We reviewed the laws of exponents and logarithms. We began be noting that there are three basic laws of exponents from which all the other laws follow. These are, for any $a, b, c > 0$, and all x, y,

$$b^{x+y} = b^x \cdot b^y,$$

$$(b^x)^y = b^{xy},$$

and

$$(\frac{ab}{c})^x = \frac{a^x b^x}{c^x}.$$

Remember, we refer to b as the **base** and x as the **exponent** in the expression b^x.

Thus, if we set $c = b^0$, then

$$c^2 = (b^0)^2 = b^{0 \cdot 2} = b^0 = c,$$

so whatever c is, it has the property that it equals its own square. But there are only two real numbers with this property, namely zero and one. If we were to suppose that $c = 0$, then we would have for every real number x,

$$b^x = b^{0+x} = b^0 \cdot b^x = c \cdot b^x = 0 \cdot b^x = 0,$$

and this would mean in particular that $b = b^1 = 0$, for every positive number b, which is certainly a contradiction, so it must be the case that $c = 1$, and therefore

$$b^0 = 1.$$

We noted that if $b > 1$, then the graph of $y = b^x$ increases explosively from left to right whereas if $0 < b < 1$, then $y = b^x$ is simply the reflection of $y = (1/b)^x$ through the vertical axis. The fact that $b^0 = 1$ means that the $y-$intercept of any exponential function $y = b^x$ is simply $y = 1$. Of course, if $b = 1$, then $b^x = 1$, for all x, so the graph becomes simply a horizontal line at height one in this trivial case.

We call any function of the form

$$y = f(x) = A \cdot b^x$$

an **EXPONENTIAL FUNCTION** with base b. We see that as $b^0 = 1$, the coefficient A is the value of f at $x = 0$,

$$f(0) = y(0) = A.$$

We defined the **LOGARITHM FUNCTION** with base b, denoted $\log_b$, as the inverse function to the exponential function f_b where $f_b(x) = b^x$. Keep in mind that the domain of f_b is the set of all real numbers, whereas the range of f_b is the set of all positive real numbers. Thus the domain of $\log_b$ is the set of all positive real numbers and the range of $\log_b$ is the set of all real numbers. Thus

$$\log_b(b^x) = x, \ all\ x,$$

and

$$b^{log_b x} = x, \ all\ x > 0.$$

Put another way, the equations

$$y = b^x$$

and

$$\log_b y = x$$

are equivalent. Thus the fact that $b^0 = 1$ for every positive number b says also that

$$\log_b(1) = 0,\ all b > 0.$$

These two equations can be used to turn each law of exponents into a corresponding law of logarithms. For instance, the law of exponents that says $bx + y = b^x b^y$ gives the law of logarithms

$$\log_b(X \cdot Y) = \log_b(X) + \log_b(Y),\ X > 0,\ Y > 0.$$

to see this, setting $x = \log_b X$ and $y = \log_b Y$ we have

$$b^x = X$$

and

$$b^y = Y$$

so

$$X \cdot Y = b^x \cdot b^y = b^{x+y},$$

so

$$X \cdot Y = b^{x+y}$$

and this equation is equivalent to

$$\log_b(X \cdot Y) = x + y = \log_b(X) + \log_b(Y).$$

The law of exponents that says $(b^x)^y = b^{xy}$ gives the law of logarithms which says that for any numbers p, X,

$$\log_b(X^p) = p \cdot \log_b(X).$$

To see this, again with $x = \log_b X$, we have

$$b^x = X,$$

so

$$X^p = (b^x)^p = b^{xp} = b^{px},$$

or

$$X^p = b^{px}$$

which is equivalent to

$$\log_b(X^p) = p \cdot x = p \log_b(X).$$

We also have

$$\log_b(1/X) = -\log_b(X)$$

as a consequence of these laws of logarithms, since

$$0 = \log_b(1) = \log_b(X \cdot [1/X]) = \log_b(X) + \log_b(1/X).$$

When working with or applying exponential functions keep in mind that always the rate of change is proportional to the amount that is there. Recall that the special exponential function with base $b = e$ is its own derivative:

$$(e^x)' = \frac{d}{dx}(e^x) = e^x.$$

The natural logarithm function is $\ln = \log_e$, the logarithm to base e. Thus, we can convert any logarithm function to base e to differentiate it:

$$f(x) = A \cdot b^x = Ae^{\ln(b^x)} = A \cdot e^{x \ln b},$$

so by the chain rule,

$$f'(x) = A \cdot e^{x \ln b} \cdot \ln b = [f(x)] \cdot \ln b.$$

Therefore,

$$f'(x) = [\ln b] \cdot f(x).$$

Thus, the rate of change divided by the value at any x always tells us the natural logarithm of the base:

$$\frac{f'(x)}{f(x)} = \ln b.$$

The laws of exponents and logarithms also show that we can always change bases to whatever is convenient. In fact, it is the case that

$$(\log_a b)(\log_b c) = \log_a c, \ \ \textit{for } a, b, c \textit{ all positive.}$$

To see this, let $x = \log_a b$ and $p = \log_b c$, so

$$b = a^x$$

and

$$c = b^p,$$

so substituting,

$$c = b^p = (a^x)^p = a^{xp},$$

which is equivalent to

$$\log_a c = xp = (\log_a b)(\log_b c).$$

In other words, the symbols in quotes:"$b)(\log_b$" can simply be removed, as if they "cancel out".

We reviewed the use of doubling time and half-life to deal with applications of exponential functions to growth and decay. Thus, if D is the doubling time, then

$$y(t) = A \cdot 2^{t/D} = A \cdot (2^{1/D})^t$$

is the amount or value at time t, whereas if H is the half-life, then

$$y(t) = A \cdot (1/2)^{t/H} = A \cdot 2^{-t/H} = A \cdot (2^{-1/H})^t,$$

gives the amount or value at time t. Thus, in the growth case, we have base $b = 2^{1/D}$ whereas in the decay case we have base $b = 2^{-1/H}$. Applying our calculation for rate of change, we have in the growth case,

$$y'(t) = y(t)\ln(2^{1/D}) = y(t) \cdot \frac{\ln 2}{D},$$

whereas

$$y'(t) = y(t)\ln(2^{-1/H}) = y(t)\cdot(-\frac{\ln 2}{H}).$$

We also noted that for an exponential function you can always make any time into a "start" time. For instance, if we know that at specific time $t = k$ we have the value $y(k) = K$, then with base b we have

$$y(k+t) = K\cdot b^t.$$

To see this, since $y(t) = A\cdot b^t$, where $y(0) = A$, even without knowing what A is, we have

$$y(t) = A\cdot b^t,$$

so

$$K = y(k) = A\cdot b^k,$$

and therefore

$$y(k+t) = A\cdot b^{k+t} = A\cdot b^k\cdot b^t = K\cdot b^t.$$

For instance, if we know that the value of an investment after 3.2 years is going to be 8971 dollars, and if the doubling time is 7 years, then after 2 *more* years, that is at $t = 5.2$ years, it will be worth twice as much, or 17942 dollars, or after 1.7 more years, at $t = 4.9$ years it will be worth

$$y(4.9) = (8971)(2^{(1.7)/7}).$$

Notice that you need a calculator here. Without a calculator, we can note that if you use the derivative to estimate the amount at a later time using the tangent line at a specific time, then you will underestimate the value because the tangent line is always underneath the graph which is obvious from the shape of the graph. Likewise, in a decay situation, you will always underestimate because the graph is always above its tangent line.

Be sure to bring your calculator on Wednesday for Quiz 4.

23. **LECTURE** WEDNESDAY 3 MARCH 2010

Today we reviewed exponential functions and had QUIZ 4 in class.

24. **LECTURE** FRIDAY 5 MARCH 2010

Today we began by reviewing the derivative of the exponential function and consequences.
We begin with

$$(e^x)' = e^x,$$

and note that by the chain rule, for any function f we have

$$(e^{f(x)})' = e^{f(x)} \cdot f'(x).$$

That is to differentiate e raised to any power, you just copy the expression down, and then multiply it by the derivative of what is in the exponent. For instance,

$$(e^{x^5+7x^2})' = e^{x^5+7x^2} \cdot (5x^4 + 14x).$$

If we have two variables x and y both depending on t, then

$$\frac{d}{dt}(e^{x^2+y^2}) = e^{x^2+y^2} \cdot (2x\frac{dx}{dt} + 2y\frac{dy}{dt}).$$

Alternately, we can use

$$\frac{d}{dt}F(x,y) = \partial_x F(x,y)\frac{dx}{dt} + \partial_y F(x,y)\frac{dy}{dt},$$

and find

$$\partial_x(e^{x^2+y^2}) = e^{x^2+y^2} \cdot 2x,$$

$$\partial_y(e^{x^2+y^2}) = e^{x^2+y^2} \cdot 2y,$$

so

$$\frac{d}{dt}(e^{x^2+y^2}) = e^{x^2+y^2} \cdot 2x \cdot \frac{dx}{dt} + e^{x^2+y^2} \cdot 2y \cdot \frac{dy}{dt}.$$

Notice this last expression for the time derivative is algebraically equivalent to the first, since we could factor out the original function from the last expression to get the first answer.

Next, we recalled that the exponential function with base e is the inverse to ln, so

$$\ln(e^x) = x, \ \textit{all}\ x,$$

and

$$e^{\ln x} = x, \ \textit{all}\ , x > 0.$$

Whenever we have a pair of mutually inverse functions f and g, so

$$g(f(x)) = x, \ \textit{all } x \textit{ in domain } f,$$

and

$$f(g(x)) = x, \ \textit{all } x \textit{ in domain } g,$$

then if we know how to find $f'(x)$ we can use this result to find $g'(x)$. We just use the chain rule. For instance, the second equation above can be differentiated on both sides using the chain rule on the left hand side to get the equation

$$f'(g(x))g'(x) = x' = 1,$$

so

$$g'(x) = \frac{1}{f'(g(x))}$$

gives and expression for the derivative of g. When we use that with the logarithm and exponential functions, we differentiate both sides of the equation

$$e^{\ln x} = x$$

to find

$$e^{\ln x} \cdot \ln'(x) = 1,$$

so using the equation we differentiated to simplify the result we have

$$x \cdot \ln'(x) = 1,$$

and therefore

$$\ln'(x) = \frac{1}{x},$$

or

$$\ln'(x) = x^{-1}.$$

You might notice here, that when differentiating power functions we always get power functions as the result, but never did we get any constant multiple of x^{-1} as the result of differentiating a power function. For the power rule says

$$(x^p)' = px^{p-1},$$

so if we were to get some multiple of x^{-1} on the right hand side here, then we would have to have

$$p - 1 = -1,$$

whose only solution is

$$p = 0.$$

But if $p = 0$, then $x^p = x^0 = 1$ is constant whose derivative is zero, not $1/x$. Thus, it is impossible to have the derivative of any power function be $1/x$, we need the natural logarithm function ln for that.

When we combine the chain rule with the formula for differentiating ln we find that for any function f,

$$[\ln(f(x))]' = \frac{1}{f(x)} \cdot f'(x) = \frac{f'(x)}{f(x)},$$

so

$$\frac{d}{dx} \ln(f(x)) = \frac{f'(x)}{f(x)}.$$

Thus, to differentiate the natural log of any expression, we simply copy the expression into the denominator and then put its derivative in the numerator. For example,

$$\frac{d}{dx} \ln(x^6 + 7x^2) = \frac{6x^5 + 14x}{x^6 + 7x^2}.$$

Notice that as the domain of ln is all positive real numbers, the function we differentiated makes sense for all non-zero reals, since the powers are all even in the original expression inside the natural log. Likewise, the answer is defined for all nonzero real numbers.

Our next topic will be **AREA**, and before trying to use our differential calculus developed so far to tackle this problem of area, we need to be aware of some of the basic properties of continuous functions and of differentiable functions. These properties are also fundamental to many applications of calculus.

The first important property is the **INTERMEDIATE VALUE PROPERTY OF CONTINUOUS FUNCTIONS (IMVP)**. The IMVP says that if f is continuous and defined everywhere on an interval, then any number between two values of f must also be a value of f. A continuous function whose domain is an interval cannot just "skip" or "jump" over an intermediate value. If we think of the graph of such a function as a curve which can be drawn without lifting the pen, then there is no way to get from one height to another without going through all heights in between. The property seems obvious, but using the mathematical definition of continuity there is some proof required which we will not give here, but for anyone interested, consult my MATH-121 LECTURES for Fall 2009.

The second important property is the **OPTIMIZATION PROPERTY OF CONTINUOUS FUNCTIONS (OPCF)**. The OPCF says that if f is continuous on a closed finite interval, say $[a, b]$, then there must be points x_{min} and x_{max} in that closed interval with the property

$$f(x_{min}) \leq f(x) \leq f(x_{max}), \ all\ x\ in\ [a, b].$$

We call $f(x_{min})$ the **MINIMUM VALUE** of f, and we call $f(x_{max})$ the **MAXIMUM VALUE** of f. We call either one an **EXTREME VALUE** of f.

This property is obvious from pictures, but just like the IMVP, the OPCF takes proof using the mathematical definition of continuity, and is somewhat difficult to prove, but can be found in any standard text book in Mathematical Analysis.

This property (the OPCF) is useful for two reasons. First it says that there is a number M, which is the maximum of all the values of f and there is a number m which is the minimum value of all the values of f. But second, it also guarantees that the equations

$$M = f(x)$$

and

$$m = f(x)$$

have solutions in the given finite closed interval $[a, b]$. For instance, a business man wants to minimize cost and maximize profit. The cost and profit are functions of many variables, and generally these are the various input values to the various factory operations. But if cost and profit are continuous functions of the variables, then these properties still apply, and tell the business man that there are definite solutions to the problem of finding optimal operating strategies to give the maximum profit or to give the minimum cost. Obviously, the problem of minimizing cost is sort of "dual" to the problem of maximizing profit.

The actual method of finding the solutions to optimization problems actually depend on properties of differentiable functions as well. The first property of differentiable functions we consider is the "local" version of the (OPCF) for differentiable functions and it is known as **FERMAT'S THEOREM**. To describe it, we first need to say what we mean by a local extreme value, local minimum value, or local maximum value for a function f. If c is a point of the domain of f and if there is some small open subset of the domain of f so that $f(c)$ is the maximum value of f on that open subset, then we say that $f(c)$ is a local maximum value or simply a local maximum for f, and we say that f has a local maximum (value) at $x = c$. Similarly, if there is some small open subset of the domain of f such that f has minimum value $f(c)$ on that small open subset, then we say f has a local minimum (value) at $x = c$ and we call $f(c)$ a local minimum (value) of f. In either case, we say f has a local extreme (value) at $x = c$ and call $f(c)$ a local extreme (value) of f. The local extreme values are very easy to spot whenever you look at the graph of a function. Conversely, if you are trying to graph a function, knowing where all the local extreme values are (and plotting them) makes the job of drawing an accurate graph much easier than just plotting "whatever" points you feel like. Now we can state Fermat's Theorem.

FERMAT'S THEOREM. If f has a local extreme value at $x = c$ and if f is differentiable at $x = c$ and if c is not on the boundary of the domain of f, then

$$f'(c) = 0.$$

From Fermat's Theorem, we see that if f has an extreme value at $x = c$, then either f in not differentiable at $x = c$ or else f is differentiable at $x = c$ and $f'(c) = 0$. Since for most functions, as soon as we see the derivative we know where it is not differentiable, and such points in the domain of f are usually only finite in number, finding all points where the derivative is zero then gives us all the possible local extreme values, so checking the values at these finite number of points we can find the two extreme values of f. That is, the strategy for optimizing a function is to differentiate it, note all points of the domain where the derivative is undefined, set the derivative equal to zero and find all solutions. We can then check the values of f at all the points found to find the optimum values.

To see why Fermat's Theorem is true, since

$$f'(c) = \lim_{x \to c} \frac{f(x) - f(c)}{x - c},$$

for all x very near c we have the derivative is approximately the slope m_x of the line through the points $(x, f(x))$ and $(c, f(c))$ which is

$$m_x = \frac{f(x) - f(c)}{x - c}.$$

Now, if f has a local maximum at $x = c$, then the numerator is always negative for x near c, whereas the denominator is negative when $x < c$, and positive for $x > c$. Thus, looking at the values for this slope m_x with x to the left of c would lead us to conclude that $f'(c) \geq 0$, since negative divided by negative is positive, whereas looking at m_x for points x to the right of c would lead us to conclude that $f'(c) \leq 0$, since negative over positive is negative. Thus we have

$$0 \leq f'(c) \leq 0,$$

which means

$$f'(c) = 0.$$

If f has instead a local minimum at $x = c$, then $-f$ has local maximum at $x = c$, so $(-f')(c) = 0$, and therefore again $f'(c) = 0$.

Next, we combine the OPCF and differentiabilty to get **ROLLE'S THEOREM**.

ROLLE'S THEOREM. Suppose that f is a continuous function on the closed interval $[a, b]$ and that f is differentiable at all points in the open interval (a, b), that is at all points x with $a < x < b$. Also, suppose that

$$f(a) = 0 = f(b).$$

Then there is a point c in the open interval (a, b) with

$$f'(c) = 0.$$

To see this, we note that by the OPCF there are at least two points in $[a, b]$ where f has its extreme values. If both these points are end points of $[a, b]$, then as $f(a) = 0 = f(b)$, it follows that $f = 0$ on the whole interval, so f is constant and has derivative zero throughout the open interval and we can take c to be any point in the open interval (a, b). If one of these two extreme values is not zero, then it must be in the open interval (a, b), so by Fermat's Theorem f' must be zero at that point.

From Rolle's Theorem we conclude the very useful **MEAN VALUE THEOREM(MVT)**. The MVT is the real work horse of calculus.

MEAN VALUE THEOREM. Suppose that f and g are continuous on $[a, b]$ and both are differentiable at all x with $a < x < b$. Further suppose that

$$f(a) = g(a)$$

and that

$$f(b) = g(b).$$

Then there is a point c with $a < c < b$ such that

$$f'(c) = g'(c).$$

To prove the MVT we set $h = f - g$ and notice that h is continuous on $[a, b]$ and differentiable for all x with $a < x < b$. Notice then $h(a) = 0 = h(b)$, so Rolle's Theorem applies to tell us we can find c with $a < c < b$ such that $h'(c) = 0$, but

$$h'(x) = f'(x) - g'(x),\ all\ x\ in\ (a, b),$$

and therefore

$$0 = h'(c) = f'(c) - g'(c),$$

so

$$f'(c) = g'(c).$$

Pictorially, this says there must be two points, one on each of two curves between any two crossing points, both on the same vertical line where the tangent lines are parallel.

As an application, notice that if two cars start off at the same time and place, travel the same route and end up at the same place at the same time, then at some time during their trip both cars simultaneously had exactly the same velocity and therefore the same speed.

As a very useful special case of the MVT we can take g to be the function whose graph is the straight line connecting the points $(a, f(a))$ and $(b, f(b))$. Then the derivative of g is always the slope m of this line which is

$$g'(x) = m = \frac{f(b) - f(a)}{b - a}, \text{`}all\ x\ in\ (a, b),$$

and therefore there must be a point c with $a < c < b$, where

$$f'(c) = \frac{f(b) - f(a)}{b - a}.$$

This simply says that for such a function, if we draw a straight line L connecting any two points on the graph, then there is a point on the curve with tangent line T parallel to L.

25. **LECTURE** MONDAY 8 MARCH 2010

Today we reviewed the basic properties of continuous and differentiable functions discussed last time and their use as a strategy for finding optimum values for continuous functions. In particular, we showed that if f and g are continuous functions on the closed interval $[a, b]$ and if $f'(x) = g'(x)$, for all x with $a < x < b$, then $f - g$ is constant on $[a, b]$ and in particular, if there is a point c in $[a, b]$ with $f(c) = g(c)$, then in fact $f = g$ on $[a, b]$ which is to say,

$$f(x) = g(x), \ \textit{all } x \textit{ in } [a, b].$$

To prove this, we noted that if $h = f - g$, then h has derivative equal to zero, so it must be a constant function, since if it is not constant, then there are two points on the graph of h which have different heights and therefore the line L connecting those two points has non-zero slope. But by the Mean Value Theorem, there would then have to be a point on the curve between those two points where the tangent line is parallel to L and therefore where the tangent slope is not zero. But, the tangent slope is a value of the derivative of h which we already know to be zero, a contradiction. Thus, h must be a constant function, say $h = K$, so

$$f(x) - g(x) = K, \ \textit{all } x \textit{ in } [a, b].$$

If further there is some point c in $[a, b]$ with $f(c) = g(c)$, then

$$K = f(c) - g(c) = 0,$$

and therefore $K = 0$, so $f = g$ on $[a, b]$.

For instance, as an application, suppose that we want to find a certain function f and we know $f'(x) = x^2$, and $f(0) = 5$. We can easily see that

$$[(1/3)x^3]' = x^2,$$

so we now know that

$$f(x) = \frac{1}{3}x^3 + K,$$

for some constant K. Now, use the fact that $f(0) = 5$ to find K.

$$5 = f(0) = \frac{1}{3}0^3 + K = K,$$

so it must be that $K = 5$, and therefore,

$$f(x) = \frac{1}{3}x^3 + 5.$$

Problems like this where we have equations involving derivatives are called **DIFFERENTIAL EQUATIONS**. For instance, if our information about an unknown function tells us the second derivative, then we can work our way back to the original function stepwise. For instance, if

$$f''(x) = x^2,$$

then as f'' is the derivative of f', we can say that

$$f'(x) = \frac{1}{3}x^3 + K,$$

for some constant K. Now, we know that the derivative of Kx with respect to x is K, and therefore

$$[(1/3)(1/4)x^4 + Kx]' = \frac{1}{3}x^3 + K = f'(x),$$

so we can say that

$$f(x) = \frac{1}{12}x^4 + Kx + C,$$

where K and C are some constants. If we know that $f(0) = 5$ and $f'(0) = 7$, say, then

$$5 = f(0) = \frac{1}{12}0^4 + K \cdot 0 + C = C,$$

so we find $C = 5$, and

$$7 = f'(0) = \frac{1}{3}0^3 + K = K,$$

so $K = 7$, and this means that

$$f(x) = \frac{1}{12}x^4 + 7x + 5.$$

The answer here is partly being determined by knowing the values of f and its derivative at $x = 0$. We could have instead been given values $f(2)$, say and $f'(4)$, for instance, and we could have still solved the problem. Notice also that we can do this with any number of derivatives. For instance, if we are given the twentieth derivative of f as a function of x and if we are given the values of each lower derivative at some single point (possibly at different points for each derivative) then we can find the original f. These values of the lower derivatives are called *Initial Conditions* for the differential equation, because we often take the value $x = 0$ to be where we specify all the lower derivatives' values.

As an example, in motion problems we are often given the acceleration as a function of time for all time, from Newton's Laws, and then if we know the starting position and starting velocity, we can use the preceding method to find the position at every instant of time. That is if we know the initial position and initial velocity, and if we know the acceleration at all time, then we can find the velocity for all time and the position for all time.

We are going to apply these techniques to the problem of finding the area under a curve which is the graph of a continuous function. We will assume that areas bounded by continuous curves can be defined so as to have the properties for areas with moving boundaries discussed back in the first few lectures of the semester. Thus, if we have a region with a piece of boundary of length L which moves out at a certain instant with velocity v, then we recall that $L \cdot v$ is the rate of change of area due to that part of the moving boundary, that is to say, if $A(t)$ is the area at time t, then

$$\frac{dA}{dt} = L \cdot v.$$

Suppose now that f is a continuous function on the closed interval $[a, b]$ and that $f \geq 0$, so the graph of f never goes below the horizontal axis. The horizontal axis, the vertical lines $x = a$ and $x = b$ and the curve which is the graph of f, that is $y = f(x)$ are then four curves forming the boundary of a region R. We want to find the area of R which we denote by A_b. To do this, we consider a variable region we denote by $R(x)$ whose left boundary is the curve $x = a$, but whose right boundary is the vertical line through x, and of course the lower boundary is still the horizontal axis and upper boundary is the part of the graph of f over the interval $[a, x]$. We denote the area of $R(x)$ by $A(x)$, so A is a function of x, and clearly $A(b) = A_b$ is the area we really want to find. We now replace our problem with what at first appears to be a more difficult problem, namely finding the function A, which means finding $A(x)$ for all x in $[a, b]$. Now, imagine we allow x to move with time from left to right with velocity v. Then

$$\frac{dx}{dt} = v.$$

But, as x moves from left to right, for the region $R(x)$, we see that its right hand boundary, the vertical line through x is also moving from left to right, and its length at the instant it is located at x is precisely $f(x)$. That is if we imagine the value of x changes with time, then $R(x)$ is changing with time by having a moving boundary piece of length $L = f(x)$ which moves at velocity v. Thus by our formula for the rate of change of area due to a moving boundary, we have

$$\frac{dA}{dt} = L \cdot v = f(x) \cdot v = f(x) \cdot \frac{dx}{dt} = f(x(t)) \cdot \frac{dx}{dt}.$$

On the other hand, as A depends on x which in turn depends on t, we know by the Chain Rule,

$$\frac{dA}{dt} = A'(x) \cdot \frac{dx}{dt} = A'(x) \cdot v.$$

Therefore, we have the equation

$$A'(x) \cdot v = f(x) \cdot v.$$

Thus, provided we actually make x move, we can conclude that $v \neq 0$, and can be canceled here leaving us with

$$A'(x) = f(x).$$

Suppose next that using the techniques discussed above, we have found some function F on $[a, b]$ with the property that $F'(x) = f(x)$, for all x strictly between a and b, which is to say for all x satisfying $a < x < b$. Then we know that $A = F + C$ for some constant C. On the other hand, we know that $A(a) = 0$, so we have

$$0 = A(a) = F(a) + C,$$

and therefore

$$C = -F(a),$$

which means

$$A(x) = F(x) - F(a), \textit{ all } x \textit{ in } [a, b].$$

In particular, to find A_b, the area under the curve, we calculate

$$A_b = A(b) = F(b) - F(a).$$

This process of calculating $F(b) - F(a)$, that is the difference of two values of F comes up so much in area calculations, that we have a very useful notation for it:

$$F(x)|_a^b = F(b) - F(a).$$

As an example, suppose that we wish to find the area A under $y = x^2$ between $x = 1$ and $x = 2$. We notice that with $f(x) = x^2$, we can take

$$F(x) = \frac{1}{3}x^3$$

and have $F'(x) = f(x)$, so

$$A = \frac{1}{3}x^3|_1^2 = \frac{1}{3}2^3 - \frac{1}{3}1^3 = \frac{8}{3} - \frac{1}{3} = \frac{7}{3}.$$

In general, if $F'(x) = f(x)$, for all x in the domain of f, we say that F is an **ANTIDERIVATIVE** of f. We say "an" antiderivative here, because notice any constant can be added to F and the result is still an antiderivative of f, so as soon as we find an antiderivative of f we can find infinitely many antiderivatives of f. However, on any interval in the domain of f all antiderivatives can only differ by constants.

We can notice that if f is a power function, say $f(x) = x^p$, then we see easily

$$F(x) = \frac{x^{p+1}}{p+1}$$

is an antiderivative of f. Notice that this does not work for $p = -1$, but here we can recall that

$$\ln'(x) = \frac{1}{x}.$$

Thus, if $A(x)$ is the area under the curve $y = 1/x$ between $x = 1$ and $x = b$, with $b > 1$, then

$$A(b) = \ln(x)|_1^b = \ln(b) - \ln(1) = \ln(b),$$

and therefore, we can view ln as the function which gives the area under the curve $y = 1/x$, for $x > 1$. In fact, this can be used as the definition of the function ln, and as $1/x$ is always positive for positive x, it follows that the graph of ln always increases as you go from left to right, and therefore the graph of ln must satisfy the horizontal line test so it has an inverse function exp, which is the exponential function. This is the simplest way to actually define the log and exp functions since all the technical problems are reduced to problems of areas, which can in turn be solved rigorously.

26. **LECTURE** WEDNESDAY 10 MARCH 2010

Today we reviewed and took QUIZ 5 in class.

27. **LECTURE** FRIDAY 12 MARCH 2010

Today we discussed the **COMPLETENESS PROPERTY** of the set $\mathbb{R}$ of all real numbers, and the use of this property to solve problems of area and problems where solutions require irrational numbers.

To begin, suppose that A is any subset of $\mathbb{R}$ and that b is any real number, so b is in $\mathbb{R}$. We say that b is an **UPPER BOUND** for A provided that $a \leq b$, for every a in A. Likewise, we say that b is a **LOWER BOUND** for A provided that $b \leq a$, for every a in A. To picture this, think of $\mathbb{R}$ as represented by a horizontal line with a number scale, so that the number one is to the right of the number zero, and in this way every point on the line becomes a real number, with numbers increasing from left to right. The condition that b is an Upper Bound for A the is simply that A contains no points to the right of b. The condition that b is a Lower Bound for A says simply that A contains no points to the left of b. We say that A is **BOUNDED ABOVE** if A has an upper bound, and we say that A is **BOUNDED BELOW** if A has a lower bound. We say that b is a **LEAST UPPER BOUND(LUB)** for A provided that b is an upper bound for A with the property that if c is any upper bound for A, then $b \leq c$. Likewise, we say that b is a **GREATEST LOWER BOUND(GLB)** for A provided that b is a lower bound for A with the property tat if c is any lower bound for A, then $b \geq c$. Notice that A can have at most one LUB, since if b and c are both LUB's for A, then we have in particular that both are upper bounds for A and thus both $b \leq c$ and $c \leq b$, and therefore $b = c$. Likewise, you can easily prove that A can have at most one LUB, since if both b and c are LUB's for A, then both $b \geq c$ and $c \geq b$. Thus, if A has a LUB, then it is unique, and if A has a GLB, it is also unique.

The **COMPLETENESS PROPERTY** of $\mathbb{R}$ says that if A has an upper bound then it has a LUB. If A has a lower bound, then $-A$, the set of all negatives of numbers in A has an upper bound, so it has a LUB whose negative bcomes the GLB of A. Thus the completeness property also supplies a GLB to any subset of $\mathbb{R}$ which has a lower bound.

To see how the completeness property of $\mathbb{R}$ can be used to get the existence of certain irrational numbers, we begin with fact which we will not prove: if n is a positive integer, then either $\sqrt{n}$ is also a positive integer, or $\sqrt{n}$ is irrational. In particular, this means that $\sqrt{2}$ is irrational. When the Pythagorean Brotherhood discovered this fact, as the ancient Greeks thought of all numbers as being rational, the Pythogoreans decided to keep it a secret and threatened death to any one of their members who would disclose the fact to a non-member. In any case, to see that the equation $x^2 = 2$ has a solution in $\mathbb{R}$, we know already we cannot find a rational number which serves as a solution, so somehow the completeness property will be needed. We can form the set A consisting of all real numbers x with the property that $x^2 < 2$. We then observe that A has an upper bound. For instance, obviously ten is an upper bound for A. Let b be the LUB of A. Notice if x is in A, then we can find a number $y > x$ which is so slightly bigger than x that $y^2 < 2$ as well. This means that x cannot be an upper bound for A, that is to say, no member of A is an upper bound for A, and therefore b is not in A. This means that $b^2 \geq 2$. If it is the case that $b^2 > 2$, we could decrease b slightly finding a number $c < b$ with $c^2 > 2$. Then c must also be an upper bound for A, but as $c < b$, this would contradict the fact that b is the LUB for A. Thus, we must have $b^2 = 2$, so $\sqrt{2} = b$. Thus, the completeness property of $\mathbb{R}$ is what guarantees that all non-negative real numbers have square roots, since the preceding argument can be repeated with two replaced by any non-negative real number we choose. In particular, this shows that there are points on the geometric line which have irrational coordinates, since the square root of two can be geometrically constructed with ruler and compass. Simply construct a segment of length two on a line using a compass and construct a forty five degree angle line. Use the compass to lay off the segment on the forty five degree line and then drop a perpendicular from that two unit segment on the forty five degree line back to the original line. You have then constructed an irrational number on the number line.

We can also use the completeness property of $\mathbb{R}$ to define area of plane regions (subsets of the plane). Suppose that $\mathcal{R}$ is any subset of the geometric plane. We call T a **TRIANGULATION** in the plane if T is a finite collection of triangles which do not overlap except possibly on an edge. If T is a triangulation, then we define $Area(T)$ to be the sum of the areas of all the triangles making up T. We say that T is an **Inner triangulation** of $\mathcal{R}$ if all the triangles of T are in $\mathcal{R}$. We say that T is an **Outer Triangulation** of $\mathcal{R}$ provided that every point of $\mathcal{R}$ lies on some triangle of T. Notice that if T is an outer triangulation of $\mathcal{R}$, then $area(T) \geq 0$, so zero is a lower bound for the set of areas of outer triangulations of $\mathcal{R}$. On the other hand, if $\mathcal{R}$ can be put inside a big rectangle, then we easily find an outer triangulation, and its area serves an an upper bound for the set of all areas of inner triangulations of $\mathcal{R}$. Therefore, we can define $\mathcal{A}_{in}(\mathcal{R})$ to be the set of all areas of inner triangulations of $\mathcal{R}$ and know that $\mathcal{A}_{in}(\mathcal{R})$ has a LUB. Likewise, we can define $\mathcal{A}_{out}(\mathcal{R})$ to be the set of areas of outer triangulations of $\mathcal{R}$ and be guaranteed that $\mathcal{A}_{out}(\mathcal{R})$ has a GLB. We then define the inner and outer areas for $\mathcal{R}$ as

$$Inner\ Area(\mathcal{R}) = LUB\ of\ \mathcal{A}_{in}(\mathcal{R})$$

and

$$Outer\ Area(\mathcal{R}) = GLB\ of\ \mathcal{A}_{out}(\mathcal{R}).$$

If both the outer and inner areas of $\mathcal{R}$ are the same, then it is natural to say that $\mathcal{R}$ has an area and call that common value the area of $\mathcal{R}$ denoted $Area(\mathcal{R})$. Thus,

$$Inner\ Area(\mathcal{R}) = Area(\mathcal{R}) = Outer\ Area(\mathcal{R}),\ if\ and\ only\ if\ \mathcal{R}\ has\ area.$$

Theorem 27.1. *If $\mathcal{R}$ is a plane set whose boundary consists of a finite union of differentiable curves, then $\mathcal{R}$ has area.*

It is the case that the theorem above can be generalized. In fact $\mathcal{R}$ has area if its boundary consists of a finite union of continuous curves of a restricted type. For each curve one needs to be able to construct coordinate axes in the plane so that the curve is the graph of a continuous function. In particular, if you can find an inner triangulation such that along the outer boundary of the triangulation the boundary curves form graphs of continuous functions where the "horizontal" axis is the outer edge of a triangle, then that is good enough. Notice that to prove this fact, it is enough to prove it in case you have the graph of a continuous function and the region is simply the region underneath the graph and above the horizontal axis and between two vertical lines. This is just what we did last time using antiderivatives, but we assumed that the region had an area.

To see that there are regions which do not have area, consider two squares in the coordinate plane, one inside the other, say $\mathcal{R}_1$ and $\mathcal{R}_2$, where $\mathcal{R}_2$ denotes the larger square. Let $\mathcal{R}$ be the set of all point in $\mathcal{R}_2$ which are either in $\mathcal{R}_1$ or else have at least one rational coordinate. Obviously any rectangle has an area. Then any inner triangulation if $\mathcal{R}$ will have to have all its triangles contained in $\mathcal{R}_1$, and therefore

$$Inner\ Area(\mathcal{R}) = Area(\mathcal{R}_1).$$

On the other hand, any outer triangulation of $\mathcal{R}$ will have to cover up $\mathcal{R}_2$, so that

$$Outer\ Area(\mathcal{R}) = Area(\mathcal{R}_2).$$

As an extreme example, we could take the inner rectangle to be empty, so the inner area of $\mathcal{R}$ is zero and the outer area of $\mathcal{R}$ can be as large as we like by simply taking $\mathcal{R}_2$ very big. Notice that in these examples, the boundary of $\mathcal{R}$ is very nasty and in some sense "thick". These are not the regions we draw with pen and paper.

28. **LECTURE** MONDAY 15 MARCH 2010

Today we used the properties of continuous functions to show that if $f \geq 0$ is continuous on $[a, b]$, the with R denoting the region under the graph of f between the vertical lines $x = a$ and $x = b$, then

$$Inner\ Area(R) = Outer\ Area(R),$$

so that the area under the graph of f actually makes sense, and

$$Area(R) = \int_a^b f = \int_a^b f(x)dx.$$

In more detail, let us suppose that f is continuous on the closed interval $[a, b]$. We can form what we call a **PARTITION** of the interval $[a, b]$ by which we mean a sequence of points

$$a = x_0 < x_1 < x_2 ... < x_{n-1} < x_n = b.$$

Let

$$P = (x_0, x_1, x_2, ..., x_{n-1}, x_n)$$

be used to denote this partition. For each k with $1 \leq k \leq n$, we call the interval $[x_{k-1}, x_k]$ a **SUBINTERVAL** of P or of $[a, b]$, more specifically, it is the k^{th} subinterval of the partition. Then the length of the k^{th} subinterval is obviously $x_k - x_{k-1}$ which we denote by Δx_k, so

$$\Delta x_k = x_k - x_{k-1}.$$

We can form particular inner and outer triangulations of the region under the graph of f by using rectangles whose sides are vertical and cross the horizontal axis at each of the partition points x_k, $k = 0, 1, 2, 3, ..., n$. This gives a triangulation since any rectangle can be cut into two triangles using one of its diagonals. For the inner triangulation, we chose the rectangle over the k^{th} subinterval to have height the minimum value of f on this subinterval, which we know exists as f is continuous on this subinterval. Let m_k be the minimum value of f on the k^{th} subinterval. Then the area of the rectangle whose base is the k^{th} subinterval on the horizontal axis and whose height is m_k is obviously

$$\Delta A_k = m_k \cdot \Delta x_k.$$

If it exists, the area under the graph of f is at least as much as the sum of these areas, which we call a **LOWER SUM** for f denoted by $L(f, P)$, so

$$L(f, P) = \sum_{k=1}^{n} m_k \cdot \Delta x_k.$$

On the other hand, as f is continuous, it is also the case that f has a maximum value M_k on the k^{th} subinterval, so using the rectangle of height M_k over the k^{th} subinterval instead for each $k \leq n$ gives the corresponding **UPPER SUM** for f denoted $U(f, P)$ so

$$U(f, P) = \sum_{k=1}^{n} M_k \cdot \Delta x_k.$$

Thus if the area under the graph of f exists it must be no more than $U(f, P)$. More generally, we can choose any "sampling of P" that is choose "sample" points x_k^* so that x_k^* is in the k^{th} subinterval for each $k \leq n$ and we have

$$m_k \leq f(x_k^*) \leq M_k,$$

and therefore

$$L(f, P) \leq \sum_{k=1}^{n} f(x_k^*)\Delta x_k \leq U(f, P).$$

We call such a sum a **RIEMANN SUM** for f on $[a, b]$ or for the partition P. More specifically, if we let

$$x^* = (x_1^*, x_2^*, x_3^*, ..., x_n^*),$$

then this Riemann sum is denoted $R(f, P, x^*)$, so

$$R(f, P, x^*) = \sum_{k=1}^{n} f(x_k^*)\Delta x_k.$$

Obviously then, we have

$$U(f, P) \leq R(f, P, x^*) \leq U(f, P),$$

no matter how x^* is chosen, as long as it is a sampling of P.

Let us denote by $|P|$ the maximum length of all the subintervals of the partition P. Notice that as f is continuous, as $|P|$ approaches zero, we must have m_k and M_k becoming close to each other so the lower and upper sums appear to have the same limit as $|P| \to 0$. We can think of this as what happens as $n \to \infty$, even though we really need to be careful here, because unless some rule is enforced which keeps the subintervals of somewhat similar length, then we can have n going to infinity but $|P|$ not going to zero. For instance, if we partition the interval into equal subintervals, then Δx_k is simply $(b - a)/n$ for each k, and if we denote the resulting partition as P_n, then $|P_n| = (b - a)/n$ which certainly goes to zero as $n \to \infty$.

Here is the crucial property of continuous functions which allows us to prove that the inner area and outer area are the same for the region under the graph of f. it is called **UNIFORM CONTINUITY** and it says that if $c > 0$ is any positive number, no matter how small, then there is another positive number $d > 0$ so that if x and y are any points of $[a, b]$ with

$$|x - y| < d,$$

then in fact

$$|f(x) - f(y)| < c.$$

Notice this means that if $|P| < d$, then $M_k - m_k < c$ for each k and therefore

$$U(f, P) - L(f, P) \leq \sum_{k=1}^{n} c \cdot \Delta x_k = c \cdot \sum_{k=1}^{n} \Delta x_k = c \cdot (b - a).$$

This means that as $|P| \to 0$, we can take c as small as we like forcing the upper and lower sums to converge to the same number. Thus the inner and outer areas of the region under the graph of f must be equal which means the area under the graph of f exists. In particular, this means that if R denotes the region under the graph of f, then

$$Area(R) = \lim_{|P| \to 0} R(f, P, x^*).$$

More generally, our arguments with upper and lower sums showing that the limit of Riemann sums exists makes perfectly good sense even if f has some negative values, we just do not have a region under the graph of f, although we do have a region trapped between the graph of f and the horizontal axis. For any function f on $[a, b]$ we define

$$\int_a^b f(x)\ dx = \lim_{|P| \to 0} R(f, P, x^*),$$

provided of course that the limit exists. When it does, we say the function is Riemann integrable on $[a, b]$. Thus we have shown that any continuous function on $[a, b]$ is Riemann integrable on $[a, b]$.

We say that the partition P of $[a,b]$ is finer than the partition Q of $[a,b]$ if every subinterval of P is contained in a some subinterval of Q. Thus, the boundary points of the subintervals for Q are also boundary points for subintervals of P. We see easily that if f is any function on $[a,b]$, then

$$L(f,Q) \leq L(f,P),\ P\ finer\ than\ Q,$$

and

$$U(f,P) \leq U(f,Q),\ P\ finer\ than\ Q.$$

In particular, if we take any two P and Q, then we can form a partition S using all the subinterval boundary points from both partitions, and S will be finer than both P and Q. This means

$$L(f,P) \leq L(f,S) \leq U(f,S) \leq U(f,Q),$$

from which we conclude that ANY lower sum for f cannot exceed any upper sum. That is, every upper sum is an upper bound for the set of all lower sums and every lower sum is a lower bound for the set of all upper sums. Thus, the set of all upper sums has a greatest lower bound which we call the upper integral of f on $[a,b]$ and denoted

$$\overline{\int_a^b} f(x)dx = GLB_P\ U(f,P).$$

And, the set of all lower sums has a least upper bound which we call the lower integral of f on $[a,b]$ which we denote

$$\underline{\int_a^b} f(x)dx = LUB_P\ L(f,P).$$

Since every lower sum is a lower bound for all upper sums and as the upper integral is the greatest lower bound, this means that no lower sum can exceed the upper integral, so

$$L(f,P) \leq \overline{\int_a^b} f(x)dx,\ any\ P,$$

and therefore the upper integral is also an upper bound for all lower sums, which meas that the upper integral cannot be less than the least upper bound for all lower sums, that is the upper integral is at least as much as the lower integral, so

$$\underline{\int_a^b} f(x)dx \leq \overline{\int_a^b} f(x)dx.$$

We can therefore say that f is Riemann integrable on $[a,b]$ if and only if the upper and lower integrals agree, in which case their common value is the Riemann integral of f on $[a,b]$. Thus, if f is Riemann integrable on $[a,b]$, then

$$\underline{\int_a^b} f(x)dx = \int_a^b f(x)dx = \overline{\int_a^b} f(x)dx.$$

Notice that in case that $f \geq 0$, we must have for the region R under the graph of f that

$$\underline{\int_a^b} f(x)dx \leq Inner\ Area(R) \leq Outer\ Area(R) \leq \overline{\int_a^b} f(x)dx,$$

so if f is Riemann integrable, then the inner and outer areas must be equal and therefore R has area, and

$$Area(R) = \int_a^b f(x)dx.$$

It is easy to see that if f and g are Riemann integrable on $[a, b]$, then, by considering Riemann sums, so is their sum $f + g$ and

$$\int_a^b [f(x) + g(x)]dx = \int_a^b f(x)dx + \int_a^b g(x)dx.$$

Also for any constant C it is even more obvious that if f is Riemann integrable on $[a, b]$ then so is $C \cdot f$ and

$$\int_a^b C \cdot f(x)dx = C \cdot \int_a^b f(x)dx.$$

Another fact which is very useful in computations is the fact that if $a < b < c$, and if f is Riemann integrable on $[a, c]$, then it is also Riemann integrable on both $[a, b]$ and $[b, c]$ and as well, if f is Riemann integrable on each of the intervals $[a, b]$ and $[b, c]$, then f is Riemann integrable on $[a, c]$. And in fact,

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

To see this, we can note that if P is a partition of $[a, b]$ and Q is a partition of $[b, c]$ then we can put the two partitions together to form a partition of $[a, c]$ so this means that $L(f, P) + L(f, Q)$ is a lower sum for f on $[a, c]$ and $U(f, P) + U(f, Q)$ is an upper sum for f on $[a, c]$. It follows that

$$\underline{\int_a^b} f(x)dx + \underline{\int_b^c} f(x)dx \leq \underline{\int_a^c} f(x)dx \leq \overline{\int_a^c} f(x)dx \leq \overline{\int_a^b} f(x)dx + \overline{\int_b^c} f(x)dx.$$

But, if f is Riemann integrable on both the intervals $[a, b]$ and $[b, c]$, then the first and last terms of the inequalities must be the same so in fact all are equal which forces f to be Riemann integrable on $[a, c]$.

29. **LECTURE** WEDNESDAY 17 MARCH 2010

Today we used properties of continuous functions to prove

Theorem 29.1. THE FUNDAMENTAL THEOREM OF CALCULUS. *If f is continuous on $[a, b]$ and then f has an antiderivative F on $[a, b]$ meaning that F is continuous on $[a, b]$ and differentiable on the open interval (a, b) with $F' = f$ on (a, b). Moreover, if F is any antiderivative for f on $[a, b]$, then it can be used to compute the Riemann integral of f on $[a, b]$ and in fact*

$$\int_a^b f(x)\ dx = F(b) - F(a),\ if\ F'(x) = f(x),\ for\ all\ x\ satisfying\ a < x < b.$$

Here

$$\int_a^b f(x)\ dx = \lim_{|P|\to 0} R(f, P, x^*),$$

is the Riemann integral of f on the interval $[a, b]$ which then gives the area under f if $f \geq 0$, provided the limit on the right hand side of the equation actually exists. In general, if f is continuous, then we showed last time that the limit actually exists when $f \geq 0$, using pictures, and in fact it exists as long as f is continuous, and the condition $f \geq 0$ is not needed. If the region between the graph of f on $[a, b]$ and the horizontal axis has regions below the horizontal axis, then obviously those regions will be counted negatively by the Riemann integral. For instance, if the graph of f on $[a, b]$ consists of the three regions R_1, R_2, R_3 where R_1 is above the horizontal axis, R_2 below the axis, and R_3 above, then

$$\int_a^b f(x)\ dx = area(R_1) - area(R_2) + area(R_3).$$

It is also customary to refer to the Riemann integral of f on $[a, b]$ as the **DEFINITE INTEGRAL** of f and a is called the **LOWER LIMIT OF INTEGRATION**, whereas b is called the **UPPER LIMIT OF INTEGRATION**. We also write for short sometimes

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

for this definite integral. When the limits are left off, the integral is called an **INDEFINITE INTEGRAL** and in that case is merely a symbol for any antiderivative of f. Thus, with C an arbitrary constant,

$$\int f(x)\ dx = F(x) + C.\ if\ and\ only\ if,\ F'(x) = f(x),$$

or with shorter notation,

$$\int f = F + C,\ if\ and\ only\ if,\ F' = f.$$

We discussed the Mean Value Theorem for integrals as a consequence of the intermediate value theorem and used that to prove the Fundamental Theorem of Calculus.

Theorem 29.2. MEAN VALUE THEOREM FOR INTEGRALS. *If f is continuous on $[a, b]$ then there is at least one point c in the interval $[a, b]$ with the property that*

$$\int_a^b f(x)dx = f(c) \cdot (b - a).$$

Notice that the mean value theorem for integrals says that one of the Riemann sums is already exactly equal to the Riemann integral of f on $[a, b]$, for in fact $f(c) \cdot (b - a)$ is the Riemann sum where the partition has only one subinterval, namely $[a, b]$ itself, and c is then the only sample point. If you get lucky in your choice of sample point, you might get the exact value of the integral with only one subinterval in the partition. To see that this must be true, notice that as f is continuous on $[a, b]$, in fact, f has a minimum value m and a maximum value M on $[a, b]$ by the optimization theorem for continuous functions. But then

$$m \cdot (b - a) \leq \int_a^b f \leq M \cdot (b - a),$$

so dividing both sides by $(b - a)$ we find

$$m \leq \frac{1}{b - a} \int_a^b f \leq M.$$

But, then by the intermediate value theorem for continuous functions, we know that there is some number c in $[a, b]$ with

$$f(c) = \frac{1}{b - a} \int_a^b f.$$

In general, it is customary to call the right hand side of the previous equation the average value of f on the interval $[a, b]$. But when we multiply both sides of this equation by $(b - a)$, we get the mean value theorem for integrals.

To prove the Fundamental Theorem of Calculus, if f is continuous on $[a, b]$ then for any x in $[a, b]$ we know that f is Riemann integrable on the interval $[a, x]$, so we define the function G by the rule

$$G(x) = \int_a^x f.$$

Thus, in case that $f \geq 0$, the function G is giving the area under the graph from a to x. As x moves to the right, we know from our intuitive arguments from the beginning lectures that as the right edge boundary has length $f(x)$, if x moves to the right with velocity $v > 0$, then

$$\frac{dG}{dt} = f(x) \cdot v.$$

On the other hand, if we can prove that G is differentiable, then, as

$$\frac{dx}{dt} = v,$$

by the chain rule

$$\frac{dG}{dt} = G'(x)\frac{dx}{dt} = G'(x) \cdot v.$$

Putting these facts together gives us

$$G'(x) \cdot v = \frac{dG}{dt} = f(x) \cdot v,$$

so as $v > 0$, we have

$$G'(x) = f(x).$$

Now this is really just an "intuitive" argument, and here we can actually use our properties of continuous functions so far to give this a more rigorous proof. We note that if we change x by an amount Δx, then the slope of the line through the two corresponding points on the graph of G is rise over run, where the run is Δx but the rise is

$$G(x+\Delta x) - G(x) = \int_a^{x+\Delta x} f - \int_a^x f = \int_a^x f + \int_x^{x+\Delta x} f - \int_a^x f = \int_x^{x+\Delta x} f.$$

But then by the mean value theorem for integrals, there is a point x^* between x and $x + \Delta x$ with

$$\int_x^{x+\Delta x} f = f(x^*) \cdot \Delta x.$$

This means that the slope we want is rise over run where the run is Δx and the rise is $f(x^*) \cdot \Delta x$. This in turn means that the slope or rise over run is $f(x^*)$. Now as $\Delta x \to 0$, since x^* is always between x and Δx, it must be that $x^* \to x$. But as f is continuous, as $x^* \to x$ this causes $f(x^*) \to f(x)$. Thus, the limit slope here must have value $f(x)$, which is to say that G is differentiable at x and

$$G'(x) = f(x), \ \textit{any } x \textit{ with } a < x < b.$$

Thus, we have shown that f has an antiderivative on $[a, b]$. Now, suppose that F is any antiderivative of f on $[a, b]$, meaning that F is continuous on $[a, b]$, differentiable on the open interval (a, b), and with $F'(x) = f(x)$ for all x in the open interval (a, b). Then we now know that $G = F + C$, for some constant C as both F and G have the same derivative on (a, b). We notice that we also have

$$\int_a^b f = G(b),$$

so to compute the integral using F instead of G, we only need to figure out C. But, notice that $G(a) = 0$, so

$$0 = G(a) = F(a) + C,$$

and therefore

$$C = -F(a),$$

forcing

$$G(x) = F(x) - F(a), \ \textit{all } x \textit{ in } [a, b],$$

and in particular, giving

$$G(b) = F(b) - F(a).$$

Thus finally,

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

The beauty and usefulness of this result is that we can use any antiderivative we can find to solve the problem of computing the Riemann integral and thereby circumvent the process of calculating limits of Riemann sums.

As a useful computational procedure here, we denote

$$F(x)|_a^b = F(b) - F(a),$$

so the computation of the integral using the fundamental theorem proceeds by first finding the antiderivative and then evaluating between the two limits of integration:

$$\int_a^b f(x)dx = F(x)|_a^b = F(b) - F(a).$$

As we read from left to right, we see the first step is to write down the antiderivative and copy the limits of integration on the vertical line drawn after the antiderivative function. Then the second step is to substitute the limits of integration into the antiderivative to calculate the value of the integral.

Keep in mind that any differentiation formula can be turned into a result about antiderivatives.

We have then

$$\int x^p dx = \frac{x^{p+1}}{p+1} + C,\ p \neq -1,$$

$$\int \frac{1}{x} dx = \ln x + C,\ x > 0,$$

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

just to give a few examples. We can note that if $x < 0$, then $|x| > 0$ and $|x| = -x$, so as

$$[\ln(-x)]' = \frac{1}{-x}(-1) = \frac{1}{x},$$

we see that

$$\int \frac{1}{x} dx = \ln |x| + C,\ x \neq 0.$$

We worked examples showing how to apply the fundamental theorem to the calculation of integrals and areas.

30. **LECTURE** FRIDAY 19 MARCH 2010

We began reviewing for TEST 2. We reviewed some antidifferentiation formulas:

$$\int x^p dx = \frac{x^{p+1}}{p+1} + C, \; p \neq -1,$$

$$\int \frac{1}{x} dx = \ln|x| + C, \; x \neq 0,$$

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

$$\int \ln x dx = x \ln x - x + C, \; x > 0.$$

We also reviewed calculations with exponential functions as applied in situations of growth and decay. For instance, if

$$Y(t) = A \cdot e^{rt},$$

then the rate of change at time t is simply

$$Y'(t) = Y(t) \cdot r,$$

as an immediate consequence of the chain rule.

31. LECTURE MONDAY 22 MARCH 2010

We reviewed the answers to problems on PRACTICE TEST 2.

32. LECTURE WEDNESDAY 24 MARCH 2010

TEST 2 IN CLASS

33. LECTURE FRIDAY 26 MARCH 2010

We went over the answers to TEST 2.

34. LECTURE MONDAY 29 MARCH 2010

NO CLASS-SPRING BREAK

35. LECTURE WEDNESDAY 31 MARCH 2010

NO CLASS-SPRING BREAK

36. LECTURE FRIDAY 2 APRIL 2010

NO CLASS-SPRING BREAK

37. LECTURE MONDAY 5 APRIL 2010

NO CLASS-SPRING BREAK

38. LECTURE WEDNESDAY 7 APRIL 2010

We reviewed and discussed the method of integration by substitution which is the result of the applying the chain rule to anti-differentiation.

39. **LECTURE** FRIDAY 9 APRIL 2010

40. **LECTURE** MONDAY 12 APRIL 2010

41. **LECTURE** WEDNESDAY 14 APRIL 2010

42. **LECTURE** FRIDAY 16 APRIL 2010

43. **LECTURE** MONDAY 19 APRIL 2010

44. **LECTURE** WEDNESDAY 21 APRIL 2010

45. **LECTURE** FRIDAY 23 APRIL 2010

46. **LECTURE** MONDAY 26 APRIL 2010

47. **LECTURE** WEDNESDAY 28 APRIL 2010

48. **LECTURE** FRIDAY 30 APRIL 2010