MATH-115 (DUPRÉ) FALL 2009 LECTURES

1. **LECTURE** MONDAY 24 AUGUST 2009

Today we discussed the two basic problems of calculus and their relationship through the Fundamental Theorem of Calculus. The first problem is the problem of finding the tangent line to a curve at a point on the curve. The second problem is to find the area of a region bounded by a curve. The Fundamental Theorem of Calculus is based on a simple but subtle trick. One must go beyond the static area problem and consider rates of change of areas of variable regions. First, one can easily see intuitively with pictures that if we spill some ink on a table top, then (if area even makes sense) the area A changes with time. The solution to the tangent problem provides the solution to the problem of rate of change in general. If X is a quantity which changes with time t, then $\dot{X}$ denotes the rate of change of X with respect to time t. The question as to whether rate of change makes sense was deeply troubling for Greek mathematicians and prevented them from developing calculus. But, in our modern time, anyone who drives a car understands the idea of speed and feels comfortable reading a speedometer, even if they have never taken calculus. Technically, the rate of change of position is called velocity and is a vector quantity since it involves magnitude as well as direction. In the car, the speedometer tells the speed (magnitude of velocity) and the view through the windshield tells the direction. To successfully pilot any moving vehicle or aircraft requires a good intuitive feeling for velocity. The windshield is not necessary if instruments are available to tell direction, and aircraft pilots must be able to fly on instruments alone. If we return to the ink spill, we see that at the instant when the moving boundary has length L and all its points are moving out in a direction perpendicular (or normal) to the boundary curve at velocity v, then the small change in area, ΔA, of the ink spill during a very small amount of elapsed time Δt is to a very high degree of accuracy simply $Lv\Delta t$ since the region of the increase is a very thin region with curved boundaries almost "parallel" of length L and width $v\Delta t$. Since the region is so thin, the fact that its thickness is small by comparison to its curvature means its area to good approximation is just length times width. For instance, if we were to try to paint a six inch wide median stripe down a ten mile length of mountain road, we know the area we must paint is for all practical purposes just (10)(5280)(1/2) square feet. The curvature of the road is of no practical significance. For our moving ink spill we then have to good approximation

$$\Delta A \doteq Lv\Delta t$$

and therefore, as we also have to good approximation $\Delta A \doteq \dot{A}\Delta t$, we find very approximately

$$\dot{A}\Delta t \doteq Lv\Delta t.$$

Canceling Δt on both sides gives the approximate equality

$$\dot{A} \doteq Lv,$$

with accuracy of this approximation naturally increasing as Δt goes to zero. Since the Δt does not appear in the last equation, the inescapable conclusion is that the equation is exactly true: the rate of change of area is simply the length of the moving boundary multiplied by the normal velocity at which it moves: $\dot{A} = Lv$. If the ink spill is contained along part of its boundary and L is merely the length of the moving part of the spill boundary, then clearly now we still have

$$\dot{A} = Lv.$$

In particular, if the spill has two moving parts, at the instant the first moving part has length L_1 and the second L_2, and if we see the normal velocity of the the first moving boundary is v_1 whereas for the second it is v_2, then $\dot{A}\Delta t \doteq \Delta A \doteq L_1 v_1 \Delta t + L_2 v_2 \Delta t$, and the same argument now tells us

$$\dot{A} = L_1 v_1 + L_2 v_2.$$

More generally still, if the normal velocity varies along the boundary, then we could chop up the boundary into little pieces so small that on each piece the normal velocity is approximately constant, and add up all the little products, and this would lead to a type of integral over the boundary.

The simplest case of two moving boundary pieces is the case of a rectangle with sides L and W. If we fix two perpendicular edges of the rectangle and allow the other two edges to move, we have the case of $A = LW$ with L and W being time dependent. The normal velocity of the moving edge of length L is obviously $\dot{W}$ whereas the normal velocity of the edge of length W is obviously $\dot{L}$. This means we now have

$$(LW)^{\cdot} = \dot{A} = \dot{L}W + L\dot{W},$$

which gives us a product rule for calculating rates of change. Of course, the fact that

$$(L + W)^{\cdot} = \dot{L} + \dot{W}$$

is intuitively clear say from considering a car moving on a moving platform such as a train flatcar. If $L = c$ is a constant, then obviously $\dot{L} = 0$. Then the product rule reduces simply to the rule

$$(cW)^{\cdot} = c\dot{W}.$$

We therefore have some basic rules for calculating rates of change, merely based on the assumption that rate of change and area both make sense.

If we have simply $L = t$ for every time t, then obviously $\Delta L = \Delta t$ so their ratio is one: $\dot{t} = 1$. If $L = t^2$, then by the product rule we have $\dot{L} = (tt)^{\cdot} = \dot{t}t + t\dot{t} = 2t$. If $L = t^3$, then

$$\dot{L} = (t^2 t)^{\cdot} = 2tt + t^2 = 3t^2.$$

Obviously we find in general that in case $L = t^n$ we have $\dot{L} = nt^{n-1}$. That is, we find the power rule for computing rates of change:

$$(t^n)^{\cdot} = nt^{n-1}.$$

The same considerations can be made for the relationship between volumes of varying solid regions and the normal velocity of moving boundary pieces. For instance if a potato develops a blister on a region of skin having area A at a certain instant, and if at this instant the normal velocity of the boundary region of the blister is v, then the rate of change of volume due to the growing blister is Av. Notice, a thin shell of area A and thickness d has volume very approximately Ad so during time Δt the volume change is approximately $\Delta V \doteq Av\Delta t$ and this likewise leads to

$$\dot{V} = Av.$$

Similarly, if the potato has two blisters, the rate of change of volume is simply the sum of the rates of change of volume due to each:

$$\dot{V} = A_1 v_1 + A_2 v_2.$$

2. **LECTURE** WEDNESDAY 26 AUGUST 2009

Today we discussed set theory in general as the foundation of mathematics, union and intersection of sets and the general definition of a function $f : X \longrightarrow Y$ and its graph $Graph(f) \subset X \times Y$. Here $X \times Y$ denotes the set of all possible ordered pairs (x, y) where $x \in X$ and $y \in Y$. We denote by $\mathbb{R}$ the set of all real numbers which is pictured as the geometric line, so $\mathbb{R} \times \mathbb{R}$ can naturally be pictured as the plane. We just think of each ordered pair of numbers as corresponding to the rectangular coordinates of a point in the plane. If S is a subset of the cartesian product $X \times Y$, then it is not necessarily the graph of a function. In order for S to be the graph of a function, it must be the case that each member of X is the first member of some ordered pair in S and it also must be the case that if two ordered pairs in S have the same first entry, then their second entries are also the same, so they are the same ordered pair. In case that $X \subset \mathbb{R}$ and $Y \subset \mathbb{R}$, then $X \times Y$ is a subset of the plane, and the two conditions on S can easily be pictured as saying that any vertical line through X on the horizontal axis must cross S exactly once. If L is the vertical line through $(a, 0)$ where $a \in X$, then the intersection of L with S contains exactly one ordered pair (x, y). As the line is vertical and contains $(a, 0)$ we see $x = a$ and therefore the function f defined by S has $f(a) = y$. We discussed the idea that in arbitrary sets we cannot in general think of the members as points in some "space". In order for a set to be thought of as a set of points in a space, there must be some notion of boundary for subsets. That is, if A is the the set of all cars with broken radios and X is the set of all cars, there does not seem to be any reasonable boundary of A in X whereas, if X is the set of all points in the plane, and if A is a subset of X, then it is natural to consider the boundary of A as a new subset B which somehow divides the plane into two parts, one part being A and the other part being the complement of A in X. The important sets in this situation are the subsets which are disjoint from their boundary and these are called open sets. If $X = \mathbb{R}$ is the real line, if $a, b \in \mathbb{R}$, and if A is the set of points x satisfying the inequality $a < x < b$, then either A is empty, or, in case $a < b$, the subset A is an interval of points whose boundary is the set $\{a, b\}$ which is obviously disjoint from A. We say the interval is open because it does not contain any of its boundary points. If X is any set where we have a notion of boundary, then a subset A would be called open if it is disjoint from its own boundary, that is, if it contains none of its boundary points. In such a situation, with reasonable assumptions on the open sets, we say that the set has a topology, and it is in this situation that it is natural to think of the members of the set as points in some kind of space. It is also in this situation that we can formulate the concept of a limit of a function.

3. **LECTURE** FRIDAY 28 AUGUST 2009

Today we discussed methods of combining functions to get new functions. If X, Y, Z are sets with $f : X \to Y$ and $g : Y \longrightarrow Z$ both functions, then we can form the **composite** function $g \circ f : X \longrightarrow Z$ whose rule is simply $(g \circ f)(x) = g(f(x))$. It is often useful to think of a function as an input-output device, the domain is the set of allowable inputs and the outputs must land in the codomain. For $f : X \longrightarrow Y$, we call X the domain and Y the codomain. More generally, if $f : W \longrightarrow X$ and $g : Y \longrightarrow Z$, then we can define the composite function to still be given with the rule $(g \circ f)(x) = g(f(x))$, but now the domain of $g \circ f$ consists only of the set $\{x \in W : f(x) \in Y\}$. For any set X, we can define the **identity** function, denoted id_X by the rule $id_X(x) = x$. Thus for an identity function, the output is always simply identical to the input. Obviously, if $f : X \longrightarrow Y$, then

$$id_Y \circ f = f = f \circ id_X.$$

We say that the functions $f : X \longrightarrow Y$ and $g : Y \longrightarrow X$ are mutually inverse to each other provided that both $g \circ f = id_X$ and $f \circ g = id_Y$. When f and g are **mutually inverse**, each determines the other uniquely, so we write $g = f^{-1}$ and $f = g^{-1}$.

If $f : X \longrightarrow Y$ and $A \subset X$, then $f|_A : A \longrightarrow Y$ is a new function called the **restriction** of f to A, and its rule is given by $(f|_A)(x) = f(x)$. If $g = f|_A$, we say the g is a restriction of f, and likewise, we call f an **extension** of g. In case that $A \subset X \subset \mathbb{R}$ and $Y \subset \mathbb{R}$, then when we picture the graph of f, we can form the graph of $f|_A$ by simply erasing the part of the graph of f that does not lie over the subset A. Likewise, and extension of f is just formed by "drawing" more points to the graph over points not already in the domain of f. Suppose that the graph of $f : A \longrightarrow \mathbb{R}$ is a continuous curve whose domain is a closed interval $A = [a, b] = \{x \in \mathbb{R} : a \leq x \leq b\}$. Suppose that $c \in A$ and we form the function g whose graph is the result of removing the point $(c, (f(c))$ from the graph of f. Thus, g is a restriction of f and f is an extension of g. But obviously g has many extensions with domain A besides f. Somehow, the "continuity of the curve giving the graph of g is demanding that the point $(c, f(c))$ be put back in place. That is, somehow, f is a "natural" extension of g to include c in the domain. This is because, when we look at the graph of g and consider values $g(x)$ for x near c, as x gets nearer and nearer to c, the values $g(x)$ are getting nearer and nearer to $f(c) = L$, the original value given by the point removed from the graph of f. In this situation, we say that L is the **limit** of g as x approaches c, written $\lim_{x \to c} g = L$, or $\lim_{x \to c} g(x) = L$. Thus, to compute the limit of g at a point, we find a function whose graph is a continuous curve and which is defined at the point where we are trying to compute the limit.

If $f, g : X \longrightarrow \mathbb{R}$, then we can form the algebraic combinations $f+g$ and fg where $(f+g)(x) = f(x) + g(x)$ and $[fg](x) = [f(x)][g(x)]$, so to add functions add the values and to multiply functions, multiply the values. Likewise, if $f : A \longrightarrow \mathbb{R}$ and $g : B \longrightarrow \mathbb{R}$, then the domain of $f+g$ and the domain of fg in both cases is the set $A \cap B$, that is, the intersection of the domain of f with that of g. We can also form the quotient f/g whose domain is $(A \cap B) \setminus \{x \in B : g(x) = 0\}$, and whose rule is $(f/g)(x) = f(x)/g(x)$. In general, if we write $f(x) = etcblahblahblah$, where $etcblahblahblah$ stands for some expression involving x, then we take the domain to be the largest set for which the expression makes sense. For instance, when we consider the function

$$h(x) = \frac{x^2 + x - 6}{x^2 - 9},$$

we see that the numerator and denominator are functions defined on all of $\mathbb{R}$, but the domain of h is only $\mathbb{R} \setminus \{-3, 3\}$. That is, if $f(x) = x^2 + x - 6$ and $g(x) = x^2 - 9$, then $h = f/g$. On the other hand, if we factor the numerator and the denominator, then we find $f(x) = (x-3)(x+2)$ whereas $g(x) = (x-3)(x+3)$. In the expression for h, we can therefore cancel the common factor $(x-3)$. We are then replacing the numerator and denominator with new functions $F(x) = x+2$ and $G(x) = x + 3$, and forming the new function F/G. Notice the domain of F/G is $\mathbb{R} \setminus \{-3\}$. Thus, F/G is an extension of $h = f/g$. But examination of the graph of F/G indicates it to be a continuous curve, so that it gives us the desired extension of h having 3 in its domain. Thus, $\lim_{x\to 3} g(x) = F(3)/G(3) = 5/6$. Notice that we have not really dealt with the problem of what continuity means. To make everything here precise requires that the limit concept be defined in a way that does not rely on a prior meaning of continuity. We will postpone this to later and proceed to apply the ideas here to the calculation of tangent lines.

In calculus, the tangent problem is to find the tangent line T to a given curve C at a given point $P \in C$. To make this precise, we begin by saying that the information as to what the curve C is must be specified by an equation, and the point P of tangency we are interested in must be specified by the rectangular coordinates of a point on the curve. Thus the rectangular coordinates of the point must provide a particular solution to the equation. Our problem is then to give the equation of the tangent line. To start, we assume a simplified form of equation, namely an equation of the form $y = f(x)$ where $f : D \longrightarrow \mathbb{R}$ is a function and $D \subset \mathbb{R}$. The information as to what point of tangency we are dealing with can then simply be specified by stating a point c in the domain of f, since then we know the point of tangency on the graph is $(c, f(c))$. We then consider a nearby point Q on the curve C with coordinates (u, v). As Q is on C this means $v = f(u)$. Since the two points P and Q determine a line L called a **secant line**, it is routine to write down the equation of the line through P and Q. For instance, as Q is on the graph of f, it follows that $v = f(u)$ so the slope is

$$slope_L = m_L = \frac{\Delta y}{\Delta x} = \frac{v - f(c)}{u - c} = \frac{f(u) - f(c)}{u - c}.$$

Obviously, $y = m_L(x - c) + f(c)$ is the equation of a line passing through the point $(c, f(c))$ and having slope m_L. In order to get the equation of the tangent line, T, as we know it passes through $(c, f(c))$, we only need to find its slope. To do this, we suppose that as Q slides along the graph of f toward P, the secant line L turns into the tangent line T, and we get Q approaching P by having u approach c. We therefore need to compute the limit

$$m_T = \lim_{u\to c} m_L = \lim_{u\to c} \frac{f(u) - f(c)}{u - c} = \lim_{z\to c} \frac{f(z) - f(c)}{z - c}.$$

For computing this limit in practice, it is sometimes easier to replace $z = c + h$ and then $\Delta x = z - c = h$ so the limit is

$$m_T = \lim_{z\to c} \frac{f(z) - f(c)}{z - c} = \lim_{h\to 0} \frac{f(c + h) - f(c)}{h}.$$

If this limit actually exists and makes sense, it is completely determined by the function f and the point c in the domain of f, so we denote this by writing $f'(c) = m_T$ or

$$f'(c) = \lim_{z\to c} \frac{f(z) - f(c)}{z - c} = \lim_{h\to 0} \frac{f(c + h) - f(c)}{h}.$$

The two forms of limit above are mathematically completely equivalent, but in practice can present different algebraic manipulation problems. Generally, for polynomial functions the second form is preferable since it the first form will involve factorization whereas the second form simply involves forming powers of binomials and cancelling. With other forms of functions, the first form will sometimes be superior since it involves fewer symbols.

4. **LECTURE** MONDAY 31 AUGUST 2009

Mr. Vonk, the teaching assistant covered examples of finding tangent lines to the graphs of functions.

5. **LECTURE** WEDNESDAY 2 SEPTEMBER 2009

Today we discussed the fact that to calculate tangent lines requires the calculation of limits, so we began by discussing the computation of limits for general functions. We begin with a function $f : D \longrightarrow \mathbb{R}$ with $D \subset \mathbb{R}$.

The expression

$$\lim_{x\to c} f = L$$

means roughly, that as x gets ever closer to c, it is the case that $f(x)$ gets ever closer to L. Of course, the equation above in more detail could be written

$$\lim_{x\to c} f(x) = L,$$

which reminds us that as x approaches c, or in symbols, as $x \to c$, we must have x in the domain of the function f. In other words, to speak of the limit of f as $x \to c$, it must be the case that x can approach c within the domain of f. Of course, for the limits we need for calculating tangent slopes, the point c will generally not be in the domain of f. If c is in the domain of f, we want the limit of f as $x \to c$ to be determined by the points in the domain different from c. Thus, when we speak of the limit of f as $x \to c$, we should keep in mind this only makes sense in case c is "infinitely close" to $D \setminus \{c\}$. Such points are called **limit points** of D. Recall that an **open subset** of the space X is a set which is disjoint from its boundary. A set which contains its boundary is said to be **closed**. Thus, the interval $J = \{3 \leq x \leq 5\}$ is closed because it contains its boundary, since the boundary is clearly just the set $\{3, 5\} \subset J$. The set $\{3 < x < 5\}$ obviously has the same boundary as J, but is now disjoint from its boundary, so is open. Drawing pictures, we can easily see that c is "infinitely close to $D \setminus \{c\}$ provided that every open subset which contains the point c also intersects $D \setminus \{c\}$. Such points are limit points of D, and it is only for c being a limit point of D that it even makes sense to ask about $\lim_{x\to c} f$. A point of the domain D which is not a limit point of D is called an **isolated point** of D. It is easy to see that if $A \subset B$, then every limit point of A is also a limit point of B. We can actually distinguish the boundary of a set using open subsets. The point c is in the boundary of the subset A provided that every open subset which contains c also intersects both A and its complement. That is every open subset which contains c must contain a point of A and a point not in A. This means that both A and its complement have exactly the same boundary. If c is a limit point of the set A and if c is not in A, then c must be a boundary point of A. To see this, if U is any open subset and if $c \in U$, then since c is a limit point of A, it follows that U intersects A, whereas, we see that c itself is a point of U not in A. If B is the boundary of A and if L is the set of limit points of A, then we now see $A \cup B = A \cup L$. If c is a point of the boundary of $A \cup B$, and if U is an open set containing c, then U intersects $A \cup B$ and its complement, and therefore intersects the complement of A. If $c \in A$, then U intersects A by virtue of the fact that c is also in U. If $c \in B \setminus A$, then as c is in the boundary of A itself therefore U must intersect A. If c is not in $A \cup B$, then as c is in the boundary of $A \cup B$, it follows that U must intersect $A \cup B$. Thus U either must contain a point of A or a point of B. But, if U contains the point $b \in B$, then $b \in U$ and b is a boundary point of A, and therefore again, U must intersect A. Thus in all cases, we conclude that every open subset containing the point c must also intersect both A and its complement. We have shown that the boundary of $A \cup B$ is contained in the boundary of A, that is to say, it is contained in B. This shows that $A \cup B$ is a closed set, because it contains its own boundary. Any set union its own boundary is closed. Thus, any time we adjoin the set of limit points or the set of boundary points to a set, we arrive at a closed set. If we form $A \setminus B$, we arrive at an open set. This is because any set has the same boundary as its complement. For here, this means that the boundary of $A \setminus B$ is the same as the boundary of its complement which is $C \cup B$, where C denotes the complement of A. But, as C is the complement of A, it follows that B is also the boundary of C and therefore $C \cup B$ is closed and contains its own boundary. Now, $C \cup B$ is the complement of $A \setminus B$, so has the same boundary, which is entirely contained in $C \cup B$. Thus, $A \setminus B$ is disjoint from its own boundary and is therefore open. That is to say, whenever we remove the boundary of a set, what is left is an open set.

With this proviso on c being a limit point of the domain, the sum, difference, product and quotient rules for limits given in your textbook hold. For instance, if $f : D_1 \longrightarrow \mathbb{R}$ and $g : D_2 \longrightarrow \mathbb{R}$ are both functions, if c is a limit point of $D_1 \cap D_2$, then c is a limit point of D_1, and c is a limit point of D_2, and if

$$\lim_{x \to c} f = L$$

and if

$$\lim_{x \to c} g = M,$$

then the limits $\lim[f \pm g]$ and $\lim fg$ as $x \to c$ all exist and in fact,

$$\lim_{x \to c} [f \pm g] = L \pm M,$$

and

$$\lim_{x \to c} (fg) = LM.$$

With the quotient f/g we have to be more careful. Remember that if K is the set of points in the domain of g on which g is zero, then the domain of f/g is $D_1 \cap D_2 \setminus K$. If c is a limit point of the domain of f/g, then it is a limit point of D_1 and of D_2, and in this case we can say that $\lim(f/g)$ as $x \to c$ exists and is simply L/M, provided that the denominator is not zero. That is

$$\lim_{x \to c} \frac{f}{g} = \frac{L}{M}, \ M \neq 0.$$

Certainly under any reasonable meaning of limit, it must be the case that if $f = k$ is a constant function with domain $D = \mathbb{R}$ and with value k, then $\lim f = k$ as $x \to c$ for any $c \in \mathbb{R}$. Likewise, if $f(x) = x$ on the domain $D = \mathbb{R}$, then $\lim f = c$ as $x \to c$, for any $c \in \mathbb{R}$. The sum and product rules then say immediately that if f is any polynomial function, then $\lim_{x \to c} f = f(c)$.

More generally, we say the the function $f : D \longrightarrow \mathbb{R}$ is **continuous at the point** $c \in D$ provided that if c is also a limit point of D, then

$$\lim_{x \to c} f = f(c).$$

Notice that if c is an isolated point of D, then f is automatically continuous at c. We say that f is **continuous** provided that it is continuous at each point of its domain. Thus all polynomials are continuous, and all rational functions are continuous, by the sum, product, and quotient rules for limits. More generally, we will see that all trigonometric functions, all inverse trigonometric functions, all exponential functions, and all logarithmic functions are continuous. In fact, all power functions are continuous, because if $f(x) = x^p$ where p is some real number, then $f(x) = e^{p \log(x)}$, but this requires another additional limit theorem which tells what is required in order for the composition of continous functions to be continuous.

These results essentially mean that in many cases, the computation of the $\lim_{x \to c}$ just boils down to plugging c into the function. In case this simple minded approach leads to zero over zero, our theory so far would break down. There are two more limit theorems which are needed. The first is the **SQUEEZE THEOREM**.

Theorem 5.1. *SQUEEZE THEOREM. If f, g, h are all functions defined on D with*

$$f \leq g \leq h$$

on D, and if c is a limit point of D and $\lim_{x \to c} f$ and $\lim_{x \to c} h$ both exist and are equal, then $\lim_{x \to c} g$ exists and in fact,

$$\lim_{x \to c} f = \lim_{x \to c} g = \lim_{x \to c} h.$$

For instance, the squeeze theorem tells us that if $f(x) = x^2 sin(1/x)$, then even though f is undefined at $x = 0$, since $-x^2 \leq f(x) \leq x^2,\ x \neq 0$ and as $\lim_{x\to 0}(-x^2) = 0 = \lim_{x\to 0} x^2$, it follows that $\lim_{x\to 0} f = 0$.

The last fact is the **RESTRICTION THEOREM**, which is most used in elementary calculus without explicit mention.

Theorem 5.2. *RESTRICTION THEOREM. Suppose* $f : D \to \mathbb{R}$ *and* $A \subset D$, *and* $g = f|_A$. *If* c *is a limit point of* A, *and if* $\lim_{x\to c} f$ *exists, then so does* $\lim_{x\to c} g$ *and the two limits are equal:*

$$\lim_{x\to c} f = \lim_{x\to c} g.$$

For instance, if

$$g(x) = \frac{x^2 - x - 2}{x^2 - 4},$$

then we see that the domain A of g is $A = \mathbb{R} \setminus \{-2, 2\}$. This means we cannot use simple continuity of f to calculate $\lim_{x\to 2} g$, because 2 is not in the domain of g. However, we also see that as both the numerator and denominator vanish on replacing x with 2, it follows that $x - 2$ must be a common factor of both the numerator and the denominator. After carrying out the factorization and canceling common factors, we see that

$$g(x) = \frac{x+1}{x+2},\ x \neq 2.$$

Notice that we need the proviso $x \neq 2$ above because 2 is not in the domain of g. If we define the function f by setting

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

then f has domain $D = \mathbb{R} \setminus \{-2\}$. Thus, $g = f|_A$ and as 2 is a limit point of A, the restriction theorem tells us that $\lim g$ as $x \to 2$ must exist and be identical to $\lim f$ as $x \to 2$. But, since f is rational and therefore continuous and 2 is in the domain D of f, it follows that the limit of f as $x \to 2$ is simply $f(2)$. Thus,

$$\lim_{x\to 2} g = \lim_{x\to 2} f = f(2).$$

6. **LECTURE** FRIDAY 5 SEPTEMBER 2009

The teaching assistant Mr. Vonk reviewed limits.

7. **LECTURE** MONDAY 7 SEPTEMBER 2009

NO LECTURE. LABOR DAY

8. **LECTURE** WEDNESDAY 9 SEPTEMBER 2009

Today we reviewed limits and trigonometric functions. We defined radian measure and observed that the equation for arc length s of the arc on a circle of radius r subtended by an angle of θ radians is simply $s = r\theta$. We reviewed the graphs of the trig functions and the function cofunction relation for trig functions. We also discussed the inverse trig functions and their graphs.

We discussed the general theorem on limits of composite functions. If $f(x) \to L$ as $x \to c$ and $g(y) \to M$ as $y \to L$, then it seems reasonable that we should have $g(f(x)) \to M$ as $x \to c$. Unfortunately, there is a minor technical difficulty because limits are determined by the behavior of the function near the limit point but not at the limit point. Here is the theorem, more general than that given in the textbook.

Theorem 8.1. *Suppose that $f : A \longrightarrow \mathbb{R}$ and $g : B \longrightarrow \mathbb{R}$ are functions, that $\lim_{x\to c} f = L$ and $\lim_{y\to L} g = M$. Then the limit $\lim_{x\to c} g \circ f$ exists and is given by*

$$\lim_{x\to c} f(g(x)) = M,$$

provided either g is continuous at L or, for all x sufficiently close to c and with $x \in A \setminus \{c\}$ it is the case that $f(x) \neq L$.

As an example, the basic limit that is necessary for the differentiation of trigonometric functions is

$$\lim_{x\to 0} \frac{\sin x}{x} = 1.$$

If we try to use this to show that

$$\lim_{x\to 0} \frac{\sin x^2}{x^2} = 1,$$

we have a problem in that $g(y) = \sin y/y$ is undefined at $y = 0$, so it is certainly not continuous. On the other hand, it is certainly the case that if $x \neq 0$, then also $x^2 \neq o$, so the theorem still applies to tell us that

$$\lim_{x\to 0} \frac{\sin x^2}{x^2} = 1.$$

It is the case that we can define the function g on all of $\mathbb{R}$ by setting $g(0) = 1$ to make a continuous function and apply the theorem to that, but it is technically not the same limit. To make the situation clearer, consider the function $f(x) = x\sin(1/x)$ for $x \neq 0$. With g as originally defined, we have an example where $L = 0$ in the theorem and the closer we get to 0 the more points $x \neq 0$ we find where $f(x) = 0$, so clearly

$$\lim_{x\to 0} \frac{\sin(x\sin(1/x))}{x\sin(1/x)}$$

is problematic. If $h = g \circ f$, then the composite has a domain D which consists of all the non-zero values of x for which $\sin(1/x) \neq 0$. It is not hard to see that 0 is a limit point of D. We have therefore no problem now, as on D we never have $f(x) = 0$, so the theorem applies. We can also express this in an alternate theorem.

Theorem 8.2. *Suppose that A and B are both subsets of $\mathbb{R}$ with $f : A \longrightarrow B$ and $g : B \longrightarrow \mathbb{R}$. Suppose that*

$$\lim_{x \to b} f = c$$

and

$$\lim_{x \to c} g = M$$

and $c \in \mathbb{R} \setminus B$. Then

$$\lim_{x \to b} g \circ f$$

exists and equals M, so

$$\lim_{x \to b} g \circ f = \lim_{x \to c} g = M.$$

We also discussed the meaning of infinite limits

$$\lim_{x \to \infty} f,$$

$$\lim_{x \to -\infty} f,$$

$$\lim_{x \to c} f = \infty,$$

and

$$\lim_{x \to c} f = -\infty.$$

We say that the functions f and g are **asymptotic** at plus infinity provided that

$$\lim_{x \to \infty} [f - g] = 0.$$

Likewise we say they are asymptotic at negative infinity provided their difference has limit zero as x approaches negative infinity. For instance, $\arctan x$ is asymptotic to $\pi/2$ at plus infinity whereas it is asymptotic to $-\pi/2$ at negative infinity. The basic fact about infinite limits is that

$$\lim_{x \to 0} f = \infty$$

if

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

for all $x > 0$, and $p > 0$, and also for even integer p on the domain $x < 0$, whereas for odd integer $p > 0$ we have

$$\lim_{x \to 0} g = -\infty$$

if

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

for all $x < 0$.

For limits at infinity, if $p > 0$, we have

$$\lim_{x \to \pm\infty} \frac{1}{x^p} = 0.$$

9. **LECTURE** FRIDAY 11 SEPTEMBER 2009

Today we discussed one-sided limits and their relation to two-sided limits as well as left hand and right hand derivatives. Remember that if $f : D \longrightarrow \mathbb{R}$, if $A \subset D$, if $g = f|_A$, and if c is a limit point of A, then c is also a limit point of D. Moreover, if

$$\lim_{x \to c} f = L,$$

then necessarily $\lim_{x \to c} g$ exists and also equals L. That is to say

$$\lim_{x \to c} f = \lim_{x \to c} f|_A$$

when the limit on the left side of the equation exists. However, if the limit on the right hand side of the equation exists, the limit on the left side of the equation may not exist. As an example, we have one-sided limits. If $A^+ = \{x \in D : x > c\}$ and if $A^- = \{x \in D : x < c\}$, then we define

$$\lim_{x \to c^+} f = \lim_{x \to c} f|_{A^+}$$

and call this the **Right Hand Limit** of f at c when it exists. Likewise, replacing $+$ with $-$ above gives the **Left Hand Limit**,

$$\lim_{x \to c^-} f = \lim_{x \to c} f|_{A^-}.$$

Thus whenever f has a limit L at c then both left and right hand limits exist and are equal to L, that is all three limits must agree. In the converse direction, if both left and right hand limits at c exist and are equal, then the limit (two-sided) of f exists at c and again therefore all three limits are equal. For instance, if $f(x)$ is given by the expression $x^2 + 3x - 5$ for $x < 2$ whereas $f(x)$ is given by $2x + 5$ for $x > 2$, then

$$\lim_{x \to 2^-} f = \lim_{x \to 2^-} (x^2 + 3x - 5) = \lim_{x \to 2} (x^2 + 3x - 5) = 2^2 + 3(2) - 5 = 5.$$

It is instructive to notice carefully the reasons here. It actually works from the right back to the left in the above equations. The expression $x^2 + 3x - 5$ is continuous, so its limit exists at 2 and is simply given by plugging 2 into the expression. Since the limit of the expression exists, it is the same as the left hand limit of the expression, but on the left side of 2, the expression is the same as $f(x)$ by definition of f. Therefore, the left hand limit of the expression must be the same as the left hand limit of f as x approaches 2 from the left, and in particular,therefore the left hand limit of f at 2 exists and simply equals the value of the expression $x^2 + 3x - 5$ at $x = 2$. Likewise, the right hand limit must exist and is simply given by plugging $x = 2$ into the expression $2x + 5$ which is 9. Notice that the two one-sided limits exist but are not equal and therefore f cannot have a two-sided limit at 2.

We also discussed examples of left hand and right hand derivatives which are defined using the limit definition of the derivative on replacing the two-sided limit with one-sided limits. For instance, the left hand derivative of f above at $x = 2$ is given by differentiating the expression $x^2 + 3x - 5$ at $x = 2$. As this is $2x + 3$ evaluated at $x = 2$, the result is that the left hand derivative of f at $x = 2$ is 7. similarly, the right hand derivative exists and is 2. When the function f is differentiable $x = c$, it follows from the preceding limit results that both left and right hand derivatives of f exist at $x = c$ and are equal. Conversely, if both one sided derivatives exist $x = c$ and are equal, then the function is differentiable at $x = c$ and $f'(c)$ is the common value of the one-sided derivatives.

10. **LECTURE** MONDAY 14 SEPTEMBER 2009

REVIEW FOR TEST 1

11. **LECTURE** WEDNESDAY 16 SEPTEMBER 2009

Today we began by discussing the precise definition of the limit and what the equation

$$\lim_{x \to c} f = L$$

actually means. If $A \subset X$, if $c \in X$ is a limit point of A, and if $f : A \longrightarrow Y$, with $L \in Y$, then

$$\lim_{x \to c} f = L$$

means that for every open subset $V \subset Y$ with $L \in V$, we can find an open subset U_V of X containing the point c with the property that if

$$x \in A \cap U_V \setminus \{c\},$$

then

$$f(x) \in V.$$

Of course the open subset U_V can be chosen in many ways, but it does depend on V in the sense that if we make V smaller, then U_V will usually have to be smaller. With X and Y subsets of $\mathbb{R}$, the open sets can be taken to be open intervals, so for $\epsilon > 0$, we can take $V = V_\epsilon = (L - \epsilon, L + \epsilon)$ and then find U_V of the form $U_V = (c - \delta, c + \delta)$, for $\delta > 0$ sufficiently small. Thus, in terms of ϵ and δ, the precise definition reads:

for every positive $\epsilon > 0$ there is a positive number $\delta = \delta_{(\epsilon, f, c)} > 0$ such that if

$$0 < |x - c| < \delta,$$

with x in the domain of f, then

$$|f(x) - L| < \epsilon.$$

In general, the problem of analyzing a function to determine a choice of $\delta_\epsilon > 0$ for each $\epsilon > 0$ is difficult, but it can be broken down in a useful way that will also give proofs of the limit rules and differentiation rules. For any function f, for any limit point c of the domain of f, and for any proposed limit L, let

$$\Delta f = \Delta f(c, L, h) = f(c + h) - L,$$

so Δf is the deviation of f from its proposed limit when h is the deviation of input from c. In these terms, the precise definition of

$$\lim_{x \to c} f = L$$

is that for any $\epsilon > 0$, there is a $\delta > 0$ such that if

$$0 < |h| < \delta,$$

with $c + h \in A$ then

$$\Delta f < \epsilon.$$

Here A is the domain of f. We could use the symbol Δx for h, and then the condition becomes that for every $\epsilon > 0$ there is $\delta > 0$ such that if $c + \Delta x \in A$ and

$$0 < |\Delta x| < \delta,$$

then

$$|\Delta f(c, L, \Delta x)| < \epsilon.$$

For short here we have $0 < |\Delta x| < \delta$ insures $|\Delta f| < \epsilon$. Notice that we also have

$$f(c + h) = L + \Delta f(c, L, h) = L + \Delta f.$$

Moreover, we can say that

$$\lim_{x \to c} f = L$$

is exactly equivalent to

$$\lim_{h \to 0} \Delta f(c, L, h) = 0.$$

For instance, suppose that we propose that f has limit L at c and g has limit M at c. Then we can form the sum $f+g$, and obviously, if we are proposing these limits for f and g individually, then it is natural to propose the sum $L+M$ for the limit of $f+g$ at c. But now we can see

$$\Delta(f+g)(c, L+M, h) = (f+g)(c+h) - (L+M) = [f(c+h) - L] + [g(c+h) - M],$$

so

$$\Delta(f+g)(c, L+M, h) = \Delta f(c, L, h) + \Delta g(c, M, h).$$

This means that if we solve the $\epsilon - \delta$ problem for each of the individual functions, then for $f+g$ we can choose $\delta_{(\epsilon, f+g)}$ to be the smaller of $\delta_{(\epsilon/2, f)}$ and $\delta_{(\epsilon/2, g)}$. Then for $c+h$ in the domain of $f+g$ with

$$0 < |h| < \delta_{(\epsilon, f+g)}$$

we have both

$$|\Delta f(c, L, h)| < \epsilon/2$$

and

$$|\Delta g(c, M, h)| < \epsilon/2.$$

Since $\Delta(f+g) = \Delta f + \Delta g$, it follows that

$$|\Delta(f+g)| \leq |\Delta f| + |\Delta g| \leq \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.$$

This shows that if

$$\lim_{x \to c} f = L$$

and if

$$\lim_{x \to c} g = M,$$

then $f+g$ has a limit at c given by

$$\lim_{x \to c}(f+g) = \lim_{x \to c} f + \lim_{x \to c} g,$$

the *Sum Rule for Limits.*

We can also note that by definition of the derivative, that f is differentiable at c provided that c is in the domain of f and

$$\lim_{h \to 0} \frac{f(c+h) - f(c)}{h} = \lim_{\Delta x \to 0} \frac{\Delta f(c, f(c), \Delta x)}{\Delta x}$$

exists in which case the value of the limit is the derivative of f at c, denoted $f'(c)$, so

$$f'(c) = \lim_{\Delta x \to 0} \frac{\Delta f(c, f(c), \Delta x)}{\Delta x} = \lim_{\Delta x \to 0} \frac{\Delta f}{\Delta x}.$$

Since $\Delta(f+g) = \Delta f + \Delta g$, it follows that

$$\frac{\Delta(f+g)}{\Delta x} = \frac{\Delta f}{\Delta x} + \frac{\Delta g}{\Delta x}$$

and therefore by the Sum Rule for limits we have that if both f and g are differentiable at c, then so is $f+g$ and moreover

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

the *Sum Rule for Differentiation.*

Another easy case to deal with is the case where we multiply f by a constant k. Here it is easy to see that

$$\Delta(kf)(c, kL, h) = k\Delta f(c, L, h).$$

This means that if f has a limit L at c, then kf has limit kL at c. This is the *Constant Multiple Rule for Limits.* Moreover, if f is differentiable at c, then so id kf and

$$(kf)'(c) = k[f'(c)],$$

because

$$\frac{\Delta(kf)}{\Delta x} = k\frac{\Delta f}{\Delta x},$$

so the *Constant Multiple Rule for Differentiation* is an immediate consequence of the Constant Multiple Rule for Limits.

The case for products of functions is a little more complicated but is useful many places in calculus, so it is instructive to go over the details here. We have

$$\Delta(fg)(c, LM, h) = (fg)(c+h) - LM = f(c+h)g(c+h) - LM.$$

Here we use a standard trick for dealing with differences of products. We add and subtract the same product formed by mixing the factors. For instance, we can add and subtract $Lg(c+h)$. We then have

$$\Delta(fg)(c, LM, h) = (fg)(c+h) - LM = f(c+h)g(c+h) - LM$$

$$= f(c+h)g(c+h) - Lg(c+h) + Lg(c+h) - LM$$

$$= [f(c+h) - L]g(c+h) + L[g(c+h) - M] = [\Delta f]g(c+h) + L\Delta g$$

$$= [\Delta f](M + \Delta g) + L[\Delta g] = [\Delta f]M + L[\Delta g] + [\Delta f][\Delta g].$$

We finally have here,

$$\Delta(fg)(c, LM, h) = [\Delta f(c, L, h)]M + L[\Delta g(c, M, h)] + [\Delta f(c, L, h)][\Delta g(c, M, h)].$$

We can write this for short as

$$\Delta(fg) = (\Delta f)M + L(\Delta g) + (\Delta f)(\Delta g).$$

Notice that this gives us the way to calculate the limit of the product fg, since to show that

$$\lim_{x\to c}(fg) = LM$$

is equivalent to showing that

$$\lim_{h\to 0}\Delta(fg) = 0,$$

but from the sum rule and constant multiple rule for limits applied to the equation for $\Delta(fg)$ now gives

$$\lim_{h\to 0}\Delta(fg) = [\lim_{h\to 0}\Delta f]M + L[\lim_{h\to 0} g] + [\lim_{h\to 0}\Delta f][\lim_{h\to 0}\Delta g].$$

It is easy to show that the last term has limit zero directly from the definition of a limit whereas from the first two terms we have

$$\lim_{h\to 0}\Delta(fg) = (0)M + L(0) = 0,$$

and therefore fg has limit LM at c if f has limit L at c and g has limit M at c. We have proven the *Product Rule for Limits:*

$$\lim_{x\to c}(fg) = (\lim_{x\to c} f)(\lim_{x\to c} g).$$

We can easily see from the definition of limit that

$$\lim_{h\to 0} h = 0 = \lim_{\Delta x\to 0}\Delta x.$$

From this we conclude that if f is differentiable at c, as

$$\Delta f(c, f(c), \Delta x) = \Delta x \frac{\Delta f}{\Delta x},$$

by the product rule for limits, we have

$$\lim_{\Delta x \to 0} \Delta f(c, f(c), \Delta x) = (0)(f'(c)) = 0,$$

and therefore at c it is the case that f has limit $f(c)$. We have proven that if f is differentiable at c, then f is continuous at c.

Applying this to the case where f and g are differentiable at c, we have $L = f(c)$ and $M = g(c)$, so on dividing the equation for $\Delta(fg)$ through by Δx, we have

$$\frac{\Delta(fg)}{\Delta x} = \frac{\Delta f}{\Delta x} g(c) + f(c)\frac{\Delta g}{\Delta x} + [\Delta f]\frac{\Delta g}{\Delta x}.$$

Since f is continuous at c, we have

$$\lim_{\Delta x \to 0} f = 0,$$

so now by the sum and product rules for limits we have

$$(fg)'(c) = \lim_{\Delta x \to 0} \frac{\Delta(fg)}{\Delta x}$$

$$= f'(c)g(c) + f(c)g'(c) + (0)g'(c) = (f'g + fg')(c),$$

so

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

which is the *Product Rule for Differentiation*.

For composite functions, the same type of calculations apply. Suppose that f has proposed limit c at $x = b$ and that g has proposed limit M at $y = c$. Then

$$\Delta(g \circ f)(b, M, \Delta x) = g(f(b + \Delta x)) - M$$

$$= g(c + \Delta f(b, c, \Delta x)) - M = \Delta g(c, M, \Delta f(b, c, \Delta x)).$$

There is a very subtle problem here however. Remember, in $\Delta f(c, L, h)$ we can never use $h = 0$, so likewise here, we must restrict $\Delta g(c, M, \Delta f)$ to cases where $\Delta f \neq 0$. We can define the function G by setting $G(h) = \Delta g(c, M, h)$ in case $h \neq 0$, and $G(0) = 0$. We can say

$$\Delta(g \circ f)(b, M, \Delta x) = G(\Delta f(b, c, \Delta x))$$

makes sense even if $\Delta f = 0$. We can also notice that g has limit M at c if and only if G is continuous at 0. If f has limit c at $x = b$, then the Δf can be made arbitrarily small for x sufficiently near b, and thus Δg can be made arbitrarily small. To use the limit definition directly, if $\epsilon > 0$ is given, then we can find $\lambda > 0$ such that if $|\Delta y| < \lambda$, then $|\Delta g(c, M, \Delta y)| < \epsilon$. We can also find $\delta > 0$ so that if $|\Delta x| < \delta$, then $|\Delta f| < \lambda$. Thus, when $0 < |\Delta x| < \delta$, we have $|\Delta f| < \lambda$, and therefore $|G(\Delta f)| < \epsilon$, and therefore

$$\lim_{\Delta x \to 0} G(\Delta f(b, c, \Delta x)) = 0.$$

This shows that if g has limit M at c and f has limit c at b, then $g \circ f$ has limit M at b, provided that b is a limit point of the domain of $g \circ f$ and provided that if c is in the domain of g, that $g(c) = M$. The reason is that if c is in the domain of g and if $g(c) \neq M$, then $\Delta g(b, c, \Delta f)$ is actually defined and non-zero when Δf is zero, but does not agree with $G(\Delta f)$ which is zero when Δf is zero. Thus, in this case, the vanishing of

$$\lim_{\Delta x \to 0} G(\Delta f)$$

does not guarantee the vanishing of

$$\lim_{\Delta x \to 0} (g \circ f)(b, c, \Delta x).$$

In particular, we see that if g is continuous at c, then we can write

$$\lim_{x \to b} (g \circ f) = g(\lim_{x \to b} f).$$

Moreover, if f is continuous at b, then $c = f(b)$ and we have

$$\lim_{x \to b} (g \circ f) = g(f(b)) = (g \circ f)(b).$$

We have proven that if f is continuous at b and g is continuous at $f(b)$, then $g \circ f$ is continuous at b. Thus *the composite of continuous functions is again continuous.*

For differentiating the composite function, we can note that if f is differentiable at b and g is differentiable at $c = f(b)$, then f is continuous at b and g is continuous at c, so we can try to write (with $\Delta f = \Delta f(b, c, \Delta x)$),

$$\Delta(g \circ f) = \Delta g(c, g(c), \Delta f) = \Delta f(b, c, \Delta x) \frac{\Delta g(c, g(c), \Delta f)}{\Delta f}.$$

Again, the problem is that Δf may be zero so we are dividing by zero. To get around this problem, we define the new function

$$H(u) = \frac{\Delta g(c, g(c), u)}{u},$$

if $u \neq 0$, and $H(0) = g'(c)$. Then as g is differentiable at c, it follows that H is continuous at c. Also, we have now

$$\Delta(g \circ f)(b, g(c), \Delta x) = \Delta f(b, c, \Delta x) H(\Delta f(b, c, \Delta x)),$$

even if Δf is zero. Now we can divide through by Δx to get

$$\frac{\Delta(g \circ f)(b, g(c), \Delta x)}{\Delta x} = \frac{\Delta f(b, c, \Delta x)}{\Delta x} H(\Delta f(b, c, \Delta x)).$$

Since

$$\lim_{u \to 0} H(u) = g'(c) = H(0),$$

by our rule for composite limits, we have

$$\lim_{\Delta x \to 0} H(\Delta f(b, c, \Delta x)) = H(\lim_{\Delta x \to 0} \Delta f) = H(0) = g'(c).$$

Therefore we have by the product rule for limits

$$(g \circ f)'(b) = \lim_{\Delta x \to 0} \frac{\Delta(g \circ f)}{\Delta x} = \lim_{\Delta x \to 0} \frac{\Delta f}{\Delta x} H(0) = f'(b) g'(c) = g'(f(b)) f'(b),$$

which means finally that $g \circ f$ is differentiable at b and the derivative is given by the *Chain Rule for Differentiation:*

$$(g \circ f)'(b) = g'(f(b)) f'(b).$$

12. **LECTURE** FRIDAY 18 SEPTEMBER 2009

Today we discussed the differentiation rules for sums, products, composites, quotients, and powers. We observed that the quotient rule follows from the *Chain Rule* for differentiating composite functions:

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

We observed that as all trig functions can be expressed as expressions in $\sin x$ and $\cos x$, and as

$$\cos x = \sin(\frac{\pi}{2} - x)$$

the questions of differentiability of trig functions all boil down to the question of differentiability of the sine function. Using the addition formula for $\sin(x + h)$, we observed that if sine and cosine are differentiable at zero, then sine is differentiable. We observed that for the exponential function

$$\exp(x) = e^x,$$

the same is true, it is differentiable everywhere as soon as it is proven differentiable at zero. We observed that for the infinite degree polynomial

$$f(x) = 1 + x + \frac{x^2}{2} + \frac{x^3}{(3)(2)} + \frac{x^4}{(4)(3)(2)} + ...,$$

we have if we are allowed to differentiate termwise,

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

which is just like for the exponential function with base e. Introducing the number $i = \sqrt{-1}$, we observed that

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

where the "polynomials" g and h obey the rules

$$h' = g$$

and

$$g' = -h,$$

just like for sine and cosine having the rules

$$\sin' = \cos$$

and

$$\cos' = -\sin.$$

We therefore suspect that it is the case that

$$e^{ix} = \cos x + i \sin x.$$

This formula is very useful for remembering many facts about trig functions, as simple consequences of facts about the exponential function such as laws of exponents and derivatives of exponential functions. For instance, the derivatives of the sine and cosine follow from differentiating both sides of the equation $e^{ix} = \cos x + i \sin x$, using the chain rule on the left and the sum rule on the right. Also the addition formulas for sine and cosine follow from the law of exponents.

To see how the addition rules for sine and cosine can be easily found from the law of exponents, we simply write

$$\cos(A + B) + i \sin(A + B) = e^{i(A+B)} = e^{iA}e^{iB} = (\cos A + i \sin A)(\cos B + i \sin B)$$
$$= (\cos A \cos B - \sin A \sin B) + i(\sin A \cos B + \cos A \sin B).$$

Equating real and imaginary parts, we find

$$\cos(A + B) = \cos A \cos B - \sin A \sin B$$

and

$$\sin(A+B) = \sin A \cos B + \cos A \sin B.$$

We observed that if $|x| < \pi/2$, then

$$|\sin x| < |x|$$

and therefore

$$\lim_{x \to 0} \sin x = 0 = \sin 0$$

by the Squeeze Theorem for limits. We observed that for $|x| < \pi/2$, we also have $\cos x > 0$ and therefore

$$\cos x = \sqrt{1 - \sin^2 x}, \;\; |x| < \pi/2.$$

Since $g(x) = \sqrt{1-x^2}$ is a continuous function, it follows from the theorem on limits for composite functions, that

$$\lim_{x \to 0} \cos x = \sqrt{1 - [\lim_{x \to 0} \sin x]^2} = \sqrt{1 - 0^2} = 1 = \cos 0.$$

It follows from the addition formula for sine that it is continuous everywhere, since we just observed that sine and cosine are continuous at zero. In more detail,

$$\lim_{z \to x} \sin z = \lim_{h \to 0} \sin(x+h) = \lim_{h \to 0} (\sin x \cos h + \cos x \sin h)$$

$$= (\sin x) \lim_{h \to 0} \cos h + (\cos x) \lim_{h \to 0} \sin h = (\sin x)(1) + (\cos x)(0) = \sin x,$$

showing sine is a continuous function. As cosine is the co-function of sine, it follows from the composite function rule for limits that cosine is also continuous. Specifically, we have

$$\cos x = \sin(\frac{\pi}{2} - x),$$

showing that cosine is the composite of continuous functions and is therefore continuous.

The exact same technique can be used to show the differentiability of sine and cosine. Thus, if sine is differentiable, then so is cosine as it is the cofunction of sine so the above formula would exhibit cosine as the composite of differentiable functions and the chain rule would tell us the cosine is differentiable. The addition formula shows that if sine and cosine are differentiable at zero, then sine is differentiable (everywhere). The formula

$$\cos x = \sqrt{1 - \sin^2 x}, \;\; |x| < \frac{\pi}{2}$$

together with the chain rule shows that if sine is differentiable at zero, then so is the cosine differentiable at zero. Thus the problem of differentiability of the trig functions all boils down to showing that sine is differentiable at zero, which is the same as showing that

$$\sin' 0 = \lim_{h \to 0} \frac{\sin h - \sin 0}{h} = \lim_{h \to 0} \frac{\sin h}{h}$$

actually exists. We will do this in the next lecture with a little geometry, and in fact, we will find that the value of the limit is simply

$$\sin' 0 = \lim_{h \to 0} \frac{\sin h}{h} = 1.$$

13. **LECTURE** MONDAY 21 SEPTEMBER 2009

We began by reviewing the trig functions briefly and the results of the last lecture showing that
$$|\sin x| \le |x|, \ |x| \le \frac{\pi}{2}$$
and as
$$|\cos x| = \sqrt{1-\sin^2 x},$$
and as
$$\cos x \ge 0, \ |x| \le \frac{\pi}{2},$$
it follows that we can write the Pythagorean Identity for cosine
$$\cos x = \sqrt{1 - sin^2 x}, \ |x| \le \frac{\pi}{2}.$$
On the other hand, as cosine is the cofunction of sine, we can write in general,
$$\cos x = \sin(\frac{\pi}{2} - x).$$
The first inequality and the Squeeze Theorem for limits gives us
$$\lim_{x\to 0} \sin x = 0 = \sin 0$$
which tells us that sine is continuous at zero. The Pythagorean Identity for cosine and the composite rule for limits then gives us
$$\lim_{x\to 0} \cos x = 1 = \cos 0$$
which tells us that cosine is continuous at zero. The addition formula
$$\sin(x+h) = \sin x \cos h + \cos x \sin h$$
and our limit rules now tell us
$$\lim_{z\to x} \sin x = \lim_{h\to 0} \sin(x+h) = (\sin x)(\cos 0) + (\cos x)(\sin 0) = \sin x$$
and therefore sine is continuous. Next, the cofunction relation and the composite function rule for limits tells us that cosine must also be continuous. Notice how everything gets boiled down to the single limit $\lim_{h\to 0} \sin h$, which is simply the question of continuity for the sine function at zero.

For differentiability the same thing happens. It all boils down to the differentiability of sine at zero. For instance, if we assume that sine is differentiable at zero, then the Pythagorean Identity for cosine tells us that cosine is differentiable at zero by the Chain Rule for differentiating composite functions. An easy computation using the chain rule and the differentiation rules shows that
$$\cos'(0) = \frac{2\sin(0)\sin'(0)}{2\sqrt{1-\sin^2(0)}} = 0,$$
as $\sin(0) = 0$, so we conclude that if sine is differentiable at zero, then so is cosine and in fact its derivative must be zero at zero. We can also see this from the graph of the cosine, since it has a maximum value at zero, namely 1, and wherever a function has a maximum value, if it is differentiable, it looks like the tangent must be horizontal, a fact which we will easily prove later on. From the sum formula we can calculate the limit for the derivative of sine directly:
$$\sin' x = \lim_{h\to 0} \frac{\sin(x+h) - \sin x}{h} = \lim_{h\to 0} \frac{\sin x \cos h + \cos x \sin h - \sin x}{h}$$
$$= (\sin x) \lim_{h\to 0} \frac{\cos h - 1}{h} + (\cos x) \lim_{h\to 0} \frac{\sin h}{h}.$$

but,

$$\lim_{h\to 0}\frac{\cos h - 1}{h} = \lim_{h\to 0}\frac{\cos(0+h)-\cos(0)}{h} = \cos'(0).$$

And

$$\lim_{h\to 0}\frac{\sin h}{h} = \lim_{h\to 0}\frac{\sin h - \sin(0)}{h} = \sin'(0).$$

therefore, as soon as we show that this last limit exists, we then know that sine is differentiable and as this would mean $\cos'(0) = 0$, we would have

$$\sin' x = \sin'(0)\cos x.$$

So again, everything has boiled down to showing sine is differentiable at zero and calculating

$$\sin'(0) = \lim_{h\to 0}\frac{\sin h}{h}.$$

We will now show that this limit exists and is in fact 1, so we have $\sin'(0) = 1$ and therefore in general, we have

$$\sin' x = \cos x.$$

The Chain Rule and the cofunction relation for cosine in terms of sine then tells us cosine is differentiable, and differentiating gives

$$\cos' x = [\sin(\frac{\pi}{2} - x)]' = [\cos(\frac{\pi}{2} - x)](-1) = -\sin x.$$

To calculate the limit for differentiating sine at zero, we already know that

$$|\sin x| \leq |x|, \ |x| \leq \frac{\pi}{2},$$

and therefore

$$|\sin x| < |x|, \ |x| < \frac{\pi}{2},$$

since the possibility of equality can only happen for $x = 0$. We can therefore see that

$$\frac{|\sin x|}{|x|} < 1, \ 0 < |x| \leq \frac{\pi}{2}.$$

This fact is simple geometry-the shortest distance from a point to a line is the straight line segment length for the segment hitting the line perpendicularly. In order to apply the Squeeze Theorem here, we need an estimate from below on this fraction. Consider the picture of the unit circle in the plane centered at the origin O and let A be the point with rectangular coordinates $(1, 0)$. Let P be the point of the unit circle subtended by the angle x, so the arc length along the circle from A to P is $|x|$. Let R be the ray from O through P and let Q be the point on this ray R where the vertical line through A intersects R. Thus, we have a right triangle with vertices O, A, Q and its vertical side has length $|\tan x|$. Of course, the base has length 1, so the area of the triangle is

$$area_T = \frac{1}{2}|\tan x|.$$

On the other hand, the area of the sector of the circle subtended by $|x|$ is simply

$$area_C = \frac{1}{2}|x|.$$

To see this more generally, notice that the area subtended by angle θ on a circle of radius r is proportional to θ. For if we double the angle, the area doubles. If we triple the angle the area triples. We can therefore write the area as a function of the angle θ which is simply

$$a(\theta) = k\theta$$

for some constant k. Since we are expressing angles in radian measure, $a(2\pi)$ is the area of a circle which is πr^2. Therefore, we have

$$\pi r^2 = a(2\pi) = k(2)\pi = 2\pi k,$$

and this in turn means that

$$k = \frac{1}{2}r^2.$$

This gives us the simple formula for the area of a sector of a circle of radius r and angle θ as

$$\textit{sector area} = \frac{1}{2}\theta r^2.$$

In our case we have $\theta = |x|$ and $r = 1$, so the sector area is simply

$$area_C = \frac{1}{2}|x|,$$

as was claimed above. Since the sector area is obviously less than the triangle area as the sector sits inside the triangle, it follows that

$$\frac{1}{2}|x| < \frac{1}{2}|\tan x|.$$

Cancelling and using

$$\tan x = \frac{\sin x}{\cos x},$$

we have

$$|x| < |\tan x| = \frac{|\sin x|}{|\cos x|},\ 0 < |x| \leq \frac{\pi}{2}.$$

Next, we can divide through by $|x| \neq 0$, and multiply both sides by $|\cos x|$ to get

$$|\cos x| < \frac{|\sin x|}{|x|},\ 0 < |x| \leq \frac{\pi}{2}.$$

Now in the range $|x| \leq \pi/2$ we know that $\sin x$ and x agree in sign and cosine is not negative, so we can remove the absolute values and have

$$\cos x < \frac{\sin x}{x},\ 0 < |x| \leq \frac{\pi}{2}.$$

Finally we have then altogether

$$\cos x < \frac{\sin x}{x} < 1,\ 0 < |x| \leq \frac{\pi}{2},$$

but already cosine is continuous at zero so $\lim_{x \to 0} \cos x = \cos(0) = 1$, and the Squeeze Theorem now gives

$$\sin'(0) = \lim_{x \to 0} \frac{\sin x}{x} = 1,$$

as promised. Thus we now know that sine and cosine are differentiable functions and

$$\sin' = \cos$$

and

$$\cos = -\sin.$$

Since

$$\tan = \frac{\sin}{\cos},$$

the quotient rule for differentiation tells us tan is differentiable, and then the cofunction relations tell us all the trig functions are differentiable.

The last thing we did today was to use the rules and formulas to differentiate all the trig functions and observe the patterns which make the formulas easy to remember. The Pythagorean Identities are simple consequences of the definition of the trig functions:

$$\sin^2 + \cos^2 = 1,$$
$$1 + \tan^2 = \sec^2,$$

and

$$1 + \cot^2 = \csc^2,$$

the last two being consequences of the first relating sine and cosine. We also find that

$$\tan' = \sec^2$$

by the quotient rule, so using

$$\cot x = \tan(\frac{\pi}{2})$$

we have

$$\cot' x = [\sec^2(\frac{\pi}{2} - x)](-1) = -csc^2 x.$$

Again using the quotient rule applied to the reciprocal relation

$$\sec x = \frac{1}{\cos x}$$

we calculate

$$\sec' x = \sec x \tan x.$$

Again applying a cofuntion relation

$$\csc x = \sec(\frac{\pi}{2} - x)$$

gives

$$\csc' x = [\sec(\frac{\pi}{2} - x)\tan(\frac{\pi}{2} - x)](-1) = -\csc x \cot x.$$

We see that the Pythagorean Identities link the sine and cosine and their differentiation formulas are also linked. The Pythagorean Identities link the tangent and secant and their differentiation formulas are linked, and likewise, the pythagorean Identities link the cotangent and cosecant and their differentiation formulas are linked with the same pattern as that for tangent and secant except for the negative sign.

Trig functions make for good examples for practicing the chain rule. We worked several examples in class.

14. **LECTURE** WEDNESDAY 23 SEPTEMBER 2009

Today we began by discussing the area function $A(x)$ defined for the area under the graph of a continuous non-negative function f. To be more precise, we assume that the domain of f is the closed interval $[a, b]$ and for $a \leq x \leq b$, let $R(x)$ be the set of points in the plane under the graph of f above the line $y = 0$ and having horizontal coordinate in the interval $[a, x]$. Thus

$$R(x) = \{(t, y) \in \mathbb{R}^2 : \ a \leq t \leq x, \ 0 \leq y \leq f(t)\},$$

and the area function A is defined by

$$A(x) = area[R(x)], \ a \leq x \leq b.$$

Recall that when we showed that sine is differentiable at zero, we used the area of a sector of a circle. We assume that the area makes sense. In general now, we will assume that if B is any region of the plane bounded by continuous curves and contained is some large rectangle, then it makes sense to speak of the area of B which we denote by $area(B)$. Moreover, we assume that if B and C are both such regions, then

$$area(B) \leq area(C),$$

if

$$B \subset C.$$

Finally, if B is a rectangle having sides of length x and y, then we assume $area(B) = xy$. Also, if B and C have only boundary points in common, then

$$area(B \cup C) = area(B) + area(C).$$

Our first concern is to show that A is differentiable on the open interval $a < x < b$. Recall that at the beginning of the course, we observed intuitively that if a region grows due to a moving boundary piece of length L moving at velocity v, then the rate of change of the area is Lv. Suppose that we allow the right edge marker x for $R(x)$ to move with velocity v to the right. The length of the moving boundary is $L = f(x)$ at the instant the moving boundary is at position x in $[a, b]$. Thus our intuitive rule would tell us the rate of increase of area with respect to time is $vf(x)$ in this case. If we take position as a function of time to be simply $x(t) = t$, then $x'(t) = 1 = v$, so the rate of increase of area in this case is simply $f(x)$ at the instant the boundary is at position x. On the other hand, the area as a function of time is $A(x(t))$, so by the chain rule,

$$f(x) = \frac{d}{dt}A(x(t)) = A'(x(t))x'(t) = A'(x).$$

Today we will prove this rigourously using properties of continuous functions. We want to differentiate A as a function of $x \in [a, b]$ on the open interval $a < x < b$. Thus we have to show that the limit

$$\lim_{h \to 0} \frac{A(x + h) - A(x)}{h}$$

actually exists and is given by

$$f(x) = \lim_{h \to 0} \frac{A(x + h) - A(x)}{h}.$$

Notice that

$$A(x + h) - A(x) = area[R(x + h) \setminus R(x)], \ h > 0,$$

whereas

$$A(x) - A(x + h) = area[R(x) \setminus R(x + h)], \ h < 0,$$

with h small enough that $x + h$ is also in $[a, b]$. The difference here is the area of a vertical strip $S(x, h)$, so we can write

$$A(x + h) - A(x) = \frac{h}{|h|} area(S(x, h)),$$

and we notice that for very small h, the strip is very thin so should have area very close to $|h|f(x)$, which in turn means the difference $A(x+h) - A(x)$ is very close to $hf(x)$. Precisely, the strip $S(x,h)$ is

$$S(x,h) = R(x+h) \setminus R(x),\ h > 0,$$

and

$$S(x,h) = R(x) \setminus R(x+h),\ h < 0.$$

This means that we need to be able to precisely estimate the area of a thin strip when the top boundary curve is continuous. There are two very important properties of continuous functions which will allow us to do this.

The first property is the OPTIMUM PROPERTY OF CONTINUOUS FUNCTIONS:

Theorem 14.1. OPTIMUM VALUE THEOREM. *If f is continuous on the finite closed interval J, then there are points x_m and x_M in J such that*

$$f(x_m) \leq f(x) \leq f(x_M),\ x \in J,$$

that is to say the inequality holds for all $x \in J$. Thus $f|[a,b]$ has both a minimum value and a maximum value.

The second property of continuous functions is the INTERMEDIATE VALUE PROPERTY OF CONTINUOUS FUNCTIONS:

Theorem 14.2. INTERMEDIATE VALUE THEOREM *If f is continuous on the interval J, is $x, z \in J$, and if y is between $f(x)$ and $f(z)$, then there is $t \in J$, with $y = f(t)$. A function on an interval cannot have gaps in its range.*

Notice that by the Intermediate Value Theorem, the the point $t \in J$ can be chosen to be between x and z. In fact, for definiteness, if we assume that $x < z$, then $[x,z] \subset J$, so we can apply the theorem to $f|[x,z]$ which then tells us we can find $t \in [x,z]$ with $y = f(t)$.

Lets apply these theorems to the strip $S(x,h)$. Let J be the base of the strip, which is either $[x, x+h] \subset [a,b]$ or $[x+h, x] \subset [a,b]$ according to whether h is positive or negative. the strip is the region under the graph of $f|J$ here, so according to the Optimum Theorem, we can find points $x_m, x_M \in J$ such that

$$f(x_m) \leq f(t) \leq f(x_M),$$

for all $x \in J$. Notice that the rectangle with base J and height $f(x_m$ is entirely contained in $S(x,h)$ and has area $|h|f(x_m)$ so we conclude that

$$|h|f(x_m) \leq area(S(x,h)).$$

On the other hand, the rectangle with height $f(x_M)$ and base J contains the strip $S(x,h)$ and therefore

$$area(S(x,h)) \leq |h|f(x_M).$$

combining these two inequalities we have

$$|h|f(x_m) \leq area(S(x,h)) \leq |h|f(x_M).$$

Dividing through by $|h|$ gives the inequality

$$f(x_m) \leq \frac{area(S(x,h))}{|h|} \leq f(x_M).$$

Now we use the Intermediate Value Theorem to assert the existence of a point t_h between x and $x+h$ with the property that

$$f(t_h) = \frac{S(x,h)}{|h|}.$$

Of course, the point t_h depends on the value of h, so in fact it is really a function defined on a small interval $[-\delta, \delta]$ where $\delta > 0$ is chosen so small that

$$[x-\delta, x+\delta] \subset [a,b].$$

Thus, we can write $t_h = t(h)$ for $|t| \leq \delta$. Since $t(h)$ is in fact always between x and $x+h$, it must be the case that

$$\lim_{h \to 0} t(h) = x,$$

by the Squeeze Theorem for limits. We therefore now have

$$A(x+h) - A(x) = \frac{h}{|h|} area(S(x,h)) = h\frac{area(S(x,h))}{|h|} = hf(t(h)),$$

and consequently,

$$\frac{A(x+h) - A(x)}{h} = f(t(h)).$$

Since f is assumed to be continuous, by the Composite Limit Theorem we have

$$\lim_{h \to 0} \frac{A(x+h) - A(x)}{h} = \lim_{h \to 0} f(t(h)) = f(\lim_{h \to 0} t(h)) = f(x).$$

This shows that A is differentiable on the open interval $a < x < b$, and that the derivative of A is given simply by

$$A'(x) = f(x), \ a < x < b.$$

Suppose that we now allow the boundary edge at x to move in time, so that $x = g(t)$, for some function g. Assume that g is differentiable, so we know then x moves with velocity $v = g'(t)$. The area of the region at time t is just $A(g(t))$, so by the Chain Rule for differentiation,

$$\frac{d}{dt}[A(g(t))] = A'(g(t))g'(t) = f(x)v,$$

at the instant t where $x = g(t)$ and $g'(t) = v$. This gives us a rigorous proof of the intuitive idea that the rate of change of area when a boundary piece moves is simply the length of the moving boundary piece multiplied by the velocity at which it moves. Notice that if the boundary moves to the left, then this is the case where v is negative. So for instance, the rate of change as the boundary moves to the left at unit speed is therefore $-A'(x) = -f(x)$, so we still consistent with $A'(x) = f(x)$. Thus, if we let the right edge be fixed and allow the left edge to move left at unit speed, the same argument as above shows that the rate of change of area is again $f(x)$, whereas it is $-f(x)$ as the left edge moves to the right.

As an important example, we can take $f(x) = 1/x$ on the open interval $x > 0$. Let us define the function L by setting $L(x)$ equal to the area under $f|[1, x]$ if $x \geq 1$, and by setting $L(x)$ equal to the negative of the area under $f|[x, 1]$, in case $x \in (0, 1]$. Obviously $L(1) = 0$. We now know that L is differentiable except possibly at $x = 1$, but the left and right hand derivatives both exist and are equal to $1/1 = 1$, so that L is differentiable for $x > 0$. Thus,

$$L'(x) = \frac{1}{x}, \ x > 0.$$

In general, we say that F is an antiderivative of f provided that $F' = f$. We have shown that if $f \geq 0$ on $[a, b]$, then f has an antiderivative on the open interval $a < x < b$ given by the area function A. In particular, L is an antiderivative of $1/x$ on the interval $x > 0$. Now if we take any power function and use the power rule to differentiate, we have

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

so if the result we to be proportional to $1/x$, it would be the case where $p - 1 = -1$, and this in turn would mean that $p = 0$, so that $(x^p)' = 0$. That is the power rule tells us that no power function can be an antiderivative of $1/x$. Thus, we need the function L to get an antiderivative for $1/x$.

The last topic of the day was the Inverse Function Theorem.

Theorem 14.3. INVERSE FUNCTION THEOREM. *Suppose that f is differentiable on the open interval J, and that $c \in J$ with $f'(c) \neq 0$. Let $d = f(c)$. Then there is an open interval $U \subset J$ with $c \in U$ such that $f|U$ has an inverse function, $(f|U)^{-1}$ whose domain is an open interval containing d, and $(f|U)^{-1}$ is differentiable at d, with derivative given by the formula*

$$((f|U)^{-1})'(d) = \frac{1}{f'(c)} = \frac{1}{f'(f^{-1}(d))}.$$

It is easy to see that the formula is true from the Chain Rule for differentiation, so the real meat of the theorem is the fact that the inverse actually exists and is differentiable. If g is the inverse function to f, and if we know how to differentiate f, then to find the derivative of g, just use the fact that on the domain of g we have

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

Differentiating both sides of this equation we have by the Chain Rule,

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

so therefore

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

We worked the example using Arcsine, the inverse of the sine function restricted to the interval $[-\pi/2, \pi/2]$. We found that

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

15. **LECTURE** FRIDAY 25 SEPTEMBER 2009

Today we discussed the method of implicit differentiation and its use in differentiating inverse trig functions. To find the tangent line equation for the tangent to the general plane curve determined by the equation

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

it is useful to think of moving along the curve in time so that $x = x(t)$ and $y = y(t)$, which is to say that both variables become themselves functions of time t. For simplicity here, it is useful to write

$$\dot{x} = \frac{dx}{dt}$$

and

$$\dot{y} = \frac{dy}{dt},$$

or more generally, putting a dot over any quantity signifies its time rate of change. We will see that we do not have to care about the particulars as to how x and y depend on t. We begin then by differentiating both sides of the curve equation with respect to time using the Chain Rule, and then factoring out all terms involving $\dot{x}$ and all terms involving $\dot{y}$. The result can be arranged in the form

$$F(x, y)\dot{y} = G(x, y)\dot{x},$$

where F and G are some functions of x and y alone. This means we can write

$$\frac{dy}{dx} = \frac{dy/dt}{dx/dt} = \frac{\dot{y}}{\dot{x}} = \frac{G(x, y)}{F(x, y)}.$$

At any specific point (a, b) on the curve, the slope, m_T, of the tangent line, T, is simply

$$m_T = \frac{F(a, b)}{G(a, b)}.$$

Thus the tangent line equation is

$$y = m_T(x - a) + b,$$

that is for the line tangent to the curve at the specific point (a, b). Notice, we do not have to solve for y as a function of x for this to work. Also, if we find a point, say (a, b), on the curve, we can use the tangent line to approximate the curve near that point and so find a new point which we can think of as being nearly on the curve, and then use this point to find a new tangent line. Doing this for a sequence of points gives an approximation to the curve.

The method of implicit differentiation comes from realizing that we can choose to travel along the curve in any manner as long as we do have $\dot{x} \neq 0$. In particular, if the curve actually is locally the graph of a function near the point (a, b), that is if a small piece of the curve through (a, b) is the graph of a function, then near the point (a, b) we can simply move so that $x(t) = t$. In this case, we have $\dot{x} = 1$, and $x = t$. This means also $dx = dt$, so

$$\dot{y} = \frac{dy}{dt} = \frac{dy}{dx} = y'.$$

Thus, the method of implicit differentiation comes from simply treating y as some function of x and using the chain rule to differentiate both sides of the curve equation with respect to x.

In the case of inverse functions, if g is the inverse function to f, and if we know how to differentiate f, then as the equation for the graph of g is the equation $y = g(x)$, we can apply the inverse f to both sides and arrive at the equation $x = f(y)$ for the equation of the graph of g, now expressed using the inverse function f. It follows on differentiating both sides of this last equation that

$$1 = f'(y)y',$$

so

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

which of course is the same result as would be obtained using the Inverse Function Theorem. With trig functions, the derivative of each trig function is a simple expression involving trig functions which can be expressed in terms of the original function using trig identities. This means that when applying this method to differentiating an inverse trig function, the result will be an algebraic expression in x without the need for trig functions. As an example, consider the case of arctan which is the inverse function to $\tan$. The equation for the graph of $\arctan x$ is $y = \arctan x$ which is the same as $x = \tan y$, so differentiating both sides we find that

$$1 = (\sec^2 y)y',$$

and therefore,

$$\arctan' x = y' = \frac{1}{\sec^2 y}.$$

Next we use the Pythagorean Identity

$$\sec^2 = 1 + \tan^2,$$

to get

$$y' = \frac{1}{1 + (\tan y)^2} = \frac{1}{1 + x^2}.$$

Thus,

$$\arctan' x = \frac{1}{1 + x^2},$$

and we see the final result does not involve trig functions.

16. **LECTURE** MONDAY 28 SEPTEMBER 2009

NO CLASS BECAUSE OF YOM KIPPUR.

17. **LECTURE** WEDNESDAY 30 SEPTEMBER 2009

Today we reviewed implicit differentiation and exponential and logarithmic functions and their derivatives.

We reviewed logarithmic differentiation as a method for differentiating positive functions with complicated expressions in both the base and exponent. For any function positive f, we have
$$[\ln f]' = \frac{f'}{f},$$
so any time we see a fraction with the derivative of the denominator sitting in the numerator, we know we are looking at the derivative of the natural log of the denominator. Multiplying out the f in the denominator in the above equation we have
$$f' = f[\ln f]'$$
which gives the method of logarithmic differentiation. For instance, if $f > 0$ and g is any function, then
$$(f^g)' = [f^g][\ln(f^g)]' = [f^g][g \ln f]' = [f^g][g' \ln f + g\frac{f'}{f}]$$
$$= [f^g]\frac{g' f \ln f + gf'}{f} = (f^{g-1})[g' f \ln f + gf'].$$
The final result here then is a very general rule for exponential expressions involving variables in both the base and exponential position simultaneously,
$$(f^g)' = (f^{g-1})[g' f \ln f + gf'].$$
In particular, if p is any constant, taking $g = p$ and $f(x) = x$ gives the General Power Rule for differentiation, as for this case, $f' = 1$ and $g' = 0$, so
$$(x^p)' = x^{p-1}[(0)x \ln x + p1] = px^{p-1}.$$

We showed that
$$e = \lim_{x\to 0}(1 + x)^{1/x}$$
and as a consequence, for any z we have
$$e^z = \lim_{x\to 0}(1 + zx)^{1/x}.$$
To see this last equation, notice that
$$\lim_{x\to 0}(1 + zx)^{1/x} = \lim_{x\to 0}((1 + zx)^{1/zx})^z = [\lim_{x\to 0}(1 + zx)^{1/zx}]^z = [\lim_{w\to 0}(1 + w)^{1/w}]^z = e^z.$$
In particular, taking $x = 1/n$ we see that as $n \to \infty$, we have $x \to 0$, so we therefore have
$$e^z = \lim_{n\to\infty}(1 + \frac{z}{n})^n.$$
Of course, when we see n we are thinking of positive integer values, but as $x = 1/n$, we can think of n as being any real number, that is the expression we are taking the limit of makes sense for all positive values of n, and is in fact a differentiable function of n. Let us use logarithmic differentiation to differentiate
$$f(n) = (1 + \frac{z}{n})^n$$
with respect to n. The result, using our previous formula for logarithmic differentiation is
$$\frac{df}{dn} = (1 + \frac{z}{n})^{n-1}[1(1 + \frac{z}{n})\ln(1 + \frac{z}{n}) + n(-\frac{z}{n^2})],$$

so

$$\frac{df}{dn} = (1+\frac{z}{n})^{n-1}[1(1+\frac{z}{n})\ln(1+\frac{z}{n}) + (-\frac{z}{n})]$$
$$= (1+\frac{z}{n})^{n-1}[\ln(1+\frac{z}{n}) + (\frac{z}{n})([\ln(1+\frac{z}{n})] - 1)].$$

We want to see that this is actually positive. If so, this means that $f(n)$ is always increasing with n, that is as n gets larger so does $f(n)$. To do this, put

$$h = \frac{z}{n}$$

so we have

$$\frac{df}{dn} = (1+h)^{n-1}[\ln(1+h) + h([\ln(1+h)] - 1)] = (1+h)^{n-1}[(1+h)\ln(1+h) - h].$$

For this to be positive, as $f > 0$, it is the same as having

$$(1+h)\ln(1+h) > h.$$

But this inequality is the same as

$$\frac{h}{1+h} < \ln(1+h).$$

Now remember, we showed that $\ln(1+h)$ is the area of the region of the plane between the $x-$axis and curve $y = 1/x$ and between the vertical lines $x = 1$ and $x = 1+h$. The height of the curve over the point $x = 1+h$ is $1/(1+h)$ and therefore the rectangle with base the interval $[1, 1+h]$ and height $1/(1+h)$ is entirely under the curve $y = 1/x$ so has area less than $\ln(1+h)$. But the area of the rectangle is

$$Area = \frac{h}{1+h}$$

which proves the inequality. Thus we now know that for any n, we have

$$(1+\frac{z}{n})^n < e^z.$$

Another way to see this more simply is to notice this last inequality is equivalent to

$$n\ln(1+\frac{z}{n}) < z$$

which is in turn equivalent to the inequality

$$\ln(1+h) < h.$$

This last inequality is clear from the picture of the graph and the fact that the curve $y = \ln(1+x)$ has a convex shape and tangent line at $x = 0$ which always stays above the curve $y = \ln(1+x)$. Alternately, the region under $y = 1/x$ between $x = 1$ and $x = 1+h$ is entirely contained in the rectangle with height 1 having base $[1, 1+h]$ and this rectangle has area h so

$$\ln(1+h) < h$$

when $h > 0$, and for $h < 0$, the rectangle is contained in the region under the curve, but $\ln(1+h)$ is the negative of the area in this case, and the area of the rectangle is $-h$, so the inequality still holds. That is,

$$\ln(1+h) < h,$$

whenever $h > -1$.

We discussed applications of exponential functions to growth and decay. In the case of Banking, if we invest an initial amount P, called in banking terms, the *Principal*, at an annual *simple* interest rate r, then after one year the interest earned is Pr, so the total value of the account if the interest is not withdrawn would be

$$B_1 = P + Pr = P(1+r).$$

If the interest and principal are left in the account for another year, at the end of two years the balance is then

$$B_2 = B_1(1+r) = P(1+r)^2.$$

Clearly after t years, the balance is

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

as the interest itself begins drawing interest, an effect called *compounding*. On the other hand, one might surmise that if you are leaving the account alone and if the interest rate for a year is r, then for a half of a year it should be $r/2$ so the interest $Pr/2$ should have been earned by the end of the first half year which would clearly lead to the interest at the end of the year being

$$B_1 = P(1+\frac{r}{2})^2$$

and thus after t years the balance would be

$$B_t = P(1+\frac{r}{2})^{2t}.$$

In this case, we say the interest is being compounded semi-annually. Of course, the same argument could be made for the interest earned over a single day or even a single second. In general, for n compounding periods per year, the interest rate for each period is r/n so the balance would be

$$B_1 = P(1+\frac{r}{n})^n$$

at the end of a single year, and after t years it is

$$B_t = P[(1+\frac{r}{n})^n]^t.$$

Financiers actually always consider values of accounts in terms of *continuous* compounding which is obtained as the limit as $n \to \infty$ for this compounding process. Thus, the balance B is a function of time $B = B(t)$ given by

$$B(t) = \lim_{n\to\infty} P[(1+\frac{r}{n})^n]^t = P[\lim_{n\to\infty}(1+\frac{r}{n})^n]^t = P(e^r)^t = Pe^{rt},$$

or

$$B(t) = Pe^{rt} = Pb^t,$$

where $b = e^r$. Thus, the balance at time t is an exponential function. Notice that

$$B'(t) = rPe^{rt} = rB(t),$$

which says that at each instant of time, the rate of increase of the Balance is proportional to the Balance itself. Whenever we have situations where the rate of increase of a quantity at each instant is proportional to the amount there at each instant, the time dependence will be exponential.

We can think of a more simplistic way of viewing the continuously compounded interest in terms of the derivative directly. In an infinitessimally small amount of time dt years, the interest rate is rdt so the interest earned from time t to time $t+dt$ would be very approximately

$$dB = B(t)rdt.$$

Dividing both sides by dt gives the equation

$$B'(t) = \frac{dB}{dt} = rB(t).$$

Thus, the rate of change is proportional to the full balance at each instant. This guarantees that

$$B(t) = Pe^{rt} = Pb^t,$$

with $b = e^r$. We call b the base of the exponential function, and given b, we see that the interest rate for continuous compounding is $r = \ln b$.

We can also notice that in the case of compounding with n periods, the balance B_t after t years can be expressed
$$B_t = P[(1 + \frac{r}{n})^n]^t = Pb_n^t,$$
where $b_n = (1 + \frac{r}{n})^n$. We can thus think of the balance as being a continuous function given by
$$B(t) = Pb_n^t$$
in this case too. Notice that in this case, the continuous compounding rate $r_n = \ln b_n$ is less than r. For we know always
$$\ln(1 + h) < h,$$
so
$$r_n = \ln b_n = \ln[(1 + \frac{r}{n})^n] = n \ln(1 + \frac{r}{n}) < n\frac{r}{n} = r.$$

18. **LECTURE** FRIDAY 2 OCTOBER 2009

Today we discussed applications of exponential and logarithmic functions to problems of growth and decay as well as to cooling problems using Newton's Law of Cooling. Any time we have a quantity X which changes in time, we write for convenience,

$$\dot{X} = \frac{dX}{dt},$$

for the time rate of change of X. If X changes according to the law

$$\dot{X} = kX,$$

where k is a constant, then the unique solution is

$$X(t) = Ae^{kt},$$

for some constant A. Thus, this is the solution whenever the rate of change of X is proportional to X at each instant, for some constants A and k.

Notice that $A = X(0)$, so that the constant A is the initial value of X, or the value at time zero. In population growth problems, it is often convenient to give the essential information in terms of the *Doubling Time* denoted D, which is the time it takes the population to double in size. In this case, we have

$$2A = X(D) = Ae^{kD},$$

so we can cancel the A's and have an equation in D alone. Thus

$$2 = e^{kD},$$

$$kD = \ln 2,$$

$$k = \frac{\ln 2}{D},$$

giving the constant k in terms of the doubling time D.

When we put this value into the original expression for $X(t)$, we find then

$$X(t) = Ae^{(t\ln 2)/D} = A(e^{\ln 2})^{t/D} = A[2^{t/D}].$$

That is, in terms of the doubling time D we have very simply

$$X(t) = X(0)[2^{t/D}].$$

This can actually be done more quickly by simply noting that from

$$2 = e^{kD}$$

we can conclude

$$e^k = 2^{1/D},$$

so

$$X(t) = X(0)e^{kt} = X(0)[e^k]^t = X(0)[2^{1/D}]^t = X(0)[2^{t/D}].$$

In particular, we see that

$$\frac{X(t)}{X(0)} = 2^{t/D}.$$

We are also seeing that the exponential function $X(t)$ is determined by knowing $X(0)$ and $X(t_1)$ for any specific time t_1. More generally, it is the case that these exponential functions are determined by any two points on their graph. In this sense, they are like lines. Two points determine a line, and similarly, two points are all that is needed to determine an exponential function. For suppose that we are not given A or k but we are given the values

$$X(t_1) = B_1$$

and

$$X(t_2) = B_2.$$

We the notice we have the two equations

$$B_2 = X(t_2) = Ae^{kt_2},$$
$$B_1 = X(t_1) = Ae^{kt_1},$$

so dividing the top equation by the bottom equation allows us to cancel A's and find

$$\frac{B_2}{B_1} = e^{k(t_2-t_1)}.$$

Taking logarithms of both sides now gives

$$\ln B_2 - \ln B_1 = k(t_2 - t_1)$$

and therefore the constant k is found to be

$$k = \frac{\ln B_2 - \ln B_1}{t_2 - t_1}.$$

Thus the constant k is sort of a "log slope" of the exponential curve. To find A, we can use either of the two original equations $B_i = Ae^{kt_i}$, as now k is known and t_i and B_i are known. Thus,

$$B_1 e^{-kt_1} = A = B_2 e^{-kt_2}.$$

In case of radioactive decay, the useful simple minded characterization is the *Half Life* denote H. Thus H is the time required for half the substance to decay away. Thus,

$$\frac{1}{2}X(0) = X(H) = X(0)e^{kH},$$

$$\frac{1}{2} = e^{kH},$$

so

$$e^k = (\frac{1}{2})^{1/H}$$

and therefore

$$X(t) = X(0)(e^k)^t = X(0)[(\frac{1}{2})^{1/H}]^t = X(0)(\frac{1}{2})^{t/H}.$$

We discussed Related Rates problems as applications of the method of implicit differentiation where we view all variables as changing with time in some unspecified way. In general, if we have an equation

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

where c is a constant, then on differentiating with respect to time we find and equation of the form

$$F(x, y)\dot{x} = G(x, y)\dot{y}.$$

Given a specific point (x_0, y_0) satisfying the equation, if we are told that at the instant of arrival at that point we have a given value say $\dot{x}_0$ for $\dot{x}$, then we can find the value of $\dot{y}_0$ at that instant from the equation

$$F(x_0, y_0)\dot{x}_0 = G(x_0, y_0)\dot{y}_0.$$

Notice that we have two equations in the 4 unknowns $x, y, \dot{x}, \dot{y}$, and generally we need 4 equations to determine values for 4 unknowns. Also, the equation involving $\dot{x}$ and $\dot{y}$ is homogeneous in these two variables-that is each term is simply of degree one in these variables, so you will have to be given at least one of these two to arrive at a solution. On the other hand, the basic equation $f(x, y) = c$ allows you to find the value of either of x or y from the other. So if you are given in addition to the equation $f(x, y) = c$, say a value for one of the variables x, y and a value for one of the variables $\dot{x}, \dot{y}$ then you have really three equations and differentiating $f(x, y) = c$ with respect to time gives the needed fourth equation.

We worked examples of a boat being pulled to a dock, a water balloon being filled with water, and a camera tracking a rocket. One can notice that sometimes the equations predict that a velocity will become infinite in certain limits. For instance, for the boat being pulled toward the dock, the equations show that if the rate at which line is pulled in remains constant, the boat will hit the dock with infinite speed, an absurd conclusion. However, in applications, such results indicate that some of the basic equations and assumptions will have to break down. In the case of the boat, as it nears the dock, the bow will begin to raise up out of the water destroying the assumed relations. Such observations in applied situations can often be as valuable as the solutions which are well behaved.

We also discussed the differential as an attempt to make sense of the old fashioned notation

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

The symbols dx and dy were originally thought of as "infinitesimal" numbers which had the property that any product was zero or at least infinitesimally smaller than either infinitesimal factor. A curve was then thought of as being infinitesimally straight at the point $(x_0, f(x_0))$ with slope m_0 where

$$m_0 = \frac{dy}{dx}$$

gave the rise over run along the curve for an infinitesimal move along the curve $y = f(x)$. Philosophers later pointed out that such notions could not be given a firm logical foundation, and it was not until the nineteenth century that the $\epsilon - \delta$ methods were developed to put calculus on a firm logical foundation. The notion of the *differential* of a function is an attempt to make some sense out of these "infinitesimal" ideas. The differential of f at x_0 is by definition the line through the origin parallel to the tangent line to the graph of f at the point $(x_0, f(x_0))$, and where we simply use new variables dx and dy to express the equation of this line. Of course, since it passes through the origin $(dx, dy) = (0, 0)$, this means that the equation must be

$$dy = m_0 dx,$$

where m_0 is the slope of the tangent line to the graph of f at $(x_0, f(x_0))$. Since $m_0 = f'(x_0)$, this means the equation is

$$dy = f'(x_0)dx,$$

and therefore, in terms of these new variables dx and dy we do have the interpretation of the equation

$$f'(x_0) = \frac{dy}{dx}$$

as giving an actual fraction, as long as $dx \neq 0$. Since the variables dx and dy are thought of as new variables distinct from x and y, we can say in general the equation

$$f'(x) = \frac{dy}{dx}$$

is equivalent to the equation

$$dy = f'(x)dx.$$

thus, the dy is really dependent on both x and dx, so to make this appear explicitly, we write the differential of the function f as a function of both the variables x and dx as

$$df(x, dx) = f'(x)dx.$$

Of course, if we actually think of using a finite non-zero value for $dx = \Delta x$, then as the tangent line only approximates the actual graph of the function near the point of tangency, and as

$$\Delta f = f(x + dx) - f(x) = f(x + \Delta x) - f(x),$$

it follows that we have for small dx

$$\Delta f \doteq df.$$

Here $\doteq$ means approximately equal. In application, to approximate $f(x + \Delta x)$ when you know $x, f(x)$ and Δx, you simply use the fact that $\Delta f \doteq df$, so

$$f(x + \Delta x) = f(x + dx) = f(x) + \Delta f \doteq f(x) + df.$$

This is easily seen to be simply equivalent to using the tangent line equation to approximate values for f near known values. Thus, at a particular x_0 in the domain of f, the equation of the tangent line is

$$y = f(x_0) + f'(x_0)(x - x_0),$$

so using this equation to give a value for y as an approximation to $f(x)$ is the same as using

$$f(x_0) + df(x_0, dx)$$

as an approximation of $f(x)$ if $dx = \Delta x = x - x_0$ since

$$df(x_0, dx) = f'(x_0)dx = f'(x_0)(x - x_0).$$

This method of approximating values of a function using values nearby where both the function and its derivative are known is also called the *Linear Approximation* or the *Tangent Line Approximation.* As an instructive example, when applied to the function

$$f(x) = \sqrt{x},$$

as

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

we see that for instance

$$df(81, dx) = \frac{1}{18}dx,$$

and therefore to good approximation for instance

$$\sqrt{82} = 9 + \frac{1}{18}$$

and

$$\sqrt{83} = 9 + \frac{1}{9}.$$

It is instructive to draw the graph of f accurately and notice that as the curve has very small curvature near $(81, 9)$, the tangent line approximation is very accurate in this case.

19. **LECTURE** MONDAY 5 OCTOBER 2009

Today we reviewed for TEST 2 to be given in Lab tomorrow. We noted that in related rates or implicit differentiation, if we have an equation in two variables $f(x, y) = c$, where c is a constant and f is a function of the two variables x and y, then assuming that x and y are functions of time t, we can get the equation relating their time derivative $\dot{x}$ and $\dot{y}$ by differentiating the equation with respect to time and the result can always be put in the form

$$F(x, y)\dot{x} + G(x, y)\dot{y} = 0.$$

We also noted that the functions F and G can always be obtained quickly by noticing that if we differentiate f with respect to x as if f is a function of x alone treating y as if it is constant, the result is $F(x, y)$, whereas if we differentiate f with respect to y as if f is a function of y alone treating x as a constant, then the result is $G(x, y)$. Once F and G are found, then

$$\frac{dy}{dx} = -\frac{F(x, y)}{G(x, y)}$$

20. **LECTURE** WEDNESDAY 7 OCTOBER 2009

Today we started chapter 4. We began by reviewing the fundamental properties of continuous functions, the OPTIMUM THEOREM and the INTERMEDIATE VALUE THEOREM. We discussed the fact that these theorems guarantee the existence to solutions to certain types of problems frequently encountered. However, in case of differentiable functions, we have additional theorems which facilitate the actual solution of the problems. There is often a big difference between knowing a problem has a solution and actually finding the solution. Of course, when searching for a solution with any method, it certainly helps when you know a solution exists. Moreover, when you can see a solution cannot exist, it certainly saves you a lot of time and effort which you might otherwise waste looking for the non-existent solution.

We observed that as far as looking for optimal values (minimum values or maximum values) of a function, when we examine how differentiability enters the problem, pictorially we see that local optimum values all happen at places where either the derivative does not exist or where it is zero or at the boundary of the domain. In fact, we observed that if f is a function defined on an open set containing a point c where f has a local extreme (maximum or minimum) value, then one sided derivatives of f at c must have differing sign. The assumption that f is differentiable at c guarantees that both one sided derivatives are the same. The only way a single number can be both non-negative and non-positive is to be zero. Thus we have the second optimum theorem

Theorem 20.1. SECOND OPTIMUM THEOREM. *Suppose that f has a local extreme value at the point c and that f is differentiable at the point c which is in the interior of the domain of f. Then*

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

As an immediate consequence, we see that if we are trying to optimize a function on a closed interval $[a, b]$, then since the optimum values will happen where f has a local extreme value, to find the optimum value besides boundary points we need only check the values of f at places where either $f' = 0$ or f' fails to exist.

These considerations lead to the following definition. We define c in the interior of the domain of f to be a *critical point* if either $f'(c) = 0$ or c is not in the domain of f', which is to say that f is not differentiable at c. Thus, if $f(x) = x^2$, then $f'(x) = 2x$, so we see that the only critical point of f is $c = 0$, whereas if $f(x) = |x|$, then $f'(x) = x/|x|$, so we see again the only critical point of f is $c = 0$, but now because f fails to be differentiable at $x = 0$. Thus the general strategy for optimizing a function on a closed interval is to just check all the critical values and boundary values, since one of these will have to be the optimum value.

The next property of differentiable functions we discussed is called the MEAN VALUE THEOREM, one of the must useful theorems in calculus. We proved a generalization:

Theorem 20.2. *GENERAL MEAN VALUE THEOREM. Suppose that f and g are continuous on the closed interval $[a, b]$ and that their boundary values agree, meaning both $f(a) = g(a)$ and $f(b) = g(b)$. Suppose that both f and g are differentiable in the interior of $[a, b]$, that is, in the open interval (a, b). Then there exists a point c with $a < c < b$ such that*

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

Notice that this theorem can be viewed as guaranteeing the existence of a solution to the equation $f'(x) = g'(x)$ as soon as we check to see the two functions have the same boundary values, which in practice is certainly easy to spot when true.

To prove the general mean value theorem, we observed that it is an easy consequence of a very special case known as Rolle's Theorem.

Theorem 20.3. *ROLLE'S THEOREM. Suppose h is continuous on the closed interval $[a, b]$ and differentiable in its interior. Further, suppose that $h(a) = 0 = h(b)$. Then there is a point c in the interior of $[a, b]$ with $h'(c) = 0$.*

To see this theorem is true, just notice that either h is constant or it is not. If it is, then $h'(x) = 0$ for any x in the interior, whereas if not, then the optimum theorem guarantees there is a point c where f has a local extreme value which is different from zero. This point cannot be on the boundary since h vanishes on the boundary. But, then by the Second Optimum Theorem, we know $h'(c) = 0$.

To prove the General Mean Value Theorem, we just apply Rolle's Theorem to the function $h = f - g$. Since f and g have the same boundary values under the hypothesis, it follows that $h(a) = 0 = h(b)$, so there is a point c in the interior of the domain interval with

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

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

It is interesting to interpret the General Mean Value Theorem in the case of two objects moving along the same path. If we let $f(t)$ be the distance from the starting point for the first object and $g(t)$ be the distance from the start along the path for the second object, and if we assume that they both start at the same place at the same time and arrive at the end of the path at the same time, then that means the boundary values of f and g agree at the start time and the end time. Thus there is some time strictly between the start and end times at which both objects simultaneously have exactly the same velocity.

The theorem known as the MEAN VALUE THEOREM is a special case of the general mean value theorem

Theorem 20.4. *MEAN VALUE THEOREM. Suppose that f is a continuous function on the close interval $[a, b]$ and that f is differentiable in its interior. Then there is a point c in the interior of $[a, b]$ at which we have*

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

To see how this follows from the mean value theorem, just take g to be the linear function which passes through the two points $(a, f(a))$ and $(b, f(b))$. Then g is certainly differentiable and has the same boundary values as f. By the general mean value theorem, we therefore can find a point c in the interior of $[a, b]$ at which $f'(c) = g'(c)$. But since g is the linear function passing through the two points $(a, f(a))$ and $(b, f(b))$, it follows that $g'(c)$ is just the slope m of the line through these two points, and that is obviously

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

Thus

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

The Mean Value Theorem has very useful consequences for graphing and optimizing functions. To make sure we agree in terminology, we say the function f is *increasing* on the set S contained in its domain provided that $x_1 < x_2$ always implies $f(x_1) \leq f(x_2)$, for x_1, x_2 both in the set S. Thus, when we look from left to right at the graph of $f|S$, we see it never goes down. Likewise, we say that f is *decreasing* on S if $x_1 < x_2$ always implies that $f(x_1) \geq f(x_2)$, for both x_1, x_2 in the set S. Thus when f is decreasing on S, as we scan the graph of $f|S$ from left to right, we can never see it rise. We say that f is *strictly increasing* on S provided that $x_1 < x_2$ always implies that $f(x_1) < f(x_2)$, for both x_1, x_2 in S. If f is strictly increasing on S, when we look at the graph of $f|S$ we see that the graph must continually rise as we scan from left to right, it can never even just remain constant even for the slightest little bit. Similarly, we say that f is *strictly decreasing* on S provided that $x_1 < x_2$ always implies that $f(x_1) > f(x_2)$, for both x_1, x_2 in S. So when f is strictly decreasing on S the graph of $f|S$ must always fall as we scan from left to right, it can never even just remain constant for the slightest little bit. It is useful to notice that f is increasing on S if and only if $-f$ is decreasing on S whereas f is strictly increasing on S if and only if $-f$ is strictly decreasing on S.

Theorem 20.5. *MONOTONICITY THEOREM. Suppose that f is continuous on the closed interval $[a, b]$ and differentiable in its interior, the open interval (a, b). If $f' \geq 0$ on the open interval (a, b), then f is increasing on the closed interval $[a, b]$. If $f' \leq 0$ on the open interval (a, b), then f is decreasing on the closed interval $[a, b]$. If $f' > 0$ on the open interval (a, b), then f is strictly increasing on the closed interval $[a, b]$. If $f' < 0$ on the open interval (a, b), then f is strictly decreasing on the closed interval $[a, b]$.*

To see that this theorem is true, we note that once we have proven it for the increasing cases, the decreasing cases follow from replacing f by its negative. If $f' \geq 0$ on the open interval (a, b) and if f is not increasing on $[a, b]$, then there must be two points x_1, x_2 in $[a, b]$ with $x_1 < x_2$ but with $f(x_1) > f(x_2)$. But then by the Mean Value Theorem, there is a point c in the open interval (x_1, x_2) with $f'(c) < 0$, because the line connecting the two points $(x_1, f(x_1))$ and $(x_2, f(x_2))$ has a negative slope. But, as $x_1 < c < x_2$, it follows that c is in the open interval (a, b). This contradicts the hypothesis that $f \geq 0$ on the open interval (a, b). Next, suppose that $f' > 0$ on the open interval (a, b). If f is not strictly increasing, then we can find x_1, x_2 in $[a, b]$ with $x_1 < x_2$, but with $f(x_1) \leq f(x_2)$. Now the line connecting the two points $(x_1, f(x_1))$ and $x_2, f(x_2))$ must have non-positive slope and by the Mean Value Theorem there is a point c in the open interval (x_1, x_2) where $f'(c)$ is this non-positive slope, that is $f'(c) \leq 0$. But, c must be in the open interval (a, b) and this contradicts the assumption that $f' > 0$ on the open interval (a, b).

We used the Monotonicity Theorem to observe that when graphing the function f we can solve $f'(x) = 0$ and from the solutions we find all critical points and where f is increasing and decreasing and where f is strictly increasing and strictly decreasing.

We observed that for graphing a function f, if we think of the independent variable as time, so $f(x)$ is position at time x on a vertical line, then f' is velocity and f'' is acceleration. Thus, where we see local maxima it is similar to the graph of the flight of a ball thrown in the air. It goes up and at the top of its flight it seems to momentarily stop, since in fact its velocity is zero at the top of its flight by the Second Optimization Theorem. But, in fact, the graph is roughly a downward opening parabola, since near the surface of the Earth, gravitational acceleration is approximately constant. On the other hand, near a local minimum value, the graph of the function looks like the motion of a ball in an upside down world. If gravity pulls upward, so acceleration is upward, and if we then throw a ball downward, it goes down until gravity overpowers it and at the bottom of its flight it stops momentarily, and then comes back up. Thus, typically, where we have a positive second derivative, the shape of the graph is similar in form to that of a parabola which opens up (flight of the ball in an upside down world) whereas when the second derivative is negative, the shape of the graph is similar in form to that of a downward opening parabola, like the graph of the flight of a ball thrown up in our ordinary world. This shape we call *concave down*, whereas for the shape like the upward opening parabola, we call the shape *concave up*. Notice that for the concave up shape, as we scan the graph from left to right we see that the tangent line is ever more tilted upward, that is f' is strictly increasing, which will be the case if $f'' > 0$. Likewise, for the concave down shape like the flight of the ball in an ordinary world, the velocity is always strictly decreasing which is the case if $f'' < 0$. If c is a point where the concavity changes, we call c in *inflection point*. Thus, if f'' has opposite signs on the opposite sides of c, then c is an inflection point. Of course, then if $f''(c)$ exists, then it is necessarily zero. Thus, to find inflection points, we set the second derivative of f equal to zero. This means that when graphing the function f, we begin by differentiating twice, finding both f' and f''. We then solve the equations

$$f'(x) = 0$$

and

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

We want to make sure that no solution of these equations is overlooked, since that can greatly confuse the whole analysis. In fact, that is one of the useful things about this procedure-the fact that if some mistake is made in one of the calculations, there will often be fairly obvious inconsistencies when we try to sketch the graph. We then calculate the values of the function at all the points which are solutions of the two equations, since they will be the likely local extreme values and inflection points. Between any two inflection points, the concavity cannot change, and between any two critical points, the monotonicity cannot change.

We worked an example involving a cubic function and noticed the typical behavior. Such a function often has a local minimum and a local maximum and in this case it always has a single inflection point which is the midpoint of the two critical points. When the cubic curve is graphed, the inflection point on the graph will be the midpoint of the line segment joining the local extreme points on the graph. Knowing this can save time in calculations or alternately serve as a check on calculations.

21. **LECTURE** FRIDAY 9 OCTOBER 2009

Today we worked examples of graphing functions using the first derivative to find critical points and monotonicity and using the second derivative to determine concavity and inflection points.

22. **LECTURE** MONDAY 12 OCTOBER 2009

Today we began with an example of using first and second derivatives to find critical points and inflection points and their use in graphing. We discussed the case where a function continuous on an interval has only a single critical point. In this case, the global extreme value must occur at that critical point. We went over the second derivative test for local extreme values. We observed that the underlying principle in all this kind of analysis is the simple fact that if a function f vanishes at $x = c$ but the derivative does not, then f must change sign at $x = c$. Applied to the second derivative, this gives us the second derivative test. This says that if the derivative vanishes at $x = c$ but the second derivative does not, then the first derivative changes sign at $x = c$, which means that f must have a local extreme value at $x = c$, since f changes monotonicity at $x = c$. If the second derivative vanishes at $x = c$ but the third derivative does not, then likewise the second derivative must change sign at $x = c$ which means f must have an inflection point at $x = c$. The converse is not true. If f vanishes and changes sign at $x = c$, then it still may be the case that the derivative of f vanishes at $x = c$. For instance, if $f(x) = x|x|$, then $f(0) = 0$, and $f'(0) = 0$, but f definitely changes sign at $x = c$. Likewise, if $f(x) = x^4$, then $f'(0) = 0$ and $f''(0) = 0$, but f definitely has a minimum value at $x = 0$. On the other hand, if $g(x) = x^3$, then both the first and second derivatives vanish at $x = 0$, but f has no minimum at $x = 0$ and f does have an inflection point at $x = 0$. Thus, when c is critical for f and $f''(c) = 0$, we cannot tell whether or not f has a local extreme value at $x = c$ without other considerations. In this case, we say the second derivative test fails. The second derivative test is only useful when the second derivative is non-zero at the critical point.

We discussed the example of the bell curve with mean μ and standard deviation σ, which is the graph of the function f where

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}.$$

We notice that $f > 0$ always and that

$$\lim_{x\to\pm\infty} f(x) = 0.$$

When we differentiate, we find

$$f'(x) = \frac{1}{\sigma\sqrt{2\pi}}[-\frac{1}{2}(2)(\frac{x-\mu}{\sigma})^1\frac{1}{\sigma}]e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2},$$

so

$$f'(x) = \frac{1}{\sigma\sqrt{2\pi}}[-(\frac{x-\mu}{\sigma^2}]e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2},$$

or

$$f'(x) = [-(\frac{x-\mu}{\sigma^2})]f(x).$$

Since f is always positive, we see from this that f has only the single critical point at $x = \mu$. But, this means that f must have its absolute maximum at $x = \mu$, since f can never attain its minimum, as it goes asymptotically to zero at $\pm\infty$. Taking advantage of the relation between f and its derivative, we can differentiate again and find

$$f''(x) = [-\frac{1}{\sigma^2}]f(x) + [-(\frac{x-\mu}{\sigma^2})][-(\frac{x-\mu}{\sigma^2})]f(x)$$

or

$$f''(x) = [(\frac{x-\mu}{\sigma^2})^2 - \frac{1}{\sigma^2}]f(x) = \frac{1}{\sigma^2}[(\frac{x-\mu}{\sigma})^2 - 1]f(x).$$

Again as f is never zero, we see the only way the second derivative can vanish is for $(x-\mu)^2 = \sigma^2$ which means only for

$$x = \mu \pm \sigma.$$

Therefore this curve has two inflection points symmetrically spaced about the axis of symmetry $x = \mu$, and each is at a distance σ from the axis of symmetry. This means that whenever we look at a bell curve indicating a distribution of probability for some population, the axis of symmetry is the mean and the most likely value whereas the distance from the mean to the inflection point gives the standard deviation. This function is one of the most important in mathematics and has many applications in engineering and physics beyond its application to probability theory.

We went over L'Hospital's Rule for computing limits of the form $0/0$ or of the form ∞/∞. We noted that if the limit is of the form $0 * \infty$, then we can convert it to either of the two previous forms, but usually one way will be much easier than the other. We noted that all of the elementary limits of the form zero over zero can be quickly calculated using L'Hospital's Rule, for instance, limits involving the special trigonometric limits we encountered in the differentiation of the trig functions.

We finished by defining the hyperbolic trig functions and observing some of the relations which mirror the trig function relations. We recalled that we had previously notice useful exponential form for the trig functions:

$$e^{ix} = \cos x + i \sin x$$

which makes it easy to derive the angle addition and double angle formulas for sine and cosine. Here $i = \sqrt{-1}$. We treat i just like any other constant, except we always use $i^2 = -1$ to simplify expressions when convenient. Notice that since $\cos(-x) = \cos x$ and $\sin(-x) = -\sin x$, we have

$$e^{-ix} = \cos x - i \sin x$$

and therefore,

$$\cos x = \frac{e^{ix} + e^{-ix}}{2}$$

and

$$\sin x = \frac{e^{ix} - e^{-ix}}{2i}.$$

We can now take the attitude that turn about is fair play. Let us consider what happens if we consider $\sin(ix)$ and $\cos(ix)$. From the preceding formulas we see

$$\cos(ix) = \frac{e^{-x} + e^{x}}{2},$$

whereas

$$i \sin(ix) = \frac{e^{-x} - e^{x}}{2},$$

so

$$-i \sin(ix) = \frac{e^{x} - e^{-x}}{2}.$$

We define hyperbolic cosine and hyperbolic sine functions by

$$\cosh x = \cos(ix) = \frac{e^{x} + e^{-x}}{2}$$

and

$$\sinh x = -i \sin(ix) = \frac{e^{x} - e^{-x}}{2}.$$

Obviously then,

$$\cosh x + \sinh x = e^x.$$

We can easily find the derivatives,

$$\cosh' x = [\cos(ix)]' = i[-\sin(ix)] = \sinh x,$$

and

$$\sinh' x = -i[i \cos(ix)] = \cos(ix) = \cosh x.$$

Or more directly,

$$\cosh' x = [\frac{e^{x} + e^{-x}}{2}]' = \frac{e^{x} - e^{-x}}{2} = \sinh x,$$

and
$$\sinh' x = [\frac{e^x - e^{-x}}{2}]' = \frac{e^x + e^{-x}}{2} = \cosh x.$$
Thus, we have more simply than with ordinary trig functions,
$$\sinh' = \cosh$$
and
$$\cosh' = \sinh .$$
If we accept the Pythagorean identity holds using imaginary angle inputs, we have
$$\cos^2(ix) + \sin^2(ix) = 1.$$
But,
$$\sinh^2 x = (-i\sin(ix))^2 = (-1)^2 i^2 \sin(ix) = -\sin(ix),$$
so substituting we find the hyperbolic Pythagorean Identity
$$\cosh^2 x - \sinh^2 x = 1.$$
This can also be verified directly using the exponential formulas for the hyperbolic trig functions. The rest of the trig identities are pretty much simply carried over by definition. Thus we define the hyperbolic tangent as
$$\tanh x = \frac{\sinh x}{\cosh x}$$
and likewise we define the hyperbolic secant and cosecant by
$$\text{sech } x = \frac{1}{\cosh x}$$
and
$$\text{csch } x = \frac{1}{\sinh x}.$$

23. **LECTURE** WEDNESDAY 14 OCTOBER 2009

Today we discussed the hyperbolic functions, their derivatives, and their graphs. As I got confused on the plus and minus signs during the lecture, I will straighten them out here. We had earlier discussed them briefly, and the formulas all in terms of cosh and sinh are the same as for the ordinary trig functions. We can then use the formulas

$$\cosh x = \cos(ix)$$

and

$$\sinh x = -i\sin(ix) = \frac{1}{i}\sin(ix)$$

to find formulas for the other hyperbolic trig functions in terms of the ordinary trig functions of ix. Keep in mind that since $i^2 = -1$, it follows that $1/i = -i$. We have

$$\tanh x = \frac{\cosh x}{\sinh x} = \frac{\sin(ix)}{i\cos(ix)} = \frac{1}{i}\tan(ix),$$

which means that

$$\tanh x = \frac{1}{i}\tan(ix).$$

Likewise, we have

$$\coth x = \frac{\cosh x}{\sinh x} = \frac{i\cos(ix)}{\sin(ix)} = i\cot(ix),$$
$$\text{sech } x = \frac{1}{\cosh x} = \frac{1}{\cos(ix)} = \sec(ix),$$

and

$$\text{csch } x = \frac{1}{\sinh x} = \frac{i}{\sin(ix)} = i\csc(ix).$$

Once you remember these facts, the derivatives of the hyperbolic trig functions can easily be found from the ordinary trig functions. We have

$$\cosh' x = [\cos(ix)]' = -i\sin(ix) = \sinh x,$$
$$\sinh' x = (\frac{1}{i}\sin(ix))' = \frac{1}{i}i\cos(ix) = \cos(ix) = \cosh x,$$
$$\tanh' x = (\frac{1}{i}\tan(ix))' = \frac{1}{i}i\sec^2(ix) = \text{sech}^2\ x,$$
$$\coth' x = (i\cot(ix))' = i^2(-\csc^2(ix)) = -(i\csc(ix))^2 = -\text{csch}^2\ x,$$
$$\text{sech}'\ x = [\sec(ix)]' = i\sec(ix)\tan(ix) = -[\text{sech } x]\frac{1}{i}\tan(ix) = -\text{sech } x\tanh x,$$
$$\text{csch}'\ x = (i\csc(ix))' = i^2(-\csc(ix)\cot(ix)) = -(i\csc(ix))(i\cot(ix)) = -\text{csch } x\coth x.$$

We also discussed the graphs, their critical points and inflection points. Finally, we discussed the use of long division of polynomials to find oblique asymptotes for rational functions when the degree of the numerator is one more than the denominator. Remember, in general, we say that f and g are asymptotic at ∞ (respectively $-\infty$) provided that the limit of their difference as x approaches infinity (respectively negative infinity) is zero. Thus if

$$\lim_{x\to\infty}[f(x) - g(x)] = 0,$$

then f and g are asymptotic at ∞ whereas if

$$\lim_{x\to-\infty}[f(x) - g(x)] = 0,$$

then f and g are asymptotic at $-\infty$. We then observed that $\cosh x$ and $\sinh x$ are asymptotic to $e^x/2$ at positive infinity and to $\pm e^{-x}/2$ at negative infinity, plus for cosh and minus for sinh .

24. **LECTURE** FRIDAY 16 OCTOBER 2009

NO LECTURE TODAY BECAUSE OF FALL BREAK.

25. **LECTURE** MONDAY 19 OCTOBER 2009

Today we reviewed the hyperbolic trig functions, their graphs, their derivatives, their inverse functions, their graphs, and their derivatives. The logarithmic formulas for the inverse hyperbolic trig functions are given in the textbook in the case of cosh, sinh, and tanh. The other hyperbolic trig function inverses can be given as

$$\coth^{-1} x = \frac{1}{2}\ln(\frac{x+1}{x-1}),$$

$$\operatorname{sech}^{-1} x = \ln(\frac{1+\sqrt{1-x^2}}{x}),$$

and

$$\operatorname{csch}^{-1} x = \ln(\frac{1+\sqrt{1+x^2}}{x}).$$

However, for finding the derivatives of the inverse hyperbolic trig functions, the method of implicit differentiation is the most efficient method. It is also helpful to notice that besides the hyperbolic forms of the Pythagorean Identities,

$$\cosh^2 - \sinh^2 = 1,$$

and

$$\tanh^2 + \operatorname{sech}^2 = 1,$$

it is also true that

$$\cosh + \sinh = \exp,$$

which of course means

$$\cosh x + \sinh x = e^x.$$

Since $\cosh \geq 1$, it follows that always we have by the hyperbolic Pythagorean Identity

$$\cosh = \sqrt{1+\sinh^2}.$$

This means that the equations following are all equivalent:

$$y = \sinh^{-1} x,$$

$$x = \sinh y,$$

$$e^y = \sinh y + \cosh y,$$

$$e^y = \sinh y + \sqrt{1+\sinh^2 y},$$

$$e^y = x + \sqrt{1+x^2},$$

$$y = \ln(x+\sqrt{1+x^2}),$$

$$\sinh^{-1} x = \ln(x+\sqrt{1+x^2}).$$

Keep in mind that as sinh is one-to-one on the whole real line and with range the whole real line, the same must be true for its inverse function.

However, for cosh, the domain is the whole real line whereas the range is the interval $[1,\infty)$. It is not one-to-one, but is when restricted to $[0,\infty)$. This means that $\cosh^{-1}$ has domain $[1,\infty)$ and range $[0,\infty)$. For the logarithmic expression for $\cosh^{-1}$, notice that if

$$y = \cosh^{-1} x,$$

then

$$x = \cosh y,$$

with $y \geq 0$ and $x \geq 1$. Then $\sinh y \geq 0$ and $\cosh y \geq 1$, so again by the hyperbolic Pythagorean Identity,

$$\sinh y = \sqrt{\cosh^2 y - 1} = \sqrt{x^2 - 1}.$$

We therefore now have

$$e^y = \cosh y + \sinh y = x + \sqrt{x^2 - 1}$$

which means

$$y = \ln(x + \sqrt{x^2 - 1})$$

so

$$\cosh^{-1} x = \ln(x + \sqrt{x^2 - 1}).$$

For the inverse hyperbolic tangent, we notice that tanh is one-to-one on the whole real line but its range is the open interval $(-1, 1)$. This means that $\tanh^{-1}$ has domain the open interval $(-1, 1)$ and range the whole real line. When we look at the graphs, we notice that tanh roughly resembles arctan whereas the graph of $\tanh^{-1}$ roughly resembles the graph of tan. Assume that

$$y = \tanh x.$$

Then

$$x = \tanh y.$$

Using the equation $e^y = \sinh y + \cosh y$ and dividing through by $\cosh y$, we find

$$1 + x = 1 + \tanh y = \frac{e^y}{\cosh y} = \frac{2e^y}{e^y + e^{-y}} = \frac{2}{1 + e^{-2y}}.$$

Thus,

$$1 + e^{-2y} = \frac{2}{1+x},$$

$$e^{-2y} = \frac{2}{1+x} - 1 = \frac{1-x}{1+x},$$

$$e^{2y} = \frac{1+x}{1-x},$$

$$y = \frac{1}{2}\ln(\frac{1-x}{1+x}),$$

$$\tanh^{-1} x = \frac{1}{2}\ln(\frac{1-x}{1+x}).$$

The logarithmic expressions for the remaining inverse hyperbolic trig functions can be found using the reciprocal relations. For instance, if we have $y = \coth^{-1} x$, then $x = \coth y$, so $\tanh y = 1/x$, and therefore

$$y = \tanh^{-1}(\frac{1}{x}) = \frac{1}{2}\ln(\frac{1 + (1/x)}{1 - (1/x)}) = \frac{1}{2}\ln(\frac{x+1}{x-1}),$$

and therefore

$$\coth^{-1} x = \frac{1}{2}\ln(\frac{x+1}{x-1}).$$

In the same manner, if we consider $y = \mathrm{sech}^{-1}\, x$, then $x = \mathrm{sech}\, y$, so

$$\frac{1}{x} = \cosh y$$

and

$$y = \cosh^{-1}(\frac{1}{x}) = \ln(\frac{1}{x} + \sqrt{\frac{1}{x^2} - 1}) = \ln(\frac{1 + \sqrt{1 - x^2}}{x}),$$

so

$$\operatorname{sech}^{-1}\, x = \ln(\frac{1+\sqrt{1-x^2}}{x}).$$

For $y = \operatorname{csch}^{-1}\, x$ we have $x = \operatorname{csch}\, y$ and therefore

$$\frac{1}{x} = \sinh y,$$

so

$$y = \sinh^{-1}(\frac{1}{x}) = \ln(\frac{1}{x} + \sqrt{\frac{1}{x^2}+1}) = \ln(\frac{1+\sqrt{1+x^2}}{x}),$$

and therefore finally,

$$\operatorname{csch}^{-1}\, x = \ln(\frac{1+\sqrt{1+x^2}}{x}).$$

26. **LECTURE** WEDNESDAY 21 OCTOBER 2009

Today we worked applied optimization problems. Typically in these problems, there is an expression $f(x,y)$ which we want to either maximize or minimize, but the variables x and y are subject to a constraining equation of the form $g(x,y) = constant$. In this situation, it is common to refer to the function f as the *objective* and refer to g as the *constraint*. If we think of x and y as varying with time t, then f and g become functions of time as composite functions, and we can differentiate both with respect to time. We know that we will have

$$\dot{f} = \frac{df}{dt} = F_1(x,y)\dot{x} + F_2(x,y)\dot{y},$$

and

$$\dot{g} = \frac{dg}{dt} = G_1(x,y)\dot{x} + G_2(x,y)\dot{y},$$

for some functions F_1, F_2, G_1, G_2 of the variables x and y. Since $g = 0$ is the constraint, it follows that $\dot{g} = 0$ always, but at the optimum value of x and y where f has a local extreme value, it also has a local extreme value as a function of time t, and therefore at the optimum point (x_0, y_0), we must have $f(x_0, y_0) = 0$. Thus, the strategy is to solve the pair of equations

$$\dot{f}(x,y) = 0,$$

$$\dot{g}(x,y)$$

simultaneously and then examine the solutions to find the optimum. For instance, if there is a unique solution, then it is typically the point where f either has a maximum value or where it has a minimum value. If there are two solutions, then typically one is where f has its maximum and the other is where f has its minimum.

As examples, we worked several problems involving trying to maximize area under constraints or maximize volume under constraints. For instance, if we want the cylinder of maximum volume contained in a sphere of radius R, then obviously the cylinder's top and bottom should be circular arcs on the sphere as otherwise the cylinder's volume could be increased simply by extending the cylinder until it touches the sphere. Moving the axis of the cylinder in line with a diameter of the sphere will also increase its volume, so we can assume the axis of symmetry of the cylinder is a diameter of the sphere. The "outline of this sphere is a circle of radius R, and when we align the axis of symmetry of the cylinder to be vertical, then its top edge touches the outline circle in a point (x,y) with $x^2 + y^2 = R^2$. The radius of the top of this cylinder is then obviously x and the height is $2y$, so the volume is

$$V = (\pi x^2)(2y) = 2\pi x^2 y.$$

Notice that for finding the optimum value of x and y here, the factor of 2π is irrelevant, so we can replace V in this problem by the simpler $f(x,y) = x^2 y$. Our constraint function is

$$g(x,y) = x^2 + y^2.$$

differentiating with respect to t in both these expressions and setting $\dot{f} = 0$ gives the two equations

$$2x\dot{x}y + x^2\dot{y} = 0$$

and

$$x\dot{x} + y\dot{y} = 0.$$

Thus we have the first equation of the pair can be written

$$0 = 2y(x\dot{x}) + x^2\dot{y},$$

and we can use second equation of the pair to get $x\dot{x} = -y\dot{y}$ so

$$0 = 2y(-y\dot{y}) + x^2\dot{y} = (x^2 - 2y^2)\dot{y}.$$

Now, this is all independent of how y depends on time, so we can assume we move y so as to make $\dot{y} \neq 0$, and this means we now must have

$$x^2 - 2y^2 = 0$$

at the optimum values for the variables x and y. It follows that for the optimum we must have

$$x^2 = 2y^2.$$

Since we must obviously have positive values for x and y to get the maximum volume, we must have

$$x = \sqrt{2}\, y.$$

From the constraint equation we have

$$R^2 = x^2 + y^2 = 2y^2 + y^2 = 3y^2,$$

and therefore

$$y = \frac{R}{\sqrt{3}},$$

and

$$x = \frac{\sqrt{2}}{\sqrt{3}} R.$$

It follows that the maximum volume of a cylinder inside the sphere of radius R must be

$$V_{max} = 2\pi x^2 y = 2\pi(\frac{\sqrt{2}}{\sqrt{3}}R)^2(\frac{R}{\sqrt{3}}) = (\frac{4}{3}\pi R^3)\frac{1}{\sqrt{3}}.$$

In particular, we see here that the cylinder takes up the fraction $1/\sqrt{3}$ of the volume inside the sphere. This method of using two equations is often simpler than reducing the objective function to a function of a single variable using the constraint equation to start with, since it typically makes for a complicated expression in the single variable, so the differentiation becomes complicated and therefore one is more likely to make a mistake in all the algebra.

If we look for the maximum volume circular pyramid (cone) inscribed in a sphere, then it is simplest to imagine the cone is upside down with its vertex at the "south pole" of the sphere and its axis of symmetry being the vertical diameter of the sphere. The base circle then touches the outline circle of the sphere at a point (x, y) so the base has radius x and the height is $R+y$. The volume of the pyramid is therefore

$$V = \frac{1}{3}\pi x^2(R + y)$$

and the constraint function is again

$$g = x^2 + y^2,$$

and the constraint equation is $g = R^2$. Again, we may as well dispense with the factor $\pi/3$ and replace V by the objective

$$f = x^2(R + y).$$

Differentiating with respect to time now gives the pair of equations

$$2x\dot{x}(R + y) + x^2\dot{y} = 0,$$

$$x\dot{x} + y\dot{y} = 0.$$

We can notice that the first term of the first equation contains the factor $x\dot{x}$ which by the second equation of the pair must equal $-y\dot{y}$. We therefore have

$$0 = 2(-y\dot{y})(R + y) + x^2\dot{y} = (x^2 - 2y[R + y])\dot{y}.$$

Thus, assuming we choose $\dot{y} \neq 0$, we must have

$$x^2 - 2y[R + y] = 0.$$

Using the constraint equation we have $x^2 = R^2 - y^2$, so making this replacement gives

$$R^2 - y^2 - 2y[R + y] = 0.$$

This is obviously equivalent to the quadratic equation

$$3y^2 + 2Ry - R^2 = 0.$$

Using the quadratic formula we find

$$y = \frac{-2R \pm \sqrt{(2R)^2 - 4(3)(-R^2)}}{2(3)} = \frac{-R \pm 2R}{3}.$$

This gives two possible values where V is extreme, namely $y = -R$ where $V = 0$, and $y = R/3$. Obviously this second solution $y = R/3$ must be the value of y which gives the maximum volume. From the constraint equation we have

$$x = \sqrt{R^2 - y^2} = \frac{\sqrt{8}}{3}R,$$

so the maximum volume is

$$V_{max} = \frac{1}{3}\pi\frac{8}{9}R^2[R + (R/3)] = \frac{8}{27}(\frac{4}{3}\pi R^3),$$

which shows also that the maximum volume pyramid contained in a sphere takes up the fraction $8/27$ of the volume contained in the sphere.

We also observed that in general, for problems of minimizing distance from a point to a curve with equation $g(x, y) = \mathit{constant}$, that the point on the curve closest to the given point is the one for which the line segment from the given point to the curve is perpendicular to the tangent to the curve at that point. That is to minimize the distance from the point (a, b) to the curve, the objective function is

$$d(x, y) = \sqrt{(x - a)^2 + (y - b)^2},$$

and we noted that since squaring of non-negative numbers is order preserving, we may as well replace the objective with its square $f = d^2$. Then if (x_0, y_0) is the point on the curve closest to (a, b), then the slope of the tangent to the curve at the point (x_0, y_0) is the negative reciprocal of

$$\frac{y_0 - b}{x_0 - a}.$$

As this last is the slope of the line segment from (a, b) to $x_0, y_0)$, this means the tangent line is perpendicular to the segment.

27. **LECTURE** FRIDAY 23 OCTOBER 2009

Today we continued working optimization problems. These problems often come in the form: maximize the area of something subject to a constraint on the size, or maximize the volume of something subject to a constraint on its size. For instance, find the maximum volume pyramid that can fit inside a sphere of given radius, of find the maximum area rectangle that can fit inside a given circle. However, another type of problem consists of finding the maximum area of a certain type of geometric object given a constraint on its perimeter or the maximum volume of a certain type of object given a constraint on its surface area. Notice that there is a duality here. For instance, finding the maximum volume cylinder of given surface area should have the same proportions as the minimum surface area cylinder of given volume. To see this more clearly, if we have two geometric parameters x and y for the geometric figure, and we have perimeter and area expressed in terms of these parameters

$$A = A(x, y)$$

and

$$P = P(x, y),$$

then we know if we seek to maximize A subject to fixed $P = constant$, then allowing x and y to change with time t, we have

$$\dot{P} = 0,$$

because P is constant, and we are seeking to maximize A, so we know at the optimal values of x and y we will have

$$\dot{A} = 0,$$

since then A has a local maximum as a function of t at these optimal values of x and y. So to find the optimal values of x and y we simultaneously solve the pair of equations

$$\dot{P} = 0,$$

$$\dot{A} = 0.$$

On the other hand, if we seek to minimize the perimeter P subject to a fixed area A, then it is A that is constant, so $\dot{A} = 0$ and thus at the optimal x and y we have $\dot{P} = 0$ as these give a local extreme value for P as a function of t. Thus, we need to solve the same pair of equations

$$\dot{P} = 0,$$

$$\dot{A} = 0,$$

for this second problem. That is, both problems are really the same problem.

28. **LECTURE** MONDAY 26 OCTOBER 2009

Today we discussed Newton's Method for finding roots of differentiable functions and we also discussed antiderivatives and their use for finding areas under curves.

If f is a continuous function on an interval I, and we seek a *root* of f, that is a solution to the equation

$$f(x) = 0,$$

a very simple minded method is to begin by trying to find a pair of points x_0 and x_1 in I such that $f(x_0) < 0$ and $f(x_1) > 0$. Then by the Intermediate Value Theorem, we know there is a root between x_0 and x_1. This means that the maximum distance of either of these two points from a root is the distance from x_0 to x_1. Call this distance D. Next, define

$$x_2 = \frac{x_0 + x_1}{2},$$

so x_2 is the midpoint of the segment from x_0 to x_2. This is likely to be closer to the root, so we check to see whether $f(x_2)$ is positive or negative. If it is negative, we know a root lies between x_2 and x_1, whereas if it is positive, we know there is a root between x_0 and x_2. In either case, we find a new pair of points where the maximum distance of either of the two points to a root is half of what it was for the initial pair, that is $D/2$. Continuing in this way, we arrive at a sequence of points (x_n) where the distance from x_n to a root is no more than $D/(2^n)$, so the sequence must converge to a root. This simple-minded method is known as the Bisection Method for obvious reasons. It is guaranteed to work, but it may be very slow to converge to a root.

In case of a differentiable function, a better method is known as Newton's Method or the Newton-Raphson Method. The idea is that you start by choosing any initial point x_0 and if $f(x_0) \neq 0$, then follow that tangent line to where it crosses the $x-$axis. Pictorially it appears this often converges very quickly to a root. Thus, having found x_n, we choose x_{n+1} to be the point where the tangent to the graph of f at the point $(x_n, f(x_n))$ crosses the $x-$axis. Since the equation of the tangent line at $(x_n, f(x_n))$ is

$$y = f'(x_n)[x - x_n] + f(x_n),$$

we set the left side equal to zero and the solution is x_{n+1}. Thus

$$0 = f'(x_n)[x_{n+1} - x_n] + f(x_n),$$
$$f(x_n) = f'(x_n)[x_n - x_{n+1}],$$
$$\frac{f(x_n)}{f'(x_n)} = x_n - x_{n+1},$$

and therefore

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}.$$

Let the function F be defined by

$$F(x) = x - \frac{f(x)}{f'(x)}.$$

Notice its domains is all points in the domain of f except for critical points. We start by choosing x_0 arbitrarily, and then we have

$$x_1 = F(x_0),$$
$$x_2 = F(x_1),$$
$$\vdots$$
$$x_{n+1} = F(x_n),$$

and so on. In many cases this converges rapidly to a root of the original function f. How fast depends heavily on the choice of starting point x_0. Obviously, if we are lucky enough to start near a root, then pictorially we see that convergence to a root should be rapid. However, there are exceptional cases where there is no convergence at all. For instance, if we draw a small parallelogram with a pair of opposite vertices on the $x-$axis and with a pair of vertical sides, then any function whose graph is tangent to the other two vertices will have a pair of starting points with the property that $F(x_1) = x_0$ and thus the method simply "goes around in circles". Clearly, such a geometry for a graph of a function is exceptional and unusual.

We can also notice that if we are lucky and find a point x_0 with the property that $F(x_0) = x_0$, then x_0 is the root of f. Indeed, if $F(x_0) = x_0$, then

$$x_0 = F(x_0) = x_0 - \frac{f(x_0)}{f'(x_0)},$$

so

$$\frac{f(x_0)}{f'(x_0)} = 0$$

and therefore

$$f(x_0) = 0.$$

We see what is really going on is that roots of f are the same as solutions of the equation $F(x) = x$. If F is any function, a point x in the domain of F with the property $F(x) = x$ is called *Fixed Point* of the function. The Newton-Raphson Method merely converts the root problem for f into the problem of finding a fixed point for F. The root problem is converted to a fixed point problem. Of course, if we have a fixed point problem for some function F it can be converted back to a root problem since setting $g(x) = x - F(x)$, we see x is a fixed point of F if and only if it is a root of g. If we have a useful method of finding fixed points, then that gives a useful method for finding roots via Newton's method.

Suppose now that F is any continuous function whose range is included in its domain. If we seek a fixed point of F, one simple method is to simply pick any point x_0 in the domain of F and form the sequence

$$\begin{aligned} x_1 &= F(x_0), \\ x_2 &= F(x_1), \\ &\ \ . \\ &\ \ . \\ &\ \ . \\ x_{n+1} &= F(x_n), \end{aligned}$$

and so on. We call this the *Iteration Sequence* for F. This is just what we did with the F in the Newton-Raphson Method, for the case where

$$F(x) = x - \frac{f(x)}{f'(x)}.$$

Suppose this sequence converges to a limit x_∞. Then by continuity we have

$$F(x_\infty) = F(\lim_{n\to\infty} x_n) = \lim_{n\to\infty} F(x_n) = \lim_{n\to\infty} x_{n+1} = x_\infty,$$

and therefore x_∞ is a fixed point of F.

Of course, if F has no fixed point, then it is useless to search for one, so it is important to have methods that guarantee existence of fixed points. In this direction we have the following theorem

Theorem 28.1. *FIXED POINT THEOREM. If F is a continuous function from the closed interval $[a, b]$ with range contained in the closed interval $[a, b]$, then F must have a fixed point.*

To see why this must be true, consider the function g defined by $g(x) = x - F(x)$. If $F(a) = a$, then a is a fixed point, so if a itself is not a fixed point, the $F(a) > a$, and therefore $g(a) < 0$. If $F(b) = b$, then b is a fixed point, then $b > F(b)$ and therefore $g(b) > 0$. Thus if neither a nor b is a fixed point, then $g(a) < 0$ and $g(b) > 0$, so by the Intermediate Value Theorem, there is some x_0 between a and b with $g(x_0) = 0$. But then $F(x_0) = x_0$, so x_0 is a fixed point for the function F.

If the function F has also the property that

$$|F(x) - F(y)| < K|x - y|,$$

where K is a positive constant with $K < 1$, then it can be shown that the sequence will converge to a fixed point know matter how x_0 is chosen. Notice that the iteration sequence points are getting closer and closer together in this case.

In general, if we have a differentiable function f and we use Newton's method, then if we can restrict to a closed interval I contained in the domain of f where there are no critical points of f and where the range of F is contained in I, then the Fixed Point Theorem guarantees a fixed point of F in I and therefore a root of f in I.

As a simple application, we noted that to apply Newton's method to find the square root of a positive number a, we can view $\sqrt{a}$ as a root of the function

$$f(x) = x^2 - a.$$

Then

$$F(x) = x - \frac{f(x)}{f'(x)} = x - \frac{x^2 - a}{2x} = \frac{2x^2 - (x^2 - a)}{2x}$$
$$= \frac{x^2 + a}{2x} = \frac{x + \frac{a}{x}}{2}.$$

Thus

$$F(x) = \frac{x + \frac{a}{x}}{2}$$

is just the average of x and a/x. That is for the iteration sequence, with any positive number $x_0 < a$, just average x_n with $a/(x_n)$ to get x_{n+1}. This is a very common sense approach, since if x is smaller than $\sqrt{a}$, then a/x is larger than $\sqrt{a}$, whereas if x is larger than $\sqrt{a}$, then a/x is smaller, and therefore either way, $\sqrt{a}$ is between x and a/x. Thus the average of x and a/x should be even closer to $\sqrt{a}$. In fact this method was known in ancient Babylon.

Our next topic is finding antiderivatives for functions. We say that F is an *Antiderivative* of f provided that $F' = f$. We observed that if $F' = 0$, then by the Intermediate Value Theorem F must be constant on each interval. In particular, if F_1 and F_2 are both antiderivatives of f on the interval I, then $F_1 - F_2 = C$ for some constant C. We denote the general antiderivative F of f by

$$F = \int f,$$

or

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

and must keep in mind that the general antiderivative is only defined up to an additive constant on each interval. Alternately, we remind ourselves of this by writing

$$\int f = F + C$$

or

$$\int f(x)dx = F(x) + C,$$

whenever F is any specific antiderivative of f that we have found. For instance, we see easily that

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

so we have found a particular antiderivative of x^p, and therefore

$$\int x^p dx = \frac{x^{p+1}}{p+1} + C.$$

Since $\ln' x = 1/x$, we also have

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

The expression

$$\int f$$

is also called the *Indefinite Integral* of f because of the role that antiderivatives play in finding areas under curves. If f is a positive function on the interval $[a, b]$ and we wish to find the area B under the graph of f, then we can define $A(x)$ to be the area under the graph of $f|[a, x]$, the restriction of f to the interval $[a, x]$. Thus $B = A(b)$. If we allow x to move to the right at positive velocity $\dot{x}$, then we know that by the Chain Rule we have

$$\frac{d}{dt}[A(x)] = A(x)\dot{x},$$

but from the geometric version of the Fundamental Theorem of Calculus that we discussed in our first lecture at the beginning of the semester, we know that the rate of change of the area with time must be the length of the moving boundary multiplied by the velocity at which it moves. As x moves, the length of the moving boundary is $f(x)$ at each instant, so

$$\frac{d}{dt}[A(x)] = f(x)\dot{x}.$$

Thus

$$A'(x)\dot{x} = \frac{d}{dt}[A(x)] = f(x)\dot{x}.$$

Since $\dot{x} \neq 0$, it follows that

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

This means that $A(x)$ is a particular antiderivative of f. Notice that $A(a) = 0$ and $A(b)$ is the area under the graph of f which we were originally looking for. This leads to the following method of finding the area under the graph of a function on an interval $[a, b]$. We begin by finding any antiderivative F for f on the interval $[a, b]$. We know then that $A - F = C$ is constant on $[a, b]$. Since $A(a) = 0$, we have

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

so

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

and therefor the original area B we are seeking is simply

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

In terms of the integral notation, it is useful to denote this area as

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

To smooth out computations it is useful to have the notation

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

We then have for any antiderivative F of f, that

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

As an example of the use of this notation, to find the area under the graph of the function $f(x) = x^3$ on the interval $[1, 2]$, we have

$$Area\ = \int_1^2 x^3 dx = \frac{x^4}{4}|_1^2 = \frac{2^4}{4} - \frac{1}{4} = \frac{4^2}{4} - \frac{1}{4} = \frac{15}{4}.$$

We are often required to find an antiderivative F of f on a given interval with a particular value at a specified point, say $F(x_0) = y_0$. In this case, we know that if G is any antiderivative of f on the given interval, then $F - G = C$ for some constant C. Thus

$$C = F(x_0) - G(x_0),$$

so

$$F = G + C = G + F(x_0) - G(x_0).$$

That is, for any x in the given interval, it is the case that

$$F(x) = G(x) + y_0 - G(x_0) = y_0 + G(x) - G(x_0).$$

Notice we also have with our notation for the integral, as G is an antiderivative of f,

$$F(x) = F(x_0) + \int_{x_0}^x f = y_0 + \int_{x_0}^x f.$$

29. **LECTURE** WEDNESDAY 28 OCTOBER 2009

Today we discussed the problem of defining areas of planar regions and lengths of continuous curves. We had observed in the first lecture that if we assume these definitions can be reasonably made, then a variable region, which changes by virtue of allowing part of its boundary to move, must have area A depending on time in such a way that

$$\dot{A} = Lv$$

at the instant when the moving boundary has length L if it moves out with velocity v at that same instant. In the last lecture we discussed antiderivatives and noted that by the Mean Value Theorem, if f and g have the same derivative on an interval, then their difference is a constant. In particular, this means that if $f' = g'$ on the open interval I and if $f(x_0) = g(x_0)$ for some single point $x_0 \in I$, then $f = g$ everywhere on I. Because, we can write

$$f - g = C$$

where C is a constant, but then

$$C = f(x_0) - g(x_0) = 0,$$

so $C = 0$ and therefore $f - g = 0$, so $f = g$ on I. We combine these two facts to find the area of a circle. Assuming the boundary of a circle has length L which depends on the radius r, we see this function is linear in r and must vanish when $r = 0$. Therefore, $L(r) = kr$ for some constant k, and by definition, $k = 2\pi$. Thus $L(r) = 2\pi r$ gives the circumference of a circle of radius r. Given this, the area inside the circle of radius r is some function A depending on r, and clearly $A(0) = 0$. On the other hand, consider a circle that grows by virtue of having its boundary move out with velocity v. Then we must have

$$\dot{A} = L(r)v.$$

But, we know if the boundary circle is growing so as to move with velocity v, then $\dot{r} = v$. On the other hand, assuming that A is a differentiable function of r, by the Chain Rule we must have

$$\dot{A} = A'(r)\dot{r},$$

so combining these facts we have

$$A'(r)\dot{r} = \dot{A} = L(r)\dot{r} = 2\pi r\dot{r},$$

and therefore assuming $\dot{r} = v \neq 0$, we can cancel and find

$$A'(r) = 2\pi r = (\pi r^2)'.$$

Since $g(r) = \pi r^2$ is also a differentiable function of r which vanishes for $r = 0$, this means $A = g$, so

$$A(r) = \pi r^2.$$

We therefore see that we can find the formula for the area enclosed by a circle provided that we assume that it actually makes sense to say the circle has a length and encloses a region for which it makes sense to say it actually has an area.

We next discussed the problem of defining the length of a curve C. We assume the curve to be oriented so the curve has a beginning and an end, and this orders all the points on the curve. If we we chose any finite set of points on the curve, arranging them in order as they appear along the curve, we can connect them with straight lines forming a *polygonal path* connecting the beginning of C to the end of C. We define the length of a polygonal path to be the sum of the lengths of the straight line segments making it up. Since the shortest distance between two points is the length of the straight line segment joining them, if C has a length in any reasonable sense, it must be longer than the length of such a polygonal path, all of whose vertices are on C. For any polygonal path P, let $L(P)$ denote its length. If we can say that C has a length, we call it $L(C)$. Therefore, we know if $L(C)$ exists, then $L(P) \leq L(C)$, for any polygonal path all of whose vertices lie on C in the same order. We say that the curve C has *Bounded Variation* if

$$\mathcal{P}_C = \{L(P) : P \textit{ is a polygonal path with vertices on } C \textit{ in order}\} \subset [0, B]$$

for some positive number B. More generally, if S is any set of numbers, we call B an Upper Bound of S provided that $S \subset (-\infty, B]$, and likewise we say that B is a *Lower Bound* of S provided that $S \subset [B, \infty)$. Thus, if $L(C)$ exists, then certainly $L(C)$ is an upper bound for $\mathcal{P}_C$, so conversely, if $\mathcal{P}_C$ has no upper bound, then C cannot have a length. On the other hand, if C has a length, then certainly $L(C) \leq B$, for any B which is an upper bound for $\mathcal{P}_C$. That is, $L(C)$ is what is called the least upper bound for $\mathcal{P}_C$. In general, we say that L is the *Least Upper Bound* for the set $S \subset \mathbb{R}$, if it is an upper bound and S has no upper bound smaller than L. The *Completeness Property* of the real number system is the property that every set which has an upper bound in fact has a least upper bound. If $S \subset \mathbb{R}$ is bounded above (meaning it has an upper bound), then we denote its least upper bound by $LUB(S)$. Thus, we have shown that if C has a length in any reasonable sense, then it must be the case that

$$L(C) = LUB(\mathcal{P}_C),$$

so we take this as the definition of the length of any curve with bounded variation. Thus, if C has bounded variation, then $\mathcal{P}_C$ is bounded so by the completeness property of $\mathbb{R}$ it must have a least upper bound, and so we define

$$L(C) = LUB(\mathcal{P}_C)$$

for any curve C of bounded variation. It can be proved that if C is the graph of a function which is continuous on the closed interval $[a, b]$ and differentiable on the open interval (a, b), then C is of bounded variation. Clearly, we can put together curves of bounded variation to get new curves of bounded variation whenever the end of one of them is the beginning of the other. In this case, if the pieces are C_1 and C_2, then

$$L(C_1 \cup C_2) = L(C_1) + L(C_2).$$

Thus plane curves which can be formed by piecing together graphs of differentiable functions all have length. In particular, the circle of radius r centered at $(0, 0)$ can be formed as the union of the upper half and the lower half circles and each of these semi-circles is the graph of a continuous function on $[-r, r]$ which is differentiable on the open interval $(-r, r)$. This shows that the circle has length. Obviously whenever you magnify a region by the scale factor r, all lengths of polygonal paths are multiplied by that same scale factor r, and therefore so is the length of any curve in the region. Thus, the circle of radius r must have a length which is r multiplied by the length of a circle of radius one. We define π to be the length of the semi-circle of radius one, so the circle of radius one has length 2π, and therefore the circle of radius r must have length $2\pi r$.

Our discussion of length depended on the completeness property of $\mathbb{R}$. If $S \subset \mathbb{R}$, and S has a lower bound, say L, then

$$-S = \{-x : x \in S\}$$

has upper bound $-L$, so it has a least upper bound M, and therefore $-M = G$ is a lower bound which is greater than any other lower bound, so we call it the *Greatest Lower Bound*, and we define

$$GLB(S) = G.$$

Thus, by the completeness property of $\mathbb{R}$, any set of numbers having a lower bound has a greatest lower bound.

Next, we discussed the problem of defining the area enclosed by a curve, or more generally, the area of any subset of the plane. If $\mathcal{R}$ is a region in the plane, and if $\mathcal{R}$ is the union of two regions $\mathcal{R}_1$ and $\mathcal{R}_2$ which overlap at most in boundary points, and if areas make sense for these regions, then clearly we should have

$$Area(\mathcal{R}) = Area(\mathcal{R}_1 \cup \mathcal{R}_2) = Area(\mathcal{R}_1) + Area(\mathcal{R}_2).$$

In particular, if $\mathcal{R}$ is the union of a set of triangles which only overlap on their boundaries, then $Area(\mathcal{R})$ should be the sum of the areas of the triangles. However, as intuitively clear as this is, it is very difficult to prove that for two different such sets of triangles making up $\mathcal{R}$ the sums of their areas are the same. We call such a decomposition of $\mathcal{R}$ a triangulation. Clearly, if the boundary of $\mathcal{R}$ is a polygonal path, then $\mathcal{R}$ has a triangulation, but there are many, and we need to see that all such triangulation have the same total areal. If two such triangulations are supper-imposed, then each triangle of one gets chopped up into polygonal regions by the other triangulation, and then the polygonal regions can be triangulated to arrive at a final triangulation for which each triangle in either of the first two triangulations is a union of triangles of this last triangulation. Thus, to see that the area of a polygonal region is independent of the choice of triangulation, it is enough to prove that for any triangulation of a triangle, that the sum of the areas of the triangles making up the triangulation gives the area of the triangle. You can see that to actually prove such a statement could be a problem. We will assume that fact can be proven and leave it at that.

Once we have settled the problem of areas of polygonal regions, for any plane region $\mathcal{R}$ we define $Inn\mathcal{P}(\mathcal{R})$ to be the set of all polygonal regions contained in $\mathcal{R}$ and $Out\mathcal{P}(\mathcal{R})$ to be the set of all polygonal regions which contain $\mathcal{R}$. Then we define the *Inner Area* of $\mathcal{R}$, denoted $InnArea(\mathcal{R})$ by the formula

$$InnArea(\mathcal{R}) = LUB(\{Area(P) : P \in Inn\mathcal{P}(\mathcal{R})\})$$

and likewise, we define the *Outer Area* of $\mathcal{R}$, denoted $OutArea(\mathcal{R})$, by the formula

$$OutArea(\mathcal{R}) = GLB(\{Area(P) : P \in Out\mathcal{P}(\mathcal{R})\}).$$

Notice that zero is a lower bound in the second case for dealing with outer area, so the outer area always exists as soon as the region can be contained in a polygonal region, say P. On the other hand, if this is the case, then certainly any polygonal region contained in $\mathcal{R}$ is contained in P and therefore the inner area exists. That is, if $\mathcal{R}$ can be contained in some polygonal region, say a big square, then both $InnArea(\mathcal{R})$ and $OutArea(\mathcal{R})$ exist. Notice that if $P_0 \subset \mathcal{R} \subset P_1$, where P_0 and P_1 are polygonal regions, then $Area(P_0) \leq Area(P_1)$, and therefore

$$Area(P_0) \leq GLB(\{Area(P) : P \in Out\mathcal{P}(\mathcal{R})\}) = OutArea(\mathcal{R}),$$

and so

$$InnArea(\mathcal{R}) \leq OutArea(\mathcal{R}),$$

always. Certainly if $\mathcal{R}$ is a polygonal region then the outer and inner areas are the same, so in general, we say that $\mathcal{R}$ has area if both its inner area and outer area agree, and if so we denote this by $Area(\mathcal{R})$. Thus, if $Area(\mathcal{R})$ exists, then

$$InnArea(\mathcal{R}) = Area(\mathcal{R}) = OutArea(\mathcal{R}).$$

We gave a picture argument that for the region inside a circle, that the inner and outer areas coincide and have the value πr^2. The first basic useful fact here is that if $\mathcal{R}$ is the region bounded by a continuous curve, then $Area(\mathcal{R})$ exists. To see this in the case of the region under the graph of a function, we take a continuous function defined on the interval $[a, b]$ and we partition the interval with a sequence of points $a = x_0, x_1, ..., x_n = b$, with

$$x_0 < x_1 < ... < x_n.$$

We call an interval of the form $[x_{k-1}, x_k]$ the k^{th} subinterval of the partition and we define

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

Choose a *Sample Point* in the k^{th} subinterval denoted t_k, so

$$x_{k-1} \leq t_k \leq x_k.$$

The rectangle of height $f(t_k)$ and base the k^{th} subinterval has area $f(t_k)\Delta x_k$, and all such rectangles taken together form a polygonal region which approximates the region under the curve. If we choose all t_k to be the point where $f|[x_{k-1}, x_k]$ has a minimum value, then all such rectangles are inside the region under the curve so their total area is called a *Lower Sum* for f, whereas if we always choose maximum values, then the total area is called an *Upper Sum* for f. Clearly the greatest lower bound for the upper sums must exceed the least upper bound for the lower sums, just as in the case of inner and outer area. In fact the least upper bound of the lower sums coincides with the inner area of the region under the graph of f and the greatest lower bound of the upper sums coincides with the outer area. Thus, this gives a criterion for the region under the graph of f to have area in terms of the upper and lower sums.

The main fact here is that for a continuous function the region under the graph has an area so the greatest lower bound of the upper sums coincides with the least upper bound of the lower sums. The general sum with arbitrarily chosen sample points is called a *Riemann Sum*, and clearly, for a given partition, any Riemann sum is between the lower and upper sums for that partition.

30. **LECTURE** FRIDAY 30 OCTOBER 2009

Today we discussed the definition of the *Riemann Integral* and Riemann sums. Suppose f is a function on the interval $[a, b]$. Suppose $\mathcal{P} = (x_0, x_1, ..., x_n$ is a partition of $[a, b]$ so $a = x_0 < x_1 < ... < x_n = b$. Then we call $[x_{k-1}, x_k]$ the k^{th} subinterval of the partition, and its length is

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

We next can choose a sequence of sample points

$$\mathbf{t} = (t_1, t_2, ..., t_n)$$

with t_k in the k^{th} subinterval for each $k \leq n$. We define

$$R(f, \mathcal{P}, \mathbf{t}) = \sum_{k=1}^{n} f(t_k)\Delta x_k,$$

and we call the the Riemann sum for the given partition and given sample points. In case f is continuous, we can choose t_k to be the point where $f|[x_{k-1}, x_k]$ has its minimum value, then the resulting sum is a special case of a Riemann sum called a lower sum and which we denote by $L(f, \mathcal{P})$. On the other hand, if we choose t_k to be the point where $f|[x_{k-1}, x_k]$ has its maximum value, we get a special Riemann sum called an upper sum denoted $U(f, \mathcal{P})$. Obviously,

$$L(f, \mathcal{P}) \leq R(f, \mathcal{P}, \mathbf{t}) \leq U(f, \mathcal{P}).$$

We define $|\mathcal{P}|$ to be the length of the longest subinterval of the partition. We say that f is *Riemann integrable* on $[a, b]$ provided that

$$L = \lim_{|\mathcal{P}| \to 0} R(f, \mathcal{P}, \mathbf{t})$$

exists for all choices of sample points. Precisely, this means that if $\epsilon > 0$ is given, then there is a $\delta > 0$ such that if $|\mathcal{P}| \leq \delta$, then

$$|L - R(f, \mathcal{P}, \mathbf{t})| < \epsilon.$$

no matter how $\mathbf{t}$ is chosen or how the particular points forming the partition are chosen. In that case we write this limit L as

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

That is

$$\int_a^b f = \lim_{|\mathcal{P}| \to 0} R(f, \mathcal{P}, \mathbf{t}).$$

We noted that for continuous f, the lower sums allow us to define the lower integral as the least upper bound of all lower sums and the upper sums allow us to define the upper integral as the greatest lower bound of all upper sums. We showed that if we have two partitions $\mathcal{P}_1$ and $\mathcal{P}_2$ where the second partition is finer than the first, meaning every point of the first is already in the second, then we have

$$L(f, \mathcal{P}_1) \leq L(f, \mathcal{P}_2) \leq U(f, \mathcal{P}_2) \leq U(f, \mathcal{P}).$$

But then, if $\mathcal{P}_1$ and $\mathcal{P}_2$ are any two partitions, we showed we could find a third partition $\mathcal{P}_3$ finer than either of the first two partitions, but then

$$L(f, \mathcal{P}_1) \leq L(f, \mathcal{P}_3) \leq U(f, \mathcal{P}_3) \leq U(f, \mathcal{P}_2).$$

This means any Lower sum is less than or equal to any upper sum for f. Thus, if f is integrable, the upper and lower integrals must coincide and both equal the Riemann integral of f on $[a, b]$, and conversely, if the upper and lower integral coincide, then f is Riemann integrable on $[a, b]$.

We also noted various properties of the integral and the Mean Value Theorem for integrals of continuous functions and used it to rigorously prove the Fundamental Theorem of Calculus which states that if f is continuous, then

$$F(x) = \int_a^x f$$

defines a continuous function which is differentiable on the open interval (a, b) and

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

Finally, we discussed the technique for finding the area between the graphs of f and g on $[a, b]$ by finding all crossing points that is solutions of $f(x) = g(x)$ in the interval $[a, b]$ and integrating the difference of the two functions between successive crossing points. Simply take the absolute value of each of these integrals and add them all up to get the area between the two graphs.

31. **LECTURE** MONDAY 2 NOVEMBER 2009

Today we discussed integration by substitution and calculation of areas and volumes. Keep in mind that as anti-differentiation or integration is the reverse of differentiation, any rules for differentiation can be turned around to give useful rules for anti-differentiation. Thus the sum rule for differentiation tells us that

$$\int (f+g) = \int f + \int g$$

and the constant multiple rule for differentiation tells us that

$$\int kf = k \int f, \ k \ \textit{constant}.$$

But, the most powerful differentiation rule is the Chain Rule which says that

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

If we set $g = G'$, then G is an antiderivative of g and the chain rule says

$$\int g(f(x)f'(x)dx = G(f(x)) + C,$$

which means that we really only have to anti-differentiate g in this situation. This leads to a useful computational procedure. We use the notation

$$\left[G(u)|_{u=f(x)}\right] = G(f(x)).$$

As

$$\int g = G + C,$$

this means we have

$$\int g(f(x))f'(x)dx = \int g(u)du|_{u=f(x)}.$$

Thus, once the substitution is made, the integration problem is simplified, and once done, we go back and substitute for u in terms of x via the equation $u = f(x)$. In the case of a definite integral, the results are even simpler because we will see that the limits can be changed according to the substitution and then the original substitution can be forgotten. Specifically, with $G' = g$, we have

$$\int_a^b g(f(x)f'(x)dx = G(f(x)|_a^b = G(f(b)) - G(f(a)) = G(u)|_{f(a)}^{f(b)} = \int_{f(a)}^{f(b)} g(u)du.$$

We worked examples of integration using substitution. In practice this means that if you decide to substitute $u = f(x)$ when you see $g(f(x))$ in the integrand, then you compute

$$\frac{du}{dx} = f'(x)$$

or

$$du = f'(x)dx.$$

We can then solve this equation for dx and find

$$dx = \frac{du}{f'(x)},$$

so substituting these into the integrand gives

$$\int g(f(x)h(x)dx = \int g(u)\frac{h(x)}{f'(x)}du.$$

Thus, if $h(x)/f'(x)$ can be expressed in terms of u, then the substitution has been accomplished. However, it is only if the resulting integral in terms of u can actually be done that the substitution is really successful. In general, there are no rules here as to when substitution is successful. You must learn through practice, trial, and error. Sometimes substitution helps and sometimes it does not. Moreover, in many problems there are more than one substitution to try, each leading to a different integration problem, some solvable, some very difficult, and some unsolvable. Because integration is an inverse process, it is not like differentiation where you have rules that cover all possibilities. With integration, we develop an arsenal of techniques to try. It is easy to write down integrals which cannot be solved, specifically, it is easy to write down functions for which we have no way to write down the antiderivative function. As an easy example, there is no expression for the antiderivative of

$$f(x) = e^{-x^2}.$$

We observed that the area A trapped between the graphs of two functions f and g between the limits $x = a$ and $x = b$ is found in steps. We notice first that

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

but we have no rule for integrating an absolute value of a function in terms of the original function. Here are the steps.

STEP 1: Find all solutions of $f(x) = g(x)$ which lie in the interval $[a, b]$, and include the interval endpoints whether or not they are solutions. List them as a sequence in increasing order, say

$$a = x_0 < x_1 < x_2 < ... < x_n = b.$$

STEP 2: For each successive pair of points x_{k-1}, x_k found in step 1 calculate the integral A_k given by

$$A_k = \int (f(x) - g(x)) dx,$$

without worrying which of the functions is bigger than the other.

STEP 3: Now just add up the absolute values of the integrals calculated in step 2. Thus,

$$A = \int_a^b |f(x) - g(x)| dx = |A_1| + |A_2| + |A_3| + ... + |A_n|.$$

In particular, this applies to the calculation of $\int_a^b |f|$ by simply taking $g = 0$.

Notice that $|f(x) - g(x)|$ is the cross-sectional distance width of the region between the two curves f and g. Thus our formula says that the area of a region is the integral of its cross-sectional width. That is, if we have a region $\mathcal{R}$ in the plane for which we wish to find the area, we begin by imagining a straight line L in that plane as an $x-$axis. Now imagine a movable line perpendicular to L, which intersects L at the number value x. Call this perpendicular line $P(x)$. Then for each x we know $P(x)$ intersects the region $\mathcal{R}$ in a bunch of line segments whose total length we calculate and call it $w_{\mathcal{R}}(x)$. Thus $w_{\mathcal{R}}(x)$ is the cross-sectional width of the region $\mathcal{R}$ where the line perpendicular to the axis L meets L at x. We take a to be the minimum value of x such that $w_{\mathcal{R}}(x) > 0$ and take b to be the maximum value for which $w_{\mathcal{R}}(x) > 0$. Then

$$Area(\mathcal{R}) = \int_a^b w_{\mathcal{R}}(x) dx.$$

We can similarly apply the same idea to finding volumes of regions in 3-D space. Imagine an axis L which as a number line is the $x-$axis. Then for each point x on the number line L imagine a plane $P(x)$ perpendicular to the number line which intersects the region $\mathcal{R}$ in a planar region $\mathcal{R}(x)$, called the cross-section through x in L. Suppose that we can calculate the function

$$A_{\mathcal{R}}(x) = Area(\mathcal{R}(x))$$

which we call the cross-sectional area function. If $A_{\mathcal{R}}(x) = 0$ for all $x < a$ and for all $x > b$, then the volume of the region is $Vol(\mathcal{R})$ given by

$$Vol(\mathcal{R}) = \int_a^b A_{\mathcal{R}}(x)dx.$$

As examples, we calculated the area of a quarter circle using substitution and we calculated the volume of a hemisphere using the cross-sectional area function.

32. **LECTURE** WEDNESDAY 4 NOVEMBER 2009

Today we worked examples in preparation for TEST 3, including some of the problems on the practice test posted on my website.

33. **LECTURE** FRIDAY 6 NOVEMBER 2009

Today we reviewed for TEST 3.

34. **LECTURE** MONDAY 9 NOVEMBER 2009

Today we reviewed for TEST 3.

35. **LECTURE** WEDNESDAY 11 NOVEMBER 2009

Today we began discussing the computation of volume using area cross-section functions. In particular, we note that if a solid region $\mathcal{R}$ lies between two parallel planes and if we have an $x-$ axis perpendicular to these planes, then one will be say at $x = a$ and the other at say $x = b$, and then the parallel plane at a general x slices the region in a planar region $\mathcal{R}(x)$ whose area we denote $A(x) = Area(\mathcal{R}(x))$. Then the volume of the solid region $\mathcal{R}$ is simply

$$Vol(\mathcal{R}) = \int_a^b A(x)dx.$$

In particular, this shows that if two solids have equal cross-sectional area functions, then they have the same volume, a fact which was known to the Greeks at the time of Archimedes. We discussed the use of this to give integral representations for volumes of solids of revolution. We also discussed the related method of cylindrical shells and illustrated the use of both to compute the volume of a sphere.