Math 18.01 Lecture Summaries

Homework. These are the problems from the assigned Problem Set which can be completed using the material from that date's lecture.

Practice Problems. Practice problems are not to be written up or turned in. These are assigned only for practice, and are entirely voluntary. Problems listed as "1B-1", for example, are taken from Section E of the 18.01 course reader.

Lecture 1.	Sept.	8	Velocity and derivatives
Lecture 2.	Sept.	9	Limits
Lecture 3.	Sept.	13	Rules of differentiation
Lecture 4.	Sept.	15	The chain rule and implicit differentiation
Lecture 5.	Sept.	16	The derivatives of exponential and logarithm functions
Lecture 6.	Sept.	20	The derivatives of trigonometric functions
Lecture 7.	Sept.	22	Review for Exam 1
Lecture 8.	Sept.	27	Linear and quadratic approximations
Lecture 9.	Sept.	29	Sketching curves
Lecture 10.	Sept.	30	Applied maximum/minimum problems
Lecture 11.	Oct.	4	Related rates problems
Lecture 12.	Oct.	6	Newton's method
Lecture 13.	Oct.	13	Antidifferentiation
Lecture 14.	Oct.	14	Riemann integrals
Lecture 15.	Oct.	18	The Fundamental Theorem of Calculus
Lecture 16.	Oct.	20	Properties of the Riemann integral
Lecture 17.	Oct.	21	Separable ordinary differential equations
Lecture 18.	Oct.	25	Numerical integration
Lecture 19.	Oct.	28	Applications of integration to volumes
Lecture 20.	Nov.	1	Averages and volumes by shells
Lecture 21.	Nov.	3	Parametric equation curves and arc length
Lecture 22.	Nov.	4	Area of a surface of revolution and polar coordinate curves
Lecture 23.	Nov.	8	Tangent lines, arc length and areas for polar curves
Lecture 24.	Nov.	15	Inverse trigonometric functions and hyperbolic functions
Lecture 25.	Nov.	17	Inverse hyperbolic functions and inverse substitution
Lecture 26.	Nov.	18	Partial fraction decomposition
Lecture 27.	Nov.	22	Integration by parts

Lecture 28.	Dec.	1	L'Hospital's rule
Lecture 29.	Dec.	2	Improper integrals
Lecture 30.	Dec.	6	Sequences and series
Lecture 31.	Dec.	8	Power series and Taylor series
Lecture 32.	Dec.	8	More Taylor series and review

Lecture 1. September 8, 2005

Homework. Problem Set 1 Part I: (a)–(e); Part II: Problems 1 and 2.

Practice Problems. Course Reader: 1B-1, 1B-2
Textbook: p. 68, Problems 1–7 and 15.

1. Velocity. Displacement is $s(t)$. *Increment* from t_0 to $t_0 + \Delta t$ is,

$$\Delta s = s(t_0 + \Delta t) - s(t_0).$$

Average velocity from t_0 to $t_0 + \Delta t$ is,

$$v_{\text{ave}} = \frac{\Delta s}{\Delta t} = \frac{s(t_0 + \Delta t) - s(t_0)}{\Delta t}.$$

Velocity, or *instantaneous velocity*, at t_0 is,

$$v(t_0) = \lim_{\Delta t \to 0} v_{\text{ave}} = \lim_{\Delta t \to 0} \frac{s(t_0 + \Delta t) - s(t_0)}{\Delta t}.$$

This is a *derivative*, $v(t)$ equals $s'(t) = ds/dt$. The derivative of velocity is **acceleration**,

$$a(t_0) = v'(t_0) = \lim_{\Delta t \to 0} \frac{v(t_0 + \Delta t) - v(t_0)}{\Delta t}.$$

Example. For $s(t) = -5t^2 + 20t$, first computed velocity at $t = 1$ is,

$$v(1) = \lim_{\Delta t \to 0} 10 - 5\Delta t = \boxed{10.}$$

Then computed velocity at $t = t_0$ is,

$$v(t_0) = \lim_{\Delta t \to 0} -10t_0 + 10 - 5\Delta t = \boxed{-10t_0 + 20.}$$

Finally, computed acceleration at $t = t_0$ is,

$$a(t_0) = \lim_{\Delta t \to 0} -10 = \boxed{-10.}$$

2. Derivative. Let $y = f(x)$ be a *dependent variable* depending on an *independent variable* x, varying freely. The *increment* of y from x_0 to $x_0 + \Delta x$ is,

$$\Delta y = f(x_0 + \Delta x) - f(x_0).$$

The *difference quotient* or *average rate-of-change* of y from x_0 to $x_0 + \Delta x$ is,

$$\frac{\Delta y}{\Delta x} = \frac{f(x_0 + \Delta x) - f(x_0)}{\Delta x}.$$

The *derivative* of y (or $f(x)$) with respect to x at x_0 is,

$$\lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} = \lim_{\Delta x \to 0} \frac{f(x_0 + \Delta x) - f(x_0)}{\Delta x}.$$

3. Examples in science and math.

(i) Economics. *Marginal cost* is the derivative of cost with respect to some other variable, for instance, the quantity purchased.

(ii) Thermodynamics. The ideal gas law relating pressure p, volume V, and temperature T of a gas is,

$$pV = nRT.$$

Under isothermal conditions, T is a constant T_0 so that,

$$p(V) = \boxed{\tfrac{nRT_0}{V}.}$$

Under adiabatic conditions (i.e., no transfer of heat), pV^γ is a constant K. Using this to eliminate p gives,

$$T(V) = \boxed{\tfrac{K}{nR}\tfrac{1}{V^{\gamma-1}}.}$$

As this illustrates, the independent variable, dependent variable and constants in an equation very much depend on the problem to be solved.

(iii) Biology. Exponential population growth models the population $N(t)$ after t years as,

$$N(t) = N_0 e^{rt},$$

where e^x is the exponential function, N_0 is initial population, and r is a growth factor. Later we will see, $N'(t) = rN(t)$, i.e., the population grows at a rate proportional to the size of the population.

(iv) Geometry. The volume of a right circular cone is,

$$V = \frac{1}{3} A \times h.$$

where A is the base area of the cone and h is the height of the cone. The radius r of the base is proportional to the height,

$$r(h) = ch,$$

for some constant c. Since $A = \pi r^2$, this gives,

$$V(h) = \frac{\pi}{3}c^2h^3.$$

The derivative is,

$$\frac{dV}{dh} = \pi c^2 h^2 = \pi r^2 = \boxed{A.}$$

This is very reasonable. In some sense, this explains the classical formula for the volume of a cone.

Lecture 2. September 9, 2005

Homework. Problem Set 1 Part I: (f)–(h); Part II: Problems 3.

Practice Problems. Course Reader: 1C-2, 1C-3, 1C-4, 1D-3, 1D-5.

1. Tangent lines to graphs. For $y = f(x)$, the equation of the *secant line* through $(x_0, f(x_0))$ and $(x_0 + \Delta x, f(x_0 + \Delta x))$ is,

$$y = \frac{f(x_0 + \Delta x) - f(x_0)}{\Delta x}(x - x_0) + f(x_0).$$

In the limit, the equation of the *tangent line* through $(x_0, f(x_0))$ is,

$$y = f'(x_0)(x - x_0) + y_0.$$

Example. For the parabola $y = x^2$, the derivative is,

$$y'(x_0) = \boxed{2x_0.}$$

The equation of the tangent line is,

$$y = 2x_0(x - x_0) = \boxed{2x_0x - x_0^2.}$$

For instance, the equation of the tangent line through $(2, 4)$ is,

$$y = \boxed{4x - 4.}$$

Given a point (x, y), what are all points (x_0, x_0^2) on the parabola whose tangent line contains (x, y)? To solve, consider x and y as constants and solve for x_0. For instance, if $(x, y) = (1, -3)$, this gives,

$$(-3) = 2x_0(1) - x_0^2,$$

or,

$$x_0^2 - 2x_0 - 3 = 0.$$

Factoring $(x_0 - 3)(x_0 + 1)$, the solutions are x_0 equals -1 and x_0 equals 3. The corresponding tangent lines are,

$$y = -2x - 1,$$

and

$$y = 6x - 9.$$

For general (x, y), the solutions are,

$$x_0 = \boxed{x \pm \sqrt{x^2 - y}.}$$

2. Limits. Precise definition is on p. 791 of Appendix A.2. Intuitive definition: $\lim_{x \to x_0} f(x)$ equals L if and only if all values of $f(x)$ can be made arbitrarily close to L by choosing x sufficiently close to x_0. One interpretation is the "microscope/laser illuminator" analogy: An observer focuses a microscopes field-of-view on a thin strip parallel to the x-axis centered on $y = L$. The goal of the illuminator is to focus a laser-beam centered on x_0 parallel to the y-axis (but with the line $x = x_0$ deleted) so that only the portion of the graph in the field-of-view is illuminated. If for every magnification of the microscope, the illuminator can succeed, then the limit is defined and equals L.

There is a beautiful Java applet on the webpage of Daniel J. Heath of Pacific Lutheran University,

`http://www.plu.edu/~heathdj/java/calc1/Epsilon.html`

If you use this, try $a = -1$.

For left-hand limits, use a laser that illuminates only to the left of x_0. For right-hand limits, use a laser that illuminates only to the right of x_0.

3. Continuity. A function $f(x)$ is **continuous** at x_0 if $f(x_0)$ is defined, $\lim_{x \to x_0} f(x)$ is defined, and $\lim_{x \to x_0} f(x)$ equals $f(x_0)$. Also, $f(x)$ is continuous on an interval if it is continuous at every point of the interval. The types of discontinuity are: removable discontinuity, jump discontinuity, infinite discontinuity and essential discontinuity.

Lecture 3. September 13, 2005

Homework. Problem Set 1 Part I: (i) and (j).

Practice Problems. Course Reader: 1E-1, 1E-3, 1E-5.

1. Another derivative. Use the 3-step method to compute the derivative of $f(x) = 1/\sqrt{3x + 1}$ is,

$$f'(x) = \boxed{-3(3x + 1)^{-3/2}/2}.$$

Upshot: Computing derivatives by the definition is too much work to be practical. We need general methods to simplify computations.

2. The binomial theorem. For a positive integer n, the *factorial*,

$$n! = n \times (n-1) \times (n-2) \times \cdots \times 3 \times 2 \times 1,$$

is the number of ways of arranging n distinct objects in a line. For two positive integers n and k, the *binomial coefficient*,

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} = \frac{n(n-1)\cdots(n-k+2)(n-k+1)}{k(k-1)\cdots 3 \cdot 2 \cdot 1},$$

is the number of ways to choose a subset of k elements from a collection of n elements. A fundamental fact about binomial coefficients is the following,

$$\binom{n}{k-1} + \binom{n}{k} = \binom{n+1}{k}.$$

This is known as Pascal's formula. This link is to a webpage produced by MathWorld, part of Wolfram Research.

The *Binomial Theorem* says that for every positive integer n and every pair of numbers a and b, $(a+b)^n$ equals,

$$a^n + na^{n-1}b + \cdots + \binom{n}{k}a^{n-k}b^k + \cdots + nab^{n-1} + b^n.$$

This is proved by *mathematical induction*. First, the result is very easy when $n = 1$; it just says that $(a+b)^1$ equals $a^1 + b^1$. Next, make the *induction hypothesis* that the theorem is true for the integer n. The goal is to deduce the theorem for $n+1$,

$$(a+b)^{n+1} = a^{n+1} + (n+1)a^n b + \cdots + \binom{n+1}{k}a^{n+1-k}b^k + \cdots + (n+1)ab^n + b^{n+1}.$$

By the definition of the $(n+1)^{\text{st}}$ power of a number,

$$(a+b)^{n+1} = (a+b) \times (a+b)^n.$$

By the induction hypothesis, the second factor can be replaced,

$$(a+b)(a+b)^n = (a+b)\left(a^n + \cdots + \binom{n}{k}a^{n-k}b^k + \cdots + b^n\right).$$

Multiplying each term in the second factor first by a and then by b gives,

$$\begin{array}{ccccccccccccccc} a^{n+1} & + & na^n b & + & \dots & + & \binom{n}{k}a^{n+1-k}b^k & + & \binom{n}{k+1}a^{n-k}b^{k+1} & + & \dots & + & ab^n & & \\ & + & a^n b & + & \dots & + & \binom{n}{k-1}a^{n+1-k}b^k & + & \binom{n}{k}a^{n-k}b^{k+1} & + & \dots & + & nab^n & + & b^{n+1} \end{array}$$

Summing in columns gives,

$$a^{n+1} \;+\; (n+1)a^n b \;+\; \dots \;+\; (\tbinom{n}{k} + \tbinom{n}{k-1})a^{n+1-k}b^k \;+\; (\tbinom{n}{k+1} + \tbinom{n}{k})a^{n-k}b^{k+1} \;+\; \dots \;+\; (1+n)ab^n$$

Using Pascal's formula, this simplifies to,

$$a^{n+1} \quad + \quad (n+1)a^nb \quad + \quad \dots \quad + \quad \binom{n+1}{k}a^{n+1-k}b^k \quad + \quad \binom{n+1}{k+1}a^{n-k}b^{k+1} \quad + \quad \dots \quad + \quad (n+1)ab^n \quad + \quad b^{n+1}.$$

This proves the theorem for $n+1$, assuming the theorem for n.

Since we proved the theorem for $n = 1$, and since we also proved that for each integer n, the theorem for n implies the theorem for $n+1$, the theorem holds for every integer.

3. The derivative of x^n. Let $f(x) = x^n$ where n is a positive integer. For every a and every h, the binomial theorem gives,

$$f(a+h) = (a+h)^n = a^n + na^{n-1}h + \dots + \binom{n}{k}a^{n-k}h^k + \dots + h^n.$$

Thus, $f(a+h) - f(a)$ equals,

$$(a+h)^n - a^n = na^{n-1}h + \dots + \binom{n}{k}a^{n-k}h^k + \dots + h^n.$$

Thus the difference quotient is,

$$\frac{f(a+h) - f(a)}{h} = na^{n-1} + \binom{n}{2}a^{n-2}h + \dots + \binom{n}{k}a^{n-k}h^{k-1} + \dots + h^{n-1}.$$

Every summand except the first is divisible by h. The limit of such a term as $h \to 0$ is 0. Thus,

$$\lim_{h\to 0}\frac{f(a+h) - f(a)}{h} = na^{n-1} + 0 + \dots + 0 = na^{n-1}.$$

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

3. Linearity. For differentiable functions $f(x)$ and $g(x)$ and for constants b and c, $bf(x) + cg(x)$ is differentiable and,

$$(bf(x) + cg(x))' = bf'(x) + cg'(x).$$

This is often called *linearity* of the derivative.

4. The Leibniz rule/Product rule. For differentiable functions $f(x)$ and $g(x)$, the product $f(x)g(x)$ is differentiable and,

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

The crucial observation in proving this is rewriting the increment of $f(x)g(x)$ from a to $a+h$ as,

$$f(a+h)g(a+h) - f(a)g(a) = f(a+h)[g(a+h) - g(a)] + f(a+h)g(a) - f(a)g(a) = f(a+h)[g(a+h) - g(a)] + [f(a+h)$$

5. The quotient rule. Let $f(x)$ and $g(x)$ be differentiable functions. If $g(a)$ is nonzero, the quotient function $f(x)/g(x)$ is defined and differentiable at a, and,

$$(f(x)/g(x))' = [f'(x)g(x) - f(x)g'(x)]/g(x)^2.$$

One way to deduce this formula is to set $q(x) = f(x)/g(x)$ so that $f(x) = q(x)g(x)$, and the apply the Leibniz formula to get,

$$f'(x) = q'(x)g(x) + q(x)g'(x) = q'(x)g(x) + f(x)g'(x)/g(x).$$

Solving for $q'(x)$ gives,

$$q'(x) = [f'(x) - f(x)g'(x)/g(x)]/g(x) = [f'(x)g(x) - f(x)g'(x)]/g(x)^2.$$

6. Another proof that $d(x^n)/dx$ equals nx^{n-1}. This was mentioned only very briefly. The product rule also gives another induction proof that for every positive integer n, $d(x^n)/dx$ equals nx^{n-1}. For $n = 1$, we proved this by hand. Let n be some specific positive integer, and make the induction hypothesis that $d(x^n)/dx$ equals nx^{n-1}. The goal is to deduce the formula for $n + 1$,

$$\frac{d(x^{n+1})}{dx} = (n + 1)x^n.$$

By the Leibniz rule,

$$\frac{d(x^{n+1})}{dx} = \frac{d(x \times x^n)}{dx} = \frac{d(x)}{dx}x^n + x\frac{d(x^n)}{dx} = (1)x^n + x\frac{d(x^n)}{dx}.$$

By the induction hypothesis, the second term can be replaced,

$$\frac{d(x^{n+1})}{dx} = x^n + x(nx^{n-1}) = x^n + nx^n = (n + 1)x^n.$$

Thus the formula for n implies the formula for $n + 1$. Therefore, by mathematical induction, the formula holds for every positive integer n.

Lecture 4. September 15, 2005

Homework. No new problems.

Practice Problems. Course Reader: 1F-1, 1F-6, 1F-7, 1F-8.

1. Product rule example. For $u = \sqrt{3x + 1}$, what is $u'(x)$? Since $u \cdot u = 3x + 1$, $(u \cdot u)' = (3x + 1)' = 3$. By the product rule, $(u \cdot u)' = u' \cdot u + u \cdot u' = 2uu'$. Thus solving,

$$u'(x) = 3/(2u) = \boxed{3(3x + 1)^{-1/2}/2}.$$

2. The derivative of u^n. From above, $(u^2)'$ equals $2uu'$. By a similar computation, $(u^3)'$ equals $3u^2u'$. This suggests a pattern,

$$\frac{d(u^n)}{dx} = nu^{n-1}\frac{du}{dx}.$$

This can be proved by induction on n. For $n = 1, 2$ and 3, it was checked. Let n be a particular integer (for instance, 70119209472933054321). For that integer, suppose the result is known,

$$\frac{d(u^n)}{dx} = nu^{n-1}\frac{du}{dx}.$$

The goal is to prove the result for $n+1$, that is,

$$\frac{d(u^{n+1})}{dx} = (n+1)u^n\frac{du}{dx}.$$

Let $v = u^n$. Then u^{n+1} equals uv. So, by the product rule,

$$\frac{d(u^{n+1})}{dx} = \frac{d(uv)}{dx} = \frac{du}{dx}v + u\frac{dv}{dx}.$$

Plugging in $v = u^n$, this is,

$$\frac{d(u^{n+1})}{dx} = \frac{du}{dx}\cdot(u^n) + u\frac{d(u^n)}{dx}.$$

By the induction hypothesis, $d(u^n)/dx$ equals $nu^{n-1}(du/dx)$. Plugging in,

$$\frac{d(u^{n+1})}{dx} = \frac{du}{dx}\cdot(u^n) + u(nu^{n-1}\frac{du}{dx}).$$

This simplfies to,

$$\frac{d(u^{n+1})}{dx} = u^n\frac{du}{dx} + nu^n\frac{du}{dx} = (n+1)u^n\frac{du}{dx}.$$

Thus, the result for $n+1$ follows from the result for n. By induction, the result holds for every n.

3. The derivative of x^a, a a fraction. Let a be a fraction m/n and let $u(x)$ be x^a. Then u^n equals x^m. Thus,

$$\frac{d(u^n)}{dx} = \frac{d(x^m)}{dx},$$

which equals mx^{m-1}. By the above, $d(u^n)/dx$ equals $nu^{n-1}(du/dx)$. Thus,

$$nu^{n-1}\frac{du}{dx} = mx^{m-1}.$$

Solving for du/dx,

$$\frac{du}{dx} = \frac{mx^{m-1}}{nu^{n-1}} = \frac{mx^{m-1}}{n(x^{m/n})^{n-1}}.$$

One of the basic rules of exponents is that $(a^b)^c$ equals a^{bc}. Thus the denominator $n(x^{m/n})^{n-1}$ equals $nx^{m/n(n-1)}$, which equals $nx^{m-m/n}$. Thus,

$$\frac{du}{dx} = \frac{mx^{m-1}}{nx^{m-m/n}} = \frac{m}{n}x^{m-1}\cdot x^{m/n-m}.$$

Another basic rule of exponents is that $a^b \cdot a^c$ equals a^{b+c}. Thus,

$$\frac{du}{dx} = \frac{m}{n} x^{(m-1)+(m/n-m)} = \frac{m}{n} x^{m/n-1}.$$

Remembering that m/n is just a, and $u(x)$ is x^a, this finally gives,

$$\frac{d(x^a)}{dx} = ax^{a-1}.$$

4. The chain rule. Let y be a function of x, $y = f(x)$, and let u be a function of y, $u = g(y)$. Then u is a function of x, $u = g(f(x))$. This function is a **composite function**, and is denoted by,

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

What is the derivative of a composite function? The claim is that,

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

This is often easier to remember in the form,

$$\frac{du}{dx} = \frac{du}{dy} \cdot \frac{dy}{dx}.$$

This also suggests the proof,

$$(g \circ f)'(x_0) = \lim_{\Delta x \to 0} \frac{\Delta u}{\Delta x} = \lim_{\Delta x \to 0} \frac{\Delta u}{\Delta y} \cdot \frac{\Delta y}{\Delta x},$$

where y_0 equals $f(x_0)$, u_0 equals $g(y_0) = g(f(x_0))$, Δy equals $f(x_0+\Delta x) - f(x_0) = f(x_0+\Delta x) - y_0$, and Δu equals $g(y_0+\Delta y) - g(y_0) = g(f(x_0+\Delta x)) - g(f(x_0))$. So long as Δy is nonzero, the fraction in the limit is defined. And, as Δx approaches 0, also Δy approaches 0. Thus the limit breaks up as,

$$(g \circ f)'(x_0) = \lim_{\Delta y \to 0} \frac{\Delta u}{\Delta y} \cdot \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} = g'(y_0) \cdot f'(x_0).$$

Thus $(g \circ f)'(x_0)$ equals $g'(f(x_0))f'(x_0)$.

Example. Let $y(x)$ equals $1 + x^2$, and let $u(y)$ equal $1/y = y^{-1}$. Then $y'(x) = 0 + 2x = 2x$ and $u'(y) = -y^{-2}$. Thus, by the chain rule,

$$\frac{d}{dx}\left(\frac{1}{1+x^2}\right) = \frac{-1}{y^2}(2x) = \frac{-2x}{(1+x^2)^2}.$$

5. Implicit differentiation. This method has already been used many times. Given a function $y(x)$ satisfying some equation involving both x and y, formally differentiate each side of the equation with respect to x and then try to solve for y'.

Lecture 5. September 16, 2005

Homework. Problem Set 2 Part I: (a)–(e); Part II: Problem 2.

Practice Problems. Course Reader: 1I-1, 1I-4, 1I-5

1. Example of implicit differentiation. Let $y = f(x)$ be the unique function satisfying the equation,

$$\frac{1}{x} + \frac{1}{y} = 2.$$

What is slope of the tangent line to the graph of $y = f(x)$ at the point $(x, y) = (1, 1)$?

Implicitly differentiate each side of the equation to get,

$$\frac{d}{dx}\left(\frac{1}{x}\right) + \frac{d}{dx}\left(\frac{1}{y}\right) = \frac{d(2)}{dx} = 0.$$

Of course $(1/x)' = (x^{-1})' = -x^{-2}$. And by the rule $d(u^n)/dx = nu^{n-1}(du/dx)$, the derivative of $1/y$ is $-y^{-2}(dy/dx)$. Thus,

$$-x^{-2} - y^{-2}\frac{dy}{dx} = 0.$$

Plugging in x equals 1 and y equals 1 gives,

$$-1 - 1y'(1) = 0,$$

whose solution is,

$$y'(1) = \boxed{-1}.$$

In fact, using that $1/y$ equals $2 - 1/x$, this can be solved for every x,

$$\frac{dy}{dx} = (x^{-2})/(y^{-2}) = \frac{1}{x^2} \cdot \frac{1}{(2 - 1/x)^2} = \frac{1}{(2x - 1)^2}.$$

2. Rules for exponentials and logarithms. Let a be a positive real number. The basic rules of exponentials are as follows.

Rule 1. If a^b equals B and a^c equals C, then a^{b+c} equals $B \cdot C$, i.e.,

$$a^{b+c} = a^b \cdot a^c.$$

Rule 2. If a^b equals B and B^d equals D, then a^{bd} equals D, i.e.,

$$(a^b)^d = a^{bd}.$$

If a^b equals B, the *logarithm with base a* of B is defined to be b. This is written $\log_a(B) = b$. The function $B \to \log_a(B)$ is defined for all positive real numbers B. Using this definition, the rules of exponentiation become rules of logarithms.

Rule 1. If $\log_a(B)$ equals b and $\log_a(C)$ equals c, then $\log_a(B \cdot C)$ equals $b + c$, i.e.,

$$\log_a(B \cdot C) = \log_a(B) + \log_a(C).$$

Rule 2. If $\log_a(B)$ equals b and B^d equals D, then $\log_a(D)$ equals $d \log_a(B)$, i.e.,

$$\log_a(B^d) = d \log_a(B).$$

Rule 3. Since $\log_B(D)$ equals d, an equivalent formulation is $\log_a(D)$ equals $\log_a(B) \log_B(D)$, i.e.,

$$\log_a(D) = \log_a(B) \log_B(D).$$

3. The derivative of a^x. Let a be a positive real number. What is the derivative of a^x? Denote the derivative of a^x at $x = 0$ by $L(a)$. It equals the value of the limit,

$$L(a) = \lim_{h \to 0} \frac{a^h - 1}{h}.$$

Then for every x_0, the derivative of a^x at x_0 equals,

$$\lim_{h \to 0} \frac{a^{x_0+h} - a^{x_0}}{h}.$$

By Rule 1, a^{x_0+h} equals $a^{x_0}a^h$. Thus the limit factors as,

$$\lim_{h \to 0} \frac{a^{x_0}a^h - a^{x_0}}{h} = a^{x_0} \lim_{h \to 0} a^h - 1h.$$

Therefore, for every x, the derivative of a^x is,

$$\frac{d(a^x)}{dx} = L(a)a^x.$$

What is $L(a)$? To figure this out, consider how $L(a)$ changes as a changes. First of all,

$$L(a^b) = \lim_{h \to 0} \frac{(a^b)^h - 1}{h}.$$

By Rule 2, $(a^b)^h$ equals a^{bh}. So the limit is,

$$L(a^b) = \lim_{h \to 0} \frac{a^{bh} - 1}{h} = b \lim_{h \to 0} \frac{a^{bh} - 1}{bh}.$$

Now, inside the limit, make the substitution that k equals bh. As h approaches 0, also k approaches 0. So the limit is,

$$L(a^b) = b \lim_{k \to 0} \frac{a^k - 1}{k} = bL(a).$$

This is very similar to Rule 2 for logarithms.

Choose a number a_0 bigger than 1, say $a_0 = 2$. Then for every positive real number a, $a = a_0^b$ where $b = \log_{a_0}(a)$. Thus,

$$L(a) = L(a_0^b) = bL(a_0) = L(a_0)\log_{a_0}(a).$$

So, with a_0 fixed and a allowed to vary, $L(a)$ is just the logarithm function $\log_{a_0}(a)$ scaled by $L(a_0)$. Looking at the graph of $(a_0)^x$, it is geometrically clear that $L(a_0)$ is positive (though we have not *proved* that $L(a_0)$ is even defined). Thus the graph of $L(a)$ looks qualitatively like the graph of $\log_{a_0}(a)$. In particular, for a less than 1, $L(a)$ is negative. The value $L(1)$ equals 0. And $L(a)$ approaches $+\infty$ and a increases. Therefore, there must be a number where L takes the value 1. By long tradition, this number is called e;

$$L(e) = \lim_{h\to 0} \frac{e^h - 1}{h} = 1.$$

This is the definition of e. It sheds very little light on the decimal value of e.

Because e is so important, the logarithm with base e is given a special name: the **natural logarithm**. It is denote by,

$$\ln(a) = \log_e(a).$$

So, finally, $L(a)$ equals,

$$L(a) = \log_e(a)L(e) = \ln(a)(1) = \ln(a).$$

This leads to the formula for the derivative of a^x,

$$\frac{d(a^x)}{dx} = \boxed{\ln(a)a^x.}$$

In particular,

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

In fact, e^x is characterized by the property above and the property that e^0 equals 1.

4. The derivative of $\log_a(x)$ and the value of e. By the chain rule,

$$\frac{d(a^u)}{dx} = \ln(a)a^u\frac{du}{dx}.$$

For $u = \log_a(x)$, a^u equals x. Thus,

$$\frac{d(a^u)}{dx} = \frac{d(x)}{dx} = 1.$$

Thus,

$$\ln(a)a^u\frac{du}{dx} = 1.$$

Solving gives,

$$\frac{d\log_a(x)}{dx} = \frac{1}{\ln(a)}\frac{1}{a^u} = \boxed{1/(\ln(a)x)}.$$

In particular, for $a = e$, this gives,

$$\frac{d\ln(x)}{dx} = \boxed{1/x}.$$

What is the derivative of $\ln(x)$ at $x = 1$? On the one hand, since the derivative of $\ln(x)$ equals $1/x$, the derivative at $x = 1$ is $1/1 = 1$. On the other hand, the definition of the derivative gives,

$$\lim_{h\to 0} \frac{\ln(1+h) - \ln(1)}{h}.$$

Of course, $\ln(1)$ equals 0, so this simplifies to,

$$\lim_{h\to 0} \frac{1}{h}\ln(1+h).$$

Using Rule 2 for logarithms, this gives,

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

Since $\ln(y)$ is continuous, the limit equals,

$$\ln[\lim_{h\to 0}(1+h)^{1/h}].$$

So the natural logarithm of the inner limit equals 1. But e is the unique number whose natural logarithm equals 1. This leads to the formula,

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

Making the substitution $n = 1/h$ leads to the more familiar form,

$$\lim_{n\to +\infty}(1+1/n)^n = \boxed{e}.$$

This can be used to compute e to arbitrary accuracy. The first few digits of e are 2.718281828459045...

5. Logarithmic differentiation. There is a method of computing derivatives of products of functions that is often useful. If y is a product of n factors, say $f_1(x) \cdot f_2(x) \cdot \dots \cdot f_n(x)$, the derivative of y can be computed by the product rule. However, it seems to be a fact that multiplication is more error-prone than addition. Thus introduce,

$$u = \ln(y) = \ln(f_1(x)) + \ln(f_2(x)) + \dots + \ln(f_n(x)).$$

The derivative of u is,
$$\frac{du}{dx} = \frac{d}{dx}(\ln(f_1(x))) + \cdots + \frac{d}{dx}(\ln(f_n(x))).$$
Using the chain rule, this is,
$$\frac{du}{dx} = \frac{f_1'(x)}{f_1(x)} + \cdots + \frac{f_n'(x)}{f_n(x)}.$$
Thus, far fewer multiplications are needed to compute u'. This is good, because also,
$$\frac{du}{dx} = \frac{d\ln(y)}{dx} = \frac{1}{y}\frac{dy}{dx}.$$
Therefore the derivative of y can be computed as,
$$y' = yu' = (f_1(x) \cdot \cdots \cdot f_n(x)) \left(\frac{f_1'(x)}{f_1(x)} + \cdots + \frac{f_n'(x)}{f_n(x)}\right).$$

Example. Let y be,
$$\frac{(1+x^3)(1+\sqrt{x})}{x^{3/7}}.$$
Then,
$$u = \ln(y) = \ln(1+x^3) + \ln(1+\sqrt{x}) - \frac{3}{7}\ln(x).$$
By the chain rule, $\ln(1+x^3)' = 3x^2/(1+x^3)$ and $\ln(1+\sqrt{x})' = (\sqrt{x})'/(1+\sqrt{x}) = (1/2x^{-1/2})/(1+\sqrt{x})$. Thus, u' equals,
$$u' = \frac{3x^2}{(1+x^3)} + \frac{1}{2\sqrt{x}(1+\sqrt{x})} - \frac{3}{7x}.$$
So, finally,
$$y' = yu' = \frac{(1+x^3)(1+\sqrt{x})}{x^{3/7}}\left(\frac{3x^2}{(1+x^3)} + \frac{1}{2\sqrt{x}(1+\sqrt{x})} - \frac{3}{7x}\right).$$

Lecture 6. September 20, 2005

Homework. Problem Set 2 Part I: (f)–(j); Part II: Problems 1, 3 and 4.

Practice Problems. Course Reader: 1J-1, 1J-2, 1J-3, 1J-4

1. Trigonometric functions. What is *angle*? For a sector of a unit circle (a circle of radius 1), the *angle* of the sector equals both the length of the arc of the sector and 1/2 the area of the sector. Although we have as yet *general definitions* of neither arc length nor area, this can be used to give a rigorous definition of angle. We *can* divide any sector in two equal pieces: simply bisect the chord of the sector. We also know how to add two angles, by laying the sectors in adjacent positions. Denoting the area of a unit circle by the symbol π (which happens to be the familiar π), these 2 operations produce every angle of the form $m\pi/2^n$, with m and n integers. Every angle can be approximated arbitrarily well by such angles. Thus, for every *continuous* function of an angle, every value of the function can be computed.

The basic functions are $\sin(\theta)$, $\cos(\theta)$, $\tan(\theta)$, $\sec(\theta)$, $\csc(\theta)$ and $\cot(\theta)$. Full descriptions of these are in §9.1 of the textbook by Simmons. The same information is contained in the webpage on Trigonometry produced by MathWorld, part of Wolfram Research.

2. Trigonometric identities. For today, the most important identities are the *angle addition formulas*,

$$\sin(\alpha+\beta) = \boxed{\sin(\alpha)\cos(\beta)+\cos(\alpha)\sin(\beta),}$$

$$\cos(\alpha+\beta) = \boxed{\cos(\alpha)\cos(\beta)-\sin(\alpha)\sin(\beta).}$$

Other important identities are,

(i) $\cos(-\theta)$ equals $\cos(\theta)$, i.e., $\cos(\theta)$ is an *even function*,

(ii) $\sin(-\theta)$ equals $-\sin(\theta)$, i.e., $\sin(\theta)$ is an *odd function*,

(iii) $\sin(\theta+\pi/2)$ equals $\cos(\theta)$,

(iv) $\cos(\theta+\pi/2)$ equals $-\sin(\theta)$, and

(v) $\sin^2(\theta)+\cos^2(\theta)$ equals 1 for every θ.

3. Some trigonometric limits. In computing trigonometric limits, the following limit is crucial,

$$\lim_{\theta\to 0}\frac{\sin(\theta)}{\theta} = \boxed{1.}$$

As explained in class, this is essentially the statement that as $\theta \to 0$, the quotient of the arc length by the chord length tends to 1. This was not proved in lecture, nor is it proved in your textbook in §2.1 (despite the author's claim). However, it is geometrically reasonable. And, of course, it can be proved.

This limit implies another limit,

$$\lim_{\theta\to 0}\frac{\cos(\theta)-1}{\theta} = \boxed{0.}$$

To see this, rewrite the term as,

$$\frac{\cos(\theta)-1}{\theta}\frac{\cos(\theta)+1}{\cos(\theta)+1} = \frac{\cos^2(\theta)-1}{\theta\cdot(\cos(\theta)+1)}.$$

By Identity (v), $\cos^2(\theta)-1$ equals $-\sin^2(\theta)$, so the term equals,

$$\frac{-\sin^2(\theta)}{\theta\cdot(\cos(\theta)+1)} = -\frac{\sin(\theta)}{\theta}\frac{1}{\cos(\theta)+1}\sin(\theta).$$

As $\theta \to 0$, this limit tends to,

$$-(1) \times (1/2) \times 0 = 0.$$

By a similar computation,

$$\lim_{\theta \to 0} \frac{\cos(\theta) - 1}{\theta^2} = \frac{-1}{2}.$$

4. Derivatives of $\sin(x)$ **and** $\cos(x)$**.** To compute the derivative of $y = \sin(x)$ at $x = a$, use the angle addition formulas to write,

$$\sin(a + h) = \sin(a)\cos(h) + \cos(a)\sin(h).$$

This gives,

$$\sin(a + h) - \sin(a) = \sin(a)(\cos(h) - 1) + \cos(a)\sin(h).$$

Thus the difference quotient equals,

$$\frac{\sin(a + h) - \sin(a)}{h} = \sin(a)\frac{\cos(h) - 1}{h} + \cos(a)\frac{\sin(h)}{h}.$$

Taking the limit gives,

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

Using the limits from above, this gives,

$$\sin'(a) = \sin(a) \times 0 + \cos(a) \times 1 = \cos(a).$$

Thus the derivative of $\sin(x)$ equals,

$$\frac{d\sin(x)}{dx} = \boxed{\cos(x).}$$

An entirely similar computation gives,

$$\frac{\cos(a + h) - \cos(a)}{h} = \cos(a)\frac{\cos(h) - 1}{h} - \sin(a)\frac{\sin(h)}{h},$$

which leads to,

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

Thus the derivative of $\cos(x)$ equals,

$$\frac{d\cos(x)}{dx} = \boxed{-\sin(x).}$$

5. Derivatives of other trigonometric functions. Using the quotient rule,

$$\frac{d\tan(x)}{dx} = \frac{1}{\cos^2(x)}(\cos(x)\times\cos(x) - \sin(x)(-\sin(x))) = \frac{\cos^2(x)+\sin^2(x)}{\cos^2(x)} = \frac{1}{\cos^2(x)}.$$

Therefore, the derivative of $\tan(x)$ equals,

$$\frac{d\tan(x)}{dx} = \boxed{\sec^2(x).}$$

In a similar manner,

$$\frac{d\cot(x)}{dx} = \boxed{-\csc^2(x),}$$

$$\frac{d\sec(x)}{dx} = \boxed{\sec(x)\tan(x),}$$

and

$$\frac{d\csc(x)}{dx} = \boxed{-\csc(x)\cot(x).}$$

Lecture 7. September 22, 2005

Review for Exam 1. No new material was presented. There were no practice problems from the course reader.

Lecture 8. September 27, 2005

Homework. Problem Set 2 all of Part I and Part II.

Practice Problems. Course Reader: 2A-1, 2A-4, 2A-9, 2A-11, 2A-12.

1. Linear approximations. For a differentiable function $f(x)$, the *linear approximation* or *linearization* of $f(x)$ at $x = a$ is the linear function,

$$f(a) + f'(a)(x-a).$$

In a precise sense, this is the best approximation of $f(x)$ by a linear function near $x = a$. For x close to a, the value of $f(x)$ is close to the value of the linearization. The notation for this is,

$$f(x) \approx f(a) + f'(a)(x-a) \text{ for } x \approx a.$$

Example. The linearization of,

$$f(x) = e^{-3x}\sin(2\pi x) + 5e^{-3x}\cos(2\pi x),$$

near $x = 0$ is,

$$f(x) \approx 5 - (15 - 2\pi)x \text{ for } x \approx 0.$$

In particular, for $x = 0.02$, this gives the approximate answer,

$$f(0.02) \approx 5 - (15 - 2\pi)(0.02) \approx \boxed{4.8.}$$

The actual value is approximately 4.71.

2. Basic approximations. Some linear approximations occur so often, they should be committed to memory. Each of the following is the linear approximation for $x \approx 0$, together with the terms in the quadratic and higher approximations.

$$\begin{aligned}
\tfrac{1}{1-x} &\approx \boxed{1 + x} + x^2 + x^3 + \dots, \\
(1+x)^r &\approx \boxed{1 + rx} + \tbinom{r}{2}x^2 + \tbinom{r}{3}x^3 + \dots, \\
\sin(x) &\approx \boxed{x} - x^3/3! + x^5/5! + \dots, \\
\cos(x) &\approx \boxed{1} - x^2/2! + x^4/4! + \dots, \\
e^x &\approx \boxed{1 + x} + x^2/2! + x^3/3! + \dots, \\
\ln(1+x) &\approx \boxed{1 - x} + x^2/2 - x^3/3 + \dots
\end{aligned}$$

3. Combining basic approximations. The basic approximations can be combined to get new linear approximations.

(i) The linear approximation of $f(x)$ for $x \approx a$ can be converted to a linear approximation at 0 by setting $g(u) = f(a + u)$. In symbols,

$$\boxed{f(a) + f'(a)(x - a) = g(0) + g'(0)u.}$$

This is equivalent to the formula,

$$\frac{d}{dx}(f(x-a)) = \frac{df}{dx}(x-a).$$

(ii) The linear approximation of $f(cx)$ for $x \approx a$ is obtained from the linear approximation of $f(u)$ for $u \approx ca$ by substituting $u = cx$,

$$\boxed{f(cx) \approx f(ca) + f'(ca)(cx - ca).}$$

This is equivalent to the formula,

$$\frac{d}{dx}(f(cx)) = c\frac{df}{dx}(x).$$

(iii) The linear approximation of $cf(x)$ for $x \approx a$ is c times the linear approximation of $f(x)$ for $x \approx a$,

$$\boxed{cf(x) \approx cf(a) + cf'(a)(x-a)}.$$

This is different than the previous rule. Also, the linear approximation of $f(x) + g(x)$ for $x \approx a$ is the sum of the linear approximations of $f(x)$ and $g(x)$,

$$\boxed{(f+g)(x) \approx f(a) + g(a) + (f'(a) + g'(a))(x-a).}$$

Together, these two rules are equivalent to the formulas,

$$\frac{d}{dx}(cf(x)) = c\frac{df}{dx}(x), \quad \frac{d}{dx}(f(x) + g(x)) = \frac{df}{dx}(x) + \frac{dg}{dx}(x).$$

(iv) The linear approximation of $f(x)g(x)$ for $x \approx a$ is the product of the linear approximations, disregarding all quadratic terms,

$$f(x)g(x) \approx (f(a) + f'(a)(x-a))(g(a) + g'(a)(x-a)),$$

which simplifies to,

$$\boxed{f(x)g(x) \approx f(a)g(a) + (f'(a)g(a) + f(a)g'(a))(x-a).}$$

This is equivalent to Leibniz's rule,

$$\frac{d}{dx}(f(x)g(x)) = \frac{df}{dx}(x)g(x) + f(x)\frac{dg}{dx}(x).$$

(v) The linear approximation of $f(x)/g(x)$ for $x \approx a$ is the quotient of the linear approximations, using the linear approximation $1/(1-x) \approx 1 + x$,

$$\frac{f(x)}{g(x)} \approx \frac{f(a) + f'(a)(x-a)}{g(a) + g'(a)(x-a)} = (f(a) + f'(a)(x-a))\frac{1}{g(a)}\frac{1}{1 - (-g'(a)(x-a)/g(a))} \approx$$

$$(f(a) + f'(a)(x-a))\frac{1}{g(a)}(1 - g'(a)(x-a)/g(a)) = \frac{1}{g(a)^2}(f(a) + f'(a)(x-a))(g(a) - g'(a)(x-a)).$$

This simplifies to,

$$\boxed{f(x)/g(x) \approx f(a)/g(a) + (1/g(a)^2)(f'(a)g(a) - f(a)g'(a))(x-a).}$$

This is equivalent to the quotient rule,

$$\frac{d}{dx}(f(x)/g(x)) = \frac{1}{g(x)}(\frac{df}{dx}(x)g(x) - f(x)\frac{dg}{dx}(x)).$$

(vi) The linear approximation of $g(f(x))$ for $x \approx a$ is obtained from the linear approximation of $g(u)$ for $u \approx f(a)$ by substituting in for u the linear approximation of $f(x)$ for $x \approx a$ and ignoring quadratic terms,

$$u = f(x) \approx f(a) + f'(a)(x-a),$$

$$g(f(x)) = g(u) \approx g(f(a)) + g'(f(a))(u - f(a)) \approx g(f(a)) + g'(f(a))((f(a) + f'(a)(x-a)) - f(a)).$$

This simplifes to,

$$g(f(x)) \approx g(f(a)) + g'(f(a))f'(a)(x-a).$$

This is equivalent to the chain rule,

$$\frac{d}{dx}(g(f(x))) = \frac{dg}{dx}(f(x))\frac{df}{dx}(x).$$

Together, these 6 rules account for all the general rules we have regarding differentiation. So every rule of differentiation has an equivalent formulation in terms of linear approximations.

Example. Using the rules, the linear approximation for,

$$f(x) = e^{-3x}\sin(2\pi x) + 5e^{-3x}\cos(2\pi x),$$

for $x \approx 0$ is given by,

$$(1 + (-3x))(2\pi x) + 5(1 + (-3x))(1) = 2\pi x + 5 - 15x,$$

which simplifies to,

$$f(x) \approx 5 - (15 - 2\pi)x.$$

4. Quadratic approximations. Sometimes the linear approximation is not good enough. One example is the linear approximation of $\cos(x)$ as 1 for $x \approx 0$. The linear approximation gives no idea whether $\cos(x)$ is greater than 1, less than 1, concave up, concave down, etc. This is remedied by the *quadratic approximation*,

$$f(x) \approx f(a) + f'(a)(x-a) + \tfrac{1}{2}f''(a)(x-a)^2 \text{ for } x \approx a.$$

Each of the basic approximations has an analogous quadratic approximation. Each of the rules for combining linear approximations has an analogous rule for quadratic approximations.

5. The mean value theorem. This was discussed only very briefly. If a function $f(x)$ is differentiable on the interval having a and b as endpoints, then there is a point c strictly between a and b so that the slope of the tangent line to $y = f(x)$ at $x = c$ equals the slope of the secant line to $y = f(x)$ containing $(a, f(a))$ and $(b, f(b))$,

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

This is sometimes useful for bounding $f(b) - f(a)$, if a bound on the derivative of $f(x)$ is known.

Lecture 9. September 29, 2005

Homework. Problem Set 2 all of Part I and Part II.

Practice Problems. Course Reader: 2B-1, 2B-2, 2B-4, 2B-5.

1. Application of the Mean Value Theorem. A real-world application of the Mean Value Theorem is *error analysis*. A device accepts an input signal x and returns an output signal y. If the input signal is always in the range $-1/2 \leq x \leq 1/2$ and if the output signal is,

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

what precision of the input signal x is required to get a precision of $\pm 10^{-3}$ for the output signal?

If the ideal input signal is $x = a$, and if the *precision* is $\pm h$, then the actual input signal is in the range $a - h \leq x \leq a + h$. The precision of the output signal is $|f(x) - f(a)|$. By the Mean Value Theorem,

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

for some c between a and x. The derivative $f'(x)$ is,

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

For $-1/2 \leq x \leq 1/2$, this is bounded by,

$$|f'(x)| \leq \frac{3(1/2)^2 + 2(1/2) + 1}{[1 + (-1/2) + (-1/2)^2 + (-1/2)^3]^2} = 7.04.$$

Thus the Mean Value Theorem gives,

$$|f(x) - f(a)| = |f'(c)||x - a| \leq 7.04|x - a| \leq 7.04h.$$

Therefore a precision for the input signal of,

$$h = \boxed{10^{-3}/7.04} \approx 10^{-4}$$

guarantees a precision of 10^{-3} for the output signal.

2. First derivative test. A function $f(x)$ is *increasing*, respectively *decreasing*, if $f(a)$ is less than $f(b)$, resp. greater than $f(b)$, whenever a is less than b. In symbols, f is increasing, respectively decreasing, if

$$f(a) < f(b) \text{ whenever } a < b, \text{ resp. } f(a) > f(b) \text{ whenever } a < b.$$

If $f(a)$ is less than or equal to $f(b)$, resp. greater than or equal to $f(b)$, whenever a is less than b, then $f(x)$ is *non-decreasing*, resp. *non-increasing*. If $f(x)$ is increasing, the graph rises to the right. If $f(x)$ is decreasing, the graph rises to the left.

If $f'(a)$ is positive, the *First Derivative Test* guarantees that $f(x)$ is increasing for all x sufficiently close to a. If $f'(a)$ is negative, the First Derivative Test guarantees that $f(x)$ is decreasing for all x sufficiently close to a.

Example. For the function $y = x^3 + x^2 - x - 1$, determine where y is increasing and where y is decreasing.

The derivative is,

$$y' = 3x^2 + 2x - 1 = (3x - 1)(x + 1).$$

Thus the derivative of y changes sign only at the points $x = -1$ and $x = 1/3$. By testing random elements, y' is positive for $x > 1/3$, it is negative for $-1 < x < 1/3$, and it is positive for $x < -1$. Therefore, by the First Derivative Test, y is increasing for $x < -1$, y is decreasing for $-1 < x < 1/3$, and y is increasing for $x > 1/3$.

3. Extremal points. If $f(x) \leq f(a)$ for all x near a, then x is a *local maximum*. If $f(x) \geq f(a)$ for all x near a, then x is a *local minimum*. Because of the First Derivative Test, if $f'(a) > 0$ and f is defined to the right of a, the graph of f rises to the right of a. Thus a is not a local maximum. Similarly, if $f'(a) < 0$ and f is defined to the left of a, the graph of f rises to the left of a. Thus a is not a local maximum. In particular, if f is defined to both the right and left of a, if $f'(a)$ is defined, and if a is a local maximum, then $f'(a)$ equals 0. Similarly, if f is defined to both the right and left of a, if $f'(a)$ is defined, and if a is a local minimum, then $f'(a)$ equals 0.

A point a where $f'(a)$ is defined and equals 0 is a *critical point*. By the last paragraph, if $x = a$ is a local maximum of f, respectively a local minimum of f, then one of the following holds.

(i) The function $f(x)$ is discontinuous at a.

(ii) The function $f(x)$ is continuous at a, but $f'(a)$ is not defined.

(iii) The point a is a left endpoint of the interval where f is defined, and $f'(a) \leq 0$, resp. $f'(a) \geq 0$.

(iv) The point a is a right endpoint of the interval where f is defined, and $f'(a) \geq 0$, rexp. $f'(a) \leq 0$.

(v) The function f is defined to the left and right of a, and $f'(a)$ equals 0. In other words, a is a critical point of f.

Example. For the function $y = x^3 + x^2 - x - 1$, the critical points are $x = -1$ and $x = 1/3$. By examining where y is increasing and decreasing, $x = -1$ is a local maximum and $x = 1/3$ is a local minimum.

The plurals of "maximum" and "minimum" are "*maxima*" and "*minima*". Together, local maxima and local minima are called *extremal points*, or *extrema*. These are points where f takes on an extreme value, either positive or negative. A point where f achieves its maximum value among *all* points where f is defined is a *global maximum* or *absolute maximum*. A point where f achieves its minimum value among *all* points where f is defined is a *global minimum* or *absolute minimum*.

4. Concavity and the Second Derivative Test. For a differentiable function f, every "interior" extremal point is a critical point of f. But not every critical point of f is an extremal point.

Example. The function $f(x) = x^3$ has a critical point at $x = 0$. But $f(x)$ is everywhere increasing, thus $x = 0$ is not an extremal point of f.

When is a critical point an extremal point? When is it a local maximum? When is it a local minimum? This is closely related to the *concavity* of f. A function $f(x)$ is *concave up*, respectively *concave down*, if no secant line segment to $f(x)$ crosses below the graph of f, resp. above the graph of f. In symbols, f is concave up, resp. concave down, if

$$(f(c) - f(a))/(c - a) \leq (f(b) - f(a))/(b - a) \text{ whenever } a < c < b,$$

$$\text{resp. } (f(c) - f(a))/(c - a) \geq (f(b) - f(a))/(b - a) \text{ whenever } a < c < b.$$

For a differentiable function f, this equation is close to,

$$f'(c) \leq f'(b) \text{ whenever } a < c < b,$$

$$\text{resp. } f'(c) \geq f'(b) \text{ whenever } a > c > b.$$

This precisely says that f' is non-decreasing, resp. f' is non-increasing. If f' is non-decreasing, resp. non-increasing, then f is concave up, resp. concave down. Applying the First Derivative Test to determine when f' is increasing, resp. decreasing, gives the *Second Derivative Test*: If $f''(a) > 0$, then f is concave up near $x = a$; if $f''(a) < 0$ then f is concave down near $x = a$.

If f is concave up near a critical point, the critical point is a local minimum. If f is concave down near a critical point, the critical point is a local maximum. Combined with the Second Derivative Test, this gives a test for when a critical point is a local maximum or local minimum: If $f'(a)$ equals 0 and $f''(a) < 0$, then $x = a$ is a local maximum. If $f'(a)$ equals 0 and $f''(a) > 0$, then $x = a$ is a local minimum.

Example. For $y = x^3 + x^2 - x - 1$, the second derivative is $y'' = 6x + 2$. Since $y''(-1) = -4$ is negative, the critical point $x = -1$ is a local maximum. Since $y''(1/3) = 4$ is positive, $x = 1/3$ is a local minimum.

5. Inflection points. If f is differentiable, but for every neighborhood of a, f is neither concave up nor concave down on the entire neighborhood, then a is an *inflection point*. If $f''(a)$ is defined, the Second Derivative Test says that $f''(a)$ must equal 0. Except in pathological cases, an inflection point is a point where f is concave up to one side of f, and concave down to the other side of f.

Example. For $y = x^3 + x^2 - x - 1$, the second derivative $y'' = 6x + 2$ is negative for $x < -1/3$ and is positive for $x > 1/3$. By the Second Derivative Test, y is concave down for $x < -1/3$ and y is concave up for $x > -1/3$. Therefore $x = -1/3$ is an inflection point for y.

Lecture 10. September 30, 2005

Homework. Problem Set 3 Part I: (a)–(f). Part II: Problems 1, 2 and 3.

Practice Problems. Course Reader: 2C-5, 2C-10, 2C-12, 2D-3, 2D-4.

1. Asymptotes. An *asymptote* describes the behavior of the graph of $y = f(x)$ as it becomes unbounded, in some sense. There are two main examples. The function f has a *vertical asymptote* $x = a$ if at least 1 of the following holds,

$$\lim_{x \to a^-} f(x) = +\infty, \lim_{x \to a^-} f(x) = -\infty, \lim_{x \to a^+} f(x) = +\infty, \lim_{x \to a^+} f(x) = -\infty.$$

In each case, the graph of $y = f(x)$ becomes unbounded, and becomes arbitrarily close to the line $x = a$. If $x = a$ is a vertical asymptote, then $f(x)$ has an infinite discontinuity at $x = a$.

The function f has a *horizontal asymptote* $y = b$ if at least 1 of the following holds,

$$\lim_{x \to +\infty} f(x) = b, \lim_{x \to -\infty} f(x) = b.$$

In other words, the graph of $y = f(x)$ becomes arbitrarily close to the line $y = b$ as x approaches either $+\infty$ or $-\infty$.

Example. For the function $y = (x^3 + x)/(x^2 - 1) = x(x^2 + 1)/(x^2 - 1)$, the lines $x = -1$ and $x = -1$ are vertical asymptotes. There is no horizontal asymptote. However, the graph of y is *asymptotic* to the line $y = x$. This was not discussed in lecture. A pair of functions f and g are *asymptotic* to each other if the line $y = 0$ is a horizontal asymptote of $f - g$.

2. Applied maximum/minimum problems. Using the First Derivative Test, the maximum and minimum of many functions can be computed. This is very important in applications.

Example. Two long walls meet at right angles making a corner. Using a length of 10 meters of fence to form the other 2 sides of a rectangle, what is the largest area that can be enclosed in this corner?

Step 1. Identify parameters. A *parameter* is a constant or variable. The constant in this problem is 10 meters. Two variables are the length l of one side of the rectangle, and the width w of the remaining side of the rectangle.

Step 2. Draw a diagram. This was done in lecture.

Step 3. Find the quantity to be maximized or minimized. The quantity to be maximized is the area A of the rectangle. Since the area is the product of the length and width, A equals lw.

Step 4. Use the constraints to eliminate variables. The constraint is that the total length of fence is 10 meters. Thus $l + w$ equals 10. This is used to eliminate w,

$$w = 10 - l.$$

Making this substitution, A is now a function of l alone,

$$A(l) = lw(l) = l(10 - l) = -l^2 + 10l.$$

Step $4\frac{1}{2}$. Sketch a graph of the quantity to be maximized or minimized. This is not absolutely necessary. Sometimes it is impossible. When you can make a rough sketch, this will typically give a very good idea where the maximum or minimum lies. In the example above, $A(l)$ is a quadratic equation. Because both l and w must be nonnegative, $A(l)$ is only defined on the interval $0 \leq l \leq 10$. Thus the graph of $A(l)$ is a segment of a parabola opening down. The vertex of the parabola is contained in the segment. Thus the vertex is the maximum.

Step 5. Compute the derivative. In this case,

$$A'(l) = -2l + 10.$$

Step 6. Find all critical points, endpoints, discontinuity points, etc. In most cases, it suffices to find all critical points and endpoints. Occasionally it is also necessary to find all points where f' is not defined. Rarely it is necessary to also consider discontinuity points (although this is usually so obvious that it does not require a separate step). In this case, the endpoints are $l = 0$ and $l = 10$. The one critical point is $l = 5$.

Step 7. Determine the global maximum or minimum. Checking all critical points, endpoints, etc., determine the global maximum or the global minimum. In this case, $A(0)$ equals 0, $A(10)$ equals 0 and $A(5)$ equals 25. Thus $l = 5$ is the global maximum.

Step 8. Back-substitute. Plug in the value of the single remaining independent variable to determine the values of the remaining independent variables. In this case, w equals $10 - l$, which is $10 - 5 = 5$ for $l = 5$. Thus, the largest area 25 is enclosed by a square of side length 5.

Example. A swimmer is in the water at a distance b_1 meters from shore. She wants to reach a point on land b_2 meters from the water. The point is a meters parallel to the shore. If the swimmer swims v_1 meters per second and runs v_2 meters per second, at what distance x from the closest point on shore should she aim to minimize her time to the target? Mathematically, the swimmer is at point $(0, b_1)$ and wants to reach point $(a, -b_2)$, where the shore is the x-axis. At what point $(x, 0)$ should she aim?

The constants are a, b_1, b_2, v_1 and v_2. The variable is x. It is also convenient to introduce a variable d_1 for the distance from $(0, b_1)$ to $(x, 0)$, and a variable d_2 for the distance from $(x, 0)$ to $(a, -b_2)$. Although not obvious, it is also very convenient to introduce a variable θ_1 for the acute angle formed by the x-axis and the line segment joining $(0, b_1)$ to $(x, 0)$. Also introduce θ_2 for the acute angle formed by the x-axis and the line segment joining $(x, 0)$ to $(a, -b_2)$.

The time T_1 to swim to point $(x, 0)$ is,

$$T_1 = \frac{d_1}{v_1} = \frac{1}{v_1}(x^2 + b_1^2)^{1/2}.$$

The time T_2 to run from $(x, 0)$ to point $(a, -b_2)$ is,

$$T_2 = \frac{d_2}{v_2} = \frac{1}{v_2}((a - x)^2 + b_2^2)^{1/2}.$$

Thus the total time to reach the target is,

$$T = T_1 + T_2 = \frac{1}{v_1}(x^2 + b_1^2)^{1/2} + \frac{1}{v_2}((a-x)^2 + b_2^2)^{1/2}.$$

The derivative of T with respect to x is,

$$\frac{dT}{dx} = \frac{1}{v_1}\left(\frac{1}{2}(x^2 + b_1^2)^{-1/2}(2x)\right) + \frac{1}{v_2}\left(\frac{1}{2}((a-x)^2 + b_2^2)^{-1/2}(-2(a-x))\right).$$

This simplifies to,

$$\frac{dT}{dx} = \frac{x}{v_1 d_1} - \frac{a-x}{v_2 d_2}.$$

Observe that x/d_1 equals $\sin(\theta_1)$ and $(a-x)/d_2$ equals $\sin(\theta_2)$. Thus,

$$\frac{dT}{dx} = \frac{\sin(\theta_1)}{v_1} - \frac{\sin(\theta_2)}{v_2}.$$

Technically, there are no endpoints. However, it is obvious that the maximum must occur for $0 \leq x \leq a$. Thus these may be taken to be endpoints. The critical value occurs when,

$$\boxed{\frac{\sin(\theta_1)}{v_1} = \frac{\sin(\theta_2)}{v_2}.}$$

This is *Snell's Law* for refraction of light upon crossing from one medium to another. For refraction, a particle of light (perhaps fictitious) replaces the swimmer, a translucent medium of one type replaces the water, and a translucent medium of a second type replaces the land. If light travels with velocity v_1 in the first medium and with velocity v_2 in the second medium, light rays will *refract* upon crossing the boundary between media. Snell's Law describes the angles of this refraction.

Lecture 11. October 4, 2005

Homework. Problem Set 3 Part I: (g) and (h).

Practice Problems. Course Reader: 2E-4, 2E-8, 2E-9.

1. Related rates. A situation that arises often in practice is that two quantities, say x and y, depend on a third independent variable, say t. The quantities x and y are related through some constraint. Using the constraint, if the rate-of-change dx/dt is known, the rate-of-change dy/dt can be inferred.

Example. For a spring displaced x units from equilibrium, Hooke's law implies the potential energy of the spring is,

$$P = \frac{1}{2}kx^2,$$

where k is a constant with units kg/s^2. At some moment $t = T$, a spring is displaced $5cm$ from equilibrium and has velocity $5cm/s$. In terms of the spring constant k, describe the rate-of-change of the potential energy at $t = T$.

Implicitly differentiating the equation with respect to t gives, using the chain rule,

$$\frac{dP}{dt} = \frac{1}{2}k(2x)\frac{dx}{dt} = kx\frac{dx}{dt}.$$

So, at time $t = T$,

$$\frac{dP}{dt}(T) = kx(T)\frac{dx}{dt}(T) = k(5)(5)cm^2/s = \boxed{25kcm^2/s.}$$

2. Method for solving related-rates problems. Many of these steps apply to any word-problem in mathematics.

(i) Identify the independent variable. In the example, this is t.

(ii) Label all constants. In the example, k is a constant.

(iii) Label all dependent variables. In the example, x and P are dependent variables.

(iv) Draw a diagram and carefully label it.

(v) Write the given rate-of-change and the unknown rate-of-change. In the example, $dx/dt(T)$ is given as $5cm/s$, and dP/dt is unknown.

(vi) Using the diagram and any other information, find constraints among the dependent variables. In the example, this is the equation $P = kx^2/2$.

(vii) Implicitly differentiate the constraint equations with respect to the independent variable. In the example, this gives $dP/dt = kxdx/dt$.

(viii) Substitute in all known quantities and solve for the unknown rate-of-change. In the example, $dP/dt(T)$ equals $25kcm^2/s$.

Example. A state trooper waits a distance a from a highway for passing speeders. The speed limit is $60mph$. The trooper aims her radar gun at an angle of $\pi/4$ to the road. The radar registers a passing car moving away from the trooper at a speed of $50mph$. Should the trooper ticket the driver?

The independent variable is time t. The constants are the distance a and the angle $\theta = \pi/4$. Label a coordinate system with the trooper at the origin and the highway equal to the line $y = a$. Label the position of the car along the highway as x, moving in the positive direction. Denote by r the distance of the car from the trooper. Then x and r are dependent variables. The rate-of-change $dr/dt(T)$ is given as $50mph$. The unknown rate-of-change is $dx/dt(T)$. The constraint is the Pythagorean theorem,

$$r^2 = x^2 + y^2.$$

Implicit differentiation with respect to t yields,

$$2r\frac{dr}{dt} = 2x\frac{dx}{dt} + 0 = 2x\frac{dx}{dt}.$$

At time $t = T$, $x(T)$ equals a, because the angle θ is $\pi/4$. Thus $r(T)$ equals $\sqrt{2}a$. Substituting in gives,

$$2(\sqrt{2}a)50 = 2(a)\frac{dx}{dt}(T).$$

Solving gives,

$$\frac{dx}{dt}(T) = \sqrt{2}50 \approx \boxed{71mph.}$$

So the trooper should ticket the driver.

Example. A point on the x-axis moves away from the origin. There is an angle θ subtended by the point and the unit circle with equation $x^2 + y^2 = 1$. In other words, standing at the point $(x, 0)$ and staring at the circle, θ is the angle of your field-of-vision occupied by the circle. At a moment $t = T$, the point is at the position $(2, 0)$ and moving with velocity v. What is the rate-of-change of θ at $t = T$?

The independent variable is time t. There is no constant. The dependent variables are the x-coordinate of the point, $x(t)$, and the angle $\theta(t)$. The rate-of-change $dx/dt(T)$ is given to be v. The rate-of-change $d\theta/dt$ is unknown.

The constraint is somewhat tricky. There are two tangent lines to the circle containing $(x, 0)$. These are the tangent lines to points $(a, +b)$ and $(a, -b)$ on the circle. Because the tangent line to the circle at (a, b) is perpendicular to the radius through (a, b), the triangle with vertices $(0, 0)$, (a, b) and the point $(x, 0)$ is a right triangle. The angle of the triangle at $(x, 0)$ is $\theta/2$. Since the radius has length 1 and the hypotenuse has length x, the constraint is,

$$\sin(\theta) = \frac{1}{x}.$$

Implicit differentiation with respect to t gives,

$$\frac{d\sin(\theta)}{d\theta}\frac{d\theta}{dt} = \frac{d(x^{-1})}{dx}\frac{dx}{dt},$$

or,

$$\cos(\theta)\frac{d\theta}{dt} = \frac{-1}{x^2}\frac{dx}{dt}.$$

Since $x(T)$ equals 2, $\sin(\theta(T)) = 1/2$, and thus $\cos(\theta(T))$ equals $\sqrt{3}/2$. Plugging in gives,

$$\frac{\sqrt{3}}{2}\frac{d\theta}{dt}(T) = \frac{-1}{(2)^2}v = \frac{-v}{4}.$$

Solving gives,

$$\frac{d\theta}{dt}(T) = \boxed{-v/(2\sqrt{3}).}$$

3. Another applied max/min problem. As review for Exam 2, this is another applied max/min problem. A trapezoid is inscribed inside the upper unit semicircle, $x^2 + y^2 = 1, y \geq 0$. The base of the trapezoid is the diameter of the semicircle lying on the x-axis. The top of the trapezoid is parallel to the x-axis joining $(-x, y)$ to (x, y) for a point (x, y) on the unit circle in the first quadrant. What is the maximal area enclosed by such a trapezoid?

The parameters are x and y. The height of the trapezoid is y. The area of a trapezoid is the product of the height with the average of the parallel sides. Thus,

$$A = y\frac{(2+2x)}{2} = (x+1)y.$$

This is the quantity to be maximized. There is a constraint among the parameters,

$$x^2 + y^2 = 1.$$

Also, since (x, y) is in the first quadrant, $0 \leq x \leq 1$ and $0 \leq y \leq 1$.

There are at least 3 ways to proceed. The most direct is to solve for y in terms of x,

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

Substituting this into the equation for A gives,

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

Differentiating gives,

$$\frac{dA}{dx} = \sqrt{1-x^2} + (x+1)\frac{-2x}{2\sqrt{1-x^2}} = \frac{1}{\sqrt{1-x^2}}((1-x^2) - (x^2+x)) = \frac{-(2x^2+x-1)}{\sqrt{1-x^2}}.$$

Because the quadratic polynomial $2x^2 + x - 1$ factors as,

$$2x^2 + x - 1 = (2x-1)(x+1),$$

the critical points of A are $x = -1$ and $x = 1/2$. But $x = -1$ does not give a point in the first quadrant. Thus A is maximized either at one of the endpoints $x = 0$, $x = 1$ or at the critical point $x = 1/2$. Plugging in gives,

$$A(0) = 1, A(1/2) = 3\sqrt{3}/4, A(1) = 0.$$

This gives the answer,

$$\boxed{\text{A achieves its maximum } 3\sqrt{3}/4 \text{ for the point } (x, y) = (1/2, \sqrt{3}/2).}$$

Two other methods were given in lecture. The fastest among the three is to instead minimize A^2,

$$A^2 = (x+1)^2 y^2.$$

Using the constraint, $y^2 = 1 - x^2$, thus,

$$(A^2)(x) = (x+1)^2(1-x^2).$$

The derivative of this polynomial is very fast to compute, and gives the same answer as above.

Lecture 12. October 6, 2005

Homework. Problem Set 3 Part I: (i) and (j).

This was a guest lecture by Sabri Kilic. Notes from the lecture will not be posted. As always, please do the required reading in the course textbook.

Lecture 13. October 13, 2005

Homework. Problem Set 4 Part I: (a) and (b); Part II: Problem 3.

Practice Problems. Course Reader: 3A-1, 3A-2, 3A-3.

1. Differentials. An alternative notation for derivatives is *differential notation*. The differential notation,

$$dF(x) = f(x)dx,$$

is shorthand for the sentence "The derivative of $F(x)$ with respect to x equals $f(x)$." Formally, this is related to the Leibniz notation for the derivative,

$$\frac{dF}{dx}(x) = f(x),$$

which means the same thing as the differential notation. It may look like the first and second equation are obtained by dividing and multiplying by the quantity dx. It is crucial to remember that dF/dx is **not a fraction**, although the notation suggests otherwise.

In differential notation, some derivative rules have a very simple form, and are thus easier to remember. Here are a few derivative rules in differential notation.

$$\begin{array}{rcl} dF(x) & = & F'(x)dx \\ d(F(x)+G(x)) & = & dF(x) + dG(x) \\ d(cF(x)) & = & cdF(x) \\ d(F(x)G(x)) & = & G(x)dF(x) + F(x)dG(x) \\ d(F(x)/G(x)) & = & 1/(G(x))^2(G(x)dF(x) - F(x)dG(x)) \end{array}$$

The chain rule has a particularly simple form,

$$d(F(u)) = \frac{dF}{du}du = \frac{dF}{du}\frac{du}{dx}dx.$$

Example. Using differential notation, the derivative of $\sin(\sqrt{x^2+1})$ is,

$$d\sin((x^2+1)^{1/2}) = \cos((x^2+1)^{1/2})d(x^2+1)^{1/2} = \cos((x^2+1)^{1/2})(\tfrac{1}{2}(x^2+1)^{-1/2})d(x^2+1) = \cos((x^2+1)^{1/2})\tfrac{1}{2}(x^2+1)^{-1/2}(2xdx) = \boxed{x(x^2+1)^{-1/2}\cos((x^2+1)^{1/2})dx}.$$

2. Antidifferentiation. Recall, the basic problem of differentiation is the following.

Problem (Differentiation). Given a function $F(x)$, find the function $f(x)$ satisfying $\frac{dF}{dx} = f(x)$.

The bais problem of *antidifferentiation* is the inverse problem.

Problem (Antidifferentiation). Given a function $f(x)$, find a function $F(x)$ satisfying $\frac{dF}{dx} = f(x)$.

A function $F(x)$ solving the problem is called an *antiderivative of* $f(x)$, or sometimes an *indefinite integral of* $f(x)$. The notation for this is,

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

The expression $f(x)$ is called the *integrand*. It is important to note, if $F(x)$ is one antiderivative of $f(x)$, then for each constant C, $F(x)+C$ is also an antiderivative of $f(x)$. The constant C is called a *constant of integration.*

In a sense that can be made precise, the problem of differentiation has a complete solution whenever $F(x)$ is a "simple expression", i.e., a function built from the differentiable functions we have seen so far. Unfortunately, for very many simple functions $f(x)$, no antiderivative of $f(x)$ has a simple expression. In large part, this is what makes antidifferentiation difficult. Luckily, many of the most important simple functions $f(x)$ do have an antiderivative with a simple expression. One goal of this unit is to learn how to recognize when a simple antiderivative exists, and some tools to compute the antiderivative.

3. Antidifferentiation. Guess-and-check. The main technique for antidifferentiation is educated guessing.

Example. Find an antiderivative of $f(x) = x^2+2x+1$. Since the derivative of x^n is nx^{n-1}, it is reasonable to guess there is an antiderivative of the form $F(x) = Ax^3+Bx^2+Cx$. Differentiation gives,

$$\frac{dF}{dx} = 3Ax^2+2Bx+C.$$

Thus, $F(x)$ is an antiderivative of $f(x)$ if and only if,

$$3A = 1, \quad 2B = 2, \quad \text{and } C = 1.$$

This gives an antiderivative,

$$\int(x^2+2x+1)dx = \boxed{\tfrac{1}{3}x^3+x^2+x+E},$$

where E is any constant.

Guess-and-check is a game we can lose, as well as win. However, there are a few rules that better the odds in this guessing game. In fact, they are basically the same rules for derivatives in differential notation, simply written backwards.

$$\begin{array}{ccc} \int (f(x)+g(x))dx & = & \int f(x)dx + \int g(x)dx \\ \int cf(x)dx & = & c\int f(x)dx \\ \int f(u(x))u'(x)dx & = & \int f(u)du \end{array}$$

4. Antidifferentiation. Integration by substitution. The last rule above is very important, and called *integration by substitution.*

Example. Find an antiderivative of $x\sin(x^2)$. This time guess-and-check is much less effective. By roughly the same logic in the last example, we might guess an antiderivative has the form $Ax^3\sin(x^2)$. The derivative is $3Ax^2\sin(x^2) + 2Ax^4\cos(x^2)$. The first term is good, but the second term is bad. We can try to correct our guess by adding a term, $Ax^3\sin(x^2) - 2/5Ax^5\cos(x^2)$, whose derivative is now $3Ax^2\sin(x^2) + 4/5Ax^6\sin(x^2)$. This still doesn't work, and is leading in the wrong direction.

A better solution is to use integration by substitution. Observe part of $f(x)$ can be written as a function of $u(x) = x^2$. Also, the derivative $u'(x) = 2x$ occurs in $f(x)$ through $x = 1/2(2x) = u'(x)/2$. Thus,

$$x\sin(x^2) = sin(u(x))u'(x)/2, \quad u(x) = x^2.$$

Applying integration by substitution,

$$\int x\sin(x^2)dx = \int \sin(u(x))\tfrac{1}{2}u'(x)dx = \int \tfrac{1}{2}\sin(u)du = \\ \tfrac{-1}{2}\cos(u) + C = \boxed{\tfrac{-1}{2}\cos(x^2) + C.}$$

Here is a checklist for applying integration by substition to find the antiderivative of $f(x)$.

(i) Find an expression $u(x)$ so that most of the integrand $f(x)$ can be expressed as a simpler function of $u(x)$.

(ii) Compute the differential $du(x) = u'(x)dx$.

(iii) Inside the differential $f(x)dx$, try to find $du = u'(x)dx$ as a factor.

(iv) Try to write $f(x)dx$ as $g(u)du$. If you cannot do this, the method does not apply with the given choice of u.

(v) Find an antiderivative $G(u) = \int g(u)du$ for the simpler integrand $g(u)$ (if this is possible).

(vi) Back-substitute $u = u(x)$ to get an antiderivative $F(x) = G(u(x))$ for $f(x)$.

Example. Compute the antiderivative,

$$\int \sin(x)^3 \cos(x)dx.$$

Most of the integrand is a function of $\sin(x)$. So substitute $u(x) = \sin(x)$. The differential of u is $du = \cos(x)dx$. The differential $\sin(x)^3 \cos(x)dx$ contains $du = \cos(x)dx$ as a factor. The remainder of the integrand is $\sin(x)^3 = u^3$. So, according to integration by substitution,

$$\int \sin(x)^3 \cos(x)dx = \int u^3 du = \frac{1}{4}u^4 + C.$$

Finally, back-substitute $u = \sin(x)$ to get,

$$\int \sin(x)^3 \cos(x)dx = \boxed{(\sin(x))^4/4 + C.}$$

Lecture 14. October 14, 2005

Homework. Problem Set 4 Part II: Problem 2.

Practice Problems. Course Reader: 3B-1, 3B-3, 3B-4, 3B-5.

1. The problem of areas. The ancient Greeks computed the areas of triangles, quadrilaterals, and many other polygons. Their basic method was *dissection*: dissecting a polygonal region exactly into smaller regions, usually triangles, having known areas. The area of the large region is the sum of the areas of the small regions. But the ancient Greeks also knew the area of a circle, which cannot be dissected exactly into finitely many polygonal regions. Their method was *exhaustion*: finding polygonal regions *approximately* equal to the original region, and computing the limit of the areas of the polygons as the approximation improves.

Example. A regular N-sided polygon inscribed in a circle of radius r has *apothem length* $a = r\cos(\pi/N)$ and *chord length* $b = 2r\sin(\pi/N)$. Thus the area of the polygon is,

$$A = N\frac{ab}{2} = Nr^2 \sin(\pi/N)\cos(\pi/N) = r^2\frac{N}{2}\sin(2\pi/N) = \pi r^2 \frac{\sin(2\pi/N)}{2\pi/N}.$$

As N increases, $2\pi/N$ decreases to 0. Because $\lim_{t\to 0}\sin(t)/t$ equals 1, as N approaches infinity, the area of the polygon approaches,

$$\lim_{N\to\infty} \pi r^2 \frac{\sin(2\pi/N)}{2\pi/N} = \boxed{\pi r^2.}$$

A more sophisticated version of the method of exhaustion gives the *Riemann integral*. Here is the basic problem.

Problem (Area). Find the signed area between the graph of $y = f(x)$ and the x-axis over the interval $a \le x \le b$.

For a region above the x-axis, the *signed area* is simply the area. For a region below the x-axis, the signed area is the negative of the area. For a region partly above the x-axis and partly below the x-axis, the signed area is the sum of the signed area of the region above the x-axis and the signed area of the region below the x-axis.

2. Partitions. A *partition* of an interval $[a, b]$ is a finite decomposition of the interval as a union of non-overlapping subintervals,

$$[a, b] = [x_0, x_1] \cup [x_1, x_2] \cup \cdots \cup [x_{n-2}, x_{n-2}] \cup [x_{n-1}, x_n].$$

Since an interval is determined by its right and left endpoints, to specify a partition of $[a, b]$, it is equivalent to give an ordered sequence of increasing numbers,

$$a = x_0 < x_1 < x_2 < \cdots < x_{n-2} < x_{n-1} < x_n = b.$$

The k^{th} subinterval of the partition is the interval $[x_{k-1}, x_k]$, having length,

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

A partition is *fine* if the subintervals are small, and *coarse* if the subintervals are large. It may seem the number of intervals n is a good measure of fineness: since the subintervals of a fine partition are small, the number n of subintervals must be large. However, a partition into many subintervals may include a few subintervals that are quite large. For instance, the partition

$$[0, 1] = [0, 1/2n] \cup [1/2n, 2/2n] \cup [2/2n, 3/2n] \cup \cdots \cup [(n-2)/2n, (n-1)/2n] \cup [n-1/2n, n/2n] \cup [1/2, 1],$$

has n very small intervals of length $1/2n$, but has one interval, $[1/2, 1]$, of size $1/2$. The number $1/2$ may not seem large, but as n increases, it is quite large compared to $1/2n$.

Because of such examples, a better measure of fineness is mesh size: The *mesh size* of a partition is the maximal length of any subinterval in the partition,

$$\text{mesh} = \max \Delta x_k | k = 1, \ldots, n.$$

3. Riemann sums. Let $f(x)$ be a function defined on an interval $a \leq x \leq b$. Given a partition $a = x_0 < \cdots < x_n = b$ of $[a, b]$, and given a choice, for every $k = 1, \ldots, n$, of element x_k^* in the k^{th} subinterval, $x_{k-1} \leq x_k^* \leq x_k$, the curvilinear region bounded by $y = f(x)$ and the x-axis is approximated by a union of n vertical strips. The k^{th} vertical strip lies above or below the interval on the x-axis, $x_{k-1} \leq x \leq x_k$, and has height $y_k^* = f(x_k^*)$. The width of the vertical strip is Δx_k, thus the signed area is,

$$\Delta A_k = y_k^* \Delta x_k.$$

The total area of the union of vertical strips is simply the sum of the areas of individual vertical strips,

$$A = \sum_{k=1}^{n} y_k^* \Delta x_k.$$

The sum above is a *Riemann sum*. It is an approximation of the signed area of the curvilinear region.

There are many choices of partition. And for each partition, there are many choices for the numbers x_k^*. However, there are some special choices. On the k^{th} interval, the smallest value $f(x)$ takes on is denoted by,

$$y_{k,\min} = \min\{f(x)|x_{k-1} \leq x \leq x_{k+1}\}.$$

Similarly, the largest value $f(x)$ takes on is denote by,

$$y_{k,\max} = \max\{f(x)|x_{k-1} \leq x \leq x_{k+1}\}.$$

For every choice of x_k^* in the k^{th} interval, y_k^* is trapped between these two values,

$$y_{k,\min} \leq y_k^* \leq y_{k,\max}.$$

Denoting,

$$\Delta A_{k,\min} = y_{k,\min}\Delta x_k, \quad \Delta A_{k,\max} = y_{k,\max}\Delta x_k,$$

the area ΔA_k is trapped between these two values,

$$\Delta A_{k,\min} \leq \Delta A_k \leq \Delta A_{k,\max}.$$

Denoting the sums of the areas by,

$$\begin{array}{l} A_{\min} = \sum_{k=1}^n \Delta A_{k,\min} = \sum_{k=1}^n y_{k,\min}\Delta x_k, \\ A_{\max} = \sum_{k=1}^n \Delta A_{k,\min} = \sum_{k=1}^n y_{k,\min}\Delta x_k, \end{array}$$

the Riemann sum A is trapped between the two values,

$$A_{\min} \leq A \leq A_{\max}.$$

Thus, if $A_{\min}$ and $A_{\max}$ are close to each other, the value of A does not depend very much on the choices of the numbers x_k^*.

4. The Riemann integral. The method of the Riemann integral is to compute both $A_{\min}$ and $A_{\max}$ for a sequence of partitions whose mesh sizes approach 0. The mesh size measures the fineness of the partition, thus also the fit of the union of vertical strips to the curvilinear region. If the two limits,

$$\lim_{\text{mesh}\to 0} A_{\min}, \quad \lim_{\text{mesh}\to 0} A_{\max},$$

are defined and equal, it is said *the Riemann integral exists*, and the common limit is called the *Riemann integral*,

$$\int_a^b f(x)dx = \lim_{\text{mesh}\to 0} A_{\min} = \lim_{\text{mesh}\to 0} A_{\max}.$$

Also, $f(x)$ is said to be *Riemann integrable on the interval* $[a, b]$. Another name for the Riemann integral is the *definite integral*.

Example. Consider the function $f(x) = x$ on the interval $0 \leq x \leq L$, for some positive number L. Form the partition with n subintervals of equal length,

$$x_0 = 0 = 0L/n, x_1 = 1L/n, x_2 = 2L/n, \ldots, x_k = kL/n, \ldots x_n = nL/n = L.$$

Every interval has length $\Delta x_k = L/n$. So the mesh size is L/n. The minimum value of $f(x)$ on the interval $x_{k-1} \leq x \leq x_k$ is $y_{k,\text{min}} = x_{k-1} = (k-1)L/n$. The maximum value is $y_{k,\text{max}} = x_k = kL/n$. Thus,

$$A_{\text{min}} = \sum_{k=1}^{n} y_{k,\text{min}} \Delta x_k = \sum_{k=1}^{n} \frac{(k-1)L}{n}\frac{L}{n} = \frac{L^2}{n^2}\sum_{k=1}^{n}(k-1),$$

and,

$$A_{\text{max}} = \sum_{k=1}^{n} y_{k,\text{max}} \Delta x_k = \sum_{k=1}^{n} \frac{kL}{n}\frac{L}{n} = \frac{L^2}{n^2}\sum_{k=1}^{n} k.$$

To evaluate these sums, use the well-known formula,

$$\sum_{k=1}^{n} k = \frac{n(n+1)}{2}.$$

This also gives,

$$\sum_{k=1}^{n}(k-1) = \sum_{l=0}^{n-1} l = \sum_{l=1}^{n-1} l = \frac{(n-1)n}{2},$$

by making the substitution $l = k - 1$. Substituting the formula gives,

$$A_{\text{min}} = \frac{L^2}{n^2}\frac{n(n-1)}{2} = \frac{L^2}{2}(1 - \frac{1}{n}),$$

and,

$$A_{\text{min}} = \frac{L^2}{n^2}\frac{n(n+1)}{2} = \frac{L^2}{2}(1 + \frac{1}{n}).$$

Therefore,

$$\lim_{n\to\infty} A_{\text{min}} = \frac{L^2}{2}\lim_{n\to 0}(1 - \frac{1}{n}) = \frac{L^2}{2}(1-0) = \frac{L^2}{2}.$$

Similarly,

$$\lim_{n\to\infty} A_{\text{max}} = \frac{L^2}{2}\lim_{n\to 0}(1 + \frac{1}{n}) = \frac{L^2}{2}(1+0) = \frac{L^2}{2}.$$

Since the two limits are equal, $f(x) = x$ is Riemann integrable on the interval $[0, L]$, and,

$$\int_0^L x dx = \boxed{L^2/2.}$$

This agrees with the familiar result from high-school geometry: the area of a triangle equals one half of the base times the height, since both the base and height of this triangle are L.

5. Rules for Riemann integrals. There are several rules for Riemann integrals, summarized below.

$$\begin{array}{rcl} \int_a^b (f(x)+g(x))dx & = & \int_a^b f(x)dx + \int_a^b g(x)dx, \\ \int_a^b (r\cdot f(x))dx & = & r\cdot \int_a^b f(x)dx, \\ \int_a^b f(x)dx + \int_b^c f(x)dx & = & \int_a^c f(x)dx. \end{array}$$

Lecture 15. October 18, 2005

Homework. Problem Set 4 Part I: (d) and (e); Part II: Problem 2.

Practice Problems. Course Reader: 3B-6, 3C-2, 3C-3, 3C-4, 3C-6.

1. The Riemann sum for the exponential function. The problem is to compute the Riemann integral,

$$\int_0^b e^x dx,$$

using Riemann sums. Choose the partition of $[0, b]$ into a sequence of n equally-spaced subintervals of length b/n. So the partition numbers are $x_k = kb/n$. Also the length of each partition is $\Delta x_k = b/n$. Because e^x is increasing, the minimum value of e^x on the interval $[x_{k-1}, x_k]$ occurs at the left endpoint,

$$y_{k,\min} = e^{x_{k-1}} = e^{(k-1)b/n}.$$

Similarly, the maximum value occurs at the right endpoint,

$$y_{k,\max} = e^{x_k} = e^{kb/n}.$$

Thus the lower sum is,

$$A_{\min} = \sum_{k=1}^{n} y_{k,\min}\Delta x_k = \sum_{k=1}^{n} e^{(k-1)b/n}\frac{b}{n}.$$

And the upper sum is,

$$A_{\max} = \sum_{k=1}^{n} y_{k,\max}\Delta x_k = \sum_{k=1}^{n} e^{kb/n}\frac{b}{n}.$$

To evaluate each of the sums, make the substitution $c = e^{b/n}$. Then the lower sum is,

$$A_{\min} = \frac{b}{n}\sum_{k=1}^{n} c^{k-1} = \frac{b}{n}\sum_{l=0}^{n-1} c^l.$$

The sum is a geometric sum,

$$(1 + c + c^2 + \cdots + c^{n-2} + c^{n-1}) = \frac{c^n - 1}{c - 1}.$$

Plugging this in gives,

$$A_{\min} = \frac{b}{n}\frac{c^n - 1}{c - 1} = \frac{b}{n}\frac{e^{bn/n} - 1}{e^{b/n} - 1}.$$

This simplifies to,

$$A_{\text{min}} = (e^b - 1)\frac{b/n}{e^{b/n} - 1}.$$

A similar computation gives,

$$A_{\text{max}} = (e^b - 1)e^{b/n}\frac{b/n}{e^{b/n} - 1}.$$

Now make the substitution, $h = b/n$. This gives,

$$A_{\text{min}} = (e^b - 1)\frac{h}{e^h - 1},$$

$$A_{\text{max}} = (e^b - 1)e^h\frac{h}{e^h - 1}.$$

Taking the limit of A_{min}, respectively A_{max}, as n tends to infinity is the same as taking the limit as h tends to 0.

Now observe that,

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

is the difference quotient limit giving the derivative of e^x at $x = 0$. Since de^x/dx equals e^x, and since e^0 equals 1, this gives,

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

Inverting gives,

$$\lim_{h\to 0}\frac{h}{e^h - 1} = \left(\lim_{h\to 0}\frac{e^h - 1}{h}\right)^{-1} = (1)^{-1} = 1.$$

Also, because e^x is continuous,

$$\lim_{h\to 0} e^h = e^0 = 1.$$

Putting this together gives,

$$\lim_{n\to\infty} A_{\text{min}} = (e^b - 1)\lim_{h\to 0}\frac{h}{e^h - 1} = (e^b - 1)(1) = e^b - 1.$$

Similarly,

$$\lim_{n\to\infty} A_{\text{max}} = (e^b - 1)(\lim_{h\to 0} e^h)(\lim_{h\to 0}\frac{h}{e^h - 1}) = (e^b - 1)(1)(1) = e^b - 1.$$

Since the limit of A_{min} and the limit of A_{max} exist and are equal, the Riemann integral exists and equals,

$$\int_0^b e^x dx = \boxed{e^b - 1.}$$

2. The Riemann sum for x^r. Let $r > 0$ be a positive real number. The problem is to compute the Riemann integral,

$$\int_1^b x^r dx,$$

using Riemann sums. For this particular integral, a different partition than usual is more efficient. Let n be a positive integer, and let q be the real number,

$$q = b^{1/n}.$$

Choose the partition of $[1, b]$ into n subintervals with partition numbers,

$$x_k = q^k.$$

Observe that,

$$1 = x_0 < x_1 < \cdots < x_{n-1} < x_n = (b^{1/n})^n = b.$$

The length of the k^{th} subinterval is,

$$\Delta x_k = x_k - x_{k-1} = q^k - q^{k-1} = q^{k-1}(q-1).$$

Observe this increases as k increases. So this is not the partition of $[1, b]$ into n equal subintervals. The mesh size is,

$$\text{mesh} = \max(\Delta x_1, \ldots, \Delta x_n) = \Delta x_n = (q-1)b^{(n-1)/n} \leq q - 1.$$

As n tends to infinity, the mesh size tends to,

$$\lim_{n\to 0} \text{mesh} = \lim_{n\to 0} q - 1 = \lim_{n\to 0} b^{1/n} - 1 = 0.$$

Thus, even though this isn't the most obvious choice of partition, it can be used to compute the Riemann integral.

Because x^r is increasing, the minimum value of x^r on the interval $[x_{k-1}, x_k]$ occurs at the left endpoint,

$$y_{k,\min} = x_{k-1}^r = q^{(k-1)r}.$$

Similarly, the maximum value occurs at the right endpoint,

$$y_{k,\max} = x_k^r = q^{kr}.$$

Thus the lower sum is,

$$A_{\min} = \sum_{k=1}^{n} y_{k,\min}\Delta x_k = \sum_{k=1}^{n} q^{(k-1)r} \cdot q^{(k-1)}(q-1).$$

This simplifies to,

$$A_{\text{min}} = (q-1)\sum_{k=1}^{n} q^{(k-1)(r+1)}.$$

And the upper sum is,

$$A_{\text{max}} = \sum_{k=1}^{n} y_{k,\text{max}}\Delta x_k = \sum_{k=1}^{n} q^{kr}q^{(k-1)}(q-1).$$

This simplifies to,

$$A_{\text{max}} = (q-1)q^r \sum_{k=1}^{n} q^{(k-1)(r+1)}.$$

To evaluate the sum, make the substitution $c = q^{r+1}$. Then the sum is,

$$\sum_{k=1}^{n} c^{k-1} = 1 + c + c^2 + \cdots + c^{n-2} + c^{n-1}.$$

This geometric sum equals,

$$\frac{c^n - 1}{c - 1} = \frac{q^{n(r+1)} - 1}{q^{r+1} - 1}.$$

Thus the upper and lower sums simplify to,

$$A_{\text{min}} = (q-1)(q^{n(r+1)} - 1)/(q^{r+1} - 1),$$

$$A_{\text{max}} = q^r(q-1)(q^{n(r+1)} - 1)/(q^{r+1} - 1).$$

Now back-substitute $q = b^{1/n}$ to get that $q^{n(r+1)} = b^{r+1}$. Simplifying gives,

$$A_{\text{min}} = (b^{r+1} - 1)\frac{1}{(q^{r+1} - 1)/(q-1)},$$

$$A_{\text{max}} = (b^{r+1} - 1)q^r\frac{1}{(q^{r+1} - 1)/(q-1)}.$$

As n tends to infinity, the quantity $q = b^{1/n}$ tends to 1. The fraction,

$$\frac{q^{r+1} - 1}{q - 1},$$

is the difference quotient for $y = x^{r+1}$ for $x = 1$. As q tends to 1, the limit of the difference quotient is the derivative of $y = x^{r+1}$ at $x = 1$,

$$\lim_{q \to 1} \frac{q^{r+1} - 1}{q - 1} = \frac{d(x^{r+1})}{dx}|_{x=1} = ((r+1)x^r|_{x=1} = (r+1).$$

Also, since x^r is continuous,

$$\lim_{q\to 1} q^r = 1^r = 1.$$

Substituting this in gives,

$$\lim_{n\to\infty} A_{\min} = (b^{r+1}-1)\left(\lim_{q\to 1}\frac{q^{r+1}-1}{q-1}\right)^{-1} = \frac{b^{r+1}-1}{r+1},$$

$$\lim_{n\to\infty} A_{\max} = (b^{r+1}-1)\left(\lim_{q\to 1} q^r\right)\left(\lim_{q\to 1}\frac{q^{r+1}-1}{q-1}\right)^{-1} = \frac{b^{r+1}-1}{r+1},$$

Since the limit of $A_{\min}$ and the limit of $A_{\max}$ exist and are equal, the Riemann integral exists and equals,

$$\int_1^b x^r dx = (b^{r+1}-1)/(r+1).$$

3. The Fundamental Theorem of Calculus. There is a single theorem that it is at the heart of almost all applications involving Riemann integrals. The theorem answers two question simultaneously: Which functions are Riemann integrable? What is the Riemann integral of a function? The answer to the first question is: Every function you are likely to encounter is Riemann integrable. Precisely, every continuous function, and every piecewise continuous function is Riemann integrable.

The answer to the second question is more interesting. Assume $f(x)$ is a continuous function. Let $x = a$ be a fixed point where $f(x)$ is defined. Form the function,

$$F(x) = \int_a^x f(t)dt.$$

The function $F(x)$ is defined whenever $f(t)$ is defined on all of $[a, x]$. If $f(x)$ is continuous, the Fundamental Theorem of Calculus asserts $F(x)$ is differentiable and,

$$\frac{dF}{dx}(x) = \frac{d}{dx}\int_a^x f(t)dt = f(x).$$

The proof of the second part is very easy. Consider the increment in F from x to $x + \Delta x$,

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

Let $y_{\min}$ be the minimum value of $f(t)$ on the interval $[x, x+\Delta x]$. Let $y_{\max}$ be the maximum value of $f(t)$ on the interval $[x, x+\Delta x]$. Then for every choice of partition $t_0 < t_1 < \cdots < t_n$ of $[x, x+\Delta x]$, and every choice of values y_k^* on the subintervals,

$$y_{\min} \le y_k^* \le y_{\max},$$

for every k. Thus the Riemann sum is squeezed between,

$$\sum_{k=1}^{n} y_{\min}\Delta t_k \leq \sum_{k=1}^{n} y_k^* \Delta t_k \leq \sum_{k=1}^{n} y_{\max}\Delta t_k.$$

Of course the lower bound is,

$$\sum_{k=1}^{n} y_{\min}\Delta t_k = y_{\min}\sum_{k=1}^{n}\Delta t_k = y_{\min}\Delta x,$$

because the total length of the interval $[x, x+\Delta x]$ is Δx. Similarly, the upper bound is,

$$\sum_{k=1}^{n} y_{\max}\Delta t_k = y_{\max}\Delta x.$$

Thus the Riemann sum is squeezed between,

$$y_{\min}\Delta x \leq \sum_{k=1}^{n} y_k^* \Delta x_k \leq y_{\max}\Delta x.$$

Because the Riemann integral is a limit of Riemann sums, it is also squeezed,

$$y_{\min}\Delta x \leq \int_x^{x+\Delta x} f(t)dt \leq y_{\max}\Delta x.$$

Substituting in $F(x+\Delta x) - F(x)$ and dividing each term by Δx gives,

$$y_{\min} \leq \frac{F(x+\Delta x) - F(x)}{\Delta x} \leq y_{\max}.$$

The middle term is the difference quotient. Consider what happens as Δx tends to 0. Because $f(t)$ is continuous, both the maximum and minimum values of $f(t)$ on $[x, x+\Delta x]$ simply limit to the value $f(x)$. Thus,

$$\lim_{\Delta x} y_{\min} = \lim_{\Delta x} y_{\max} = f(x).$$

By the Squeezing Lemma for limits, since these two limits exist and are equals, the middle limit also exists and equals $f(x)$,

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

This is precisely what the Fundamental Theorem of Calculus asserts,

$$\frac{d}{dx}\int_a^x f(t)dt = \boxed{f(x)}.$$

4. Algorithm for computing Riemann integrals. The Fundamental Theorem of Calculus has many important applications. The most obvious is to give us a simpler method for computing Riemann integrals, under the hypothesis that we can compute the antiderivative. If $f(x)$ is a continuous function and $G(x)$ is a known antiderivative of $f(x)$, then,

$$\int_a^b f(t)dt = G(b) - G(a).$$

To see this, observe that,

$$F(x) = \int_a^x f(t)dt,$$

is also an antiderivative of $f(t)$ by the Fundamental Theorem of Calculus. Thus, since the general antiderivative is $G(x) + C$, there is a constant C such that $F(x) = G(x) + C$. But also,

$$F(a) = \int_a^a f(t)dt = 0.$$

Thus, $F(x) = G(x) - G(a)$. Now plug in $x = b$ to get,

$$\int_a^b f(t)dt = F(b) = \boxed{G(b) - G(a).}$$

Lecture 16. October 20, 2005

Practice Problems. Course Reader: 3D-1, 3D-3, 3D-7, 3E-3, 3E-4.

1. Dummy variables. Give a Riemann integrable function $f(x)$ defined on an interval $[a, b]$, the notation,

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

is shorthand for the Riemann integral of $f(x)$ over this interval. In particular, this equals the limit,

$$\lim_{n\to\infty} f(a + (b-a)k/n)\frac{b-a}{n}.$$

Observe, the variable x does not appear in this limit. It is very convenient to include the variable x in the notation for the Riemann integral; for how else are we to express the function integrated? But, since the definition of the Riemann integral does not involve x, x is really a *dummy variable*. Any variable name may be substituted for x, with the same meaning.

$$\int_a^b f(x)dx = \int_a^b f(u)du = \int_a^b f(v)dv = \int_a^b f(t)dt = \dots$$

This freedom is very useful, particularly when one or both of the limits of integration depend on some parameter. In this case, by convention, the dummy variable is chosen to be a different parameter.

$$\int_a^x f(x)dx \text{ INCORRECT}, \quad \int_a^x f(t)dt \text{ CORRECT}$$

This convention reduces the likelihood of an error.

2. Variable limits of integration. The Riemann integral is often used to define functions, particularly antiderivatives having no simpler expression.

Example. For every angle $0 \leq \theta < \pi/2$, define $f(\theta)$ to be the area above the x-axis, inside the unit circle $x^2 + y^2 = 1$, and bounded by the vertical lines, $-\cos(\theta) \leq x \leq \cos(\theta)$. This is an integral,

$$f(\theta) = \int_{-\cos(\theta)}^{\cos(\theta)} \sqrt{1 - x^2} dx.$$

The problem is to describe the rate-of-change of f, $df/d\theta$.

The integral $f(\theta)$ is beyond our current techniques of integration (though soon we will have techniques to solve it). The simplest solution is indirect. Here, first, is the direct solution. The integral $f(\theta)$ equals the area of 2 triangles and a circular sector. By high-school geometry, the area is,

$$f(\theta) = \frac{\pi - 2\theta}{2} + 2(\frac{1}{2}\sin(\theta)\cos(\theta)) = \pi/2 - \theta + \frac{1}{2}\sin(2\theta).$$

Using standard rules of differentiation, the derivative is,

$$\frac{df}{d\theta} = -1 + \cos(2\theta).$$

Notice, by the double-angle formula for cosine, this equals,

$$-1 + \cos(2\theta) = \boxed{-2\sin^2(\theta).}$$

The hardest step (hidden here) was the geometric computation of $f(\theta)$. However, this is completely unnecessary. Introduce a function,

$$G(t) = \int_0^t \sqrt{1 - x^2} dx.$$

Using symmetry through the y-axis, $f(\theta)$ equals,

$$f(\theta) = 2G(\cos(\theta)).$$

By the chain rule,

$$\frac{df}{d\theta} = 2\frac{dG}{d\theta} = 2\frac{dG}{dt}\frac{dt}{d\theta} = 2\frac{dG}{dt}\frac{d(\cos(\theta))}{d\theta}.$$

By the Fundamental Theorem of Calculus,

$$\frac{dG}{dt} = \sqrt{1 - t^2}.$$

This gives,

$$\frac{df}{d\theta} = 2\sqrt{1-\cos^2(\theta)}(-\sin(\theta)) = \boxed{-2\sin^2(\theta).}$$

The second method is indirect. The function $G(t)$ has no simple expression. Nonetheless, this method is faster. In many cases this is the only method that works.

The argument above using the chain rule and the Fundamental Theorem of Calculus is quite general. It gives the general equation,

$$\boxed{d/dx \int_{u(x)}^{v(x)} f(t)dt = f(v(x))v'(x) - f(u(x))u'(x).}$$

3. Geometric area and algebraic area. The Riemann integral is the *algebraic area*,

$$\int_a^b f(x)dx = \text{ Area above the } x\text{-axis } - \text{ Area below the } x\text{-axis }.$$

The *geometric area* is the total area, both above and below the x-axis. Although geometric area does not equal algebraic area, it has a simple expression using the Riemann integral,

$$\text{Geometric area } = \int_a^b |f(x)|dx.$$

Example. Find both the algebraic area and the geometric area bounded by the x-axis and the graph of $y = \sin(x)$ over the interval $-\pi < x < \pi$.

Because $\sin(x)$ is an odd function, the area below the x-axis for $-\pi < x < 0$ equals the area above the x-axis for $0 < x < \pi$. In the expression for the algebraic area, these areas cancel to give 0. This is borne out by computation,

$$\int_{-\pi}^{\pi} \sin(x)dx = (-\cos(x)|_{-\pi}^{\pi} = -\cos(\pi) + \cos(-\pi) = -(-1) + (-1) = 0.$$

On the other hand, the absolute value $|\sin(x)|$ equals,

$$|\sin(x)| = \begin{cases} -\sin(x), & -\pi < x \leq 0, \\ \sin(x), & 0 < x < \pi. \end{cases}$$

Thus the geometric area equals,

$$\int_{-\pi}^{0} -\sin(x)dx + \int_0^{\pi} \sin(x)dx =$$
$$(\cos(x)|_{-\pi}^{0} + (-\cos(x)|_0^{\pi} = (1-(-1)) + (-(-1)+1) = \boxed{4.}$$

Thus the geometric area does not equal the algebraic area. But computation of the geometric area reduces to a straightforward Riemann integral.

4. Estimates. For every pair of Riemann integrable functions $f(x), g(x)$ on $[a, b]$ satisfying the inequality $f(x) \leq g(x)$ for every choice of x, the following inequality holds,

$$\int_a^b f(x)dx \leq \int_a^b g(x)dx.$$

This is very useful for estimating integrals.

Example. Determine the following Riemann integral to within $\pm 10^{-4}$,

$$\int_0^{0.1} 1 + \sqrt{\sin(x)}dx.$$

The expression $\sqrt{\sin(x)}$ has no simple antiderivative. The value of the Riemann integral could be approximated well by a Riemann sum. An alternative approach is to use the estimates,

$$(1 - x^2/6)\sqrt{x} \leq \sqrt{\sin(x)} \leq \sqrt{x},$$

for small values of x. This gives,

$$\int_0^{0.1} 1 + x^{1/2} - \frac{1}{6}x^{5/2}dx \leq \int_0^{0.1} 1 + \sqrt{\sin(x)}dx \leq \int_0^{0.1} 1 + x^{1/2}dx.$$

The first and third Riemann integral follow from the Fundamental Theorem of Calculus,

$$\int_0^{0.1} 1+x^{1/2}-\frac{1}{6}x^{5/2}dx = \left(x + \frac{2}{3}x^{3/2} - \frac{1}{21}x^{7/2}\right|_0^{0.1} = 0.1+\frac{2}{3\sqrt{1000}}-\frac{1}{21\sqrt{10000000}} = 0.1210667926 \pm 10^{-10}.$$

Similarly,

$$\int_0^{0.1} 1 + x^{1/2}dx = \left(x + \frac{2}{3}x^{3/2}\right|_0^{0.1} = 0.1 + \frac{2}{3\sqrt{1000}} = 0.1210818511 \pm 10^{-10}.$$

Since these two integrals agree to within $\pm 10^{-4}$, this gives the original integral,

$$\int_0^{0.1} 1 + \sqrt{\sin(x)}dx = \boxed{0.1210 \pm 10^{-4}.}$$

5. Change of variables. After the Fundamental Theorem of Calculus, the most useful integral rule is the change of variables rule. The rule for Riemann integrals is nearly the same as the rule for antiderivatives. The additional feature for Riemann integrals is the change of the *limits of integration.*

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

Example. Find the Riemann integral,

$$\int_{\pi/4}^{\pi/3} \tan(x)dx.$$

Since $\tan(x)$ is not visibly the derivative of another function, we rewrite the integral and hope for the best.

$$\int_{\pi/4}^{\pi/3} \tan(x)dx = \int_{\pi/4}^{\pi/3} \frac{\sin(x)}{\cos(x)}dx.$$

In this form, the substitution $u = \cos(x)$ is natural,

$$\textstyle\int_{x=\pi/4}^{x=\pi/3} \frac{\sin(x)}{\cos(x)}dx,$$

$$\left.\begin{array}{r} u = \cos(x) \\ du = -\sin(x)dx \end{array}\right| \begin{array}{l} u(\pi/3) = \cos(\pi/3) = 1/2, \\ u(\pi/4) = \cos(\pi/4) = 1/\sqrt{2}. \end{array}$$

$$\textstyle\int_{u=1/\sqrt{2}}^{u=1/2} \frac{1}{u}(-du).$$

The new integral can be computed by the Fundamental Theorem of Calculus, since $1/u$ is the derivative of $\ln(u)$.

$$\int_{u=1/\sqrt{2}}^{u=1/2} \frac{-1}{u}du = (-\ln(|u|)|_{1/\sqrt{2}}^{1/2} = -\ln(1/2) + \ln(1/\sqrt{2}) = \ln(2) - \ln(\sqrt{2}).$$

This simplifies to give,

$$\int_{\pi/4}^{\pi/3} \tan(x)dx = \boxed{\ln(2)/2.}$$

It is only fair to note there is a second method. Make the same substitution to simplify the antiderivative of $\tan(x)$ to $-\ln(|u|) + C$, and then back-substitute to get,

$$\int tan(x)dx = -\ln(|\cos(x)|) + C.$$

Now use the Fundamental Theorem of Calculus with the original limits of integration. Both methods are correct. Usually the first method is faster and less error-prone; it requires no back-substitution.

6. Integrating backwards. This comes so naturally for most calculus students, it barely warrants mention. Technically, the Riemann integral,

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

is only defined if $a \leq b$. What if a is larger than b? The only possible answer consistent with the Fundamental Theorem of Calculus is the following,

$$\int_a^b f(x)dx = -\int_b^a f(x)dx, \text{ if } a > b.$$

Because of the central role of the Fundamental Theorem of Calculus, the above equation is true by convention. With this convention, the Fundamental Theorem of Calculus holds whenever a is less than b, equal to b, or greater than b.

Lecture 17. October 21, 2005

Homework. Problem Set 5 Part I: (a) and (b); Part II: Problem 1.

Practice Problems. Course Reader: 3F-1, 3F-2, 3F-4, 3F-8.

1. Ordinary differential equations. An ordinary differential equation is an equation involving a single independent variable x together with a dependent variable y and its derivatives d^ky/dx^k,

$$G\left(x, y, \frac{dy}{dx}, \frac{d^2y}{dx^2}, \dots, \frac{d^ky}{dx^k}\right) = 0.$$

The largest k for which d^ky/dx^k occurs in the equation is called *order* of the differential equation.

Examples. Here are examples of ordinary differential equations.

(i) The ordinary differential equation,

$$y - \sin(x^2) = 0,$$

has order 0, because no derivatives of y actually occur in the equation. It has a unique (and rather trivial) solution,

$$y = \sin(x^2).$$

Because the solution is unique, it depends on 0 parameters (and the order is 0).

(ii) The ordinary differential equation,

$$\frac{dy}{dx} - \frac{1}{x+1} = 0,$$

has order 1 because dy/dx occurs and no higher derivatives occur. Every solution is an antiderivative of $1/x + 1$,

$$y = \int \frac{1}{x+1}dx = \ln(|x+1|) + C,$$

Notice the solution depends on 1 parameter, C. And the order is 1.

(iii) The ordinary differential equation,

$$\frac{d^2y}{dx^2} + \omega^2 y = 0,$$

has order 2. The general solution was found in Problem Set 2, Problem 4,

$$y = A\cos(\omega x) + B\sin(\omega x).$$

The solution depends on 2 parameters, A and B. And the order is 2.

(iv) The previous equation was one particular linear ordinary differential equation. A k^{th} *order linear ordinary differential equation* has the form,

$$a_k(x)\frac{d^k y}{dx^k} + a_{k-1}(x)\frac{d^{k-1}y}{dx^{k-1}} + \cdots + a_1(x)\frac{dy}{dx} + a_0(x)y = b(x),$$

for functions $a_k(x), \ldots, a_0(x), b(x)$. If $b(x)$ is zero, the equation is *homogeneous*. Otherwise it is *inhomogeneous*. Very important is the case when all the functions $a_k(x), \ldots, a_0(x), b(x)$ are constant. Then the differential equation is called *constant coefficient*. The solution of constant coefficient linear ordinary differential equations is a main focus of Math 18.03.

2. Separable differential equations. Many differential equations arising in applications are examples of separable differential equation. A *separable ordinary differential equation* is a first-order differential equation,

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

for which $f(x, y)$ factors as,

$$F(x, y) = g(x)/h(y).$$

Example. Find the equation $y = f(x)$ of every curve with the following property: For every point (x, y) on the curve, the tangent line to the curve is perpendicular to the line joining (x, y) to the origin $(0, 0)$.

The slope of the tangent line to the curve at (x, y) is dy/dx. The slope of the line joining $(0, 0)$ and (x, y) is y/x. Since the tangent line is perpendicular to the line joining $(0, 0)$ and (x, y),

$$\frac{dy}{dx} = -x/y.$$

Thus, the equation $y = f(x)$ is a solution to this separable differential equation.

The algorithm for solving a separable differential equation is the following.

(i). Factor $f(x, y)$ as $g(x)/h(y)$. This is often the most difficult step. In the example, it is quite easy. Simply take $g(x) = -x$ and $h(y) = y$.

(ii). Rewrite the differential equation as an equality of differentials. In other words, rewrite the equation as,

$$\frac{dy}{dx} = \frac{g(x)}{h(y)} \Rightarrow h(y)dy = g(x)dx.$$

In the example, this gives,

$$\frac{dy}{dx} = \frac{-x}{y} \Rightarrow ydy = -xdx.$$

(iii). Antidifferentiate both sides of the equation. In the example, the antiderivatives

$$\int ydy = \int -xdx,$$

give,

$$\frac{1}{2}y^2 = \frac{-1}{2}x^2 + C.$$

(iv). If there is an inital value, use it to find the constant of integration. An *initial value problem* is an ordinary differential equation together with some information for an initial value x_0 of the independent variable. It is often written,

$$\begin{cases} dy/dx = F(x, y), \\ y(x_0) = y_0. \end{cases}$$

The example was not an initial value problem. However, it can easily be made an initial value problem by specifying,

$$y(1) = 1,$$

for instance. With this condition, the constant C satisfies the equation,

$$\frac{1}{2}(1)^2 = \frac{-1}{2}(1)^2 + C.$$

The solution is,

$$C = 1.$$

(v). Simplify the answer. Often it is best to solve for $y = f(x)$. Often this is unnecessary. It depends on the problem. In the example problem, the simplest answer is the implicit answer,

$$\boxed{x^2 + y^2 = 2C.}$$

So the solution of the initial value problem is,

$$\boxed{x^2 + y^2 = 2.}$$

Thus every curve satisfying the geometric property is a circle centered at the origin.

Example. Here is a somewhat different example. There is a single separable ordinary differential equation satisfied by every function,

$$y = (x - a)^3,$$

where a is an arbitrary constant. Find this differential equation, and find all its solutions.

The derivative of y is,

$$\frac{dy}{dx} = 3(x-a)^2.$$

The constant a can be eliminated by writing this as,

$$\frac{dy}{dx} = 3[(x-a)^3]^{2/3} = 3y^{2/3}.$$

This is a separable differential equation,

$$\boxed{dy/dx = 3y^{2/3}.}$$

The algorithm gives,

$$\begin{array}{c} y^{-2/3}dy = 3dx, \\ \int y^{-2/3}dy = \int 3dx, \\ 3y^{1/3} = 3x + C. \end{array}$$

Calling the constant $-3a$ gives the answer,

$$y = (x-a)^3.$$

However, there are **other** solutions. For instance, $y = 0$ is a solution. The general solution of the differential equation depends on 2 parameters, $a < b$,

$$y = \begin{cases} (x-a)^3, & x \leq a, \\ 0, & a < x \leq b, \\ (x-b)^3, & x > b \end{cases}$$

The problem is that in the step giving $dy/y^{2/3} = dx$. If y equals 0, this equation involves division by zero. Division by zero is not allowed, so the method breaks down.

Important fact. This fact will not be used in this class. However, it is often crucial in real-world applications to know the solution to an initial value problem is unique. The fact is,

$$\begin{cases} \frac{dy}{dx} = F(x,y), \\ y(x_0) = y_0, \end{cases}$$

has a unique solution for x close to x_0 if $F(x,y)$ is both continuous and differentiable at (x_0, y_0). In the previous example, $F(x,y) = 3y^{2/3}$ is continuous at $y_0 = 0$. But it is not differentiable at $y_0 = 0$. Ultimately, this is the reason for the extra solutions of the differential equation.

3. Applications. Separable differential equations come up often in applications. The most common separable differential equation is the equation for *exponential growth*,

$$\frac{dy}{dt} = ky,$$

where k is a constant.

The solution behaves differently if k is positive or negative. For k positive, this equation arises in population growth and interest on savings, among others. For k negative, this equation arises in radioactive decay, a discharging capactior in an RC-circuit, and Newton's law of cooling.

Population growth. The simplest model of population growth is that a population $N(t)$ (modeled as continuous for simplicity) grows at a rate proportional to the size of the population. This gives,

$$\frac{dN}{dt} = kN.$$

Following the method gives,

$$\begin{array}{c} dN/N = kdt, \\ \int 1/N dN = \int kdt, \\ \ln(|N|) = kt + C. \end{array}$$

Exponentiating both sides gives,

$$N(t) = \boxed{N_0 e^{kt}.}$$

Observe that $N(t)$ increases without bound as t increases. When N is very large, the ecosystem cannot support such a population. Thus the model is only valid if $N(t)$ is not too large.

A slightly more realistic model hypothesizes a constant, equilibrium population N_{equi} sustainable indefinitely. The model is that the population grows at a rate proportional both to the population N and the difference $N_{\text{equi}} - N$,

$$\frac{dN}{dt} = kN(N_{\text{equi}} - N).$$

This is again a separable differential equation. It gives the solution,

$$N(t) = \boxed{N_0 N_{\text{equi}}/(N_0 + (N_{\text{equi}} - N_0)e^{-kN_{\text{equi}}t}).}$$

The most important feature is that $N(t)$ approaches N_{equi} as t increases. This is called the *steady-state solution*. In general, to find the steady-state solution to a separable ordinary differential equation, assume the solution is constant $y = y_1$ so that dy/dt is 0. In the original model of population growth, the only steady-state solution is $N = 0$. In the new model, there are 2 steady-state solutions, $N = 0$ and $N = N_{\text{equi}}$. In Math 18.03, stability is defined, and a method is given to show the only stable steady-state solution is $N = N_{\text{equi}}$.

Radioactive decay. A radioactive isotope decays to a more stable isotope at a rate proportional to the remaining radioactive isotope. Thus the mass $m(t)$ satisfies a differential equation,

$$\frac{dm}{dt} = -km.$$

Using the method, the solution is,

$$m(t) = \boxed{m_0 e^{-kt}.}$$

An important feature in decay problems is the half-life. The *half-life* is the length of time necessary for the mass of radioactive isotope to decrease to one-half the initial mass,

$$m(T_{\text{half}}) = m_0/2.$$

Solving in the formula gives,

$$T_{\text{half}} = \boxed{\ln(2)/k.}$$

Example. The half-life of a certain radioactive isotope is 20 years. How long is required for the mass to decrease to 1% of the initial mass? Using the formula above, $k = \ln(2)/25$. Therefore the equation for the mass is,

$$m(t) = m_0 e^{-\ln(2)t/25}.$$

Thus the time t_f when the mass equals $0.01m_0$ satisfies,

$$m_0 e^{-\ln(2)t_f/25} = m_0/100,$$

or,

$$\ln(2)t_f/25 = \ln(100) = 2\ln(10).$$

Solving gives,

$$t_f = 50\ln(10)/\ln(2) = \boxed{166 \text{ years.}}$$

Newton's Law of Cooling. Isaac Newton proposed a law for the rate-of-change of the temperature T of an object placed in a large, effectively infinite, environment at a fixed ambient temperature T_{amb}. The law is that the rate-of-change of T is negatively proportional to the temperature gradient $T - T_{\text{amb}}$,

$$\frac{dT}{dt} = -k(T - T_{\text{amb}}).$$

The method gives the solution,

$$T(t) = T_{\text{amb}} + (T - T_{\text{amb}})e^{-kt}.$$

As t increases, the temperature T approaches the steady-state temperature, T_{amb}.

Lecture 18. October 25, 2005

Homework. Problem Set 5 Part I: (c).

Practice Problems. Course Reader: 3G-1, 3G-2, 3G-4, 3G-5.

1. Approximating Riemann integrals. Often, there is no simpler expression for the antiderivative than the expression given by the Fundamental Theorem of Calculus. In such cases, the simplest method to compute a Riemann integral is to use the definition. However, this is not necessarily the most *efficient* method. Often trapezoids or segments under a parabola give a better approximation to the Riemann integral than do vertical strips.

2. The trapezoid rule. The problem is to find an approximation of the Riemann integral,

$$I = \int_a^b y dx$$

for a function $y(x)$ defined on the interval $[a, b]$. Choose a partition of the interval $[a, b]$ into n equal subintervals. The points of this partition are,

$$x_k = a + \frac{(b-a)k}{n}, \quad \Delta x_k = \frac{b-a}{n}.$$

The values of these points are,

$$y_k = f(x_k).$$

The Riemann sum using always the left endpoint is,

$$I_l = \sum_{k=1}^{n} y_{k-1} \Delta x_k.$$

The Riemann sum using always the right endpoint is,

$$I_r = \sum_{k=1}^{n} y_k \Delta x_k.$$

The average of the two is,

$$I_{\text{trap}} = \sum_{k=1}^{n} \frac{y_{k-1} + y_k}{2} \Delta x_k.$$

This is usually a better approximation than either of the two approximations individually. Part of the reason is that the term $(y_{k-1} + y_k)\Delta x_k/2$ is the area of the *trapezoid* containing the points $(x_{k-1}, 0)$, (x_{k-1}, y_{k-1}), $(x_k, 0)$ and (x_k, y_k). In particular, if the graph of $y = f(x)$ is a line, this trapezoid is precisely the region between the graph and the x-axis over the interval $[x_{k-1}, x_k]$. Thus, the approximation above gives the *exact* integral for linear integrands.

Writing out the sum gives,

$$I_{\text{trap}} = \frac{b-a}{2n}((y_0 + y_1) + (y_1 + y_2) + (y_2 + y_3) + \cdots + (y_{n-2} + y_{n-1}) + (y_{n-1} + y_n)).$$

Gathering like terms, this reduces to,

$$I_{\text{trap}} = \boxed{(b-a)(y_0 + 2y_1 + 2y_2 + \cdots + 2y_{n-1} + y_n)/2n.}$$

3. Simpson's rule. Again partition the interval $[a, b]$ into n equal subintervals. For reasons that will become apparent, n must be even. So let $n = 2m$ where m is a positive integer. Again define,

$$x_k = a + \frac{(b-a)k}{n} = a + \frac{(b-a)k}{2m}, \ \Delta x_k = \frac{b-a}{n} = \frac{b-a}{2m}.$$

Pair off the intervals as $([x_0, x_1], [x_1, x_2])$, $([x_2, x_3], [x_3, x_4])$, etc. Thus the l^{th} pair of intervals is,

$$([x_{2l-2}, x_{2l-1}], [x_{2l-1}, x_{2l}]).$$

The idea is to approximate the area of the graph over the pair of intervals by the area under the unique parabola containing the 3 points $(x_{2l-2}, y_{2l-2}), (x_{2l-1}, y_{2l-1}), (x_{2l}, y_{2l})$. For notation's sake, denote $2l-1$ by k. Thus the 3 points are (x_{k-1}, y_{k-1}), (x_k, y_k), and (x_{k+1}, y_{k+1}) (this is slightly more symmetric).

The first problem is to find the equation of this parabola. Since the parabola contains the point (x_k, y_k), it has the equation,

$$y = A(x - x_k)^2 + B(x - x_k) + y_k,$$

Plugging in $x = x_{k-1}$ and $x = x_{k+1}$, and using that $x_{k+1} - x_k = x_k - x_{k-1}$ equals Δx,

$$y_{k+1} = A(\Delta x)^2 + B(\Delta x) + y_k,$$

$$y_{k-1} = A(\Delta x)^2 - B(\Delta x) + y_k.$$

Summing the two sides gives,

$$y_{k+1} + y_{k-1} = 2A(\Delta x)^2 + 2y_k.$$

Solving for A gives,

$$A = \frac{1}{2(\Delta x)^2}(y_{k-1} - 2y_k + y_{k+1}).$$

Similarly, taking the difference of the two sides gives,

$$y_{k+1} - y_{k-1} = 2B(\Delta x).$$

Solving for B gives,

$$B = \frac{1}{2(\Delta x)}(y_{k+1} - y_{k-1}).$$

Thus, the equation of the parabola passing through (x_{k-1}, y_{k-1}), (x_k, y_k) and (x_{k+1}, y_{k+1}) is,

$$y = A(x - x_k)^2 + B(x - x_k)^2 + y_k,$$
$$A = (y_{k-1} - 2y_k + y_{k+1})/2(\Delta x)^2,$$
$$B = (y_{k+1} - y_{k-1})/2(\Delta x).$$

The next problem is to compute the area under the parabola from $x = x_{k-1}$ to $x = x_{k+1}$. This is a straightforward application of the Fundamental Theorem of Calculus,

$$\int_{x_{k-1}}^{x_{k+1}} A(x - x_k)^2 + B(x - x_k) + y_k dx = \left(\frac{A}{3}(x - x_k)^3 + \frac{B}{2}(x - x_k)^2 + y_k(x - x_k)\right|_{x_{k-1}}^{x_{k+1}}.$$

Plugging in and using that $x_{k+1} - x_k = x_k - x_{k-1}$ equals Δx, this is,

$$\frac{2A}{3}(\Delta x)^3 + 2y_k(\Delta x).$$

Substituting in the formula for A and simplifying, this is,

$$\frac{\Delta x}{3}(y_{k-1} - 2y_k + y_{k+1}) + \frac{\Delta x}{3}(6y_k) = \frac{\Delta x}{3}(y_{k-1} + 4y_k + y_{k+1}).$$

Back-substituting $2l-1$ for k and $(b-a)/2m$ for Δx, the approximate area for the pair of intervals $[x_{2l-2}, x_{2l-2}]$ and $[x_{2l-1}, x_{2l}]$ is,

$$\Delta I_l = \frac{b-a}{6m}(y_{2l-2} + 4y_{2l-1} + y_{2l}).$$

Finally, summing this contribution over each choice of l gives the Simpson's rule approximation,

$$I_{\text{Simpson}} = \frac{b-a}{6m}\sum_{l=1}^{m}(y_{2l-2} + 4y_{2l-1} + y_{2l}).$$

Writing out the sum gives,

$$I_{\text{Simpson}} = \tfrac{b-a}{6m}((y_0 + 4y_1 + y_2) + (y_2 + 4y_3 + y_4) + (y_4 + 4y_5 + y_6) + \cdots + (y_{2m-4} + 4y_{2m-3} + y_{2m-2}) + (y_{2m-2} + 4y_{2m-1} + y_{2m})).$$

Gathering like terms, I_{Simpson} reduces to,

$$(b-a)(y_0 + 4y_1 + 2y_2 + 4y_3 + 2y_4 + 4y_5 + 2y_6 + \cdots + 4y_{2m-3} + 2y_{2m-2} + 4y_{2m-1} + y_{2m})/6m.$$

Example. Approximate $\ln(2)$ using a partition into 4 equal subintervals with the Trapezoid Rule and with Simpson's Rule.

The value $\ln(2)$ equals the Riemann integral,

$$\int_1^2 \frac{1}{x}dx.$$

The points of the partition are $x_0 = 4/4, x_1 = 5/4, x_2 = 6/4, x_3 = 7/4$ and $x_4 = 8/4$. The corresponding values are $y_0 = 4/4, y_1 = 4/5, y_2 = 4/6, y_3 = 4/7, y_4 = 4/8$. Thus the Trapezoid Rule gives,

$$I_{\text{trap}} = \frac{b-a}{2n}(y_0 + 2y_1 + 2y_2 + 2y_3 + y_4) = \frac{1}{8}(\frac{4}{4} + 2\frac{4}{5} + 2\frac{4}{6} + 2\frac{4}{7} + \frac{4}{8}) = \frac{1171}{1680} \approx 0.6970$$

For Simpson's Rule, because n equals 4, m equals 2. Thus,

$$I_{\text{Simpson}} = \frac{b-a}{6m}(y_0 + 4y_1 + 2y_2 + 4y_3 + y_4) = \frac{1}{12}(\frac{4}{4} + 4\frac{4}{5} + 2\frac{4}{6} + 4\frac{4}{7} + \frac{4}{8}) = \frac{1747}{2520} \approx 0.6933$$

According to a calculator, the true value is,

$$\ln(2) = \boxed{0.6931 \pm 10^{-4}}$$

Note that trapezoids overestimate the area, because $1/x$ is concave up. The approximating parabolas cross the graph of $y = 1/x$, thus the underestimation to the left of (x_k, y_k) somewhat cancels the overestimation to the right of (x_k, y_k), explaining the better approximation.

4. One review problem. This is a related rates review problem for Exam 3. A particle moves with constant speed 3 on the parabola $y = x^2$. The particle is moving away from the origin. What is the rate-of-change of the distance from the origin to the particle when the distance equals $2\sqrt{5}$?

The independent variable is time, t. The dependent variables are the x-coordinate of the particle, $x(t)$, the y-coordinate of the particle, $y(t)$, and the distance $L(t)$ from the particle to $(0, 0)$. The constant is the speed $s = 3$ of the particle. The constraints are that the point moves on the parabola,

$$y = x^2,$$

and the Pythagorean theorem,

$$L^2 = x^2 + y^2.$$

Also, since the speed is constant,

$$s^2 = \left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2.$$

This plays the role of the "known rate-of-change" in a typical related rates problem.

It is simplest to relate the dependent variables y and L to x. The first step is to determine x at the moment when L equals $2\sqrt{5}$. Plugging $y = x^2$ into the equation for L^2 gives,

$$L^2 = x^2 + y^2 = x^2 + (x^2)^2 = x^2 + x^4.$$

At the instant when L equals $2\sqrt{5}$, L^2 equals 20. Thus, at that moment,

$$x^4 + x^2 = 20.$$

This factors as,

$$(x^2 - 4)(x^2 + 5) = 0.$$

Since x^2 is nonnegative, the solution is $x^2 = 4$. Assuming the particle is in the first quadrant (this is not specified in the problem), x is positive. The other choice leads to a symmetric problem and the same final answer. So, at the moment when L equals $2\sqrt{5}$, x equals 2.

The next step is to determine the "known rate-of-change", dx/dt at the moment when L equals $2\sqrt{5}$. Implicitly differentiating the equation $y = x^2$ gives,

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

Substituting this into the equation for s^2 gives,

$$s^2 = \left(\frac{dx}{dt}\right)^2 + \left(2x\frac{dx}{dt}\right)^2 = (1+4x^2)\left(\frac{dx}{dt}\right)^2.$$

Since s is known to be 3, and x is known to be 2, this equation can be solved for dx/dt,

$$\left(\frac{dx}{dt}\right)^2 = \frac{3^2}{1+4(2)^2} = \frac{9}{17}.$$

Since the particle is in the first quadrant and moving *away* from the origin, dx/dt is positive. So, at the moment when L equals $2\sqrt{5}$, dx/dt equals $3/\sqrt{17}$.

The final step is to compute dL/dt at the moment when L equals $2\sqrt{5}$. Implicitly differentiating the equation,

$$L^2 = x^2 + x^4,$$

gives,

$$2L\frac{dL}{dt} = (2x + 4x^3)\frac{dx}{dt}.$$

Plugging in for L, x and dx/dt gives,

$$2(2\sqrt{5})\frac{dL}{dt} = (2(2) + 4(2)^3)\frac{3}{\sqrt{17}}.$$

Solving gives,

$$\frac{dL}{dt} = \boxed{27/\sqrt{85}}.$$

at the moment when L equals $2\sqrt{5}$.

Lecture 19. October 28, 2005

Homework. Problem Set 5 Part I: (d) and (e); Part II: Problems 2 and 3.

Practice Problems. Course Reader: 4A-1, 4A-3, 4B-1, 4B-3, 4B-6.

1. Differentials revisited. In a typical applied integration problem, the main difficulty is finding the integrand and the limits of integration. An unknown quantity, for instance area A, depends on some other quantity, for instance the x-coordinate. The method is to allow the independent variable x vary "infinitesimally" from x to $x + dx$ and then use geometry or science to deduce the corresponding variation dA of the unknown quantity. The deduction is usually intuitive rather than rigorous. What is important is whether the deduction leads to the correct solution. If so, the method of Riemann sums usually gives a rigorous proof of the intuitive answer. But if the solution is incorrect, no argument will prove it correct.

2. Areas between curves. Given an interval $a \leq x \leq b$ and two functions $f(x) \geq g(x)$ defined on the interval, what is the area of the region bounded by the lines $x = a, x = b$ and the curves

$y = f(x)$, $y = g(x)$? This problem can be solved directly: the area is the difference of the area between $y = f(x)$ and the x-axis and the area between $y = g(x)$ and the x-axis. Each of these is a Riemann integral.

The differential method asks, what is the infinitesimal change in the area A as x varies from x to $x + dx$? The infinitesimal region is a rectangle of base dx and height $f(x) - g(x)$. Thus the infinitesimal change in A is,

$$dA = \text{height} \times \text{base} = (f(x) - g(x))dx.$$

Integrating gives,

$$A = \int dA = \int_{x=a}^{x=b} f(x) - g(x)dx.$$

Of course this is the same answer as in the last paragraph. But for other applied integral problems, the differential method may be the only method that works.

Example. Find the area bounded by the curve $y = x(x^2 - 3)$ and a horizontal tangent line.

The horizontal tangent lines are the tangent lines to the curve at critical points. Setting the derivative equal to 0 gives,

$$\frac{dy}{dx} = 3x^2 - 3 = 3(x - 1)(x + 1).$$

Thus the critical points are $x = \pm 1$. The function is odd, so symmetry suggests the area is the same regardless of the choice of critical point. Thus, choose the critical point $x = -1$. The corresponding value of the function is,

$$y = (-1)((-1)^2 - 3) = (-1)(-2) = 2.$$

This is the equation of the horizontal tangent line. Each intersection point $(b, f(b))$ of the tangent line with $y = x(x^2 - 3)$ occurs at a solution $x = b$ of,

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

By hypothesis, $x = -1$ is a solution. Thus the polynomial factors as,

$$x^3 - 3x - 2 = (x + 1)(x^2 - x - 2) = (x + 1)^2(x - 2).$$

The remaining intersection point is $(2, 2)$. So the area bounded by the curve $y = x(x^2 - 3)$ and the tangent line $y = 2$ is,

$$\int_{x=-1}^{x=2} 2 - (x(x^2 - 3))dx = \int_{-1}^{2} -x^3 + 3x + 2dx.$$

Using the Fundamental Theorem of Calculus, this equals,

$$\left(\frac{-x^4}{4} + \frac{3x^2}{2} + 2x \right|_{-1}^{2} = \boxed{27/4.}$$

3. Volumes of solids of revolutions: the disk method. If the region in the xy-plane bounded by $x = a$, $x = b$, $y = f(x)$ and the x-axis is revolved through xyz-space about the x-axis, what is the volume of the resulting solid? The solid is called a *solid of revolution*. The *disk method* applies the method of differentials to solve this problem. As x increases from x to $x+dx$, the corresponding infinitesimal region of the solid is essentially a disk. The width of the disk is dx. The area of the disk is $\pi[f(x)]^2$. Thus the infinitesimal volume of the disk is,

$$dV = \text{Area} \times \text{width} = \pi[f(x)]^2 dx.$$

Thus the volume is,

$$V = \int dV = \int_{x=a}^{x=b} \pi[f(x)]^2 dx.$$

Example. Find the volume of a right circular cone whose base has radius R and whose vertex has height H above the base.

The cone is the solid of revolution for the plane region bounded by $x = 0$, the x-axis, and the line containing $(0, R)$ and $(H, 0)$. The equation of this line is,

$$y = \frac{(H-x)R}{H}.$$

Thus the area of an infinitesimal disk is,

$$dV = \text{Area} \times \text{width} = \pi\frac{(H-x)^2R^2}{H^2}dx,$$

and the volume is,

$$V = \int dV = \int_{x=0}^{x=H} \pi\frac{(H-x)^2R^2}{H^2}dx.$$

Making the substitution $u = H - x$, $du = -dx$ gives,

$$V = \int_{u=H}^{u=0} \pi\frac{R^2}{H^2}u^2(-du) = \pi\frac{R^2}{H^2}\left(-\frac{u^3}{3}\right|_H^0.$$

Evaluating gives the volume,

$$V = \boxed{\pi R^2 H/3.}$$

In particular, the antiderivative of u^2 is responsible for the denominator 3 in the formula for the volume.

Example. Find the volume of a sphere of radius R.

The sphere of radius R is the solid of revolution for the plane region bounded by the x-axis and the upper semicircle $y = \sqrt{R^2 - x^2}$. Thus the volume is,

$$V = \int_{x=-R}^{x=R} \pi[\sqrt{R^2-x^2}]^2 dx = \int_{-R}^{R} \pi(R^2 - x^2)dx = \pi\left(R^2x - \frac{x^3}{3}\right|_{-R}^{R}.$$

Evaluating gives the volume,

$$V = \boxed{4\pi R^3/3.}$$

4. The slice method. A generalization of the disk method is the *slice method*. The problem is to find the volume of a region bounded by the planes $x = a$ and $x = b$ given the knowledge of the area $A(x)$ of the *slice* of the solid by the plane containing $(x, 0, 0)$ parallel to the yz-plane. As x increases from x to $x + dx$, the corresponding infinitesimal region of the solid is essentially a cylinder. The width of the cylinder is dx. And the area is the area $A(x)$ of the slice. Thus the infinitesimal volume of the cylinder is,

$$dV = \text{Area} \times \text{width} = A(x)dx.$$

Thus the volume is,

$$V = \int dV = \int_{x=a}^{x=b} A(x)dx.$$

Example. Find the volume of the "corner" region bounded by the xy-plane, the xz-plane, the yz-plane, and the plane containing $(L, 0, 0)$, $(0, L, 0)$ and $(0, 0, L)$.

This region is bounded by the planes $x = 0$ and $x = L$. The x-slice of the region is a right isosceles triangle. The base and altitude of the triangle both equal $f(x)$, where $y = f(x)$ is the equation of the line passing through $(0, L)$ and $(L, 0)$. This equation is,

$$f(x) = L - x.$$

Thus the slice area is

$$A(x) = \frac{1}{2}\text{base} \times \text{altitude} = \frac{1}{2}(L - x)^2.$$

Thus the infinitesimal volume is,

$$dV = A(x)dx = \frac{1}{2}(L - x)^2 dx,$$

giving the total volume,

$$V = \int dV = \int_{x=0}^{x=L} \frac{1}{2}(L - x)^2 dx.$$

Make the substitution $u = L - x$, $du = -dx$ to get,

$$V = \int_{u=L}^{u=0} \frac{1}{2}u^2(-du) = \frac{1}{2}\left(-\frac{u^3}{3}\right|_0^L.$$

Evaluating gives,

$$V = \boxed{L^3/6.}$$

Thus the "corner" takes up one sixth of the corresponding cube of edge length L.

5. Volumes of solids of revolution: the washer method. A variation on the disk method is the *washer method*. A *washer* is the solid obtained by removing from a larger disk a concentric smaller disk of the same width. If the outer radius of the washer is r_o and the inner radius is r_i, then the net area of the washer is,

$$A = \pi r_o^2 - \pi r_i^2 = \pi(r_o^2 - r_i^2).$$

Thus the volume of the washer is,

$$dV = \text{Area} \times \text{width} = \pi(r_o^2 - r_i^2)dx,$$

giving a total volume,

$$V = \int dV = \int_{x=a}^{x=b} \pi(r_o^2 - r_i^2)dx.$$

Example. The main part of a dog dish is a solid of revolution whose radial cross-section is a triangle of height H whose base has inner radius R_i and outer radius R_o. Find the volume of material used to make the dog dish.

Note. This integral was only set-up in lecture. The derivation will be completed in recitation. Here is the complete derivation. Denote by x the height along the altitude of the triangle. Thus x varies from $x = 0$ to $x = H$. When $x = H$, the inner radius and outer radius are each equal to the average $(R_i + R_o)/2$ of the two radii. Both radii depend linearly on x.

The equation for the inner radius increases linearly from $r_i = R_i$ at $x = 0$ to $r_i = (R_i + R_o)/2$ at $x = H$. Thus,

$$r_i(x) = R_i \frac{H - x}{H} + \frac{R_i + R_o}{2}\frac{x}{H}.$$

Similarly, the equation for the outer radius decreases linearly from $r_o = R_o$ at $x = 0$ to $r_i = (R_i + R_o)/2$ at $x = H$. Thus,

$$r_o(x) = R_o \frac{H - x}{H} + \frac{R_i + R_o}{2}\frac{x}{H}.$$

Since $r_o^2 - r_i^2$ is a difference of squares, it equals,

$$r_o^2 - r_i^2 = (r_o + r_i)(r_o - r_i).$$

This is interesting in its own right. Using this factorization, the net area of the region between the 2 circles, called an *annulus*, equals

$$\pi(r_o^2 - r_i^2) = \left(2\pi\frac{r_o + r_i}{2}\right)(r_o - r_i).$$

Note the first factor is the circumference of the center of the annulus. And the second factor is the radial width of the annulus. Thus the area of an annulus is the circumference of the center times the radial width.

Back to the problem, the center of each washer is the same,

$$\frac{r_i + r_o}{2} = \frac{R_i + R_o}{2}.$$

And the radial width of each washer is,

$$r_o - r_i = (R_o - R_i)\frac{H - x}{H}.$$

Both of these make sense: The centers are constant because the radial cross-section is an isosceles triangle. And the width decreases linearly from $R_o - R_i$ at $x = 0$ to 0 at $x = H$. Thus the cross-section area at height x is,

$$A(x) = 2\pi\frac{R_i + R_o}{2}(R_o - R_i)\frac{H - x}{H}.$$

Thus the infinitesimal volume is,

$$dV = \text{Area} \times \text{width} = \pi(R_o^2 - R_i^2)\frac{H - x}{H}dx,$$

giving a total area,

$$V = \int dV = \int_{x=0}^{x=H} \pi(R_o^2 - R_i^2)\frac{H - x}{H}dx.$$

Substituting $u = H - x$, $du = -dx$ gives,

$$V = \int_{u=0}^{u=H} \pi(R_o^2 - R_i^2)\frac{u}{H}dx = \frac{\pi(R_o^2 - R_i^2)}{2H}\left(u^2\right|_0^H.$$

Thus the total volume of material to produce the dog dish is,

$$V = \boxed{\pi(R_o^2 - R_i^2)H/2.}$$

One reality check is that this is the same volume as a cylinder with the same center $(R_i + R_o)/2$ and height H as the dish, and whose (constant) radial width equals the average radial width of the dish, $(R_o - R_i)/2$.

Lecture 20. November 1, 2005

Practice Problems. Course Reader: 4C-2, 4C-6, 4D-1, 4D-4, 4D-8.

1. Average values. Given a function $f(x)$ defined on some interval $[a, b]$, what is the average value of $f(x)$? A reasonable first approximation is to choose a finite collection of points from $[a, b]$ and compute the average value over those points. Break $[a, b]$ into a union of n subintervals of length $\Delta x = (b - a)/n$. From each interval, choose a point; say x_k^* in the k^{th} interval. For the finitely many values $y_k^* = f(x_k^*)$, the average value is,

$$\text{Average} \approx \frac{1}{n}\sum_{k=1}^{n} y_k^*.$$

Multiplying and dividing by Δx gives,

$$\text{Average} \approx \frac{1}{n\Delta x}\sum_{k=1}^{n} y_k^* \Delta x.$$

Now $n\Delta x$ equals $n(a-b)/n$, which is $a-b$. So the average value is,

$$\text{Average} \approx \frac{1}{b-a}\sum_{k=1}^{n} y_k^* \Delta x.$$

The sum is a Riemann sum. To get better approximations to the true average, increase the number of points n at which $f(x)$ is "sampled". In the limit, this gives the true average,

$$\text{Average} = \frac{1}{b-a}\lim_{n\to\infty}\sum_{k=1}^{n} y_k^* \Delta x = \boxed{\int_a^b f(x)dx/(b-a).}$$

Example. Under ideal conditions, a wire-producing machine produces wire of uniform radius r_0. Because of small vibrations in the machine, the actual radius of the wire varies as a function of the length,

$$r(x) = r_0 + A\cos(\omega x).$$

The quantity A is much smaller than r_0. What is the average radius of the wire?

Because the variation is periodic, the average value over any number of periods equals the average value of one period. In other words, compute the average for the interval $0 \leq x \leq 2\pi/\omega$. The length of this interval is $2\pi/\omega$. Thus the average value is,

$$\text{Average} = \frac{1}{(2\pi/\omega)}\int_0^{2\pi/\omega} r_0 + A\cos(\omega x)dx.$$

Using the Fundamental Theorem of Calculus, this equals,

$$\frac{1}{(2\pi/\omega)}\left(r_0 x + (A/\omega)\sin(\omega x)\right|_0^{2\pi/\omega}.$$

This evaluates to,

$$\frac{1}{(2\pi/\omega)} r_0(2\pi/\omega) = r_0.$$

Thus, although the radius varies and does not usually equal its ideal value r_0, the average value is indeed,

$$\text{Average} = \boxed{r_0.}$$

2. Average values: non-uniform distribution. It often happens that the average value of $f(x)$ is desired in a situation where the values $f(x)$ are not all uniformly likely. Typically, the probability that x has value in the range from x_0 to $x_0 + \Delta x$ is approximately,

$$\text{Prob}(x_0 \leq x \leq x_0 + \Delta x) \approx \rho(x_0)\Delta x,$$

for some nonnegative continuous function $\rho(x)$. The function $\rho(x)$ is called a *probability distribution*. Assuming this approximation becomes arbitrarily good as the length Δx approaches zero, the exact probability that x has value in the range x_0 to x_1 is,

$$\text{Prob}(x_0 \leq x \leq x_1) = \int_{x_0}^{x_1} \rho(x)dx.$$

In particular, because x must take value somewhere in the interval $[a, b]$, the total probability is 1. In other words,

$$\int_a^b \rho(x)dx = 1.$$

This is called the *normalization condition*.

The average value is computed as before. But this time, each value $y_k^* = f(x_k^*)$ is weighted by the approximate probability that x takes value in the k^{th} interval, $\rho(x_k^*)\Delta x$. This gives,

$$\text{Average} \approx \sum_{k=1}^{n} f(x_k)^* \rho(x_k)^* \Delta x.$$

In the limit as n goes to ∞, this gives the exact average,

$$\text{Average} = \int_a^b f(x)\rho(x)dx.$$

It must be noted, the probability distribution $\rho(x)$ often does not satisfy the normalization condition. In this case, the formula above is wrong. But it is easily correct,

$$\text{Average} = \boxed{(\textstyle\int_a^b f(x)\rho(x)dx)/(\int_a^b \rho(x)dx).}$$

Example. A particle is fired through a slit and strikes a screen on the other side. Measuring the position on the screen so that the origin is the closest point on the screen to the slit, the probability distribution is empirically observed to be,

$$\rho(x) = Ce^{-x^2/2\sigma^2},$$

where σ is a constant determining the "width" of the probability distribution, and C is an undetermined normalization constant. What is the average distance of the particle from the center of the screen? Assume the particle lies in an interval $[-R, R]$, where R is very large.
Remark. This differs from the formula given in lecture, which was $Ce^{-x^2/2\sigma}$ for a particular choice of σ. The formula given here is more "standard". I apologize for any confusion.

The distance function is,

$$f(x) = |x| = \begin{cases} -x, & x < 0 \\ x, & x \geq 0 \end{cases}$$

According to the formula, the average value is,

$$\left(\int_{-R}^{R} f(x)\rho(x)dx\right)/\left(\int_{-R}^{R} \rho(x)dx\right).$$

The numerator is,

$$\int_{-R}^{R} |x|Ce^{-x^2/2\sigma^2}dx.$$

It is easiest to compute this by breaking it into a sum of 2 integrals,

$$\int_{-R}^{0} (-x)Ce^{-x^2/2\sigma^2}dx + \int_{0}^{R} (+x)Ce^{-x^2/2\sigma^2}dx.$$

Make the substitution $u = -x^2/2\sigma^2$, $du = (-x/\sigma^2)dx$ to reduce this to,

$$\int_{-R^2/2\sigma^2}^{0} Ce^u(\sigma^2 du) + \int_{0}^{-R^2/2\sigma^2} Ce^u(-\sigma^2 du) = 2\int_{-R^2/2\sigma^2}^{0} C\sigma^2 e^u du.$$

Using the Fundamental Theorem of Calculus, this equals,

$$2C\sigma^2 \left(e^u\right|_{-R^2/2\Sigma^2}^{0} = 2C\sigma^2(1 - e^{-R^2/2\Sigma^2}).$$

As R becomes large, the quantity $e^{-R^2/2\Sigma^2}$ becomes vanishingly small. Thus, in the limit as R tends to ∞, the numerator equals,

$$\lim_{R\to\infty} \int_{-R}^{R} |x|Ce^{-x^2/2\sigma^2}dx = 2C\sigma^2.$$

Unfortunately, this is not an answer, because the normalization constant C is unknown. The normalization condition is that,

$$C \lim_{R\to\infty} \int_{-R}^{R} e^{-x^2/2\sigma^2}dx = 1.$$

Simplify this by making the substitution, $u = x/\sigma$, $du = dx/\sigma$, and $Q = R/\sigma$ to get,

$$C \lim_{R\to\infty} \int_{-R/\sigma}^{R/\sigma} e^{-u^2/2}\sigma du = C\sigma \lim_{Q\to\infty} \int_{-Q}^{Q} e^{-u^2/2}du.$$

Notice the limit,

$$\lim_{Q\to\infty} \int_{-Q}^{Q} e^{-u^2/2}du,$$

does not depend on σ. It is simply some number. Denoting this number by $1/C_1$, the normalization condition is,

$$C\sigma/C_1 = 1.$$

The solution is that $C = C_1/\sigma$. Plugging this into the formula above, the average distance is,

$$\text{Average distance} = 2C_1\sigma,$$

where,

$$1/C_1 = \lim_{Q\to\infty}\int_{-Q}^{Q} e^{-u^2/2}du.$$

There is a beautiful argument that,

$$C_1 = 1/\sqrt{2\pi}.$$

Unfortunately, we cannot yet prove this. Taking it as true gives the final answer,

$$\text{Average distance} = \boxed{2\sigma/\sqrt{2\pi}.}$$

3. Volumes of solids of revolution: the shell method. An alternative to the disk and washer method is the shell method. A *shell* is the region between 2 cylinders of the same height. If the average radius of the cylinders is r, if the width of the region is w and if the height of the cylinders is h, then the approximate volume of the shell is,

$$\text{Volume} \approx= \text{Circumference} \times \text{height} \times \text{width} = 2\pi rhw.$$

Take the plane region bounded by $x = a$, $x = b$, the x-axis and the curve $y = f(x)$. Revolve this region about the y-axis. (**Please note:** In the disk and washer method, the region was revolved about the x-axis.) To compute the corresponding volume, approximate the region obtained from x to $x + dx$ as a shell. The radius of the shell is x. The height of the shell is $y = f(x)$. The width of the shell is dx. Therefore the differential element of volume is,

$$dV = (2\pi x)(f(x))dx.$$

Integrating gives the volume,

$$V = \int_{x=a}^{x=b} 2\pi x f(x)dx.$$

Example. The dog dish revisited. The main part of a dog dish is a solid of revolution whose radial cross-section is a triangle of height H whose base has inner radius R_i and outer radius R_o. Find the volume of material used to make the dog dish.

The volume was computed using the washer method. This time it will be computed using the shell method. The triangular region is the union of two regions. The first region is bounded by $x = R_i$, $x = (R_i + R_o)/2$, the x-axis, and the line segment,

$$y = \frac{2H}{R_o - R_i}(x - R_i).$$

The second region is bounded by $x = (R_i + R_o)/2$, $x = R_o$, the x-axis, and the line segment,

$$y = \frac{2H}{R_o - R_i}(R_o - x).$$

By the shell method, the volume of the solid of revolution obtained from the first region is,

$$V_1 = \int_{x=R_i}^{x=(R_i+R_o)/2} (2\pi x)\frac{2H}{R_o - R_i}(x - R_i)dx = \frac{4\pi H}{R_o - R_i}\int_{x=R_i}^{x=(R_i+R_o)/2} x^2 - R_i x dx.$$

This becomes simpler to deal with after the substitution $u = -x + (R_i + R_o)/2$, $du = -dx$. The new integral is,

$$\begin{aligned} V_1 &= \frac{4\pi H}{R_o - R_i}\int_{u=(R_o-R_i)/2}^{u=0} (-u + (R_o + R_i)/2)(-u + (R_o - R_i)/2)(-du) \\ &= \frac{4\pi H}{R_o - R_i}\int_{u=0}^{u=(R_o-R_i)/2} (-u + (R_o + R_i)/2)(-u + (R_o - R_i)/2)du. \end{aligned}$$

By the shell method, the volume of the solid of revolution obtained from the second region is,

$$V_2 = \int_{x=(R_o+R_i)/2}^{x=R_o} (2\pi x)\frac{2H}{R_o - R_i}(R_o - x)dx = \frac{4\pi H}{R_o - R_i}\int_{x=(R_o+R_i)/2}^{x=R_o} x(R_o - x)dx.$$

Believe it or not, this will be simpler to deal with after the substitution $u = x - (R_o + R_i)/2$, $du = dx$. The new integral is

$$V_2 = \frac{4\pi H}{R_o - R_i}\int_{u=0}^{u=(R_o-R_i)/2} (u + (R_o + R_i)/2)(-u + (R_o - R_i)/2)du.$$

Notice how similar are the integrals for V_1 and V_2. They have the same fraction in front of the integral, and they have the same limits of integration. Thus, the sum of the 2 volumes is,

$$V = V_1 + V_2 =$$

$$\frac{4\pi H}{R_o - R_i}\int_{u=0}^{u=(R_o-R_i)/2} [(-u+(R_o+R_i)/2)(-u+(R_o-R_i)/2)]+[(u+(R_o+R_i)/2)(-u+(R_o-R_i)/2)]du.$$

Since both terms in the integrand have the factor $(-u + (R_o - R_i)/2)$, this can be factored to give,

$$V = \frac{4\pi H}{R_o - R_i}\int_{u=0}^{u=(R_o-R_i)/2} [(-u + (R_o + R_i)/2) + (u + (R_o + R_i)/2)](-u + (R_o - R_i)/2)du.$$

Of course the term in square brackets is simply $R_o + R_i$. So the total volume is,

$$V = \frac{4\pi H}{R_o - R_i}\int_{u=0}^{u=(R_o-R_i)/2} (R_o + R_i)(-u + (R_o - R_i)/2)du.$$

By the Fundamental Theorem of Calculus, this equals,

$$\frac{4\pi H}{R_o - R_i}(R_o + R_i)\left(\frac{-u^2}{2} + \frac{(R_o - R_i)u}{2}\right|_0^{(R_o - R_i)/2}.$$

This evaluates to,

$$\frac{4\pi H}{R_o - R_i}(R_o + R_i)\frac{(R_o - R_i)^2}{8}.$$

This simplifies to give,

$$V = \pi H(R_o - R_i)(R_o + R_i)/2 = \boxed{\pi(R_o^2 - R_i^2)H/2.}$$

This is precisely the same answer as computed using the washer method. Please observe though, how much more effort was required for the shell method. The lesson is, if you have an alternative between the disk method and the shell method, consider carefully which method requires less effort before committing to one or the other.

Lecture 21. November 3, 2005

Homework. Problem Set 6 Part I: (a) - (e); Part II: Problem 1.

Practice Problems. Course Reader: 4E-2, 4E-5, 4E-7, 4F-1, 4F-6.

1. Parametric equations. To this point in the course, plane curves were specified in 1 of 2 ways. The *explicit form*, or *graph form* of a curve in Cartesian coordinates is the common form,

$$y = f(x), \quad a \leq x \leq b.$$

The *implicit form* of a curve in Cartesian coordinates is as the set of all solutions of an equation,

$$F(x, y) = 0.$$

Often a subset of this curve is specified by imposing extra conditions, e.g., the upper unit semicircle is the set of solutions of $x^2 + y^2 = 1$ satisfying the extra condition $y > 0$.

There is a third important way to specify a curve: using *parametric equations*. Given a *parameter* t varying in an interval $a \leq t \leq b$ and given functions $f(t)$ and $g(t)$ on this interval, the associated parametric curve,

$$\begin{cases} x &= f(t), \\ y &= g(t) \end{cases}$$

is simply the set of all pairs $(x, y) = (f(t), g(t))$ as t varies over the interval $a \leq t \leq b$. We consider only the case where $f(t)$ and $g(t)$ are piecewise differentiable functions (more advanced courses discuss some pitfalls if $f(t)$ and $g(t)$ are merely continuous functions).

Examples. A. One specification of the points on the circle of radius r centered at $(0, 0)$ is using the angle θ. This gives rise to a parametric equation with parameter θ,

$$\begin{cases} x &= r\cos(\theta), \\ y &= r\sin(\theta) \end{cases} \quad 0 \leq \theta < 2\pi.$$

B. An ellipse centered at $(0,0)$ whose axes equal the coordinate axes has a parametric equation,

$$\begin{cases} x &= a\cos(\theta), \\ y &= b\sin(\theta) \end{cases} \quad 0 \leq \theta < 2\pi.$$

C. A projectile is launched from an initial position of (x_0, y_0) with an initial velocity vector of magnitude v_0 at an angle α to the horizontal, and under the influence of constant graviational acceleration $-g$. According to Newton's laws of mechanics, the position of the projectile after time t is,

$$\begin{cases} x &= v_0\cos(\alpha)t + x_0, \\ y &= -(g/2)t^2 + v_0\sin(\alpha)t + y_0 \end{cases} \quad 0 \leq t.$$

This is a parametric equation where time t is the parameter. Even when some other quantity is the parameter, it is often useful to think of the parameter as time. Thus the curve is the trail left by a point, or perhaps better, the tip of a pen, as it moves in the plane.

2. Implicitization. Under reasonable hypotheses, it is possible to turn a portion of an implicit curve into an explicit curve. Similarly, it should be possible to turn a portion of a parametric curve into an explicit curve. It is often simpler to find an implicit equation satisfied by a parametric curve. The process of finding an implicit equation is called *implicitization.*

Examples. A. For the parametric curve in Example A above, by the Pythagorean Theorem,

$$x(\theta)^2 + y(\theta)^2 = r^2\cos^2(\theta) + r^2\sin^2(\theta) = r^2(\cos^2(\theta) + \sin^2(\theta)) = r^2.$$

Thus the parametric equation satisfies the implicit equation,

$$\boxed{x^2 + y^2 = r^2.}$$

B. For the parametric curve in Example B,

$$(x(\theta)/a)^2 + (y(\theta)/b)^2 = \cos^2(\theta) + \sin^2(\theta) = 1.$$

Thus the parametric equation satisfies the implicit equation,

$$\boxed{x^2/a^2 + y^2/b^2 = 1.}$$

C. For the parametric curve in Example C, assuming $v_0\cos(\alpha)$ is nonzero, the equation for x can be solved for t,

$$x = v_0\cos(\alpha)t + x_0 \Leftrightarrow t = \frac{x - x_0}{v_0\cos(\alpha)}.$$

This can then be substituted into the equation for y to get an explicit equation for the curve,

$$\boxed{y = -g(x - x_0)^2/(2v_0^2\cos^2(\alpha)) + \tan(\alpha)(x - x_0) + y_0.}$$

In going from a parametric equation to an implicit equation, there are 2 important warnings to keep in mind:

- A parametric equation may traverse only part of the implicit curve. The most usual reason is that the parameter t is restricted to a certain range. A closely related reason is that the functions of t are themselves somehow limited, as in the parametric curve lying in the line $y = x$,

$$\begin{cases} x = \cos(t), \\ y = \cos(t) \end{cases}$$

A more interesting reason is that the implicit curve may have more than one connected piece, as in the parametric curve,

$$\begin{cases} x = 2t/(1-t^2), \\ y = (1+t^2)/(1-t^2) \end{cases} \quad -1 < t < 1.$$

As t varies, this parametric curve sweeps out the top branch of the hyperbola $y^2 - x^2 = 1$.

- A parametric equation may sweep out all or a portion of the implicit curve multiple times. This is clear in Examples A and B: as θ is allowed to vary over the interval $0 \leq \theta < 2n\pi$, the parametric curve completes n revolutions of the implicit curve.

3. Arc length. Given a segment of curve, what is the length of the curve? Imagining the curve made of some flexible extensible material like wire, what is the length when the wire is pulled taut? The answer is called the *arc length*, s.

The method for expressing arc length is an integral is by now familiar. Break the interval $a \leq t \leq b$ into a large number n of subintervals with endpoints,

$$a = t_0 < t_1 < \cdots < t_{n-1} < t_n = b.$$

Approximate the curve on each subinterval $t_{k-1} \leq t \leq t_k$ by a line segment. The line segment runs from the point,

$$(x_{k-1}, y_{k-1}) = (x(t_{k-1}), y(t_{k-1})),$$

to the point,

$$(x_k, y_k) = (x(t_k), y(t_k)).$$

The rise and run of the line segment are,

$$\Delta x_k = x_k - x_{k-1} \approx x'(t_k)\Delta t_k,$$

$$\Delta y_k = y_k - y_{k-1} \approx y'(t_k)\Delta t_k.$$

By the Pythagorean theorem, the length of the line segment is,

$$\Delta s_k = \sqrt{(\Delta x_k)^2 + (\Delta y_k)^2} \approx \sqrt{(x'(t_k))^2 + (y'(t_k))^2}\Delta t_k.$$

The arc length of the curve is approximately the sum of the lengths of the approximating line segments,

$$s \approx \sum_{k=1}^{n} \sqrt{(x'(t_k))^2 + (y'(t_k))^2}\Delta t_k.$$

This is a Riemann sum. As the mesh of the partition tends to 0, the Riemann sums tend to a Riemann integral. This Riemann integral is the arc length,

$$\text{Arc length} = \int_{t=a}^{t=b} \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2}dt.$$

Example 1. For the parametric curve in Example A above,

$$\frac{dx}{d\theta} = -r\sin(\theta), \quad \frac{dy}{d\theta} = r\cos(\theta).$$

Therefore, the expression,

$$\left(\frac{ds}{dt}\right)^2 = \left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2,$$

equals,

$$(-r\sin(\theta))^2 + (r\cos(\theta))^2 = r^2\sin^2(\theta) + r^2\cos^2(\theta) = r^2(\sin^2(\theta) + \cos^2(\theta)) = r^2.$$

Taking square roots gives the equation,

$$\frac{ds}{d\theta} = r, \ \Leftrightarrow ds = rd\theta.$$

Thus the arc length of the arc of the circle from $\theta = a$ to $\theta = b$ is,

$$s = \int ds = \int_{\theta=a}^{\theta=b} rd\theta = \boxed{r(b-a).}$$

This is, in fact, our definition of the angle: the angle θ subtended by an arc of a circle equals the ratio of the arc length by the radius of the circle. If this logic sounds circular, it is perhaps that nobody ever told you before how to **define** the length of the arc of a circle! It is also an argument in favor of the more natural definition of the angle as the ratio of the *area* of the sector of the circle by $r^2/2$.

Example 2. This is not a single example, but a class of examples. A curve in explicit form, $y = f(x)$, $a \leq x \leq b$, can always be put in parametric form,

$$\begin{cases} x &= t, \\ y &= f(t) \end{cases} \quad a \leq t \leq b.$$

Then,

$$\frac{dx}{dt} = 1, \quad \frac{dy}{dt} = f'(t).$$

Using this,

$$ds = \sqrt{1 + (f'(t))^2}dt.$$

Thus the arc length is,

$$\text{Arc length} = \int_{t=a}^{t=b} \sqrt{1 + (f'(t))^2}dt.$$

Since the parameter t in the Riemann integral is only a dummy variable anyway, it is allowed to replace it by the variable x (so long as x plays no other role in the integral, which it does not). This gives the formula for the arc length of an explicit curve,

$$\text{Arc length} = \boxed{\int_a^b \sqrt{1 + (dy/dx)^2}dx.}$$

Example 3. Consider the explicit curve,

$$y = \frac{x^2}{4} - \frac{1}{2}\ln(x), \quad a \leq x \leq b,$$

where a is a positive real number. The derivative is,

$$\frac{dy}{dx} = \frac{1}{4}(2x) - \frac{1}{2}\frac{1}{x} = \frac{1}{2}(x - x^{-1}).$$

Thus the square of the derivative is,

$$\left(\frac{dy}{dx}\right)^2 = \frac{1}{4}(x^2 - 2 + x^{-2}).$$

Clearing denominators, $1 + (dy/dx)^2$ equals,

$$\frac{4}{4} + \frac{x^2 - 2 + x^{-2}}{4} = \frac{x^2 + 2 + x^{-2}}{4}.$$

It is easy to check this equals the square,

$$\left(\frac{x + x^{-1}}{2}\right)^2.$$

This gives the formula,

$$\sqrt{1 + \left(\frac{dy}{dx}\right)^2} = \frac{1}{2}(x + x^{-1}).$$

Therefore,

$$ds = \frac{1}{2}(x + x^{-1})dx.$$

Integrating gives the arc length,

$$s = \int ds = \int_{x=a}^{x=b} \frac{1}{2}(x + x^{-1})dx = \frac{1}{2}\left(\frac{x^2}{2} + \ln(x)\right|_a^b.$$

Evaluating gives the arc length,

$$\text{Arc length} = \boxed{(b^2 - a^2)/4 + (1/2)\ln(b/a).}$$

Lecture 22. November 4, 2005

Homework. Problem Set 6 Part I: (f)–(h); Part II: Problem 2 (a) and (c).

Practice Problems. Course Reader: 4G-1, 4G-4, 4G-6, 4H-1, 4H-3.

1. Surface area of a right circular cone. Before attacking the general problem of the surface area of a surface of revolution, consider the simplest case of the area of a right circular cone of base radius R and height H. The *slant height* of the cone is the length of any line segment from the vertex to a point on the base circle. By the Pythagorean theorem, the slant height S is,

$$S = \sqrt{R^2 + H^2}.$$

Imagine the cone is made of paper. Make an incision along a line segment from the vertex to the base circle. The resulting piece of paper may be unfolded to form a sector of a circle. The radius of the sector is the slant height s. The circumference of the sector is the circumference of the original base circle $2\pi r$. The formula for the area and circumference of a sector of a circle give the identity,

$$\text{Area of sector} = \frac{1}{2}(\text{Radius of sector}) \times (\text{Circumference of sector}).$$

Thus, the area of the cone equals,

$$A = \frac{1}{2}(S)(2\pi R) = \boxed{\pi RS}.$$

In particular, the height H is involved only indirectly (as H depends on H).

Next, consider a *conical band* obtained from a right circular cone of base radius R_1 and slant height S_1 by removing the the top part of the cone of base radius R_2 and slant height S_2. In particular, the slant height of the band is the difference,

$$s = S_2 - S_1,$$

and the *average radius* of the band is the average of R_1 and R_2,

$$r = \frac{1}{2}(R_1 + R_2).$$

By similar triangles,

$$\frac{S_1}{R_1} = \frac{S_2}{R_2}.$$

Rearranging gives,

$$R_2 S_1 = R_1 S_2.$$

Using the formula above, the area of the large cone is,

$$A_1 = \pi R_1 S_1,$$

and the area of the small cone is,

$$A_2 = \pi R_2 S_2.$$

The area A of the band is the difference,

$$A = A_1 - A_2 = \pi(R_1 S_1 - R_2 S_2).$$

Since $R_2 S_1$ equals $R_1 S_2$, the formula is unchanged by adding $\pi R_2 S_1$ and subtracting $\pi R_1 S_2$ to get,

$$A = \pi(R_1 S_1 - R_2 S_2) + \pi(R_2 S_1 - R_1 S_2) = \pi((R_1 + R_2)S_1 - (R_1 + R_2)S_2).$$

Simplifying and substituting $R_1 + R_2 = 2r$ and $S_1 - S_2 = 2$ gives,

$$A = \boxed{2\pi r s.}$$

2. Surface area of a surface of revolution. Given a segment of a parametric curve,

$$\begin{cases} x &= x(t), \\ y &= y(t) \end{cases} \quad a \leq t \leq b$$

the *surface of revolution* is the surface obtained by revolving the segment through xyz-space about the y-axis. What is the area of this surface? The answer is called the *surface area*.

The method for computing the surface area is so close to the method for computing the arc length of the curve, the details will be skipped. What is relevant is the differential element of surface area. Given a small interval from t to $t + dt$, approximate the segment of the parametric curve as a line segment. The surface obtained by revolving a line segment is precisely a band of a cone. The average radius of the cone r is $x(t)$. The slant height of the cone is ds. Thus the area of the band is,

$$dA = \boxed{2\pi r ds} = \boxed{2\pi x(t)\sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2}dt.}$$

Integrating gives the formula for the surface area of the surface of revolution,

$$A = \int dA = \int_{t=a}^{t=b} 2\pi x(t)\sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2}dt.$$

Examples. A. Consider the line segment connecting the point $(0, H)$ to the point $(R, 0)$. This has equation,

$$y = \frac{H}{R}(R - x),\ 0 \leq x \leq R.$$

The slant height of the line segment is,

$$S = \sqrt{R^2 + H^2},$$

and the differential arc length of the line segment is,

$$ds = \frac{S}{R}dx.$$

Thus the differential element of surface area is,

$$dA = 2\pi r ds = 2\pi x \frac{S}{R}dx.$$

Integrating gives,

$$A = \int dA = \int_{x=0}^{x=R} \frac{2\pi S}{R} x dx = \frac{2\pi S}{R}\left(\frac{x^2}{2}\right|_0^R = \pi RS.$$

This is the same formula obtained above by more elementary means.

B. Consider the parametrized semicircle of radius R in the first and third quadrants,

$$\begin{cases} x = R\cos(\theta), \\ y = R\sin(\theta) \end{cases} \quad \frac{-\pi}{2} \leq \theta \leq \frac{\pi}{2}.$$

Revolving about the y-axis gives the *sphere* of radius R. Thus the surface area of the surface of revolution is the surface area of the sphere of radius R.

As computed in the previous lecture, the differential element of arc length is,

$$ds = Rd\theta.$$

Thus the differential element of surface area is,

$$dA = 2\pi r ds = 2\pi x(\theta)(Rd\theta) = 2\pi(R\cos(\theta))(Rd\theta) = 2\pi R^2\cos(\theta)d\theta.$$

Integrating gives,

$$A = \int dA = \int_{\theta=-\pi/2}^{\theta=\pi/2} 2\pi R^2\cos(\theta)d\theta = 2\pi R^2 \left(\sin(\theta)\right|_{-\pi/2}^{\pi/2}.$$

This evaluates to,

$$A = 2\pi R^2(2) = \boxed{4\pi R^2.}$$

The fastest way to remember this is to observe the surface area A and the volume V of a sphere of radius R are related by,

$$A = 4\pi R^2 = \frac{dV}{dr} = \frac{d}{dr}(4\pi R^3/3).$$

C. An *astroid* is a curve,

$$x^{2/3} + y^{2/3} = a^{2/3}.$$

The part of the astroid in the first quadrant has parametric equation,

$$\begin{cases} x &= a\cos^3(t), \\ y &= a\sin^3(t) \end{cases} \quad 0 \leq t \leq \frac{\pi}{2}.$$

The derivatives are,

$$\frac{dx}{dt} = -3a\cos^2(t)\sin(t), \quad \frac{dy}{dt} = 3a\sin^2(t)\cos(t).$$

Thus,

$$\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2 = (-3a\cos^2(t)\sin(t))^2 + (3a\sin^2(t)\cos(t))^2 = 9a^2\sin^2(t)\cos^2(t)(\cos^2(t)+\sin^2(t)).$$

The square root is,

$$\sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2} = \sqrt{9a^2\sin^2(t)\cos^2(t)} = 3a\sin(t)\cos(t).$$

So the differential element of arc length is,

$$ds = 3a\sin(t)\cos(t)dt.$$

Thus the differential element of surface area of the surface of revolution is,

$$dA = 2\pi r ds = 2\pi x(t)ds = 2\pi(a\cos^3(t))(3a\sin(t)\cos(t))dt = 6\pi a^2\cos^4(t)\sin(t)dt.$$

Integrating, the surface area is,

$$A = \int dA = \int_{t=0}^{t=\pi/2} 6\pi a^2\cos^4(t)\sin(t)dt.$$

Subsitute $u = \cos(t)$, $du = -\sin(t)dt$, $u(0) = 1$, $u(\pi/2) = 1$ to get,

$$A = \int_{u=1}^{u=0} 6\pi a^2 u^4(-du) = 6\pi a^2 \int_{u=0}^{u=1} u^4 du.$$

Thus the surface area of the surface of revolution is,

$$A = 6\pi a^2 \left(u^5/5\right|_0^1 = \boxed{6\pi a^2/5.}$$

3. Polar coordinate curves. After the explicit, Cartesian form of a curve as a graph, $y = f(x)$, the next most common representation is using polar coordinates. Given a function $r = r(\theta)$ and an interval $a \leq \theta \leq b$, the associated *polar coordinate curve* is the parametric curve,

$$\left\{ \begin{array}{rcl} x & = & r(\theta)\cos(\theta), \\ y & = & r(\theta)\sin(\theta) \end{array} \right. \quad a \leq \theta \leq b.$$

For each point on the curve, the distance of the point from the origin is,

$$\text{Distance from origin} = \sqrt{x^2+y^2} = \sqrt{r^2} = |r| = \left\{ \begin{array}{ll} +r, & r \geq 0, \\ -r, & r < 0 \end{array} \right.$$

Also, assuming the point does not equal the origin, the angle of the ray from the origin to the point is,

$$\text{Angle} = \tan^{-1}(y/x) = \tan^{-1}(\tan(\theta)) = \left\{ \begin{array}{ll} \theta, & r > 0, \\ \theta + \pi, & r < 0 \end{array} \right.$$

This is one of the most confusing aspects of polar curves. The symbols $r(\theta)$ and θ are engrained in mathematical thinking as the distance and angle of a point in polar coordinates. But for a polar coordinate curve these are *simply parameters*. They are very closely related to, but often different from, the actual distance and angle. This is easiest to think about by imagining the point swerving through the origin along the radius line to the opposite ray of the ray given by θ. In other words, the point "goes negative".

Given a polar curve, it is often possible to find an implicit Cartesian curve containg the polar curve.

Examples. A. Let a be a positive constant and consider the polar curve,

$$r(\theta) = a.$$

This gives,

$$r^2 = a^2 \Leftrightarrow \boxed{x^2 + y^2 = a^2.}$$

Thus the polar curve is contained in the circle of radius a.

B. Consider the polar curve,

$$r(\theta) = \frac{a}{\sin(\theta)}.$$

Multiplying both sides by $\sin(\theta)$ gives,

$$r\sin(\theta) = a \Leftrightarrow \boxed{y = a.}$$

Thus the polar curve is contained in the horizontal line passing through $(0, a)$.

C. Consider the polar curve,

$$r = 2a\cos(\theta),$$

Multiplying both sides by r gives,

$$r^2 = 2ar\cos(\theta) \Leftrightarrow x^2 + y^2 = 2ax.$$

Simplifying this gives the equation,

$$\boxed{(x-a)^2 + y^2 = a^2.}$$

This is the equation of the circle of radius a centered at $(a, 0)$.

4. Sketching polar curves. Given a polar curve, how are we to sketch it? For definiteness, consider the polar curve,

$$r(\theta) = \cos(2\theta), \quad -\pi/4 \leq \theta \leq 7\pi/4.$$

This curves is called the *four-leaved rose.* A similar curve occurs in Part II, Problem 2 of Problem Set 6.

1. Find the range of θ. In almost every case, this will be given. In this case, the range is given as $-\pi/4 \leq \theta \leq 7\pi/4$. In some cases, the range must be determined. For instance, to sketch only the "leaf" of the rose containing $(1, 0)$, first the endpoints of this leaf must be found.

2. Determine when r is positive, zero or negative. This is easiest to keep track of with a table.

θ	r
$-\pi/4$	0
$-\pi/4 < \theta < \pi/4$	$r > 0$
$\pi/4$	0
$\pi/4 < \theta < 3\pi/4$	$r < 0$
$3\pi/4$	0
$3\pi/4 < \theta < 5\pi/4$	$r > 0$
$5\pi/4$	0
$5\pi/4 < \theta < 7\pi/4$	$r < 0$

The curve crosses the origin when $\theta = -\pi/4, \pi/4, 3\pi/4$, and $5\pi/4$. The curve "goes negative" when $\pi/4 < \theta < 3\pi/4$ and when $5\pi/4 < \theta < 7\pi/4$.

3. Find the extremal values of $|r|$. A local maximum of $|r|$ is either a point where r is positive and a local maximum or a point where r is negative and a local minimum. Similarly for local minima of $|r|$. Typically, local maxima of $|r|$ occur either at endpoints of the interval or points where $r'(\theta)$ is zero (occasionally at discontinuity points, or nondifferentiable points). Local minima occur at such points, but also occur everytime the curve crosses the origin (so that $|r|$ equals 0).

In our example, the local minima are all points where $r = 0$, enumerated above. The derivative of r is,

$$r'(\theta) = -2\sin(2\theta).$$

The critical points are $\theta = 0, \pi/2, \pi$ and $3\pi/2$. For $\theta = 0$ and $\theta = \pi$, r is positive and maximum. For $\theta = \pi/2$ and $\theta = 3\pi/3$, r is negative and minimum. Thus each critical point is a local maximum of $|r|$. The value of $|r|$ at each critical point is 1.

4. Find the asymptotes. This is a bit difficult with a polar curve. What is easier is to find a line parallel to an asymptote. Whenever,

$$\lim_{\theta \to a} r(\theta) = +\infty,$$

(or the same with a right-hand limit or left-hand limit), there is an asymptote parallel to the ray $\theta = a$. Whenever,

$$\lim_{\theta \to a} r(\theta) = -\infty,$$

there is an asymptote parallel to the ray $\theta = a + \pi$.

In our example, $r(\theta)$ never limits to $\pm\infty$. Thus there are no asymptotes. But in Example B., $r = a/\sin(\theta)$, r tends to $+\infty$ as θ tends to 0 from above and as θ tends to π from below. Thus there is an asymptote parallel to the x-axis. Since the explicit equation is $y = a$, which is a line parallel to the x-axis, this is correct.

5. Find the tangent direction at important points. This will be discussed further next time. The most important tangent directions are when the curve crosses the origin, and critical points of r. If $r(\theta) = 0$ and $r'(\theta) \neq 0$, the tangent line of the curve has angle θ. If $r'(\theta) = 0$ and $r(\theta) \neq 0$, the tangent line has angle $\theta \pm \pi/2$, i.e., the tangent line is orthogonal to the radius through the point. Both of these are consequences of a more general formula. The angle ψ between the tangent line and the radius satisfies,

$$\tan(\psi) = \frac{r(\theta)}{r'(\theta)}.$$

In the example, r' is nonzero whenever r is zero. Thus the tangent direction of the curve as it crosses the origin is just the direction of the limiting radius.

This is now ample information to sketch the four-leaved rose. Up to a rotation of $\pi/4$, the sketch is the same as in Figure 16.11 on p. 566 of the textbook (the sketch was also given in lecture).

Lecture 23. November 8, 2005

Homework. Problem Set 6 Part I: (i) and (j); Part II: Problem 2.

Practice Problems. Course Reader: 4I-1, 4I-4, 4I-6.

1. Tangent lines to parametric curves. This short section was not explicitly discussed for general parametric curves. It was discussed for polar curves, which are a special collection of parametric curves.

Given a parametric curve,

$$\begin{cases} x &= f(t), \\ y &= g(t), \end{cases}$$

what is the slope of the tangent line at $(f(a), g(a))$? The relevant differentials are,

$$dx = f'(t)dt, \quad dy = g'(t)dt.$$

If $g'(a)$ is nonzero, then the slope of the tangent line is,

$$\frac{dy}{dx}(a) = \left.\frac{f'(t)dt}{g'(t)dt}\right|_{t=a} = \frac{f'(a)}{g'(a)}.$$

In particular, for a function $r = r(\theta)$, the associated polar curve is,

$$\begin{cases} x &= r(\theta)\cos(\theta), \\ y &= r(\theta)\sin(\theta) \end{cases}$$

Thus the differentials are,

$$\begin{array}{rcl} dx &=& [r'(\theta)\cos(\theta) - r(\theta)\sin(\theta)]d\theta, \\ dy &=& [r'(\theta)\sin(\theta) + r(\theta)\cos(\theta)]d\theta. \end{array}$$

Therefore the slope of the tangent line is,

$$\frac{dy}{dx} = \frac{r'(\theta)\sin(\theta) + r(\theta)\cos(\theta)}{r'(\theta)\cos(\theta) - r(\theta)\sin(\theta)}.$$

2. Tangent lines for polar curves. Although the formula above is perfectly correct, it is a bit long to remember. There is a slightly different packaging that is much easier to remember. Define α to be the angle from the horizontal ray emanating from $(x(\theta), y(\theta))$ in the positive x-direction, and the tangent line. To be precise, there are two such angles, differing by π. The defining equation for α is,

$$\tan(\alpha) = \frac{dy}{dx}.$$

And, of course,

$$\tan(\theta) = \frac{y}{x}.$$

Define ψ to be the difference between α and θ,

$$\psi = \alpha - \theta.$$

The angle addition/subtraction formulas for $\tan(\theta)$ are,

$$\tan(\phi_1 + \phi_2) = \frac{\tan(\phi_1) + \tan(\phi_2)}{1 - \tan(\phi_1)\tan(\phi_2)}, \quad \tan(\phi_1 - \phi_1) = \frac{\tan(\phi_1) - \tan(\phi_2)}{1 + \tan(\phi_1)\tan(\phi_2)}.$$

Therefore,

$$\tan(\psi) = \tan(\alpha - \theta) = \frac{\tan(\alpha) - \tan(\theta)}{1 + \tan(\alpha)\tan(\theta)}.$$

Substituting in the equations for $\tan(\theta)$ and $\tan(\alpha)$ from above gives,

$$\tan(\psi) = \frac{(dy/dx) - (y/x)}{1 + (y/x)(dy/dx)}.$$

To simplify this, imagine multiplying both numerator and denominator by xdx and manipulate formally,

$$\tan(\psi) = \frac{xdy - ydx}{xdx + ydy}.$$

The actual justification of this is a little more involved, but the formal manipulation leads to the correct equation.

To compute the denominator in the expression, differentiate both sides of,

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

to get,

$$2rdr = 2xdx + 2ydy,$$

or equivalently,

$$xdx + ydy = r(\theta)r'(\theta)d\theta.$$

To compute the numerator in the expression, differentiate both sides of,

$$\tan(\theta) = \frac{y}{x},$$

to get,

$$\sec^2(\theta)d\theta = \frac{dy}{x} - \frac{ydx}{x^2} = \frac{1}{x^2}(xdy - ydx).$$

Now substitute $x = r\cos(\theta)$ in the denominator to get,

$$\sec^2(\theta)d\theta = \frac{1}{r^2\cos^2(\theta)}(xdy - ydx) = \frac{\sec^2(\theta)}{r^2}(xdy - ydx).$$

Cancelling $\sec^2(\theta)$ and multiplying both sides by r^2 gives,

$$xdy - ydx = r^2 d\theta.$$

Thus the fraction for $\tan(\psi)$ is,

$$\tan(\psi) = \frac{xdy - ydx}{xdx + ydy} = \frac{r^2 d\theta}{rr' d\theta}.$$

Simplifying gives,

$$\tan(\psi) = \boxed{r(\theta)/r'(\theta)}.$$

Example. Consider the cardioid, discussed in recitation,

$$r(\theta) = a(1 + \cos(\theta)).$$

The formula for ψ is,

$$\tan(\psi) = \frac{r}{r'} = \frac{a(1+\cos(\theta))}{-a\sin(\theta)} = \frac{1+\cos(\theta)}{-\sin(\theta)}.$$

To simplify this, write $\theta = 2(\theta/2)$ and use the double-angle formulas to get,

$$\frac{1+\cos(2(\theta/2))}{-\sin(2(\theta/2))} = \frac{1 + (\cos^2(\theta/2) - \sin^2(\theta/2))}{-2\sin(\theta/2)\cos(\theta/2)}.$$

Replacing $1 - \sin^2(\theta/2)$ in the numerator by $\cos^2(\theta/2)$, this simplfies to,

$$\frac{2\cos^2(\theta/2)}{-2\sin(\theta/2)\cos(\theta/2)} = -\cot(\theta/2).$$

Of course there is an identity,

$$-\cot(u) = \tan(u - \pi/2).$$

Altogether, this gives,

$$\tan(\psi) = -\cot(\theta/2) = \tan(\theta/2 - \pi/2).$$

Therefore,

$$\psi = \boxed{(\theta - \pi)/2.}$$

Since α equals $\theta + \psi$, this gives,

$$\alpha = \boxed{(3\theta - \pi)/2.}$$

In particular, the angle of the tangent line to the cardioid at $\theta = \pi/2$ is $\alpha = \pi/4$.

3. Arc length in polar coordinates. As discussed previously, the formula for arc length of a parametric curve is,

$$ds = \sqrt{(dx/dt)^2 + (dy/dt)^2}dt.$$

In the case of a parametric curve, this becomes a bit simpler. The differentials are,

$$dx = (r'(\theta)\cos(\theta) - r(\theta)\sin(\theta))d\theta,$$
$$dy = (r'(\theta)\sin(\theta) + r(\theta)\cos(\theta))d\theta.$$

Squaring gives,

$$(dx)^2 = ((r')^2\cos^2(\theta) - 2rr'\sin(\theta)\cos(\theta) + r^2\sin^2(\theta))(d\theta)^2,$$
$$(dy)^2 = ((r')^2\sin^2(\theta) + 2rr'\sin(\theta)\cos(\theta) + r^2\cos^2(\theta))(d\theta)^2.$$

Summing down columns gives,

$$(dx)^2 + (dy)^2 = [(r')^2 + r^2](d\theta)^2.$$

Taking square roots gives the differential element of arc length for a polar curve,

$$ds = \boxed{\sqrt{[r'(\theta)]^2 + [r(\theta)]^2}d\theta.}$$

Example. For the cardioid,

$$r(\theta) = a(1 + \cos(\theta)),$$

the derivative is,

$$r'(\theta) = -a\sin(\theta).$$

Thus,

$$(r')^2 + r^2 = a^2(1 + \cos(\theta))^2 + (-a\sin(\theta))^2 = a^2(1 + 2\cos(\theta) + \cos^2(\theta)) + a^2\sin^2(\theta).$$

This simplifies to,

$$2a^2(1 + \cos(\theta)).$$

To simplify this further, write $\theta = 2(\theta/2)$ and use the double-angle formula to get,

$$2a^2(1 + \cos(2(\theta/2))) = 2a^2(1 + \cos^2(\theta/2) - \sin^2(\theta/2)) = 2a^2(2\cos^2(\theta/2)) = 4a^2\cos^2(\theta/2).$$

Taking square roots gives,

$$ds = \boxed{2a\cos(\theta/2).}$$

Note, this answer is only correct for $-\pi \leq \theta \leq \pi$. Outside this range, we might have to take the other square root to get a positive number. In particular, the total arc length of the cardioid is,

$$s = \int ds = \int_{\theta=-\pi}^{\theta=\pi} 2a\cos(\theta/2)d\theta = 2a\left(2\sin(\theta/2)\right|_{-\pi}^{\pi} = 2a((2) - (-2)).$$

Simplifying, the total arc length of the cardioid is,

$$s = \boxed{8a.}$$

Surface areas of surfaces of revolution can be computed in a similar way. This was only briefly discussed in lecture. Here is a continuation of the previous problem.

Example. The top half of the cardioid,

$$r(\theta) = a(1 + \cos(\theta)), \quad 0 \leq \theta \leq \pi,$$

is revolved about the x-axis to give a fairly good approximation of the surface of an apple. What is the surface area of this apple?

Since we are revolving about the x-axis, the radius of each slice is y. Therefore the differential element of surface area is,

$$dA = 2\pi y ds.$$

Substituting in $y = r(\theta)\sin(\theta) = a(1+\cos(\theta))\sin(\theta)$, and substituting in for ds gives,

$$dA = 2\pi[a(1+\cos(\theta))\sin(\theta)](2a\cos(\theta/2)d\theta).$$

To simplify this, substitute both,

$$1+\cos(\theta) = 2\cos^2(\theta/2),$$

and,

$$\sin(\theta) = 2\sin(\theta/2)\cos(\theta/2),$$

to get,

$$dA = 4\pi a^2(2\cos^2(\theta/2))(2\sin(\theta/2)\cos(\theta/2))\cos(\theta/2)d\theta = 16\pi a^2\cos^4(\theta/2)\sin(\theta/2)d\theta.$$

Thus the total surface area is,

$$A = \int dA = \int_{\theta=0}^{\pi} 16\pi a^2\cos^4(\theta/2)\sin(\theta/2)d\theta.$$

To evaluate this integral, substitute,

$$\begin{array}{rl|l} u = & \cos(\theta/2) & u(\pi) = 0, \\ du = & -(1/2)\sin(\theta/2)d\theta, & u(0) = 1 \end{array}$$

The new integral is,

$$A = 16\pi a^2\int_{u=1}^{u=0} u^4(-2du) = 32\pi a^2\int_{u=0}^{u=1} u^4 du = 32\pi a^2\left(\frac{u^5}{5}\right|_0^1.$$

This evaluates to give the total surface area of the apple,

$$A = \boxed{32\pi a^2/5.}$$

5. Area of a region enclosed by a polar curve. What is the area of the planar region enclosed by a cardioid? By the same sort of reasoning as for volumes and arc lengths, the differential element of area of the triangular region bounded by the rays θ, $\theta + d\theta$ and the curve $r(\theta)$ is,

$$dA = \frac{r(\theta)^2}{2}d\theta.$$

Thus the area enclosed by a polar curve is,

$$A = \int dA = \int_{\theta=a}^{\theta=b}\frac{r(\theta)^2}{2}d\theta.$$

In particular, the area enclosed by the cardioid is,

$$A = \int_0^{2\pi} \frac{a^2(1+\cos(\theta))^2}{2} d\theta.$$

This expands to give,

$$\frac{a^2}{2}\int_0^{2\pi} 1 + 2\cos(\theta) + \cos(\theta)^2 d\theta.$$

To simplify the last part of the integrand, substitute,

$$\cos(\theta)^2 = \frac{1+\cos(2\theta)}{2},$$

to get,

$$\frac{a^2}{2}\int_0^{2\pi} 1 + 2\cos(\theta) + \frac{1+\cos(2\theta)}{2} d\theta = \frac{a^2}{4}\int_0^{2\pi} 3 + 4\cos(\theta) + \cos(2\theta) d\theta.$$

Using the Fundamental Theorem of Calculus, this equals,

$$\frac{a^2}{4}\left(3\theta + 4\sin(\theta) + \frac{1}{2}\sin(2\theta)\right|_0^{2\pi}.$$

Evaluating gives,

$$A = \boxed{3\pi a^2/2.}$$

Lecture 24. November 15, 2005

Practice Problems. Course Reader: 5A-1, 5A-2, 5A-3, 5A-5, 5A-6.

1. Inverse functions. Let a, b, s and t be constants. Let $y = f(x)$ be a function defined on the interval,

$$a \leq x \leq b,$$

and whose values are in the interval,

$$s \leq y \leq t.$$

Does there exist a function $x = g(y)$ defined on the interval,

$$s \leq y \leq t,$$

whose values are in the interval,

$$a \leq x \leq b,$$

satisfying the two conditions,

$$g(f(x)) = x, \quad f(g(y)) = y\ ?$$

If such a function g exists, it is called an *inverse function* of f, and it is denoted by $f^{-1}(y)$. Also, the original function $f(x)$ is called *invertible*. There is some chance of confusion with the other use of "invertible", namely that $1/f(x)$ is always defined. We will be careful to specify the meaning of "invertible".

There are 2 necessary conditions for f to have an inverse function. Assume f has an inverse function g. Let x_1, x_2 be a pair of numbers in $[a, b]$. If $f(x_1)$ equals $f(x_2)$, then also,

$$x_1 = g(f(x_1)) = g(f(x_2)) = x_2,$$

i.e., x_1 equals x_2. In other words, two distinct inputs x_1 and x_2 give two distinct outputs $f(x_1)$ and $f(x_2)$. A function satisfying this condition is called *one-to-one*, because to every output, there is at most one input. This is the first necessary condition: every invertible function is one-to-one.

Next, for every number y in $[s, t]$, there is a number x in $[a, b]$ such that $y = f(x)$. In fact, just take x to be $g(y)$; then $f(x)$ equals $f(g(y))$, which equals y. A function satisfying this condition is called *onto*. This is the second necessary condition: every invertible function is onto.

Together, this says that an invertible function is one-to-one and onto. In fact, the converse is also true: every one-to-one and onto function is invertible. This is easy to check, but we will not prove it in this class.

Remark: In checking that f is one-to-one and onto, the choice of intervals $[a, b]$ and $[c, d]$ are vital. A simple example comes from $f(x) = \sin(x)$. For the interval $[-\pi/2, \pi, 2]$ and $[-1, 1]$, the function $f(x)$ is one-to-one and onto. But for many other choices of these intervals, the function is neither one-to-one nor onto.

2. The graph of an inverse function. How should we think of an inverse function? One way is graphically. The graph of the function $y = f^{-1}(x)$ is the same as the graph of $f(y) = x$. This is simply the usual graph of $y = f(x)$ with the roles of x and y reversed. What this translates to is, the graph of f^{-1} is the same as the graph of f with the roles of the x-axis and y-axis reversed. The simplest way to get the graph of $f^{-1}(x)$ is simply to reflect the graph of $f(x)$ through the 45° line $y = x$.

3. The inverse trigonometric functions. The function $\sin(x)$ is one-to-one and onto on $[-\pi/2, \pi/2]$, taking values in $[-1, 1]$. Thus there is an inverse function $\sin^{-1}(x)$ defined on the interval $[-1, 1]$, taking values in $[-\pi/2, \pi/2]$. The graph of $\sin^{-1}(x)$ is an increasing function whose lower left endpoint is $(-1, -\pi/2)$ and whose upper right endpoint is $(1, \pi/2)$.

The function $\cos(x)$ is one-to-one and onto on $[0, \pi]$, taking values in $[-1, 1]$. Thus there is an inverse function $\cos^{-1}(x)$ defined on the interval $[-1, 1]$, taking values in $[0, \pi]$. The graph of $\cos^{-1}(x)$ is a decreasing function whose upper left endpoint is $(-1, \pi)$ and whose lower right endpoing is $(1, 0)$.

The function $\tan(x)$ is one-to-one and onto on $(-\pi/2, \pi/2)$, taking values in the whole real line. Thus there is an inverse function $\tan^{-1}(x)$ defined on the whole real line, taking values in $(-\pi/2, \pi/2)$. The graph is an increasing function that is asymptotic to the line $y = -\pi/2$ as $x \to -\infty$, and asymptotic to the line $y = +\pi/2$ as $x \to +\infty$.

4. Derivatives of inverse functions. A particular simple formulation of the chain rule is the differential formulation,

$$df(u) = f'(u)du.$$

If f has an inverse function $g(x)$, let u be $g(x)$. Then this gives,

$$df(g(x)) = f'(g(x))dg(x).$$

On the other hand, $f(g(x))$ equals x. This gives the formula,

$$dx = f'(g(x))dg(x).$$

Solving for dg/dx gives,

$$\frac{d}{dx}(g(x)) = \boxed{1/f'(g(x)).}$$

This is the formula for the derivative of an inverse function.

In fact, we have seen this formula before. It is how we computed the derivative of $\ln(x)$, the inverse function of e^x:

$$\frac{d}{dx}(\ln(x)) = \frac{1}{e^{\ln(x)}} = \frac{1}{x}.$$

5. Derivatives of the inverse trigonometric functions. Because the derivative of $\sin(x)$ is $\cos(x)$, the formula above gives,

$$\frac{d}{dx}(\sin^{-1}(x)) = \frac{1}{\cos(\sin^{-1}(x))}.$$

This isn't very useful. A simple argument makes it much more useful. Denote $\sin^{-1}(x)$ by θ. Thus $\sin(\theta) = x$. Also, the formula for the derivative is a bit simpler,

$$\frac{d}{dx}(\sin^{-1}(x)) = \frac{1}{\cos(\theta)}.$$

By the Pythagorean theorem,

$$\sin^2(\theta) + \cos^2(\theta) = 1.$$

Solving gives,

$$\cos(\theta) = \sqrt{1 - \sin^2(\theta)} = \sqrt{1 - x^2}.$$

This gives a very useful formula for the derivative of $\sin^{-1}(x)$,

$$\frac{d}{dx}(\sin^{-1}(x)) = \boxed{1/\sqrt{1 - x^2}.}$$

There is a very similar derivation that,

$$\frac{d}{dx}(\cos^{-1}(x)) = \boxed{-1/\sqrt{1 - x^2}.}$$

This looks remarkably similar to the previous formula. In particular, this gives,

$$\frac{d}{dx}(\sin^{-1}(x) + \cos^{-1}(x)) = \frac{1}{\sqrt{1-x^2}} + \frac{-1}{\sqrt{1-x^2}} = 0.$$

Therefore the sum is a constant function. Checking at $x = 0$ gives the value of this constant function,

$$\boxed{\sin^{-1}(x) + \cos^{-1}(x) = \pi/2.}$$

Finally, because the derivative of $\tan(x)$ is $\sec^2(x)$, the formula gives,

$$\frac{d}{dx}(\tan^{-1}(x)) = \frac{1}{\sec^2(\tan^{-1}(x))}.$$

Again introduce $\theta = \tan^{-1}(x)$. Then the formula for the derivative is,

$$\frac{d}{dx}(\tan^{-1}(x)) = \frac{1}{\sec^2(\theta)}.$$

But the Pythagorean theorem implies,

$$\sec^2(\theta) = 1 + \tan^2(\theta) = 1 + x^2.$$

This finally gives a very useful formula for the derivative of $\tan(x)$,

$$\frac{d}{dx}(\tan^{-1}(x)) = \boxed{1/(1+x^2).}$$

Notice, in particular, that the denominator is never zero. This is closely related to the fact that $\tan^{-1}(x)$ is defined on the entire real line.

6. Hyperbolic trigonometric functions. The trigonometric functions are very useful for discussing point on the unit circle $x^2 + y^2 = 1$, because the circle is the parametric curve,

$$\begin{cases} x = \cos(\theta), \\ y = \sin(\theta) \end{cases}$$

Are there analogous continuous functions for the points on the hyperbola $x^2 - y^2 = 1$?

At first blush, the answer is no. The problem is that the hyperbola has two parts: one part is in the half-plane where $x > 0$, and the other part is in the half-plane where $x < 0$. Because of the intermediate value theorem, a continuous function $x = f(t)$ cannot jump from $x > 0$ to $x < 0$ or vice versa without crossing $x = 0$. Thus, refine the question: Are there continuous functions for the part of the hyperbola in the half-plane where $x > 0$?

The answer to this question is yes. The corresponding functions are called *hyperbolic trigonometric functions* or, more often, simply *hyperbolic functions*. They are defined as follows,

$$\cosh(t) = \frac{1}{2}(e^t + e^{-t}),$$

$$\sinh(t) = \frac{1}{2}(e^t - e^{-t}),$$
$$\tanh(t) = \frac{\sinh(t)}{\cosh(t)} = \frac{e^t - e^{-t}}{e^t + e^{-t}},$$
$$\operatorname{sech}(t) = \frac{1}{\cosh(t)} = \frac{2}{e^t + e^{-t}},$$
$$\operatorname{csch}(t) = \frac{1}{\sinh(t)} = \frac{2}{e^t - e^{-t}},$$

and,

$$\coth(t) = \frac{1}{\tanh(t)} = \frac{\cosh(t)}{\sinh(t)} = \frac{e^t + e^{-t}}{e^t - e^{-t}}.$$

The first observation is that,

$$\cosh^2(t) = \frac{1}{4}(e^t + e^{-t})^2 = \frac{1}{4}(e^{2t} + 2 + e^{-2t}),$$
$$\sinh^2(t) = \frac{1}{4}(e^t - e^{-t})^2 = \frac{1}{4}(e^{2t} - 2 + e^{-2t}).$$

Taking the difference of these, most of the terms cancel,

$$\cosh^2(t) - \sinh^2(t) = \frac{1}{4}((2) - (-2)) = \frac{4}{4} = 1.$$

This proves that the parametric curve,

$$\begin{cases} x = \cosh(t), \\ y = \sinh(t) \end{cases}$$

is contained in the right-half of the hyperbola $x^2 - y^2 = 1$. We will see next time that there is an inverse function of $\sinh(t)$, from which it follows that *every* point in the right-half of the hyperbola occurs for *exactly* one value of t. Thus the parametric curve exactly traces out the right-half of the hyperbola.

7. The derivatives of the hyperbolic functions. The derivatives of the hyperbolic functions are straightforward. The formulas are very similar to the formulas in the trigonometric case, but slightly different. Try not to confuse them.

$$\frac{d}{dx}(\sinh(x)) = \boxed{\cosh(x).}$$
$$\frac{d}{dx}(\cosh(x)) = \boxed{\sinh(x).}$$
$$\frac{d}{dx}(\tanh(x)) = \frac{1}{\cosh^2(x)}(\cosh(x)\cdot\cosh(x) - \sinh(x)\cdot\sinh(x)) = \frac{1}{\cosh^2(x)} = \boxed{\operatorname{sech}^2(x).}$$

Lecture 25. November 17, 2005

Homework. Problem Set 7 Part I: (a)–(e)

Practice Problems. Course Reader: 5D-2, 5D-6, 5D-7, 5D-10, 5D-14

1. Inverse hyperbolic functions. There are a few other useful formulas for hyperbolic functions; for instance, the analogues of the angle-addition formulas,

$$\sinh(s+t) = \sinh(s)\cosh(t) + \cosh(s)\sinh(t),$$

$$\cosh(s+t) = \cosh(s)\cosh(t) + \sinh(s)\sinh(t).$$

These imply the double-angle formulas,

$$\sinh(2t) = 2\sinh(t)\cosh(t),$$

$$\cosh(2t) = \cosh^2(t) + \sinh^2(t) = 2\cosh^2(t) - 1 = 2\sinh^2(t) + 1.$$

From these follow the analogues of the half-angle formulas,

$$\sinh^2(t/2) = \frac{1}{2}(\cosh(t) - 1),$$

$$\cosh^2(t/2) = \frac{1}{2}(\cosh(t) + 1).$$

A beautiful feature of hyperbolic functions is that their inverse functions can be expressed in terms of simpler functions. The inverse function $\sinh^{-1}(x)$ of $\sinh(x)$ is defined on the whole real line. By definition,

$$\sinh^{-1}(x) = y \text{ if and only if } \sinh(y) = x.$$

This second equation can be written out as,

$$\frac{1}{2}(e^y - e^{-y}) = x.$$

Substituting $z = e^y$ gives,

$$\frac{1}{2}(z - z^{-1}) = x.$$

Multiplying both sides by $2z$ gives,

$$z^2 - 1 = 2xz \Leftrightarrow z^2 - 2xz - 1 = 0.$$

Completing the square gives,

$$(z-x)^2 = x^2 + 1.$$

Taking square roots gives,

$$z = x \pm \sqrt{x^2+1}.$$

Since z equals e^y, z is positive. Thus, the correct square root is,

$$z = x + \sqrt{x^2+1}.$$

Finally this gives,

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

Therefore, the formula for the inverse hyperbolic sine is,

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

The same type of argument also gives,

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

and

$$\tanh^{-1}(x) = \boxed{(1/2)\ln((1+x)/(1-x)),\ -1 < x < 1.}$$

2. Derivatives of the inverse hyperbolic functions. By the same methods used to compute the derivatives of inverse trigonometric functions, the derivatives of the inverse hyperbolic functions are,

$$d\sinh^{-1}(u) = \frac{du}{\sqrt{1+u^2}},$$

$$d\cosh^{-1}(u) = \frac{du}{\sqrt{u^2-1}},\ u \geq 1,$$

$$d\tanh^{-1}(u) = \frac{du}{1-u^2},\ -1 < u < 1.$$

These can also be computed using the formulas for the inverse functions.

3. Inverse substitution. The derivatives of inverse trigonometric and inverse hyperbolic functions allow us to compute more antiderivatives than before, e.g., $\int dx/(\sqrt{x^2-1})$ equals $\cosh^{-1}(x) + C$. Essentially this comes down to making a direct substitution of an inverse function, e.g., $u = \cosh^{-1}(x)$. However, this is logically equivalent to making an *inverse substition*, $x = \cosh(u)$. When the integrand is more complicated, inverse substitution is usually simpler and faster than direct substitution of an inverse function.

Example. Compute the following antiderivative,

$$\int \sqrt{a^2 - x^2}dx.$$

This is not quite the derivative of an inverse function above. However, it is clear that inverse substituting $x = a\sin(\theta)$ will simplify the integrand, because

$$a^2 - x^2 = a^2 - (a\sin(\theta))^2 = a^2(1 - \sin^2(\theta)) = a^2\cos^2(\theta).$$

Thus we have,

$$\int \sqrt{a^2 - x^2}dx, \left\{ \begin{array}{rcl} x & = & a\sin(\theta), \\ dx & = & a\cos(\theta)d\theta \end{array} \right. , \Rightarrow \int \sqrt{a^2\cos^2(\theta)}(a\cos(\theta)d\theta) = a^2 \int \cos^2(\theta)d\theta.$$

Using the half-angle formula, this becomes,

$$a^2 \int \frac{1}{2} + \frac{1}{2}\cos(2\theta)d\theta = a^2\left(\frac{\theta}{2} + \frac{1}{4}\sin(2\theta)\right) + C.$$

Using the double-angle formula and back-substituting gives,

$$\int \sqrt{a^2 - x^2}dx = \boxed{(1/2)(a^2\sin^{-1}(x/a) + x\sqrt{a^2 - x^2}) + C.}$$

4. Three different kinds of integrals, three kinds of inverse substitution. The type of antiderivative where inverse substitution is most successful has the form,

$$\int \frac{F(x, \sqrt{Ax^2 + Bx + C})}{G(x, \sqrt{Ax^2 + Bx + C})}dx,$$

where A, B and C are constants, and $F(x, y)$ and $G(x, y)$ are polynomial functions in the two arguments. Inverse substitution together with partial fractions solves all such antiderivative problems. The first step is to complete the square of the expression $Ax^2 + Bx + C$. This gives,

$$Ax^2 + Bx + C = A\left(x + \frac{B}{2A}\right)^2 - \frac{B^2 - 4AC}{4A}.$$

In particular, making the substition,

$$u = x + \frac{B}{2A}, \quad du = dx,$$

transforms the quadratic into one of 3 possible types,

$$\beta^2u^2 + \alpha^2, \beta^2u^2 - \alpha^2, -\beta^2u^2 + \alpha^2,$$

where,

$$\beta = \sqrt{|A|}, \quad \alpha = \sqrt{\frac{|B^2 - 4AC|}{|4A|}}.$$

Defining $a = \alpha/\beta$, finally the integral is transformed to one of 3 possible types,

$$\textbf{Type I:} \int \frac{F_I(u, \sqrt{a^2 - u^2})}{G_I(u, \sqrt{a^2 - u^2})}du,$$

$$\textbf{Type II:} \quad \int \frac{F_{II}(u, \sqrt{u^2 - a^2})}{G_{II}(u, \sqrt{u^2 - a^2})} du,$$

and

$$\textbf{Type III:} \quad \int \frac{F_{III}(u, \sqrt{a^2 + u^2})}{G_{III}(u, \sqrt{a^2 + u^2})} du.$$

For each of these types, there are 3 possible inverse substitutions: trigonometric, hyperbolic and rational. A flow chart of the 9 possible outcomes will be posted on the course webpage. Here are a couple of examples. In each example, the inverse rational substitution is given, although it was only briefly discussed in lecture.

Example. Compute the following antiderivative,

$$\int \frac{x^2}{\sqrt{a^2 - x^2}} dx.$$

The trigonometric inverse substition is,

$$x = a\sin(\theta), \quad dx = a\cos(\theta)d\theta.$$

The new antiderivative is,

$$\int \frac{a^2\sin^2(\theta)}{\sqrt{a^2 - a^2\sin^2(\theta)}}(a\cos(\theta)d\theta).$$

Simplifying gives,

$$\int a^2\sin^2(\theta)d\theta.$$

This can be simplified using the half-angle formula,

$$a^2 \int \frac{1}{2} - \frac{1}{2}\cos(2\theta)d\theta.$$

This is easily seen to be,

$$a^2\left(\frac{\theta}{2} - \frac{1}{4}\sin(2\theta)\right) + C.$$

Using the double-angle formula and back-substituting,

$$\int \frac{x^2}{\sqrt{a^2 - x^2}} dx = \boxed{(1/2)(a^2\sin^{-2}(x/a) - x\sqrt{a^2 - x^2}) + C.}$$

Alternatively, the hyperbolic inverse substitution is,

$$x = a\tanh(t), \quad dx = a\text{sech}^2(t)dt.$$

The new antiderivative is,

$$\int \frac{a^2\tanh^2(t)}{\sqrt{a^2\mathrm{sech}^2(t)}} a\mathrm{sech}^2(t)dt.$$

Simplifying gives,

$$a^2\int \tanh^2(t)\mathrm{sech}(t)dt = a^2\int \frac{\sinh^2(t)}{\cosh^3(t)}dt.$$

This can be simplified a bit by multiplying numerator and denominator by $\cosh(t)$ and then expressing in terms of $\sinh(t)$ as much as possible,

$$a^2\int \frac{\sinh^2(t)}{\cosh^4(t)}\cosh(t)dt = a^2\int \frac{\sinh^2(t)}{(1+\sinh^2(t))^2}\cosh(t)dt.$$

Make the substitution $u = \sinh(t)$, $du = \cosh(t)dt$ to get,

$$a^2\int \frac{u^2}{(1+u^2)^2}du.$$

This can be rewritten as,

$$a^2\int \frac{1}{1+u^2}du - a^2\int \frac{1}{(1+u^2)^2}du.$$

The first of these terms is just $a^2\tan^{-1}(u)$. However, the second term requires another inverse substitution. All in all, this is not a very efficient approach.

Finally, the rational inverse substitution is,

$$x = a\frac{2t}{1+t^2},\ dx = a\frac{2(1-t^2)}{(1+t^2)^2}dt.$$

The point is that,

$$a^2 - x^2 = a^2\frac{(1-t^2)^2}{(1+t^2)^2}.$$

Thus the new antiderivative is,

$$\int \frac{4a^2t^2}{(1+t^2)^2}\frac{1+t^2}{a(1-t^2)}\frac{2a(1-t^2)}{(1+t^2)^2}dt.$$

This simplifies to,

$$8a^2\int \frac{t^2}{(1+t^2)^3}dt = 8a^2\int \frac{1}{(1+t^2)^2}dt - 8a^2\int \frac{1}{(1+t^2)^3}dt.$$

Notice, these two integrals are the same type that occurred with inverse hyperbolic substitution. But they came up more quickly: rational inverse substitution is more efficient than inverse hyperbolic substitution for this problem. However, both require a further inverse trigonometric substitution. So inverse trigonometric substitution is the most efficient for this problem.

Example. Compute the following antiderivative,

$$\int \frac{x^2}{\sqrt{x^2-a^2}}dx.$$

The trigonometric inverse substitution is,

$$x = a\sec(\theta), \quad dx = a\sec(\theta)\tan(\theta)d\theta.$$

The new antiderivative is,

$$\int \frac{a^2\sec^2(\theta)}{\sqrt{a^2\sec^2(\theta)-a^2}}a\sec(\theta)\tan(\theta)d\theta.$$

Because $\sec^2(\theta)-1$ equals $\tan^2(\theta)$, simplifying gives,

$$a^2\int \sec^3(\theta)d\theta = a^2\int \frac{1}{\cos^3(\theta)}d\theta.$$

This can be simplifed by multiplying numerator and denominator by $\cos(\theta)$ and then expressing in terms of $\sin(\theta)$ as much as possible,

$$a^2\int \frac{1}{\cos^4(\theta)}\cos(\theta)d\theta = a^2\int \frac{1}{(1-\sin^2(\theta))^2}\cos(\theta)d\theta.$$

Make the substitution $u=\sin(\theta)$, $du=\cos(\theta)d\theta)$ to get,

$$a^2\int \frac{1}{(1-u^2)^2}du.$$

This can be computed using partial fractions (not yet discussed).

Alternatively, the hyperbolic inverse substitution is,

$$x = a\cosh(t), \quad dx = a\sinh(t)dt.$$

The new antiderivative is,

$$\int \frac{a^2\cosh^2(t)}{\sqrt{a^2\cosh^2(t)-a^2}}a\sinh(t)dt.$$

Since $\cosh^2(t)-1$ equals $\sinh^2(t)$, simplifying gives,

$$a^2\int \cosh^2(t)dt.$$

This can be simplified using the analogue of the half-angle formula,

$$a^2\int \frac{1}{2}+\frac{1}{2}\cosh(2t)dt.$$

This is easily seen to be,

$$a^2\left(\frac{t}{2}-\frac{1}{4}\sinh(2t)\right)+C.$$

Using the double-angle formula and back-substituting,

$$\int\frac{x^2}{\sqrt{x^2-a^2}}dx=\frac{1}{2}\left(a^2\cosh^{-1}(x/a)-x\sqrt{x^2-a^2}\right).$$

Using the formula for $\cosh^{-1}(x/a)$, this becomes,

$$\int\frac{x^2}{\sqrt{x^2-a^2}}dx=\boxed{(1/2)(a^2\ln(x+\sqrt{x^2-a^2})-x\sqrt{x^2-a^2})+C.}$$

Finally, the rational substitution is,

$$x=a\frac{1+t^2}{2t},\quad dx=a\frac{-(1-t^2)}{2t^2}dt.$$

The point is that,

$$a^2-x^2=a^2\frac{(1-t^2)^2}{(2t)^2}.$$

Thus the new antiderivative is,

$$\int\frac{a^2(1+t^2)^2}{4t^2}\frac{2t}{a(1-t^2)}\frac{-a(1-t^2)}{2t^2}dt.$$

This simplifies to,

$$-\frac{a^2}{4}\int\frac{(1+t^2)^2}{t^3}dt=-\frac{a^2}{4}\int\frac{1}{t^3}+\frac{2}{t}+tdt.$$

This evaluates to,

$$-\frac{a^2}{4}\left(\frac{-1}{2t^2}+2\ln(t)+\frac{t}{2}\right)+C.$$

This is clearly the easiest of the 3 methods for computing the antiderivative, for this problem. However, there still remains the formidable problem of solving for $t=t(x)$, back-substituting, and simplifying the resulting expression. All in all, inverse hyperbolic substitution is the most efficient for this problem.

Lecture 26. November 18, 2005

Homework. Problem Set 7 Part I: (f)–(g); Part II: Problem 1 and Problem 2 (a), (b).

Practice Problems. Course Reader: 5E-8, 5E-10, and please read through Part II of Problem Set 7.

1. Review of inverse substitution and another example. Recall the general strategy for finding an antiderivative of the form,

$$\int \frac{F(x, \sqrt{Ax^2 + Bx + C})}{G(x, \sqrt{Ax^2 + Bx + C})} dx.$$

For definiteness, consider the example,

$$\int \frac{x^2}{\sqrt{x^2 - 2ax + 2a^2}} dx,$$

where a is a constant.

Step 1. Complete the square. Complete the square of the expression $Ax^2 + Bx + C$, inside the radical. In the example,

$$x^2 - 2ax + 2a^2 = (x - a)^2 + a^2.$$

Step 2. Make a linear change of coordinates. Make a linear change of coordinates to simplify the quadratic term to one of the 3 types: $a^2 - x^2$, $x^2 - a^2$, or $x^2 + a^2$. In the example, this means making the linear change of variables,

$$u = x - a, du = dx.$$

The new quadratic term is $u^2 + a^2$, the third type. The new antiderivative is,

$$\int \frac{(u + a)^2}{\sqrt{u^2 + a^2}} du = \int \frac{u^2 + 2u + a^2}{\sqrt{u^2 + a^2}} du.$$

Step 3. Use inverse substitution to eliminate the radicals. There is a choice of inverse substitution: trigonometric, hyperbolic or rational. When starting out, it is a good idea to experiment with all 3. On an exam, usually one choice will be suggested (or even demanded). When no other guidance is given, trigonometric substitution is a good starting point (because you are already very familiar with trigonometric functions).

In the example, to eliminate the radical, the correct inverse trigonometric substitution is,

$$u = a\tan(\theta), \quad du = a\sec^2(\theta)d\theta.$$

This is because the quadratic term becomes,

$$u^2 + a^2 = a^2\tan^2(\theta) + a^2 = a^2\sec^2(\theta).$$

With this substitution, the new antiderivative is,

$$\int \frac{a^2\tan^2(\theta) + 2a^2\tan(\theta) + a^2}{\sqrt{a^2\sec^2(\theta)}} a\sec^2(\theta)d\theta.$$

This simplifies to,

$$a^2 \int (\tan^2(\theta) + 2\tan(\theta) + 1)\sec(\theta)d\theta.$$

This can be written as a sum of 3 terms,

$$a^2 \int \tan^2(\theta)\sec(\theta)d\theta + 2a^2 \int \sec(\theta)\tan(\theta)d\theta + a^2 \int \sec(\theta)d\theta.$$

Step 4. Compute the new antiderivative. If this were only as simple as it sounds, how much easier calculus would be! This step is often difficult in itself. Often it requires at least one more direct substitution. Sometimes, it also requires a partial fractions decomposition. We will return to this step below.

Step 5. Back-substitute. This is always a step for a method using direct substitution or inverse substitution. This step frequently introduces terms like $\cos(\tan^{-1}(x))$. Time-permitting (or when specifically instructed to do so), these terms should be simplifed using the right-triangle method from lecture,

$$\theta = \tan^{-1}(x), x/1 = \tan(\theta) = \text{Opposite}/\text{Adjacent}, \text{Hypotenuse} = \sqrt{1+x^2},$$

$$\cos(\theta) = \text{Adjacent}/\text{Hypotenuse} = \boxed{1/\sqrt{1+x^2}.}$$

Step 6. Check your answer. When feasible, check your answer. Since differentiation is so much faster than antidifferentiation, it is usually quite easy to check an antiderivative is correct.

Example. The tricky part is, of course, Step 4. In the example, the integral broke into 3 terms,

$$a^2 \int \tan^2(\theta)\sec(\theta)d\theta + 2a^2 \int \sec(\theta)\tan(\theta)d\theta + a^2 \int \sec(\theta)d\theta.$$

The last antiderivative was actually Problem 3(b) from Part II of Problem Set 4. It turns out to be,

$$a^2 \int \sec(\theta)d\theta = a^2 \ln(u + \sqrt{u^2+a^2}) + C = \boxed{a^2 \ln(x - a + \sqrt{x^2 - 2ax + 2a^2}) + C.}$$

The middle antiderivative is simply the derivative of $\sec(\theta) = \sqrt{1+\tan^2(\theta)}$. So the middle term is,

$$2a^2 \int \sec(\theta)\tan(\theta)d\theta = 2a^2\sec(\theta) + C = 2a\sqrt{a^2+u^2} + C = \boxed{2a\sqrt{x^2 - 2ax + 2a^2} + C.}$$

But the final term does not simplify in an obvious way. In such cases, it is best to express everything in terms of $\sin(\theta)$ and $\cos(\theta)$ to get a fresh perspective,

$$a^2 \int \tan^2(\theta)\sec(\theta)d\theta = a^2 \int \frac{\sin^2(\theta)}{\cos^3(\theta)}d\theta.$$

Multiplying numerator and denominator by $\cos(\theta)$ and expressing in terms of $\sin(\theta)$ gives,

$$a^2 \int \frac{\sin^2(\theta)}{(\cos^2(\theta))^2} \cos(\theta) d\theta = a^2 \int \frac{\sin^2(\theta)}{(1-\sin^2(\theta))^2} \cos(\theta) d\theta.$$

Now substitute for $\sin(\theta)$,

$$z = \sin(\theta), \quad dz = \cos(\theta) d\theta.$$

The new antiderivative is,

$$\int \frac{z^2}{(1-z^2)^2} dz.$$

How do we compute this antiderivative? That is the topic of partial fractions.

Remark: In lecture the solution was done a bit differently. This led to a slightly different antiderivative,

$$\int \frac{1}{(1-z^2)^2} dz.$$

Notice the difference of these 2 antiderivatives is,

$$\int \frac{(1)-(z^2)}{(1-z^2)^2} dz = \int \frac{1}{(1-z^2)} dz.$$

This was computed in Problem 3(a), Part II of Problem Set 4. Thus, computing either of the 2 antiderivatives gives both of them.

2. Antidifferentiating simple rational expressions. A *rational expression* is a fraction of polynomials, $F(x)/G(x)$. These frequently arise in Step 4 of the algorithm above. From the point of view of antidifferentiation, the simplest rational expressions are either *polynomials*,

$$q(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0,$$

or else *partial fractions*,

$$\frac{A}{(x-a)^m}.$$

There are 2 other kinds of partial fractions which were not emphasized in lecture,

$$\frac{B(x-a)}{((x-a)^2+b^2)^m} \text{ and } \frac{C}{((x-a)^2+b^2)^m}.$$

These 2 kinds come up less often than the first kind. But they do come up, for instance, when studying Laplace transforms in 18.03. Both polynomials and partial fractions are (relatively) easy to antidifferentiate. The antiderivative of a polynomial is,

$$\int q(x) dx = \boxed{\tfrac{a_n}{(n+1)} x^{n+1} + \tfrac{a_{n-1}}{n} x^n + \cdots + \tfrac{a_1}{2} x^2 + a_0 x + C.}$$

The antiderivative of the first kind of partial fraction is,

$$\int \frac{A}{(}x-a)^{-m}dx = \begin{cases} (-A/(m-1))(x-a)^{-(m-1)} + C, & m \geq 2, \\ A\ln(|x-a|) + C, & m = 1 \end{cases}$$

The second kind of partial fraction can be computed with a direct substitution $v = (x-a)^2 + b^2$, $dv = 2(x-a)dx$,

$$\int \frac{B(x-a)}{((x-a)^2+b^2)^m}dx = \frac{B}{2}\int \frac{dv}{v^m} = \begin{cases} (-B/(2m-2))((x-a)^2+b^2)^{-(m-1)} + C, & m \geq 2, \\ (B/2)\ln((x-a)^2+b^2) + C, & m = 1 \end{cases}$$

The third kind of partial fraction can be computed with an inverse substitution $x = b\tan(\theta) + a$, $dx = b\sec^2(\theta)d\theta$,

$$\int \frac{C}{((x-a)^2+b^2)^m}dx = (C/b^{2m-1})\int \cos^{2m-2}(\theta)d\theta.$$

Integration by parts gives a reduction formula for such integrals; see Problems (i) and (j), Part I of Problem Set 7.

3. Simplifying rational expressions: division and factoring Many rational expressions that come up are not of the simple kinds above. The goal is to express an arbitrary rational expression as the sum of a polynomial and partial fractions. The first step is polynomial division. Given a fraction $F(x)/G(x)$, apply polynomial division to get a factorization with remainder,

$$F(x) = q(x)G(x) + r(x),$$

where $q(x)$ is a polynomial and $r(x)$ is a polynomial of degree less than $\deg(G(x))$. This leads to the *reduced form* of a rational expression,

$$\frac{F(x)}{G(x)} = q(x) + \frac{r(x)}{G(x)}.$$

Example. I forgot the example from lecture. Here is a similar example. Find the reduced form of $(x^3+1)/(x^2+3x+2)$. The polynomial division algorithm gives,

$$x^3 + 1 = (x^2+3x+2)(x-3) + (7x+7),$$

Thus $q(x)$ is $x-3$ and $r(x)$ is $7x+7$. So the reduced form is,

$$\frac{x^3+1}{x^2+3x+2} = x - 3 + \tfrac{7x+7}{x^2+3x+2}.$$

The next step is to factor the denominator into a product of linear and irreducible quadratic factors,

$$G(x) = A(x-a_1)^{m_1}\cdot(x-a_2)^{m_2}\cdot\dots\cdot(x-a_k)^{m_k}\cdot((x-\alpha_1)^2+b_1^2)^{n_1}\cdot\dots\cdot((x-\alpha_l)^2+b_l^2)^{n_l}.$$

Here k and l are nonnegative integers and $m_1, \dots, m_k, n_1, \dots, n_l$ are positive integers. Also, $a_1, \dots, a_k$, $\alpha_1, \dots, \alpha_l$, and $\beta_1, \dots, \beta_l$ are real numbers. The last l factors were not discussed in lecture until the end of lecture. Although they are important, they do not often come up in this course.

The *Fundamental Theorem of Algebra* asserts that every polynomial with real coefficients has a factorization as above. However, finding the factorization can be very difficult. In all exercises and exam problems, either the factorization is easy, or the factorization will be given to you. Whenever possible, cancel common factors from the numerator and denominator.

Example. In the example, the quadratic formula gives the factorization,

$$x^2 + 3x + 2 = \boxed{(x+2)(x+1).}$$

The numerator $r(x)$ is $7(x+1)$. Thus the numerator and denominator have a common factor. This leads to a better reduced form,

$$\frac{x^3+1}{x^2+3x+2} = x - 3 + \frac{7}{x+2}.$$

This can now be integrated to give,

$$\int \frac{x^3+1}{x^2+3x+2}dx = \boxed{(x^2/2) - 3x + 7\ln(|x+2|) + C.}$$

4. Simplifying rational expressions: partial fraction decomposition. Using the last part, every rational expression can be written in the form,

$$\frac{F(x)}{G(x)} = q(x) + \frac{r(x)}{H(x)} = q(x) + \frac{r(x)}{(x-a_1)^{m_1} \cdot \dots \cdot (x-a_k)^{m_k} \cdot ((x-\alpha_1)^2 + b_1^2)^{n_1} \cdot \dots \cdot ((x-\alpha_l)^2 + b_l^2)^{n_l}},$$

where $q(x)$ is a polynomial, the degree of $r(x)$ is less than the degree of $H(x)$, and $r(x)$ has no common factor with $H(x)$. This can be further simplified using *partial fraction decomposition*. It is a fact that every rational expression $r(x)/H(x)$ can be written in the form,

$$\left(\frac{C_{1,1}}{x-a_1} + \frac{C_{1,2}}{(x-a_1)^2} + \dots + \frac{C_{1,m_1}}{(x-a_1)^{m_1}}\right) + \dots + \left(\frac{C_{k,1}}{x-a_k} + \dots + \frac{C_{k,m_k}}{(x-a_k)^{m_k}}\right) +$$

$$\left(\frac{D_{1,1}(x-\alpha_1)}{(x-\alpha_1)^2 + b_1^2} + \frac{E_{1,1}}{(x-\alpha_1)^2 + b_1^2} + \dots + \frac{D_{1,n_1}(x-\alpha_1)}{((x-\alpha_1)^2 + b_1^2)^{n_1}} + \frac{E_{1,n_1}}{((x-\alpha_1)^2 + b_1^2)^{n_1}}\right) + \dots$$

$$+ \left(\frac{D_{l,1}(x-\alpha_l)}{(x-\alpha_l)^2 + b_l^2} + \frac{E_{l,1}}{(x-\alpha_l)^2 + b_l^2} + \dots + \frac{D_{l,n_1}(x-\alpha_l)}{((x-\alpha_l)^2 + b_l^2)^{n_l}} + \frac{E_{l,n_1}}{((x-\alpha_l)^2 + b_l^2)^{n_l}}\right).$$

Here all the terms $C_{i,j}$, $D_{i,j}$ and $E_{i,j}$ are real constants. This sum of partial fractions is called the partial fraction decomposition of $r(x)/H(x)$. The difficulty is precisely to find the constants $C_{i,j}$, $D_{i,j}$, and $E_{i,j}$.

One approach, which always works but is quite inefficient, is simply to multiply all terms by the denominator $H(x)$, and then gather coefficients of powers of x. This will give a collection of linear equations in the unknowns $C_{i,j}$, $D_{i,j}$ and $E_{i,j}$. There is a unique solution of this set of linear equations. Methods of linear algebra, e.g., Gauss-Jordan elimination, give an algorithm for finding the solution.

Example. Find the partial fraction decomposition of,

$$\frac{1}{1-x^2}.$$

In fact this was Problem 3(a), Part II of Problem Set 4. The partial fraction decomposition will have the form,

$$\frac{1}{1-x^2} = \frac{A}{x+1} + \frac{B}{x-1}.$$

Multiplying both sides of the equation by $x^2 - 1 = (x+1)(x-1)$ gives,

$$-1 = A(x-1) + B(x+1) = (A+B)x + (B-A).$$

This gives the system of 2 linear equations in 2 unknowns,

$$\begin{cases} A + B = 0, \\ -A + B = -1 \end{cases}$$

Solving the first equation for $B = -A$ and plugging this into the second equation gives,

$$-A + (-A) = -1 \Leftrightarrow 2A = 1 \Leftrightarrow A = \frac{1}{2}.$$

Thus $B = -A = -1/2$. So the partial fraction decomposition is,

$$\frac{1}{1-x^2} = \boxed{\frac{1}{2}\frac{1}{x+1} + \frac{-1}{2}\frac{1}{x-1}.}$$

5. The Heaviside cover-up method. The Heaviside cover-up method is a method for determining many of the coefficients $C_{i,j}$. For each highest power of a linear factor occuring in $H(x)$, say $(x-a_i)^{m_i}$, cover-up that term, and substitute $x = a_i$ in the remaining polynomial. Then C_{i,m_i} equals the value,

$$C_{i,m_i} = \left.\frac{r(x)}{H(x)/(x-a_i)^{m_i}}\right|_{x=a_i}.$$

The proof is quite simple. Multiply every term in the partial fraction decomposition by $(x-a_i)^{m_i}$. One term is $(x-a_i)^{m_i}(C_{i,m_i}/(x-a_i)^{m_i}) = C_{i,m_i}$. Every other term has a factor $(x-a_i)$ that is not cancelled by the denominator. Thus plugging in $x = a_i$, every other term is 0. And the only remaining term is A_{i,m_i}.

Example. Find many of the terms in the partial fraction decomposition,

$$\frac{z^2}{(z^2-1)^2} = \frac{z^2}{(z+1)^2(z-1)^2}.$$

The partial fraction decomposition will be,

$$\frac{z^2}{(z+1)^2(z-1)^2} = \frac{C_{1,2}}{(z+1)^2} + \frac{C_{2,2}}{(z-1)^2} + \frac{C_{1,1}}{z+1} + \frac{C_{2,1}}{z-1}.$$

Using the Heaviside cover-up method,

$$C_{1,2} = \left.\frac{z^2}{(z-1)^2}\right|_{z=-1} = \frac{(-1)^2}{(-2)^2} = \frac{1}{4}.$$

Also,

$$C_{2,2} = \left.\frac{z^2}{(z+1)^2}\right|_{z=+1} = \frac{(+1)^2}{(+2)^2} = \frac{1}{4}.$$

Thus the partial fraction decomposition is,

$$\frac{z^2}{(1-z^2)^2} = \boxed{\frac{1}{4}\frac{1}{(z+1)^2} + \frac{1}{4}\frac{1}{(z-1)^2} + \frac{C_{1,1}}{z+1} + \frac{C_{2,1}}{z-1}.}$$

As this example illustrates, the Heaviside cover-up method does not always determine all coefficients. However, it reduces the number of coefficients. To find the remaining coefficients, either clear denominators, or else substitute for x some useful numbers (where $H(x)$ is nonzero), and solve the resulting linear equations.

Example. Find the full partial fraction decomposition of,

$$\frac{z^2}{(1-z^2)^2}.$$

The rational expression is unchanged by the substitution $z \leftrightarrow -z$. Thus the same is true for the partial fraction decomposition. Therefore $C_{2,1}$ equals $-C_{1,1}$. This gives,

$$\frac{z^2}{(1-z^2)^2} = \frac{1}{4}\frac{1}{(z+1)^2} + \frac{1}{4}\frac{1}{(z-1)^2} + \frac{C_{1,1}}{z+1} + \frac{-C_{1,1}}{z-1}.$$

Finally, plug in $z=0$ to get,

$$0 = \frac{1}{4}\frac{1}{(+1)^2} + \frac{1}{4}\frac{1}{(-1)^2} + \frac{C_{1,1}}{+1} + \frac{-C_{1,1}}{-1} = \frac{1}{2} + 2C_{1,1}.$$

Solving gives $C_{1,1} = -1/4$. Finally this gives the full partial fraction decomposition,

$$\frac{z^2}{(1-z^2)^2} = \boxed{(1/4)(1/(z+1)^2 + 1/(z-1)^2 - 1/(z+1) + 1/(z-1)).}$$

Using the partial fraction decomposition, the antiderivative is,

$$\int \frac{z^2}{(1-z^2)^2}dz = \frac{1}{4}\left(\frac{-1}{z+1} + \frac{-1}{z-1} + \ln(|z-1|/|z+1|)\right) + C = \frac{1}{2}\left(\frac{z}{1-z^2} + \ln(\sqrt{1-z^2}) - \ln(1+z)\right) + C.$$

This allows us to finish the computation of the antiderivative from the beginning of the lecture. This is left as an exercise.

Lecture 27. November 22, 2005

Homework. Problem Set 7 Part II: Problem 2.

Practice Problems. Course Reader: 5F-2, 5F-3, 5F-4, 5F-5.

1. Integration by parts. The differential form of the product rule is,

$$d(uv) = udv + vdu.$$

An equivalent form is,

$$udv = d(uv) - vdu.$$

This gives a very useful antidifferentiation formula,

$$\int udv = uv - \int vdu.$$

This formula is *integration by parts.*

Example. Compute the antiderivative of,

$$\int x\cos(x)dx.$$

Set u to be x and dv to be $\cos(x)dx$. Then u, v, du and dv are,

$$\begin{array}{cc} u = x & dv = \cos(x)dx, \\ du = dx & v = \sin(x) \end{array}$$

Using integration by parts,

$$\int udv = uv - \int vdu,$$

$$\int x\cos(x)dx = x\sin(x) - \int \sin(x)dx.$$

The new integral is easy to evaluate. Altogether this gives,

$$\int x\cos(x)dx = \boxed{x\sin(x) + \cos(x) + C.}$$

Because it is much easier to differentiate than the antidifferentiate, it is a good idea to check you answer.

2. How to use integration by parts. The goal of integration by parts is to replace a complicated integral, $\int udv$, by a simpler integral $\int vdu$. What this usually means is that du should be simpler than u, and v should be no more complicated than dv. This was the case in the last example. However, occasionally this is not the case.

Example. Use integration by parts to compute the antiderivative,

$$\int \ln(x)dx.$$

There is very little choice here, if we are to use only integration by parts. Set u to be $\ln(x)$ and set dv to be dx. Then u, v, du and dv are,

$$\begin{array}{rl} u = \ln(x), & dv = dx \\ du = dx/x, & v = x \end{array}$$

Using integration by parts,

$$\int udv = uv - \int vdu,$$

$$\int \ln(x)dx = x\ln(x) - \int dx.$$

The new integral is easy to evaluate. Altogether this gives,

$$\int \ln(x)dx = \boxed{x\ln(x) - x + C.}$$

Notice this example does not follow the general rule. The integral $v = x$ is strictly more complicated than $dv = dx$. However, $du = dx/x$ is much simpler than $u = \ln(x)$. So $vdu = dx$ is simpler than $udv = \ln(x)dx$. The lesson is to be flexible when antidifferentiating. Try different things, and see which one works. For example, another approach to this problem, which ultimately comes down to integration by parts again, is to make an inverse substitution,

$$x = e^t, \quad dx = e^t dt.$$

The new integral is,

$$\int \ln(x)dx = \int te^t dt.$$

Set $u = t$ and $du = e^t dt$. Then u, v, du and dv are,

$$\begin{array}{rl} u = t, & dv = e^t dt \\ du = dt, & v = e^t \end{array}$$

Using integration by parts,

$$\int u dv = uv - \int v du,$$

$$\int te^t dt = te^t - \int e^t dt.$$

The new integral is easy to evaluate. Altogether this gives,

$$\int te^t dt = te^t - e^t + C.$$

Back-substituting for x gives,

$$\int \ln(x) dx = \boxed{x\ln(x) - x + C.}$$

This agrees with the earlier answer.

2. Reduction formulas. It often happens that an integral can be computed only be repeated application of integration by parts. It sometimes happens that integration by parts gives the induction step to solve infinitely many integrals. In this case, the formula given by integration by parts is called a *reduction formula*.

Example. Use integration by parts to give a reduction formula for,

$$\int [\ln(x)]^n dx.$$

Now there is much more choice for u and dv. The simplest choice is to set $u = [\ln(x)]^n$ and $dv = dx$. Then u, v, du and dv are,

$$\begin{array}{rr} u = [\ln(x)]^n, & dv = dx \\ du = n[\ln(x)]^{n-1}/x dx, & v = x \end{array}$$

Using integration by parts,

$$\int u dv = uv - \int v du,$$

$$\int [\ln(x)]^n dx = \boxed{x[\ln(x)]^n - n\int [\ln(x)]^{n-1} dx.}$$

The new integral is simpler than the original integral. And repeated application of the formula eventually leads to a formula for the integral. Thus this is a reduction formula. For instance, this gives,

$$\int [\ln(x)]^2 dx = x[\ln(x)]^2 - 2\int \ln(x) dx.$$

The new integral was already computed. Altogether this gives,

$$\int [\ln(x)]^2 dx = \boxed{x[\ln(x)]^2 - 2x\ln(x) + 2x + C.}$$

Example. Use integration by parts to find a reduction formula for,

$$\int t^n e^t dt.$$

The simplest choice is to set $u = t^n$ and $dv = e^t dt$. Then u, v, du and dv are,

$$u = t^n, \qquad dv = e^t dt$$
$$du = nt^{n-1}dt, \qquad v = e^t$$

Using integration by parts,

$$\int u dv = uv - \int v du,$$

$$\int t^n e^t dt = \boxed{t^n e^t - n \int t^{n-1} e^t dt.}$$

Notice how similar this answer was to the answer of the previous example. The connection comes from the inverse substitution,

$$x = e^t, \quad dx = e^t dt,$$

so that,

$$\int [\ln(x)]^n dx = \int t^n e^t dt.$$

3. Advanced reduction formulas. Sometimes a reduction formula can only be obtained by repeatedly applying integration by parts or by using some other identity.

Example. Using integration by parts to find a reduction formula for,

$$\int [\sin(x)]^n dx, \quad n \geq 1.$$

One choice is to set $u = [\sin(x)]^{n-1}$ and to set $dv = \sin(x)dx$. Then u, v, du and dv are,

$$u = [\sin(x)]^{n-1}, \qquad dv = \sin(x)dx$$
$$du = (n-1)[\sin(x)]^{n-2}\cos(x)dx, \quad v = -\cos(x).$$

Using integration by parts,

$$\int u dv = uv - \int v du,$$

$$\int [\sin(x)]^n dx = -[\sin(x)]^{n-1}\cos(x) + (n-1)\int [\sin(x)]^{n-2}\cos^2(x)dx.$$

At first blush, this is more complicated than the original integral since it involves both $\sin(x)$ and $\cos(x)$. But $\cos^2(x)$ equals $1 - \sin^2(x)$. This substitution gives,

$$\int [\sin(x)]^n dx = -[\sin(x)]^{n-1}\cos(x) + (n-1)\int [\sin(x)]^{n-2}dx - (n-1)\int [\sin(x)]^n dx.$$

This certainly seems circular: the new formula for the integral involves the integral we were looking for. However, bringing like terms to one side of the equation gives,

$$\int [\sin(x)]^n dx + (n-1)\int [\sin(x)]^n = -[\sin(x)]^{n-1}\cos(x) + (n-1)\int [\sin(x)]^{n-2} dx.$$

Cleaning this up a bit gives the reduction formula,

$$\int [\sin(x)]^n dx = \boxed{-[\sin(x)]^{n-1}\cos(x)/n + (n-1)/n \int [\sin(x)]^{n-2} dx.}$$

Lecture 28. December 1, 2005

Homework. Problem Set 8 Part I: (a) and (b).

Practice Problems. Course Reader: 6A-1, 6A-2.

1. Indeterminate forms. Expressions of the form $0/0$, ∞/∞, $0 \times \infty$, $\infty - \infty$, 0^∞ and ∞^0 are called *indeterminate forms*. To be precise, none of these expressions is defined in mathematics. However, if a naive limit computation $\lim_{x\to a} F(x)$ leads to an indeterminate form, it often happens that a more careful computation using calculus eliminates the indeterminate form.

Example. Let b be any real number. Compute the limit as x approaches 0 of $F(x) = (b+1/x) - 1/x$, $x \neq 0$. If we evaluate this limit in a naive manner, we get,

$$\lim_{x\to 0} F(x) = \lim_{x\to 0}\left(b + \frac{1}{x}\right) - \left(\frac{1}{x}\right) \text{ “=” } \lim_{x\to 0} b + \frac{1}{x} - \lim_{x\to 0}\frac{1}{x} = \infty - \infty.$$

This is an indeterminate form. In other words, the computation of the limit failed to give any useful information. The reason is that the general formula,

$$\lim_{x\to a}[g(x) + h(x)] = \lim_{x\to a} g(x) - \lim_{x\to a} h(x),$$

only holds if all three limits are defined, which they are not in our case.

Of course $F(x)$ is simply the constant function with value b. Therefore,

$$\lim_{x\to 0} F(x) = \lim_{x\to 0} b = \boxed{b.}$$

Thus, a more careful computation proves the limit exists and gives its value.

2. The Mean Value Theorem revisited. Recall the Mean Value Theorem: If $f(x)$ is continuous on $[a, b]$ and differentiable on (a, b), then for some c strictly between a and b,

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

Thus, given two such functions $f(x)$ and $g(x)$ such that $g(b) - g(a)$ is nonzero, there exist two values c_1 and c_2 strictly between a and b such that,

$$\frac{f'(c_1)}{g'(c_2)} = \frac{(f(b) - f(a))/(b-a)}{(g(b) - g(a))/(b-a)} = \frac{f(b) - f(a)}{g(b) - g(a)}.$$

Is there a single value $c = c_1 = c_2$ where this equality holds?

The answer is yes. Form the function

$$F(x) = (f(b) - f(a))(g(x) - g(a)) - (g(b) - g(a))(f(x) - f(a)).$$

Since $f(x)$ and $g(x)$ are continuous on $[a, b]$, also $F(x)$ is continuous on $[a, b]$. Since $f(x)$ and $g(x)$ are differentiable on (a, b), also $F(x)$ is differentiable on (a, b). Moreover,

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

Thus, by the Mean Value Theorem, there exists a value c strictly between a and b such that $F'(c) = 0$. By a straightforward computation,

$$F'(c) = (f(b) - f(a))g'(c) - (g(b) - g(a))f'(c).$$

This proves the *Generalized Mean Value Theorem*. The main consequence of the Generalized Mean Value Theorem is the following result.

Proposition. Let $f(x)$ and $g(x)$ be continuous functions on $[a, b]$ that are differentiable on (a, b). If $g'(x)$ is nonzero on (a, b), then $g(x) - g(a)$ is nonzero for all $a < x < b$ so that the expression,

$$\frac{f(x) - f(a)}{g(x) - g(a)}$$

is defined. The right-handed limit,

$$\lim_{x \to a^+} \frac{f(x) - f(a)}{g(x) - g(a)},$$

exists if and only if the right-handed limit,

$$\lim_{x \to a^+} \frac{f'(x)}{g'(x)},$$

exists. If both limits exist, they are equal,

$$\lim_{x \to a^+} \frac{f(x) - f(a)}{g(x) - g(a)} = \lim_{x \to a^+} \frac{f'(x)}{g'(x)}.$$

A similar result holds for left-handed limits. The proof follows by applying the Generalized Mean Value Theorem to the interval $[a, x]$ to replace $(f(x) - f(a))/(g(x) - g(a))$ by $f'(c)/g'(c)$. Then x approaches a as c approaches a.

3. L'Hospital's rule. The most important case of the proposition is *L'Hospital's rule.* This is exactly the case when $f(a) = g(a) = 0$. In this case, a naive computation would give,

$$\lim_{x \to a^+} \frac{f(x)}{g(x)} \text{``=''} \frac{f(a)}{g(a)} = \frac{0}{0},$$

which is an indeterminate form. Again, the problem is that the general formula,

$$\lim_{x \to a^+} \frac{f(x)}{g(x)} = \frac{\lim_{x \to a^+} f(x)}{\lim_{x \to a^+} g(x)},$$

only holds if all three limits are defined, and the limit $\lim_{x \to a^+} g(x)$ is nonzero. Since the limit is zero, the formula does not hold.

However, if $f'(x)$ and $g'(x)$ exist, and if $g'(x)$ is nonzero, then the proposition has the following consequence, known as L'Hospital's rule,

$$\boxed{\lim_{x \to a^+} f(x)/g(x) = \lim_{x \to a^+} f'(x)/g'(x).}$$

Examples.

$$\lim_{x \to 0} \frac{\sinh(x)}{\sin(x)} = \lim_{x \to 0} \frac{\cosh(x)}{\cos(x)} = \frac{1}{1} = \boxed{1.}$$

$$\lim_{x \to 2} \frac{4x^3 - 32}{x^2 - x - 2} = \lim_{x \to 2} \frac{12x^2}{2x - 1} = \frac{12 \cdot 4}{2 \cdot 2 - 1} = \frac{48}{3} = \boxed{16.}$$

$$\lim_{x \to 0} \frac{1 - \cos(x)}{x^2} = \lim_{x \to 0} \sin(x) 2x = \lim_{x \to 0} \cos(x) 2 = \boxed{1/2.}$$

4. L'Hospital's rule for other indeterminate forms. L'Hospital's rule can be used to compute limits that naively lead to indeterminate forms other than $0/0$. For instance, if

$$\lim_{x \to a^+} f(x) = \lim_{x \to a^+} g(x) = \infty,$$

then the naive computation gives,

$$\lim_{x \to a^+} \frac{f(x)}{g(x)} \text{``=''} \frac{\infty}{\infty}.$$

Now observe that,

$$\lim_{x \to a^+} (1/f(x)) = \lim_{x \to a^+} (1/g(x)) = 0.$$

Therefore, if both $g(x)$ and $g'(x)$ are nonzero on (a, b), then L'Hospital's rule gives,

$$\lim_{x \to a^+} \frac{(1/f(x))}{(1/g(x))} = \lim_{x \to a^+} \frac{(1/f(x))'}{(1/g(x))'} = \lim_{x \to a^+} \frac{-f'(x)/f(x)^2}{-g'(x)/g(x)^2}.$$

Assuming that the limits,

$$\lim_{x\to a^+} \frac{f(x)}{g(x)}, \text{ and } \lim_{x\to a^+} \frac{f'(x)}{g'(x)}$$

are defined and nonzero, the formula above can be re-written as,

$$\left(\lim_{x\to a^+} \frac{f(x)}{g(x)}\right)^{-1} = \left(\lim_{x\to a^+} \frac{f'(x)}{g'(x)}\right) \cdot \left(\lim_{x\to a^+} \frac{f(x)}{g(x)}\right)^{-2}.$$

Solving gives,

$$\boxed{\lim_{x\to a^+} f(x)/g(x) = \lim_{x\to a^+} f'(x)/g'(x),}$$

if both limits are defined and nonzero. In fact, a better result is true (with a more subtle proof): if the second limit is defined, then the first limit is defined and the 2 are equal (whether or not they are zero).

Example.

$$\lim_{x\to\pi/2^+} \frac{\ln(x-\pi/2)}{\sec(x)} = \lim_{x\to\pi/2^+} \frac{1/(x-\pi/2)}{\sec(x)\tan(x)} = \cdots = \boxed{0.}$$

By similar arguments, other indeterminate forms can also be reduced to L'Hospital's rule. Also, limits of the form,

$$\lim_{x\to\infty} F(x)$$

giving indeterminate forms can often be reduced to L'Hospital's rule. The moral is that the formula,

$$\lim_{x\to a} \frac{f(x)}{g(x)} = \lim_{x\to a} \frac{f'(x)}{g'(x)},$$

is almost always true if $f(a)/g(a)$ is an indeterminate form. But a certain amount of care should be used, since occasionally this fails.

Lecture 29. December 2, 2005

Homework. Problem Set 8 Part I: (c), (d) and (e); Part II: Problems 1 and 2.

Practice Problems. Course Reader: 6B-7.

1. A problem with Riemann integrals. Riemann integrals are defined in very many cases. The result we use most often is that for a piecewise continuous function $f(x)$ on a bounded interval $[a,b]$, the Riemann integral,

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

exists (and equals a finite number). What if the interval is unbounded, e.g., $[a,\infty)$? Quite simply, the Riemann integral is not defined. This isn't a problem with our methods for computing integrals. It is a problem with the very definition of the Riemann integral. In fact, this is only the first of many problems with the definition of the Riemann integral. Eventually these problems led mathematicians to develop a better definition, the *Lebesgue integral*, which is studied in course 18.103. Luckily, the particular problem of defining the integral on unbounded intervals can be easily overcome using limits (with no need to use the Lebesgue integral).

2. Improper integrals of the first kind. Let $f(x)$ be defined on the interval $[a, \infty)$. If for every number $t > a$ the function $f(x)$ is Riemann integrable on $[a, t]$, and if the limit,

$$\lim_{t\to\infty} \int_a^t f(x)dx,$$

exists, then we say the *improper integral*,

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

is defined and its value is,

$$\int_a^\infty f(x)dx = \lim_{t\to\infty} \int_a^t f(x)dx.$$

Please note, this is a *new* definition. It is not a theorem about Riemann integrals.

Example. Let $p > 1$ be a real number. Then for every $t > 1$, the integral,

$$\int_1^t \frac{1}{x^p}dx,$$

exists and equals,

$$\left(-\frac{1}{(p-1)x^{p-1}}\right|_1^t = \frac{1}{p-1} - \frac{1}{(p-1)t^{p-1}}.$$

Since p is greater than 1, the limit,

$$\lim_{t\to\infty} \frac{1}{t^{p-1}},$$

exists and equals 0. Therefore,

$$\lim_{t\to\infty} \int_1^t \frac{1}{x^p}dx,$$

exists and equals,

$$\frac{1}{p-1}.$$

Therefore the improper integral exists and equals,

$$\int_1^\infty \frac{1}{x^p}dx = \boxed{1/(p-1).}$$

On the other hand, when p equals 1, then,

$$\int_1^t \frac{1}{x}dx = \ln(t).$$

Since the limit $\lim_{t\to\infty} \ln(t)$ is not defined (or more precisely, equals $+\infty$), the improper integral,

$$\int_1^\infty \frac{1}{x}dx,$$

is not defined (or more precisely, equals $+\infty$).

Example. For $t > 0$, the integral,

$$\int_0^t \cos(x)dx,$$

exists and equals $\sin(t)$. Even though all values $\sin(t)$ are defined and bounded, the limit,

$$\lim_{t\to\infty} \sin(t),$$

is not defined (essentially because it never settles down). Therefore the improper integral,

$$\int_0^\infty \cos(x)dx,$$

is not defined.

3. Improper integrals of the second kind. Here is a second problem with the Riemann integral. Let $[a, b]$ be a bounded interval. Let $f(x)$ be a function that is bounded on $[t, b]$ for every $a < t < b$, but which is unbounded on $[a, b]$. According to the definition of the Riemann integral,

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

is not defined. However, it may happen that for every $a < t < b$, the integral,

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

is defined and the limit,

$$\lim_{t\to a^+} \int_t^b f(x)dx,$$

is defined. In this case, we say the *improper integral*,

$$\int_{a^+}^b f(x)dx,$$

is defined and its value is,

$$\int_{a^+}^b f(x)dx = \lim_{t\to a^+} \int_t^b f(x)dx.$$

Similarly, if $f(x)$ is Riemann integrable on every interval $[a,t]$ for $a < t < b$, and if

$$\lim_{t \to b^-} \int_a^t f(x)dx,$$

exists, we say the *improper integral*,

$$\int_a^{b_-} f(x)dx,$$

exists and its value is,

$$\int_a^{b^-} f(x)dx = \lim_{t \to b^-} \int_a^t f(x)dx.$$

Example. Let p be a real number in the range $0 < p < 1$. Because the function $1/x^p$ is unbounded on $[0,1]$, the Riemann integral,

$$\int_0^1 \frac{1}{x^p}dx,$$

is not defined. However, for every $0 < t < 1$, the Riemann integral,

$$\int_t^1 \frac{1}{x^p}dx,$$

is defined equals,

$$\frac{1 - t^{1-p}}{1-p}.$$

Since $0 < p < 1$, the limit,

$$\lim_{t \to 0} t^{1-p},$$

exists and equals 0. Therefore,

$$\lim_{t \to 0} \int_t^1 \frac{1}{x^p}dx,$$

exists and equals $1/(1-p)$. Therefore the improper integral,

$$\int_{0^+}^1 \frac{1}{x^p}dx,$$

exists and its value is,

$$\int_{0^+}^1 \frac{1}{x^p}dx = \boxed{1/(1-p).}$$

4. The Comparison Test. When is an improper integral defined? This is equivalent to asking when a limit is defined. Therefore, every rule for convergence of a limit gives a rule for convergence of an improper integral. There are 2 basic rules for convergence of a limit.

The squeezing lemma. If $F(x) \leq G(x) \leq H(x)$ on an interval, if $\lim_{x\to a} F(x)$ and $\lim_{x\to a} H(x)$ exist, and if $\lim_{x\to a} F(x)$ equals $\lim_{x\to a} H(x)$, then $\lim_{x\to a} G(x)$ exists and equals the other 2 limits.

Monotone bounded limits. If $F(x)$ is monotone increasing and bounded above on $[a, b)$, then $\lim_{x\to b^-} F(x)$ exists. Similarly, if $F(x)$ is monotone decreasing and bounded below, then $\lim_{x\to b^-} F(x)$ exists, if $F(x)$ is monotone increasing and bounded below, then $\lim_{x\to a^+} F(x)$ exists, and if $F(x)$ is monotone decreasing and bounded above, then $\lim_{x\to a^+} F(x)$ exists.

These give the following tests for convergence of an improper integral.

Squeezing lemma. If $f(x) \leq g(x) \leq h(x)$ on the interval $[a, \infty)$, and if the improper integrals,

$$\int_a^\infty f(x)dx \text{ and } \int_a^\infty h(x)dx,$$

exist and are equal, then the improper integral,

$$\int_a^\infty g(x)dx,$$

exists and equals the other 2.

The comparison theorem. If $0 \leq f(x) \leq g(x)$ on $[a, \infty)$, and if,

$$\int_a^\infty g(x)dx,$$

converges, then

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

converges. Contrapositively, if $\int_a^\infty f(x)dx$ diverges, then $\int_a^\infty g(x)dx$ diverges.

Lecture 30. December 6, 2005

Practice Problems. Course Reader: 6C-2.

1. Sequences By definition, a *sequence* of real numbers is a rule assigning to each counting number n an associated real number a_n. The integer n is called the *index* of the sequence. Usually the index begins with $n = 1$, but occasionally it begins with another integer (sometimes 0). Sequences are often specified by giving the first few values, and letting the reader infer the rule, e.g.,

$$a_1 = \frac{1}{1},\ a_2 = \frac{1}{2},\ a_3 = \frac{1}{3}, \dots$$

It is always better to give a precise definition of each sequence, e.g.,

$$a_n = \frac{1}{n}, n = 1, 2, \dots$$

The most common notation for a sequence is $(a_n)_{n\geq 1}$.

A sequence $(a_n)_{n\geq 1}$ *converges to a limit* L if the sequence becomes arbitrarily close to L, and stays arbitrarily close to L. More precisely, the sequence converges to L if for every positive number ϵ, there exists an integer N (depending on the sequence and ϵ) such that for every integer $n \geq N$,

$$|a_n - L| < \epsilon.$$

In other words, the *tail* of the sequence $a_N, a_{N+1}, a_{N+2}, \dots$ are all numbers in the interval $(L - \epsilon, L + \epsilon)$. A sequence cannot have more than 1 limit: given 2 potential limits L_1 and L_2, simply take $\epsilon = |L_1 - L_2|/2$ in the definition above. A sequence which has a limit is said to *converge*, and the limit is denoted by,

$$L = \lim_{n\to\infty} a_n.$$

A sequence which does not have a limit is said to *diverge*.

Examples.

(i) Let L be a fixed real number. The sequence $a_n = L, n = 1, 2, \dots$ converges to L.

(ii) The sequence $a_n = n$ diverges. In a precise sense, this sequence "diverges to ∞".

(iii) The sequence $a_n = (-1)^n$ diverges, even though it is bounded (it never gets bigger than 1 or smaller than -1).

(iv) Let r be a real number. The sequence $a_n = r^n, n = 0, 1, 2, \dots$ converges to 0 if $|r| < 1$ and diverges if $|r| > 1$. There are 2 remaining cases. If $r = -1$, then $a_n = (-1)^n$ diverges. If $r = 1$, then $a_n = 1$ converges to 1.

2. Tests for convergence/divergence. One useful test for convergence is the *Squeezing Lemma*.

The squeezing lemma. Let $(a_n)_{n\geq 1}$, $(b_n)_{n\geq 1}$ and $(c_n)_{n\geq 1}$ be sequences such that for every index n,

$$a_n \leq b_n \leq c_n.$$

In other words, the sequence (b_n) is "squeezed" between the sequences (a_n) and (c_n). If (a_n) and (c_n) converge, and if,

$$\lim_{n\to\infty} a_n = \lim_{n\to\infty} c_n,$$

then also (b_n) converges and its limit equals the limit of the other 2 sequences.

Another test for convergence is the *Monotone Convergence Test*. A sequence $(a_n)_{n\geq 1}$ is called *non-decreasing* if for every index n, $a_{n+1} \geq a_n$. Similarly, a sequence (a_n) is *non-increasing* if for every index n, $a_{n+2} \leq a_n$. A sequence which is either non-decreasing or non-increasing (but not both increasing and decreasing) is called *monotone*. A sequence (a_n) is *bounded above* if there exists a real number u such that for every index n, $a_n \leq u$. The number u is an *upper bound* for the sequence. A sequence (a_n) is *bounded below* if there exists a real number l such that for every index n, $a_n \geq l$. The number l is a *lower bound* for the sequence.

Monotone Convergence Test. A non-decreasing sequence converges if and only if it is bounded above. In this case, the limit of the sequence is the least upper bound for the sequence. Similarly, a non-increasing sequence converges if and only if it is bounded below and the limit is the greatest lower bound for the sequence.

3. Series. Given a sequence $(a_n)_{n\geq 1}$, there are 2 important related sequences. The first is the *sequence of partial sums*, $(b_n)_{n\geq 1}$, defined by,

$$b_n = a_1 + a_2 + \cdots + a_n = \sum_{k=1}^{n} a_k.$$

The second is the *sequence of partial absolute sums*, $(B_n)_{n\geq 1}$, defined by,

$$B_n = |a_1| + |a_2| + \cdots + |a_n| = \sum_{k=1}^{n} |a_k|.$$

If the sequence of partial sums $(b_n)_{n\geq 1}$ converges, the limit is called the *series of* $(a_n)_{n\geq 1}$, and is denoted by,

$$\sum_{k=1}^{\infty} a_k := \lim_{n\to\infty} b_n = \lim_{n\to\infty} \sum_{k=1}^{n} a_k.$$

In this case is is said *the series* $\sum_k a_k$ *converges*. If the sequence of partial absolute sums $(B_n)_{n\geq 1}$ converges, it is said *the series* $\sum_k a_k$ *converges absolutely*. Although it is not obvious, if the series converges absolutely, then the series converges (this is a basic theorem from course 18.100). If a series converges but does not converge absolutely, sometimes it is said the series *converges conditionally*.

Examples. The *harmonic sequence* is the sequence $a_n = 1/n$. As will be shown soon, the *harmonic series* $\sum_n 1/n$ diverges to ∞. The *alternating harmonic sequence* is,

$$a_n = \frac{(-1)^n}{n}.$$

The *alternating harmonic series*,

$$\sum_{n=1}^{\infty} \frac{(-1)^n}{n},$$

does converge. This will also be shown soon. Since the sequence of partial absolute sums for the alternating sequence equals the sequence of partial sums for the harmonic sequence, the alternating harmonic series does not converge absolutely. It only converges conditionally.

As counter-intuitive as this might sound, the terms in the alternating harmonic series can be rearranged so that the sum converges to any real number you like! This sounds ridiculous: finite sums are independent of the order in which the summands are added, so how could this fail for infinite sums? The answer is quite simple. Because the harmonic series $\sum_n 1/n$ diverges, the same is true for $\sum_{1/2n}$. Thus, add it up a very large number of *only* the (positive) even terms in the alternating harmonic series to make the partial sum bigger than, say, 10^6. Now add only the first odd term $-1/2$. This has a negligible effect. Now add a large number of the remaining even terms to make the partial sum bigger than 10^7. Now add one more odd term, $-1/3$. Continuing in this way, eventually every term in the sequence contributes to one of the partial sums. But because we add positive terms with a much higher frequency than negative terms, the sequence of partial sums is diverging to $+\infty$. Similarly, we could negative terms with a very high frequency and make the partial sums diverge to $-\infty$. Now it is not so surprising that by adding the terms in a careful order, we can make the partial sums converge to any value we like.

The pathology of the preceding paragraph occurs with any conditionally convergent series. It is a very important fact that every absolutely convergent series has only a single limit, independent of the order in which terms are added. For this reason, absolutely convergent series are much more useful than conditionally convergent series.

4. Test for convergence/divergence of series. If a series $\sum_n a_n$ converges, then the sequence (a_n) converges to 0. To see this, denote by L the limit of the sequence of partial sums (b_n). For every positive real number ϵ, using $\epsilon/2$ in the definition of convergence of (b_n), there exists an integer N such that for every $n \geq N$, $|b_n - L| < \epsilon/2$. But then for $n \geq N+1$,

$$|a_n| = |b_n - b_{n-1}| = |(b_n - L) - (b_{n-1} - L)| \leq |b_n - L| + |b_{n-1} - L| < \epsilon/2 + \epsilon/2 = \epsilon.$$

Thus the sequence (a_n) converges to 0. Contrapositively, if the sequence (a_n) does not converge to 0, then the series $\sum_n a_n$ diverges. This is the most basic test for divergence of a series. For example, it immediately follows that the series $\sum_{n=1}^{\infty}(-1)^n$ diverges (arguing the opposite is a favorite pasttime of "mathematical cranks").

The most basic test for absolute convergence of a sequence follows from the monotone convergence test. The sequence of partial absolute sums,

$$B_n = \sum_{k=1}^{n} |a_k|,$$

is a non-decreasing sequence. Therefore, by the monotone convergence theorem, it converges if and only if it is bounded above. The most common technique for proving the sequence of partial absolute sums is bounded above is by comparing it to a larger series that is known to converge. This gives the following.

Comparison Test. Let $(a_n)_{n\geq 1}$ and $(b_n)_{n\geq 1}$ be sequences such that for every index n, $|a_n| \leq |b_n|$. If the series $\sum_{n=1}^{\infty} b_n$ converges absolutely, then the series $\sum_{n=1}^{\infty} a_n$ converges absolutely.

A number of common convergence tests in calculus textbooks come to nothing more than combining the comparison test with an analysis of the geometric series. Let r be a real number and let $(a_n)_{n\geq 0}$ be the geometric sequence,

$$a_n = r^n, n \geq 0,$$

(by convention, if $r = 0$, the first term a_0 is defined to be 1). By high school algebra, if $r \neq 1$, the partial sums are

$$b_n = 1 + r + \cdots + r^n = \frac{1 - r^{n+1}}{1 - r} = \frac{1}{1-r} - \frac{1}{1-r} r^{n+1}.$$

Observe this sequence depends on n only in the last term r^{n+1}, which is essentially the geometric sequence. Assuming $r \neq 1$, the geometric sequence r^{n+1} converges if and only if $|r| < 1$. In this case, the sequence of partial absolute sums,

$$B_n = 1 + |r| + |r|^2 + \cdots + |r|^n = \frac{1}{1-|r|} + \frac{1}{1-|r|} |r|^{n+1},$$

also converges. Thus, the geometric series $\sum_{n=0}^{\infty} r^n$ converges absolutely to $1/(1-r)$ if $|r| < 1$, and diverges if $|r| > 1$ or $r = -1$. The only remaining case is when $r = 1$. Then the partial sums are $b_n = n + 1$, which diverges to ∞. Altogether, the series $\sum_{n=0}^{\infty} r^n$ converges to $1/(1-r)$ if $|r| < 1$, and diverges otherwise.

The ratio test. There are two tests that allow us to compare a given sequence $(a_n)_{n \geq \infty}$ to a geometric sequence $(r^n)_{n \geq 1}$. If the following limit,

$$\lim_{n \to \infty} \left| \frac{a_{n+1}}{a_n} \right|,$$

exists, call it r. Then the sequence $(a_n)_{n \geq 1}$ can be compared to a sequence $(Cr^n)_{n \geq 1}$ for some choice of C. This leads to the *ratio test*: The series $\sum_{n=1}^{\infty} a_n$ converges absolutely if the sequence $|a_{n+1}/a_n|$ converges to a real number $r < 1$ and diverges if the sequence $|a_{n+1}/a_n|$ converges to a real number $r > 1$ (in which case, the sequence $(a_n)_{n \geq 1}$ does not converge to 0). There is no information if the sequence converges to 1 or diverges.

Similarly, if the following limit,

$$\lim_{n \to \infty} \sqrt[n]{|a_n|},$$

exists, call it r. Then the sequence $(a_n)_{n \geq 1}$ can be compared to a sequence $(Cr^n)_{n \geq 1}$. This leads to the *root test*: The series $\sum_{n=1}^{\infty} a_n$ converges absolutely if the sequence $\sqrt[n]{|a_n|}$ converges to a real number $r < 1$ and diverges if the sequence $\sqrt[n]{|a_n|}$ converges to a real number $r > 1$. There is no information if the sequence converges to 1 or diverges.

Comparison to an improper integral. The final test uses improper integrals to get useful information about a series. Let $(a_n)_{n \geq 1}$ be a sequence. Let $f(x) \geq 0$ be a function on $[1, \infty)$ such that for every integer n, $f(x) \geq a_n$ for all $n \leq x \leq n + 1$. If the improper integral,

$$\int_1^{\infty} f(x) dx,$$

converges, then the series $\sum_{n=1}^{\infty} a_n$ converges absolutely. On the other hand, let $g(x) \geq 0$ be a function on $[1, \infty)$ such that for every integer n, $g(x) \leq a_n$ for all $n \leq x \leq n + 1$. If the improper integral,

$$\int_1^{\infty} g(x) dx,$$

diverges, then the series $\sum_{n=1}^{\infty} a_n$ does not converge absolutely. For both directions, define the sequence (c_n) by,

$$c_n = \int_n^{n+1} f(x)dx, \text{ or } c_n = \int_n^{n+1} g(x)dx.$$

The absolute partial sum of the series $\sum_{k=1}^{n} c_k$ is simply,

$$\sum_{k=1}^{n} c_k = \int_1^n f(x)dx, \text{ or } \int_1^n g(x)dx.$$

The result follows.

Examples. 1. The harmonic series. Let $(a_n)_{n\geq 1}$ be the harmonic sequence,

$$a_n = \frac{1}{n}.$$

Let $g(x)$ be the function $g(x) = 1/x$ on the interval $[1, \infty)$. Then for every integer n, $g(x) \leq a_n = 1/n$ on the interval $[n, n+1]$. By the Fundamental Theorem of Calculus, the partial sums of the sequence (c_n) are,

$$\sum_{k=1}^{n} c_k = \int_1^n \frac{1}{x}dx = \ln(n).$$

As n tends to ∞, the natural logarithms $\ln(n)$ also tend to ∞ (although very slowly – $\ln(n)$ does not get bigger than a fixed real number R until n gets bigger than the much larger number e^R). Therefore the partial sums diverge. By the comparison test, the harmonic series also diverges (very slowly).

Example. 2. The Riemann zeta function. Let $s > 1$ be a real number. Define the sequence $(a_n)_{n\geq 1}$ by,

$$a_n = \frac{1}{n^s}.$$

The series $\sum_{n=1}^{\infty} 1/n^s$ equals $1 + \sum_{n=2}^{\infty} 1/n^s$, which is the same as,

$$1 + \sum_{n=1}^{\infty} \frac{1}{(n+1)^s}.$$

Let $f(x)$ be the function $f(x) = 1/x^s$. Then for each integer n, $f(x) \geq 1/(n+1)^s$ for every x in $[n, n+1]$. The partial sum of (c_n) is,

$$c_n = \int_1^n \frac{1}{x^s}dx = \left(\frac{1}{1-s}\frac{1}{x^{s-1}}\right|_1^n = \frac{1}{s-1} - \frac{1}{s-1}\frac{1}{n^{s-1}}.$$

Because s is bigger than 1, as n tends to ∞, also n^{s-1} tends to ∞. Therefore the partial sums tend to $1/(s-1)$. Therefore, by the comparison test, the series,

$$\sum_{n=1}^{\infty} \frac{1}{n^s},$$

converges absolutely to a value bounded by $1/(s-1)$. The value of this limit is called the *Riemann zeta function* at s, denoted

$$\zeta(s) := \sum_{n=1}^{\infty} \frac{1}{n^s}.$$

This function is of fundamental importance in number theory. It is also pops up in Fourier series and statistical mechanics. The values of $\zeta(s)$ when s is an even integer are known. The first couple are $\zeta(2) = \pi^2/6$ and $\zeta(4) = \pi^4/90$. There are very fundamental open problems about the Riemann zeta function. For one of these problems in particular, the Clay Mathematics Institute has offered a $1 million prize for an accepted, refereed solution.

Lecture 31. December 8, 2005

Practice Problems. Course Reader: 7B-4, 7B-6, 7C-1, 7C-5, 7D-1, 7D-2.

1. Power series. Given a real number a and a sequence of real numbers $(c_n)_{n\geq 0}$, there is an associated expression, called a *power series about* $x = a$,

$$\sum_{n=0}^{\infty} c_n(x-a)^n = c_0 + c_1(x-a) + c_2(x-a)^2 + \dots$$

For every choice of a real number x, the power series gives a usual series. In particular, for the choice $x = a$, the series has only 1 nonzero term, thus converges to c_0.

Question. Given a power series, for which real numbers x does the corresponding series absolutely converge?

Examples. 1. Consider the power series,

$$0 + 1^1x^1 + 2^2x^2 + 3^3x^3 + \cdots = \sum_{n=1}^{\infty} n^n x^n.$$

Of course this converges to 0 for $x = 0$. But for any x other than 0, the sequence $n^n x^n = (nx)^n$ diverges. Therefore the series does not converge. In other words, the series converges only for $x = 0$.

2. Consider the power series,

$$1 + x + x^2 + \cdots = \sum_{n=1}^{\infty} x^n.$$

This is a geometric series. From the last lecture, the series converges absolutely for $|x| < 1$ and diverges if $|x| \geq 1$.

3. Consider the power series,

$$1 + x + x^2/2 + x^3/3! + \cdots = \sum_{n=0}^{\infty} \frac{1}{n!}x^n.$$

The ratio of the n^{th} and $(n+1)^{\text{st}}$ terms in the series is,

$$(x^{n+1}/(n+1)!)/(x^n/n!) = \frac{x}{n+1}.$$

For fixed x, as n grows, this sequence of ratios converges to 0, which is less than 1. Therefore, by the ratio test, for every choice of x the series converges.

These 3 examples illustrate the whole range of possibilities.

Theorem. Let $\sum_{n=0}^{\infty} c_n(x-a)^n$ be a power series about $x = a$. Exactly one of the following hold.

(i) For every x different from a, the series does not converge absolutely.

(ii) There exists a real number R such that the series converges absolutely if $|x-a| < R$ and does not converge absolutely if $|x-a| > R$.

(iii) For every real number x, the series converges absolutely.

The real number R occuring in Case (ii) is called the *radius of convergence*. By convention, in Case (i) the radius of convergence is defined to be $R = 0$. By convention, in Case (iii) the radius of convergence is defined to be $R = \infty$. This allows us to replace the original question by a more precise question.

Question. Given a power series, what is the radius of convergence?

Although there is no single answer to this question, in many interesting cases the ratio or root test gives an answer.

2. Analytic functions. If the radius of convergence R of a power series $\sum c_n(x-a)^n$ is positive, then the power series defines a function on the interval $(a-R, a+R)$,

$$f(x) = \sum_{n=0}^{\infty} c_n(x-a)^n.$$

A function defined in this manner is called an *analytic function*. This is the real significance of power series: they give important examples of functions that cannot be described in a more direct manner. Analytic functions have nice analytic properties (whence the name). For instance, it is a theorem (proved in 18.100) that an analytic function $f(x)$ is differentiable and the derivative has a power series converging absolutely with the same radius R,

$$f'(x) = \sum_{n=0}^{\infty} c_n n(x-a)^{n-1} = \sum_{m=0}^{\infty} (m+1)c_{m+1}(x-a)^m.$$

We can iterate the theorem, i.e., $f'(x)$ is differentiable and $f''(x)$ has a power series converging absolutely with radius R. Iterating k times, the function $f(x)$ is k-times differentiable and its k^{th} derivative has a power series,

$$f^{(k)}(x) = \sum_{n=0}^{\infty} \frac{(n+k)!}{n!} c_{n+k}(x-a)^n.$$

In particular, every derivative of $f(x)$ is defined. A function with this property is called *infinitely differentiable* or *smooth*. Thus, every analytic function is infinitely differentiable.

This is only 1 of many useful properties of analytic functions. Which functions $f(x)$ are analytic functions? By the last paragraph, if $f(x)$ is analytic, then it is infinitely differentiable. Are there other restrictions? Can more than 1 power series about $x = a$ give rise to the same analytic function?

To answer both of these questions, consider the analytic function defined by a power series,

$$f(x) = \sum_{n=0}^{\infty} c_n (x-a)^n.$$

Plugging in $x = a$ gives the equation,

$$f(a) = c_0 + c_1(a-a) + c_2(a-a)^2 + \cdots = c_0 + 0 + 0 + \cdots = c_0.$$

Thus the first coefficient of the power series is simply,

$$c_0 = f(a).$$

Moreover, from the power series for the k^{th} derivative,

$$f^{(k)}(a) = k!c_k + (k+1)!/1!c_{k+1}(a-a) + (k+2)!/2!c_{k+2}(a-a)^2 + \cdots = k!c_k + 0 + 0 + \cdots = k!c_k.$$

Solving for c_k, the k^{th} coefficient of the power series is,

$$c_k = \boxed{f^{(k)}(a)/k!.}$$

Therefore, the power series defining $f(x)$ is,

$$f(x) = \sum_{n=0}^{\infty} \frac{f^{(n)}(a)}{n!}(x-a)^n.$$

In particular, this series is unique. This answers the second question. Two absolutely convergent power series about $x = a$ give the same analytic function if and only if the power series are themselves equal (i.e., the corresponding coefficients of the 2 series are equal).

Moreover, this gives us alot of information about the first question. For an infinitely differentiable function $f(x)$ defined at a point $x = a$, there is a very important power series, the *Taylor series expansion of* $f(x)$ *about* $x = a$,

$$\boxed{\sum_{n=0}^{\infty} \frac{f^{(n)}(a)}{n!}(x-a)^n.}$$

If $f(x)$ is analytic, then the Taylor series converges absolutely to $f(x)$. This reduces the original question to 2 new questions. Does the Taylor series have a positive radius of convergence? If so, does the analytic function defined in this way equal the original function $f(x)$?

The radius of convergence question is precisely the radius of convergence question posed earlier. As there, the answer can often be found by using the ratio or root tests. The second question is yes in every practical case. There are examples of infinitely differentiable functions where the Taylor series has a positive radius of convergence, but does not converge to the original function. However, every example is somewhat contrived; they rarely come up "in nature". Just for completeness, here is an example of one of these pathological functions,

$$f(x) = \begin{cases} e^{-1/x^2}, & x \neq 0, \\ 0 & x = 0 \end{cases}$$

3. Algorithm for computing Taylor series. The method for finding the Taylor series of a function is always the same. For definiteness, consider the Taylor series expansion of $f(x) = (1-x)^{-1}$ about the point $x = 0$.

Step 1. Compute all derivatives of $f(x)$. If this sounds like alot of work, it is! In most examples, this really comes down to finding an inductive formula for the derivatives of $f(x)$. In the example, the "zeroth derivative" is,

$$f(x) = \boxed{(1-x)^{-1}.}$$

The first derivative is,

$$f'(x) = -(1-x)^{-2}(-1) = \boxed{(1-x)^{-2}.}$$

The second derivative is,

$$f''(x) = (-2)(1-x)^{-3}(-1) = \boxed{(1-x)^{-3}.}$$

This begins to suggest a pattern: The k^{th} derivative of $f(x)$ will be,

$$f^{(k)}(x) = b_k(1-x)^{-k-1},$$

for some real number b_k. Having made this guess, it is easy to verify by induction. By computation, the result is true for $k = 0, 1$ and 2 with the corresponding real numbers $b_0 = 1$, $b_1 = 1$ and $b_2 = 2$. By way of induction, assume the result is true for $k = n$, i.e.,

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

Then the $(n+1)^{\text{st}}$ derivative is,

$$f^{(n+1)}(x) = (f^{(n)}(x))' = (c_n(1-x)^{-n-1})' = c_n(-n-1)(1-x)^{-n-2}(-1) = (n+1)c_n(1-x)^{-n-2}.$$

Thus the result is also true for $k = n+1$ where c_{n+1} satisfies the equation,

$$c_{n+1} = (n+1)c_n.$$

Thus the result is proved by induction on k.

In fact, more has been accomplished, since now there is an inductive formula for the numbers c_n,

$$c_n = nc_{n-1} = n(n-1)c_{n-2} = n(n-1)(n-2)c_{n-3} = \cdots = n(n-1)(n-2)\cdot\dots\cdot 3c_2 = n(n-1)(n-2)\cdot\dots\cdot 3\cdot 2\cdot 1.$$

This number has come up before in this class. It is the n^{th} factorial number,

$$c_n = n!.$$

This gives the final formula for the n^{th} derivative of $f(x)$,

$$f^{(n)}(x) = \boxed{n!(1-x)^{-n-1}.}$$

Step 2. Substitute $x = a$ into the derivatives. Compared to the work of finding the derivatives, this is very simple. In the example, plugging in $x = 0$ gives,

$$f^{(n)}(0) = \boxed{n!.}$$

Step 3. Compute the coefficients of the Taylor series. By definition, the n^{th} coefficient of the Taylor series is,

$$c_n = \frac{f^{(n)}(a)}{n!}.$$

In the example, this gives the coefficient,

$$c_n = \frac{n!}{n!} = \boxed{1,}$$

for every integer $n \geq 0$.

Step 4. Write the Taylor series. This is really getting into the "mind-numbing details". In the example, the Taylor series expansion for $(1-x)^{-1}$ about $x = 0$ is,

$$(1-x)^{-1} = \boxed{\sum_{n=0}^{\infty} x^n.}$$

Step 5. If possible, find the radius of convergence. In the example, the Taylor series is simply the geometric series. By the previous lecture, the geometric series converges absolutely with radius $R = 1$. Moreover, it converges absolutely to $(1-x)^{-1}$. Notice, this gives another explanation for the radius $R = 1$. Since $(1-x)^{-1}$ has a vertical asymptote at $x = 1$, the Taylor series cannot converge on any interval that contains $x = 1$. The largest interval centered at $x = 0$ not containing $x = 1$ is the interval $(-1, 1)$. This interval has radius $R = 1$.

4. More examples. What is the Taylor series expansion for $(1-x)^{-1}$ about a point $x = a$ different from $x = 1$? The fortunate fact is that Step 1 allows to compute the derivatives $f^{(n)}(a)$

for any $a \neq 1$, not just $x = 0$. This is the typical case, and it is one justification for doing the work necessary in Step 1. In this case, the answer is,

$$f^{(n)}(a) = \boxed{n!(1-a)^{-n-1}.}$$

Therefore, according the Step 3, the n^{th} coefficient in the Taylor series expansion is,

$$c_n = \frac{n!(1-a)^{-n-1}}{n!} = \boxed{(1-a)^{-n-1}.}$$

Thus, according to Step 4, the Taylor series expansion for $(1-x)^{-1}$ about $x = a$ is,

$$(1-x)^{-1} = \boxed{\sum_{n=0}^{\infty}(1-a)^{-n-1}(x-a)^n.}$$

What is the radius of convergence? The ratio of the $(n+1)^{\text{st}}$ and n^{th} terms of the series is,

$$[(1-a)^{-n-2}(x-a)^{n+1}]/[(1-a)^{-n-1}(x-a)^n] = (1-a)^{-1}(x-a).$$

This is independent of n. Thus, this constant sequence converges to its constant value $(1-a)^{-1}(x-a)$. By the ratio test, the sequence is absolutely convergent if and only if this limit has absolute value less than 1,

$$|(1-a)^{-1}(x-a)| \leq 1.$$

Rearranging, the series converges if and only if,

$$|x-a| \leq |1-a|.$$

Thus the radius of convergence is,

$$R = \boxed{|1-a|.}$$

This is perfectly reasonable. The function $(1-x)^{-1}$ has a vertical asymptote at $x = 1$. Therefore, the power series cannot converge on any interval containing $x = 1$. The largest interval centered at $x = a$ not containing $x = 1$ has radius equal to the distance from $x = a$ to $x = 1$, namely $R = |1-a|$.

Example 2. For the next example, consider the Taylor series expansion for $f(x) = e^x$ near $x = 0$. In this case, Step 1 is simple. Every derivative of $f(x)$ is simply,

$$f^{(n)}(x) = \boxed{e^x.}$$

Therefore, the n^{th} coefficient of the Taylor series expansion is,

$$c_n = \frac{f^{(n)}(0)}{n!} = \frac{e^0}{n!} = \boxed{1/n!.}$$

Therefore the Taylor series expansion is,

$$\sum_{n=0}^{\infty} \frac{1}{n!} x^n.$$

Observe this is the power series considered earlier in the lecture, whose radius of convergence is $R = \infty$. Therefore, for every x, the power series converges absolutely to e^x,

$$e^x = \boxed{\textstyle\sum_{n=0}^{\infty} x^n/n!.}$$

This equation is sometimes taken as the definition of e^x. It has certain advantages to our original definition of e^x. Importantly, it is easy for a computer to determine e^x to very high precision using this formula.

Example 3. Having computed the Taylor series expansion for e^x about $x = 0$, the next question is to compute the Taylor series expansion for e^x about $x = a$. According to the formula,

$$f^{(n)}(a) = e^a,$$

and thus the coefficient is,

$$c + n = \boxed{e^a/n!.}$$

This gives the Taylor series expansion for e^x about $x = a$,

$$\sum_{n=0}^{\infty} \frac{e^a}{n!}(x-a)^n.$$

As above, the radius of convergence is $R = \infty$. Thus, for every real number x, the power series converges absolutely to e^x,

$$e^x = \boxed{\textstyle\sum_{n=0}^{\infty} e^a(x-a)^n/n!.}$$

On the other hand, we didn't need to do any extra work to see this. We could have used the formula,

$$e^x = e^{a+(x-a)} = e^a e^{x-a}.$$

Plugging in $x - a$ for x in the power series expansion for e^x gives the power series expansion,

$$e^{x-a} = \sum_{n=0}^{\infty} \frac{1}{n!}(x-a)^n.$$

This gives the same Taylor series expansion as above,

$$e^x = e^a \sum_{n=0}^{\infty} \frac{1}{n!}(x-a)^n = \boxed{\textstyle\sum_{n=0}^{\infty} (e^a/n!)(x-a)^n.}$$

Example 3. Consider the function $f(x) = \sin(x)$. The derivatives of $f(x)$ are,

$$\begin{array}{rcl} f(x) & = & \sin(x), \\ f'(x) & = & \cos(x), \\ f''(x) & = & -\sin(x), \\ f^{(3)}(x) & = & -\cos(x), \\ f^{(n+4)}(x) & = & f^{(n)}(x) \end{array}$$

Together, these give all the derivatives of $f(x)$. Write $n = 4l$, $4l + 1$, $4l + 2$ or $4l + 3$ for some nonnegative integer l. Then the rules above give,

$$f^{(n)}(x) = \begin{cases} \sin(x) & n = 4l, \\ \cos(x) & n = 4l + 1, \\ -\sin(x) & n = 4l + 2, \\ -\cos(x) & n = 4l + 3 \end{cases}$$

In particular, plugging in $x = 0$ gives,

$$f^{(n)}(0) = \begin{cases} 0, & n = 4l, \\ 1, & n = 4l + 1, \\ 0, & n = 4l + 2, \\ -1, & n = 4l + 3 \end{cases}$$

Thus, all the even coefficients of the Taylor series are 0. For an odd coefficient, say $n = 2m + 1$, the derivative is,

$$f^{(2m+1)}(0) = (-1)^m.$$

Therefore, the coefficient is,

$$c_{2m+1} = \frac{(-1)^m}{(2m + 1)!}.$$

Plugging this in gives the Taylor series expansion for $\sin(x)$ about $x = 0$,

$$\sum_{m=0}^{\infty} \frac{(-1)^m}{(2m + 1)!} x^{2m+1}.$$

The ratio of consecutive terms in the series is,

$$[(-1)^{m+1} x^{2m+3}/(2m + 3)!]/[(-1)^m x^{2m+1}/(2m + 1)!] = -x^2/(4m^2 + 8m + 3).$$

This sequence converges to 0. Therefore, by the ratio test, the power series converges absolutely to $\sin(x)$ for every choice of x,

$$\sin(x) = \boxed{\sum_{m=0}^{\infty} (-1)^m/(2m + 1)! x^{2m+1}.}$$

There is an exactly similar formula for $g(x) = \cos(x)$,

$$g^{(n)}(x) = \begin{cases} \cos(x), & n = 4l, \\ -\sin(x), & n = 4l+1, \\ -\cos(x), & n = 4l+2, \\ \sin(x), & n = 4l+3. \end{cases}$$

This gives the values,

$$g^{(n)}(0) = \begin{cases} 1, & n = 4l, \\ 0, & n = 4l+1, \\ -1, & n = 4l+2, \\ 0, & n = 4l+3. \end{cases}$$

Therefore the Taylor series is,

$$\cos(x) = \boxed{\sum_{m=0}^{\infty}(-1)^m/(2m)!x^{2m}.}$$

Notice, we didn't really need to do this work. Since $\cos(x)$ is the derivative of $\sin(x)$, the Taylor series for $\cos(x)$ is simply the term-by-term derivative of the Taylor series for $\sin(x)$. This gives the same formula as above.

To compute the Taylor series expansions of $\sin(x)$ and $\cos(x)$ about a point $x = a$, we can follow the procedure above. However, it is much faster to use the angle addition formulas,

$$\sin(x) = \sin(a + (x-a)) = \cos(a)\sin(x-a) + \sin(a)\cos(x-a),$$

$$\cos(x) = \cos(a + (x-a)) = \cos(a)\cos(x-a) - \sin(a)\sin(x-a).$$

This gives the Taylor series expansions,

$$\sin(x) = \boxed{\sum_{m=0}^{\infty} \frac{(-1)^m \cos(a)}{(2m+1)!}(x-a)^{2m+1} + \sum_{m=0}^{\infty} \frac{(-1)^m \sin(a)}{(2m)!}(x-a)^{2m},}$$

$$\cos(x) = \boxed{\sum_{m=0}^{\infty} \frac{(-1)^m \cos(a)}{(2m)!}(x-a)^{2m} + \sum_{m=0}^{\infty} \frac{(-1)^{m+1} \sin(a)}{(2m+1)!}(x-a)^{2m+1}.}$$

Lecture 32. December 9, 2005

Practice Problems. Course Reader: RI.

1. Using power series to solve calculus problems. The reason power series are useful is because they allow us to describe functions that have no direct description. For instance, consider the function,

$$f(x) = \int_0^x e^{-t^2} dt,$$

for $x \geq 0$. By the Fundamental Theorem of Calculus, this function exists and is differentiable with derivative $f'(x) = e^{-x^2}$. Unfortunately, there is no simple expression for $f(x)$ involving only polynomials, trigonometric functions, exponential functions and logarithms (the proof of this is far beyond the scope of this class). However, it is quite easy to write down a power series expansion for $f(x)$. First of all, the Taylor series for e^{-t^2} about $t = 0$ is obtained by substituting $x = -t^2$ in the Taylor series for e^x about $x = 0$,

$$e^{-t^2} = \sum_{n=0}^{\infty}(-1)^n t^{2n}/n!.$$

Because this series converges absolutely, the integral of the series is the series of the term-by-term integrals,

$$f(x) = \int_0^x e^{-t^2}dt = \int_0^x \sum_{n=0}^{\infty} \frac{(-1)^n}{n!}t^{2n}dt = \sum_{n=0}^{\infty} \int_0^x \frac{(-1)^n}{n!}t^{2n}dt.$$

Each of these integrals can be computed quite easily. This gives,

$$f(x) = \boxed{\textstyle\sum_{n=0}^{\infty}(-1)^n/[(2n+1)\cdot n!]t^{2n+1}.}$$

This is the Taylor series expansion for $f(x)$ about $x = 0$. For instance, using this series, it is easy to estimate,

$$\int_0^1 e^{-t^2}dt \approx 0.747 \pm 10^{-3}.$$

2. Taylor series with remainder term. As demonstrated by the computation just done, in reality only finitely many terms in a Taylor series are used. What can be said in this case? In other words, how quickly does the series converge? How large is the remainder after n terms? To make all this precise, introduce the function $R_{N,a}(x)$ defined to be,

$$R_{N,a}(x) = f(x) - \sum_{n=0}^{N} \frac{f^{(n)}(a)}{n!}(x-a)^n.$$

This is precisely the remainder term so that we have,

$$f(x) = \sum_{n=0}^{N} \frac{f^{(n)}(a)}{n!}(x-a)^n + R_{N,a}(x).$$

The precise version of the questions above is, what bounds exist for $R_{N,a}(x)$?

To understand the answer, consider the simplest case where $N = 0$. Then the remainder term is simply,

$$R_{0,a}(x) = f(x) - f(a).$$

By the Mean Value Theorem, for every x there exists a real number c (depending on x) between a and x such that,

$$R_{0,a}(x) = f'(c)(x-a).$$

Iterating the Mean Value Theorem, for every integer N, for every x, there exists a real number c (depending on both N and x) between a and x such that,

$$R_{N,a}(x) = \boxed{f^{(N+1)}(c)(x-a)^{N+1}/(N+1)!.}$$

In particular, if we can bound the $(N+1)^{\text{st}}$ derivative of $f(x)$ on the interval between a and c, then we can bound $R_{N,a}(x)$.

Example. Bound the remainder in the Taylor series expansion for e^x about $x = a$. The $(N+1)^{\text{st}}$ derivative is simply e^x. Therefore, a bound for $f^{(N+1)}(c)$ for c between a and x is simply,

$$M = e^m = e^{\max(a,x)}.$$

This is independent of N. The bound on the remainder term is then,

$$|R_{N,a}(x)| \leq \boxed{M(x-a)^{N+1}/(N+1)!.}$$

By choosing N suitably large, we can make this remainder term as small as possible. For instance, if we want to compute e^x for x in the interval $(-1, 1)$, then M equals e. To make the remainder term less than 10^{-10}, it suffices to take $N = 12$.

3. Review problems. Each of the following problems was discussed in lecture. Here are the problems and answers, without the discussion.

Problem 1. Let a and b be positive real numbers. There are 2 tangent lines to the ellipse with equation,

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1,$$

containing the point (a, b). Find the equations of each of these tangent lines.

The 2 tangent lines are the line tangent to the ellipse at $(x, y) = (0, b)$ and the line tangent to the ellipse at $(x, y) = (a, 0)$. The equations of these lines are,

$$\boxed{y = b,}$$

and,

$$\boxed{x = a.}$$

Problem 2. A grain silo is designed by attaching a cylinder of fixed radius r and height a directly above a right circular cone of base radius r and height b. The silo has no top, and there is no bottom between the bottom of the cylinder and the top of the cone. For a fixed volume V, what choice of b minimizes the surface area of the grain silo?

The choice of b minimizing the surface area is,

$$b = \boxed{2\sqrt{5}r/5.}$$

Problem 3. Compute the volume of the solid obtained by revolving about the x-axis the region in the first quadrant bounded by the curve $y = x^2$ and the curve $x = y^2$.

The volume of this solid is,

$$\text{Volume} = \boxed{3\pi/10.}$$

Problem 4. Using a trigonometric substitution and a trigonometric identity, compute the antiderivative,

$$\int \frac{\sqrt{1-x^2}}{x^2}dx.$$

The antiderivative equals,

$$\int \frac{\sqrt{1-x^2}}{x^2}dx = \boxed{-\sqrt{1-x^2}/x - \sin^{-1}(x) + C.}$$

Problem 5. Using integration by parts, compute the following antiderivative,

$$\int x\sin^{-1}(x)dx.$$

The antiderivative equals,

$$\int x\sin^{-1}(x)dx = \boxed{-(1/4)(1-x^2)\sin^{-1}(x) + (1/4)x\sqrt{1-x^2} + C.}$$