CALCULUS APPENDICES

These appendices summarize some important concepts and results that are omitted from the main body of the text. It is hard to understand mathematical concepts without examples, which will be liberally supplied in the lectures. It is the *combination* of examples and theory that will help you to understand the material, if you study them together.

Other topics, especially applications of the calculus to biology, will be covered in readings from other sources, particularly during the second semester of this course.

1. Historical notes

In October of 1858 the German mathematician Richard Dedekind was teaching calculus in Zurich, when he realized that he didn't really understand what he was talking about - and neither did anyone else. What are numbers? What should they be? How do properties of the real numbers, functions, and the calculus interact to allow us to draw conclusions and solve problems?

By November, Dedekind had resolved the problem, effectively describing the real number system for the first time. Other nineteenth century mathematicians, including Cauchy, Peano, Bolzano and Weierstrass, worked on other aspects of putting the calculus on a firm theoretical foundation, especially with regard to the theory of limits. The details are part of more advanced math courses, but we want to outline the basic ideas here.

2. Fields and real numbers

Fields. A *field* is a number system that satisfies the axioms of ordinary algebra.

$$\begin{aligned} x+(y+z) &= (x+y)+z & \qquad x(yz) &= (xy)z \\ x+y &= y+x & xy &= yx \\ x+0 &= x & x1 &= x \\ x+(-x) &= 0 & x\left(\frac{1}{x}\right) &= 1 \text{ for } x \neq 0 \\ x(y+z) &= xy+xz & & \end{aligned}$$

Other properties that are consequences of these axioms also hold, e.g., $0 \cdot x = 0$ and $(x+y)^2 = x^2 + 2xy + y^2$.

Three important examples of fields are the rational numbers, denoted by $\mathbb{Q}$, the real numbers $\mathbb{R}$, and the complex numbers $\mathbb{C}$.

Date: May 8, 2007.

Approximation. Every real number can be approximated arbitrarily closely by rational numbers. We express this by saying that the rational numbers are *dense* in the reals. Of course, the finite decimal numbers are also dense in the reals, which is good for calculation, but the rationals are better for theoretical purposes. (Why?)

Order. The order on real numbers satisfies

- $x^2 \geq 0$ for all $x \in \mathbb{R}$.
- $x \leq y$ implies $x + z \leq y + z$.
- $x \leq y$ implies $xz \leq yz$ for $z \geq 0$.
- $x \leq y$ implies $xz \geq yz$ for $z \leq 0$.

Completeness. There are gaps in the rational numbers. For example, $\sqrt{2}$ is not a rational number, so there is sort of a hole in $\mathbb{Q}$ between the positive rationals with $r^2 < 2$ and those with $r^2 > 2$. The Least Upper Bound property is a technical way of saying that there are no holes in the real line.

Theorem 1. *(Least Upper Bound Property) If $x_1 \leq x_2 \leq x_3 \leq \dots$ is an increasing sequence of real numbers, and there is an upper bound B such that $x_n \leq B$ for every n, then there is a least real number L such that $x_n \leq L$ for every n.*

The least upper bound L will then be the limit of the sequence of x_n's.

3. Limits, continuity and differentiability

These topics will be discussed at great length in class, with many examples. This appendix is designed to provide a quick summary by putting the main ideas in one place for reference.

The definitions are as follows.

- $\lim_{x\to a} f(x) = L$ if for every $\epsilon > 0$ there exists $\delta > 0$ such that $|f(x) - L| < \epsilon$ whenever $0 < |x - a| < \delta$.
- The function $f(x)$ is *continuous* at $x = a$ if $\lim_{x\to a} f(x) = f(a)$.
- $f(x)$ is continuous (in general) if it is continuous at each point in its domain.
- $f(x)$ is *differentiable* at $x = a$ if $\lim\limits_{x\to a} \dfrac{f(x) - f(a)}{x - a}$ exists.
- $f(x)$ is differentiable on an interval I if it is differentiable at every point in I.

The next theorem connects two of these ideas.

Theorem 2. *If $f(x)$ is differentiable at a point $x = a$, then it is continuous at $x = a$.*

The converse if false: there are plenty of functions that are continuous but not differentiable.

4. The "value" theorems

We spend a lot of time in calculus looking for maxima and minima. There is no point in "snipe hunting," looking for something that does not exist. The following series of theorems gives conditions under which maxima and minima are guaranteed to exist, tells us how to find them, and gives some of the consequences of this fact.

Theorem 3. *(Bounded Value Theorem) If $f(x)$ is a continuous function on a closed interval $[a, b]$, then there is a number B such that $|f(x)| \leq B$ for all x with $a \leq x \leq b$.*

Theorem 4. *(Intermediate Value Theorem) If $f(x)$ is a continuous function on a closed interval $[a, b]$, and K is any number between $f(a)$ and $f(b)$, then there is at least one number $x = c$ in the interval such that $f(c) = K$.*

Theorem 5. *(Extreme Value Theorem) If $f(x)$ is a continuous function on a closed interval $[a, b]$, then $f(x)$ achieves its maximum and minimum values: there are points c and d in the interval such that $f(c) \leq f(x) \leq f(d)$ for all x in the interval.*

Theorem 6. *(Fermat's Theorem) If $f(c)$ is either the maximum or minimum value of a function on an interval $[a, b]$, then one of the following holds:*

(i) $c = a$,
(ii) $c = b$,
(iii) $f'(c)$ *is undefined, or*
(iv) $f'(c) = 0$.

Theorem 7. *(Rolle's Theorem) If $f(x)$ is a differentiable function on a closed interval $[a, b]$ and $f(a) = f(b)$, then there is a point c with $a < c < b$ such that $f'(c) = 0$.*

Theorem 8. *(Mean Value Theorem) If $f(x)$ is a differentiable function on a closed interval $[a, b]$, then there is a point c with $a < c < b$ such that*

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

In fact, for Rolle's Theorem and the MVT, we only need that $f'(x)$ exist on the open interval (a, b) and that the function be continuous at the endpoints.

5. The Fundamental Theorems

Besides the Fundamental Theorem of Calculus, there are two other basic Fundamental Theorems that you need to know, plus a host of others for more advanced topics (the Fundamental Theorem of Galois Theory, etc.). These two formalize crucial facts about elementary algebra. They are due to Euclid and Gauss, respectively.

Theorem 9. *(Fundamental Theorem of Arithmetic) Every positive integer factors uniquely into prime numbers.*

Theorem 10. *(Fundamental Theorem of Algebra) Every polynomial with real coefficients factors uniquely into a constant, linear terms* $x - a$*, and quadratic terms* $x^2 + bx + c$ *with* $b^2 - 4c < 0$.

6. Integration

A function $f(x)$ is *integrable* on the interval $[a, b]$ if the (signed) area between the lines $x = a$, $x = b$, $y = 0$ and the curve $y = f(x)$ makes sense. There are several ways to make this precise, but the following one, called the *Darboux integral*, is perhaps the most elementary version.

Given a partition P of the interval $[a, b]$, let $L(f, P)$ denote the lower estimate for the (signed) area under the curve, using the minimum y-values in each subinterval of P, and let $U(f, P)$ denote the corresponding upper estimate that uses the maximum values in each subinterval. The following lemma tells us how this works.

Lemma 11. *Every lower estimate is below every upper estimate. In fact, if* $P \subseteq Q$*, then*

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

We say that $f(x)$ is *integrable* on $[a, b]$ if the upper and lower sums can be made arbitrarily close by choosing the right partition, i.e., for every $\epsilon > 0$ there is a partition P such that $U(f, P) \leq L(f, P) + \epsilon$. In that case, we let $\int_a^b f(x)\,dx$ be the unique number between all the upper sums and all the lower sums, i.e., the unique number I such that $L(f, P) \leq I \leq U(f, P)$ for all choices of the partition P. Otherwise, the integral is undefined.

The book limits the discussion to continuous functions, and indeed we have the following fundamental result.

Theorem 12. *If* $f(x)$ *is continuous on the interval* $[a, b]$*, then it is integrable on* $[a, b]$.

However, there are lots of functions that are not continuous, but are still integrable.

7. Applications of differential equations

1. Exponential growth ($k > 0$) or decay ($k < 0$) satisfies the equation

$$y' = ky$$
$$y(0) = y_0$$

where y represents the population or amount in question (as a function of time), k is the growth or decay rate, and y_0 is the initial population/amount. It has the solution

$$y = y_0 e^{kt}.$$

The *half-life* or *doubling time* is $(\ln 2)/|k|$.

2. Newton's law of cooling satisfies the equation

$$T' = -r(T - A)$$
$$T(0) = T_0$$

where T is the temperature (as a function of time), A is the ambient temperature, and r is the cooling rate. It has the solution

$$T = A + Ce^{-rt}$$

where $T_0 = A + C$.

3. The logistic equation is

$$y' = ay - by^2$$
$$y(0) = y_0$$

where y represents the population, a is the growth rate and b is a correction factor (with generally $b \ll a$). It has the solution

$$y = \frac{a/b}{1 + Ce^{-at}}$$

where

$$y_0 = \frac{a/b}{1 + C}$$

The non-zero equilibrium *carrying capacity* is a/b.

4. The Levins metapopulation model satisfies

$$p' = cp(1 - p) - mp$$
$$p(0) = p_0$$

where p is the proportion of available sites occupied by the species, c is the colonization rate and m is the mortality rate. This transforms into the logistic equation

$$p' = (c - m)p - cp^2$$

with a stable equilibrium at $p^* = 1 - m/c$.

8. Summary of series

(1) The Taylor series expansion of a function $f(x)$ is the series

$$\sum_0^\infty c_k x^k = c_0 + c_1 x + c_2 x^2 + c_3 x^3 + \cdots$$

where $c_k = \frac{f^{(k)}(0)}{k!}$. We encode this in a table to find the series expansion of a function. The *Taylor polynomials* for $f(x)$ are given by the finite sums

$$p_n(x) = c_0 + c_1 x + \cdots + c_n x^n.$$

Under relatively weak hypotheses, the Taylor polynomials will converge to the function, at least on some interval. (See (11) below.)

(2) $\dfrac{1}{1-x} = \sum_{k=0}^\infty x^k = 1 + x + x^2 + x^3 + \cdots$ for $|x| < 1$.

(3) $\dfrac{1}{1+x} = \sum_{k=0}^\infty (-1)^k x^k = 1 - x + x^2 - x^3 + \cdots$ for $|x| < 1$.

(4) $(1+x)^r = \sum_{k=0}^\infty \binom{r}{k} x^k = 1 + rx + \frac{r(r-1)}{2!}x^2 + \frac{r(r-1)(r-2)}{3!}x^3 + \cdots$ for $|x| < 1$.

(5) $e^x = \sum_{k=0}^\infty \frac{x^k}{k!} = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots$.

(6) $\cos x = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \cdots$.

(7) $\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \cdots$.

(8) $\tan x = x + \frac{x^3}{3} + \frac{2x^5}{15} - \cdots$ for $|x| < \frac{\pi}{2}$.

(9) $\sec x = 1 + \frac{x^2}{2} + \frac{5x^4}{24} - \cdots$ for $|x| < \frac{\pi}{2}$.

(10) The Lagrange form of the remainder is given by

$$f(x) = c_0 + c_1 x + \cdots + c_n x^n + \frac{f^{(n+1)}(\theta)}{(n+1)!}x^{n+1}$$

for some number θ between 0 and x.

(11) This basic result gives a nice sufficient condition for the Taylor series to represent the function on an interval.

Theorem 13. *(Taylor's Theorem) If there is a number M such that*

$$|f^{(k)}(x)| \leq M^k$$

for all $k \geq 0$ and all x with $|x| < R$, then the series converges to the function for $|x| < R$.

9. Proof of the Lagrange remainder formula

Most calculus texts follow a version of Lagrange's original proof of the remainder formula for Taylor polynomials. This proof is ingenious, but not very intuitive, and not easily understood by students. On the other hand, the proof of the remainder formula for interpolatory polynomials uses only Rolle's theorem, and is easy to understand (see any numerical analysis text). A slight variation of the latter proof gives an easy proof of the remainder formula for Taylor polynomials.

Lemma 14. *Let $h(t)$ be a function which is n times differentiable on $[0, x]$. If $h(x) = 0$ and $h^{(k)}(0) = 0$ for $0 \leq k < n$, then there exists $\eta \in (0, x)$ such that $h^{(n)}(\eta) = 0$.*

Proof. Since $h(0) = 0 = h(x)$, by Rolle's theorem there exists $c \in (0, x)$ such that $h'(c) = 0$. Applying induction to $h'(t)$ on the interval $[0, c]$ yields the conclusion of the lemma. □

For $x < 0$, replace $[0, x]$ and $(0, x)$ by $[x, 0]$ and $(x, 0)$, respectively.

Let $p(x)$ be the $(n-1)$-st Taylor polynomial for $f(x)$. Fixing $x \neq 0$, apply the lemma to

$$h(t) = f(t) - p(t) - \left[\frac{f(x) - p(x)}{x^n}\right] t^n$$

to obtain the Lagrange remainder formula

$$f(x) - p(x) = \frac{f^{(n)}(\eta)}{n!} x^n .$$

(I found this proof while teaching numerical analysis. It is the proof given in Apostol's *Calculus*, but does not seem to be well known.)

10. Vectors

(1) The *vector* $\overline{\mathbf{ab}}$ in $\mathbb{R}^n$ goes from point $\mathbf{a}$ to point $\mathbf{b}$. Two vectors $\overline{\mathbf{ab}}$ and $\overline{\mathbf{cd}}$ are regarded as equal (or *equivalent*) if $\mathbf{b} - \mathbf{a} = \mathbf{d} - \mathbf{c}$. In that case we use $\mathbf{b} - \mathbf{a}$ to represent the vector. That is the head of the vector if the tail is at the origin. This is interpreted as saying that a vector has *magnitude* and *direction*, but not position.

(2) The *dot product* is defined by $\mathbf{x} \cdot \mathbf{y} = \sum x_i y_i$. The *length* of a vector is given by $\|\mathbf{x}\| = \sqrt{\mathbf{x} \cdot \mathbf{x}}$.

(3) The basic properties of dot products are as follows.
   (a) $\mathbf{x} \cdot \mathbf{x} > 0$ for $\mathbf{x} \neq 0$, and $\mathbf{0} \cdot \mathbf{0} = 0$.

(b) $\mathbf{x} \cdot \mathbf{y} = \mathbf{y} \cdot \mathbf{x}$
(c) $(c\mathbf{x}) \cdot \mathbf{y} = c(\mathbf{x} \cdot \mathbf{y})$.
(d) $\mathbf{x} \cdot (\mathbf{y} + \mathbf{z}) = \mathbf{x} \cdot \mathbf{y} + \mathbf{x} \cdot \mathbf{z}$

Any vector operation satisfying these properties is called an *inner product.*

(4) From the law of cosines, we see that $\mathbf{x} \cdot \mathbf{y} = \|\mathbf{x}\| \|\mathbf{y}\| \cos\theta$ where θ is the angle between the vectors.

(5) In particular, $\mathbf{x}$ is perpendicular to $\mathbf{y}$ if and only if $\mathbf{x} \cdot \mathbf{y} = 0$.

11. Matrices

(1) The rules for matrix algebra are mostly the same as the rules for the algebra of numbers. Here capitals A, B, C, etc. are matrices and lower case a, b, etc. are scalars (numbers). O denotes the zero matrix and I denotes the identity matrix.
(a) $1A = A$
(b) $c(dA) = (cd)A$
(c) $c(A + B) = cA + cB$
(d) $(c + d)A = cA + dA$
(e) $A + (B + C) = (A + B) + C$
(f) $A + B = B + A$
(g) $A + O = A$
(h) $A - A = O$
(i) $A(BC) = (AB)C$
(j) $AI = A = IA$
(k) $AA^{-1} = I = A^{-1}A$ if A has an inverse.
(l) $A(B + C) = AB + AC$ and $(A + B)C = AC + BC$

In general, however, $AB \neq BA$ and you can only cancel *invertible* matrices.

(2) The matrix A is *invertible* if and only if $\det(A) \neq 0$. If A and B are invertible, then $(AB)^{-1} = B^{-1}A^{-1}$.

(3) The *transpose* operation A^* flips a matrix about its diagonal. A matrix is symmetric if $A^* = A$. Note that $(cA)^* = cA^*$, $(A + B)^* = A^* + B^*$ and $(AB)^* = B^*A^*$.

(4) The number λ is an *eigenvalue* of A if there is a nonzero vector $\mathbf{x}$ such that $A\mathbf{x} = \lambda\mathbf{x}$. You find the eigenvalues of A by solving $\det(A - \lambda I) = 0$.

(5) A *quadratic* form is a function that can be written as $q(\mathbf{x}) = \mathbf{x}^*M\mathbf{x}$ where M is a symmetric matrix. (A few examples will make this notion comprehensible.)

(6) The quadratic form $q(\mathbf{x}) = \mathbf{x}^* M \mathbf{x}$ is *positive definite* if $q(\mathbf{x}) > 0$ for all $\mathbf{x} \neq 0$. This happens if and only if all the eigenvalues of M are positive.

(7) The quadratic form $q(\mathbf{x}) = \mathbf{x}^* M \mathbf{x}$ is *negative definite* if $q(\mathbf{x}) < 0$ for all $\mathbf{x} \neq 0$. This happens if and only if all the eigenvalues of M are negative.

(8) A quadratic form may be positive definite, negative definite, or neither.

12. Summary of differential calculus in several variables

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a function of several variables. Let $\bigtriangledown f(\mathbf{x})$ denote its gradient vector and $H(\mathbf{x})$ its Hessian matrix.

(What follows is written in very compact form. You should write out the full expressions for better understanding.)

(1) The function $f(\mathbf{x})$ is *differentiable* at the point $\mathbf{a}$ if

$$f(\mathbf{x}) = f(\mathbf{a}) + \bigtriangledown f(\mathbf{a}) \cdot (\mathbf{x} - \mathbf{a}) + r(\mathbf{x})$$

where $\lim_{\mathbf{x}\to\mathbf{a}} r(\mathbf{x})/\|\mathbf{x} - \mathbf{a}\| = 0$.

(2) The *tangent hyperplane* to $u = f(\mathbf{x})$ at the point $\mathbf{a}$ is given by

$$u = f(\mathbf{a}) + \bigtriangledown f(\mathbf{a}) \cdot (\mathbf{x} - \mathbf{a}).$$

(3) Differentiability implies continuity.

(4) The function

$$f(x, y) = \begin{cases} \frac{2xy}{x^2+y^2} & \text{if } (x, y) \neq (0, 0) \\ 0 & \text{if } (x, y) = (0, 0) \end{cases}$$

has its gradient defined at the origin, in fact $\bigtriangledown f(0, 0) = (0, 0)$, but the function is not continuous there, so certainly it is not differentiable.

(5) For a differentiable function we sometimes use the formula in terms of *differentials*:

$$df = \bigtriangledown f \cdot d\mathbf{x}$$

In particular, if $\mathbf{x}$ is a function of t, then the total derivative is given by the chain rule:

$$\frac{df}{dt} = \bigtriangledown f \cdot \frac{d\mathbf{x}}{dt}$$

(6) The quadratic approximation of a twice differentiable function is given by

$$f(\mathbf{x}) \doteq f(\mathbf{a}) + \triangledown f(\mathbf{a}) \cdot (\mathbf{x} - \mathbf{a}) + \frac{1}{2}(\mathbf{x} - \mathbf{a})^* H(\mathbf{a})(\mathbf{x} - \mathbf{a}).$$

(7) If a differentiable function has a local extremum at $\mathbf{x} = \mathbf{a}$, then $\triangledown f(\mathbf{a}) = \mathbf{0}$. If $H(\mathbf{a})$ is positive definite, then $f(\mathbf{a})$ is a local minimum. If $H(\mathbf{a})$ is negative definite, then $f(\mathbf{a})$ is a local maximum. If $H(\mathbf{a})$ is neither and zero is not an eigenvalue of $H(\mathbf{a})$, then $f(\mathbf{a})$ is a saddle point.

(8) Let $\mathbf{u}$ be a unit vector, i.e., a vector with $\|\mathbf{u}\| = 1$. Each unit vector specifies a direction. The *directional derivative* of $f(\mathbf{x})$ in the direction $\mathbf{u}$ at the point $\mathbf{a}$ is given by

$$D_{\mathbf{u}}(f)(\mathbf{a}) = \triangledown f(\mathbf{a}) \cdot \mathbf{u}.$$

This represents the slope of $f(\mathbf{x})$ as viewed from the point $\mathbf{a}$ in the $\mathbf{u}$ direction.

(9) A vector function of time, $\mathbf{x}(t)$, can be thought of as representing the position of a particle. Its derivative $\frac{d\mathbf{x}}{dt}$ then represents the velocity vector of the particle. The *chain rule* says that for a function $f(\mathbf{x})$ we have

$$\frac{df}{dt} = \triangledown f(\mathbf{x}) \cdot \frac{d\mathbf{x}}{dt}.$$

13. Least squares linear fit

Suppose you are given a collection of data points

$$(x_1, d_1),\ (x_2, d_2),\ \ldots\ (x_n, d_n)$$

and you want to find the line $y = mx + b$ that best approximates your data. If we measure the difference between the line and the data by the sum of squares

$$s(b, m) = \sum_{i=1}^{n} (mx_i + b - d_i)^2$$

then this is a minimization problem in the variables b and m. Taking the gradient $\langle \frac{\partial s}{\partial b}, \frac{\partial s}{\partial m} \rangle$ and setting it equal to the zero vector, then dividing by 2, we obtain the following equations.

$$nb + (\sum x_i)m = \sum d_i$$
$$(\sum x_i)b + (\sum x_i^2)m = \sum x_i d_i$$

That is 2 linear equations in 2 unknowns, easily solved for b and m to give the *least squares line*.

This method can be modified to fit other curves to data. What kind of curve you use to model your data depends on the shape of the data. An important variation occurs when you replace the original data by (x_i, d_i) by

$(\ln x_i, \ln d_i)$. If a straight line $\hat{y} = \hat{b} + m\hat{x}$ fits that data, then with $\hat{y} = \ln y$, $\hat{b} = \ln b$ and $\hat{x} = \ln x$ we have

$$\ln y = \ln b + m \ln x$$

which is equivalent to $y = bx^m$. While this transformation may seem awkward, in fact it comes up in many real applications; see problems 10–16 on page 354 of the textbook (Greenwell et. al.).