MATH 105: MULTIVARIABLE CALCULUS
SPRING 2011 REVIEW SHEET

STEVEN J. MILLER (SJM1@WILLIAMS.EDU)

ABSTRACT. Below is a summary of definitions and some key lemmas, theorems and concepts from multivariable calculus. I have removed .eps images as several people have had difficulty downloading and viewing the file when the images are included; if you want these let me know.

Note: you are responsible for making sure all items below are correct; if you find any mistakes please let me know for extra credit.

These notes were begun for the Spring 2010 version of the class. That year the textbook used parentheses and not angular brackets for vectors, and so in the text below vectors are displayed with parentheses.

CONTENTS

Date: March 27, 2011.

1. Part 1: Vectors, Curves and Surfaces in Space

1.1. Definitions.

Definition 1.1 (Equation of a line)**.** *The line going through the point* $\overrightarrow{P}$ *in the direction* $\overrightarrow{v}$ *is the set of all points* $(x_1, \dots, x_n)$ *such that*

$$(x_1, \dots, x_n) \;=\; \overrightarrow{P} + t\overrightarrow{v}.$$

In three dimensions, we have

$$(x, y, z) \;=\; \overrightarrow{P} + t\overrightarrow{v}.$$

If $\overrightarrow{v} = (v_1, v_2, v_3)$ *and* $\overrightarrow{P} = (P_1, P_2, P_3)$*, this is equivalent to the system of equations*

$$\begin{aligned} x &= P_1 + tv_1 \\ y &= P_2 + tv_2 \\ z &= P_3 + tv_3. \end{aligned}$$

If we have two points on the line but not the direction, we may find the direction by subtracting one point from the other.

Definition 1.2 (Equation of a plane)**.** *The plane going through the point* $\overrightarrow{P}$ *with directions* $\overrightarrow{v}$ *and* $\overrightarrow{w}$ *is all points* (x, y, z) *satisfying*

$$(x, y, z) \;=\; \overrightarrow{P} + t\overrightarrow{v} + s\overrightarrow{w}.$$

If instead we are given a normal direction $\overrightarrow{n}$*, then the plane going through* $\overrightarrow{P}$ *with normal in the direction* $\overrightarrow{n}$ *is the set of all points* (x, y, z) *such that*

$$\left((x, y, z) - \overrightarrow{P}\right) \cdot \overrightarrow{n} \;=\; 0.$$

Remark 1.3. *A common mistake is that if we are given three points* $\overrightarrow{P}, \overrightarrow{Q}, \overrightarrow{R}$ *and asked for the plane containing them to write* $\overrightarrow{P} + t\overrightarrow{Q} + s\overrightarrow{R}$*; the reason this is wrong is that* $\overrightarrow{Q}$ *and* $\overrightarrow{R}$ *are* not *the two directions. To find the directions, we choose one of the three points, say* $\overrightarrow{P}$*, as the base point, and then look at each of the other two minus that for the two directions, or say* $\overrightarrow{v} = \overrightarrow{Q} - \overrightarrow{P}$*,* $\overrightarrow{w} = \overrightarrow{R} - \overrightarrow{P}$*.*

Definition 1.4 (Determinants)**.** *The determinant of two vectors represents the signed area of the parallelogram generated by the two vectors (for three vectors it is the signed volume). If*

$$A \;=\; \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

then

$$\det(A) \;=\; ad - bc.$$

We often write $|A|$ for the determinant of A. If

$$B = \begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix}$$

then

$$\det(B) = |B| = aei + bfg + cdh - gec - hfa - idb.$$

One can remember the definition of the determinant in the 3×3 case by copying the first two columns of the matrix and looking at the three diagonals from upper left to lower right and the three diagonals from the lower left to the upper right. The first three are all added while the last three are all subtracted.

Definition 1.5 (Dot Product)**.** *If $\overrightarrow{v} = (v_1, \ldots, v_n)$ and $\overrightarrow{w} = (w_1, \ldots, w_n)$ then the dot (or inner) product is defined by*

$$\overrightarrow{v} \cdot \overrightarrow{w} = v_1 w_1 + \cdots + v_n w_n = \sum_{i=1}^{n} v_i w_i.$$

For example, if $\overrightarrow{v} = (1, 2, 3)$ and $\overrightarrow{w} = (3, 2, 1)$ then

$$\overrightarrow{v} \cdot \overrightarrow{w} = 1 \cdot 3 + 2 \cdot 2 + 3 \cdot 1 = 10.$$

Remark 1.6. *A common mistake with the dot product is to forget the result is a scalar (i.e., a number) and have the result a vector. Remember to add the sum of the product of the components; do not form a new vector whose i^{th} component is the product of the two i^{th} components.*

Definition 1.7 (Cross product)**.** *If $\overrightarrow{v} = (v_1, v_2, v_3)$ and $\overrightarrow{w} = (w_1, w_2, w_3)$ then the cross product is defined by*

$$\overrightarrow{v} \times \overrightarrow{w} = (v_2 w_3 - v_3 w_2, v_3 w_1 - v_1 w_3, v_1 w_2 - v_2 w_1).$$

One can remember this by abusing notation and computing

$$\begin{vmatrix} \overrightarrow{i} & \overrightarrow{j} & \overrightarrow{k} \\ v_1 & v_2 & v_3 \\ w_1 & w_2 & w_3 \end{vmatrix}.$$

Definition 1.8 (Polar coordinates)**.** *We have*

$$x = r\cos\theta, \quad y = r\sin\theta,$$

with $\theta \in [0, 2\pi)$ and $r \geq 0$. We may invert these relations, and find

$$r = \sqrt{x^2 + y^2}, \quad \theta = \arctan(y/x).$$

Definition 1.9 (Cylindrical coordinates)**.** *We have*

$$x = r\cos\theta, \quad y = r\sin\theta, \quad z = z$$

with $\theta \in [0, 2\pi)$ and $r \geq 0$. We may invert these relations, and find

$$r = \sqrt{x^2 + y^2}, \quad \theta = \arctan(y/x), \quad z = z.$$

Definition 1.10 (Spherical coordinates)**.** *We have*

$$x \;=\; \rho \sin\phi\cos\theta, \quad y \;=\; \rho\sin\phi\sin\theta, \quad z \;=\; \rho\cos\phi,$$

where $0 \le \phi \le \pi$ *and* $0 \le \theta < 2\pi$*. We may invert these relations, and find*

$$\rho \;=\; x^2 + y^2 + z^2, \quad \phi \;=\; \arccos(z/\rho), \quad \theta \;=\; \arctan(y/x).$$

1.2. **Theorems.**

Theorem 1.11 (Pythagorean Theorem)**.** *If we have a right triangle with sides* a *and* b *and hypotenuse* c*, then*

$$c^2 \;=\; a^2 + b^2.$$

Theorem 1.12 (Law of Cosines)**.** *Consider a triangle with sides* a, b, c *and angle* θ *opposite of the side of length* c*. Then*

$$c^2 \;=\; a^2 + b^2 - 2ab\cos\theta.$$

Key steps in proof: draw good auxiliary lines to reduce to right triangles, and use the Pythagorean Theorem.

Theorem 1.13 (Length of a vector)**.** *If* $\overrightarrow{v} = (v_1, \dots, v_n)$ *then*

$$||\overrightarrow{v}|| \;=\; \sqrt{v_1^2 + \cdots + v_n^2}.$$

The key to the proof is repeated applications of the Pythagorean Theorem.

If we want to normalize a vector, that means constructing a new vector of the same direction but of unit length. If $\overrightarrow{v}$ is not the zero vector, then

$$\overrightarrow{u} \;=\; \frac{\overrightarrow{v}}{||\overrightarrow{v}||}$$

is a unit vector in the direction of $\overrightarrow{v}$.

Theorem 1.14 (Angle formula)**.** *If* θ *denotes the angle between vectors* $\overrightarrow{v}$ *and* $\overrightarrow{w}$*, then*

$$\overrightarrow{v} \cdot \overrightarrow{w} \;=\; ||v||\;||w||\;\cos\theta.$$

Key ideas in proof: write the sides of the triangle as vectors in terms of the coordinates, and then apply the Law of Cosines to these lengths.

Theorem 1.15 (Cross product interpretation)**.** *The vector* $\overrightarrow{v} \times \overrightarrow{w}$ *is a vector perpendicular to* $\overrightarrow{v}$ *and* $\overrightarrow{w}$ *such that its length is the signed area of the parallelogram generated by* $\overrightarrow{v}$ *and* $\overrightarrow{w}$*.*

Theorem 1.16 (Cauchy-Schwarz Inequality)**.** *For any two vectors* $\overrightarrow{v}$ *and* $\overrightarrow{w}$ *we have*

$$|\overrightarrow{v} \cdot \overrightarrow{w}| \;\le\; ||\overrightarrow{v}||\;||\overrightarrow{w}||.$$

Did not do the Cauchy-Schwarz inequality in class. Will not be on any exam, but with knowing.

The cross product and the dot product have many nice relations, such as

$$\begin{aligned}\overrightarrow{P} \cdot (\overrightarrow{Q} + \overrightarrow{R}) &= \overrightarrow{P} \cdot \overrightarrow{Q} + \overrightarrow{P} \cdot \overrightarrow{R} \\ \overrightarrow{P} \times (\overrightarrow{Q} + \overrightarrow{R}) &= \overrightarrow{P} \times \overrightarrow{Q} + \overrightarrow{P} \times \overrightarrow{R}.\end{aligned}$$

While the dot product is commutative, $\overrightarrow{P} \cdot \overrightarrow{Q} = \overrightarrow{Q} \cdot \overrightarrow{P}$, the cross product is not: $\overrightarrow{P} \times \overrightarrow{Q} = -\overrightarrow{Q} \times \overrightarrow{P}$.

FIGURE 1. Plot of $\sin(x+y)$ and then the level sets of $\sin(x+y)$.

FIGURE 2. Plot of $\sin(xy)$ and then the level sets of $\sin(xy)$.

2. PART 2: DERIVATIVES AND PARTIAL DERIVATIVES

2.1. Definitions. ***Just because a quantity does not have an arrow over it should not be construed as implying it cannot be a vector. Many of the concepts have the same definition for scalars and vectors, and for brevity we typically give just one.***

Definition 2.1 (Function terminology)**.** *The* domain *is the set of inputs for the function, while the* range *is the set of possible outputs. When we write $f : \mathbb{R}^n \to \mathbb{R}^m$ we mean the function takes n inputs and gives m outputs. We typically denote this*

$$f(x_1, \dots, x_n) \ = \ (f_1(x_1, \dots, x_n), \dots, f_m(x_1, \dots, x_n)).$$

An example of an $f : \mathbb{R}^3 \to \mathbb{R}^2$ is

$$f(x, y, z) \ = \ (xy\cos(y^2 z) + e^z, 3 + 4x + 5y^2 + 6z^3).$$

Remark 2.2. *When determining the domain of a function, the two most common danger points are dividing by zero and taking a square-root of a negative number (both not allowed!).*

Definition 2.3 (Level sets)**.** *The level set of value c of a function is the set of all inputs where the function takes on the value c. Specifically, if $f : \mathbb{R}^2 \to \mathbb{R}$ then the level set of value c is*

$$\{(x, y) : f(x, y) \ = \ c\}.$$

For example, see the plots of $\sin(x+y)$ in Figure 1 and $\sin(xy)$ in Figure 2. We also show their level sets (which is frequently called a contour plot).

Definition 2.4 (Limit of a sequence)**.** *We say a sequence $\{a_n\}_{n=0}^{\infty}$ has L as a limit if as n tends to infinity we have a_n tends to L. We denote this as $\lim_{n\to\infty} a_n = L$.*

For example, consider the sequence $\{a_n\}_{n=1}^{\infty}$ where $a_n = (-1)^n/n$; thus our sequence is $\{-1, 1/2, -1/3, 1/4, \dots\}$ and its limit exists, which is 0. The sequence $\{b_n\}_{n=1}^{\infty}$ given by $b_n = (-1)^n$ has no limit, as its terms oscillate between -1 and 1.

Definition 2.5 (Limit of a function)**.** *A function $f(x)$ has L as a limit at x_0 if however x approaches x_0 we have $f(x)$ approaches L. We denote this by $\lim_{x\to x_0} f(x) = L$.* NOTE: we never have a term in any of our sequence equal to x_0; the goal is to understand what happens as x approaches x_0.

For the above definition, what we are essentially saying is that given *any* sequence x_n which approaches x_0 we have $f(x_n)$ approaching $f(x_0)$. For example, consider the

FIGURE 3. Plot of $x \sin(1/x)$.

FIGURE 4. Plot of $|y| < x$ union the origin.

function

$$f(x) = \begin{cases} \sin(1/x) & \text{if } x \neq 0 \\ 0 & \text{if } x = 0. \end{cases}$$

This function does not have a limit at the point $x_0 = 0$. To see this, consider the sequence $x_n = \frac{1}{2\pi n}$ and $\widetilde{x}_n = \frac{1}{(2\pi + \frac{1}{2})n}$. Note $f(x_n) = 0$ for every term in this sequence, but $f(\widetilde{x}_n) = 1$ for every term in this sequence. Thus there are two sequences with two different limits, and thus the function does not have a limit at 0.

Definition 2.6 (Continuity of a function)**.** *A function $f(x)$ is continuous at x_0 if the limit exists as $x \to x_0$ and that limit is $f(x_0)$. This means $\lim_{x \to x_0} f(x) = f(x_0)$.*

For example, consider

$$g(x) = \begin{cases} x \sin(1/x) & \text{if } x \neq 0 \\ 0 & \text{if } x = 0. \end{cases}$$

This function is continuous at 0; see Figure 3.

Definition 2.7 (Ball or Disk (not really covered in 2011, included for completeness / future reference))**.** *The ball or disk of radius r about a point $\overrightarrow{x_0}$ is the set of all points that are less than r units from $\overrightarrow{x_0}$. We assume $r > 0$ as otherwise the ball is empty. We denote this set by*

$$D_r(\overrightarrow{x_0}) = \{\overrightarrow{x} \text{ such that } ||\overrightarrow{x} - \overrightarrow{x_0}|| < r\}.$$

Definition 2.8 (Open Set (not really covered in 2011, included for completeness / future reference))**.** *A set U is open if for any $\overrightarrow{x_0} \in U$ we can always find an r (which may depend on the point $\overrightarrow{x_0}$ such that $D_r(\overrightarrow{x_0}) \subset U$. This means that, no matter what point we take in U, we can find a very small ball (or disk) centered at that point and entirely contained in U.*

The following sets are open (not really covered in 2011, included for completeness / future reference): $\{(x, y) : y > 0\}$, $\{(x, y, z) : x^2 + 4y^2 + 9z^2 < 1\}$. The following sets are not open: $\{(x, y) : y \geq 0\}$ and $\{(x, y, z) : x^2 + 4y^2 + 9z^2 \leq 1\}$. For another set that is not open, consider $\{(x, y) : |y| < x\} \cup \{(0, 0)\}$; this is the set of all points between the lines $y = x$ and $y = -x$ *and* the origin. See Figure 4.

Definition 2.9 (Derivative)**.** *Let $f : \mathbb{R} \to \mathbb{R}$ be a function. We say f is differentiable at x_0, and denote this by $f'(x_0)$ or df/dx, if the following limit exists:*

$$\lim_{h \to 0} \frac{f(x_0 + h) - f(x_0)}{h}.$$

We may also write this limit by

$$\lim_{x \to x_0} \frac{f(x_0 + h) - f(x_0)}{h},$$

or as

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

Definition 2.10 (Partial derivatives)**.** *Let* $f : \mathbb{R}^n \to \mathbb{R}$ *be a function of* n *variables* $x_1, \dots, x_n$. *We say the partial derivative with respect to* x_i *exists at the point* $a = (a_1, \dots, a_n)$ *if*

$$\lim_{h \to 0} \frac{f(\overrightarrow{a} + h\overrightarrow{e}_i) - f(\overrightarrow{a})}{h}$$

exists, where

$$\overrightarrow{a} + h\overrightarrow{e}_i \ = \ (a_1, \dots, a_{i-1}, a_i + h, a_{i+1}, \dots, a_n);$$

here $\overrightarrow{e}_i$ *is the unit vector in the direction of the* i^{th} *coordinate axis, which means all of its entries are 0 save the* i^{th}*, which is 1.*

For example, if $f(x, y, z) = 3x^2y + x\cos(y)$, then

$$\frac{\partial f}{\partial x} \ = \ 6xy + \cos(y), \quad \frac{\partial f}{\partial y} \ = \ 3x^2 - x\sin(y), \quad \frac{\partial f}{\partial z} \ = \ 0.$$

Note that to take a partial derivative with respect to x, we treat all the other variables as constants. A good way to test your answer at the end is to go back to the original equation and replace all variables with constants, and then see if your answer agrees with the derivative of this (when you put in constants). For example, in our case if we set $y = 3$ and $z = 5$ we get $g(x) = f(x, 3, 5) = 9x^2 + x\cos(3)$, and $dg/dx = 18x + \cos(3)$, which is exactly $\frac{\partial f}{\partial x}(x, 3, 5)$.

Remark 2.11. *It is very important to use the write notation; we use* ∂ *and not* d *or a prime for a partial derivative.*

Definition 2.12 (Tangent plane approximation)**.** *Let* $f : \mathbb{R}^2 \to \mathbb{R}$*. The tangent plane approximation to* f *at* (x_0, y_0) *is given by*

$$z \ = \ f(x_0, y_0) + \frac{\partial f}{\partial x}(x_0, y_0)(x - x_0) + \frac{\partial f}{\partial y}(x_0, y_0)(y - y_0),$$

provided of course the two partial derivatives exist.

In one variable, we write $y = f(x)$ and we write the tangent line as $y = f(x_0) + f'(x_0)(x - x_0)$. The above is the natural generalization, with now $z = f(x, y)$.

Definition 2.13 (Differentiability: two variables (not really covered in 2011, included for completeness / future reference))**.** *Let* $f : \mathbb{R}^2 \to \mathbb{R}$*. We say* f *is differentiable at* (x_0, y_0) *if the tangent plane approximation tends to zero significantly more rapidly than* $||(x, y) - (x_0, y_0)||$ *tends to 0 as* $(x, y) \to (x_0, y_0)$*. Specifically,* f *is differentiable if*

$$\lim_{(x,y) \to (x_0, y_0)} \frac{f(x, y) - f(x_0, y_0) - \frac{\partial f}{\partial x}(x_0, y_0)(x - x_0) - \frac{\partial f}{\partial y}(x_0, y_0)(y - y_0)}{||(x, y) - (x_0, y_0)||} \ = \ 0.$$

Note the above is truly the generalization of the derivative in one variable. The distance $x - x_0$ is replaced with $||(x,y) - (x_0,y_0)||$; while this is always positive, the fact that the limit must equal zero for the function to be differentiable means we could have used $|x - x_0|$ in the denominator in the definition of the derivative of one variable. Also note that the last two parts of the tangent plane approximation can be written as a dot product of two vectors:

$$\frac{\partial f}{\partial x}(x_0,y_0)(x-x_0)+\frac{\partial f}{\partial y}(x_0,y_0)(y-y_0) \;=\; \left(\frac{\partial f}{\partial x}(x_0,y_0), \frac{\partial f}{\partial y}(x_0,y_0)\right)\cdot(x-x_0, y-y_0).$$

Definition 2.14 (Gradient)**.** *The gradient of a function $f : \mathbb{R}^n \to \mathbb{R}$ is the vector of the partial derivatives with respect to each variable. We write*

$$\mathrm{grad}(f) \;=\; \nabla f \;=\; Df \;=\; \left(\frac{\partial f}{\partial x_1}, \dots, \frac{\partial f}{\partial x_n}\right).$$

If $f(x,y,z) = 3x^2y + x\cos(y)$, then $\nabla f = (6xy + \cos(y), 3x^2 - x\sin(y), 0)$.

Definition 2.15 (Differentiability: several variables but only one output (not really covered in 2011, included for completeness / future reference))**.** *Let $f : \mathbb{R}^n \to \mathbb{R}$. We say f is differentiable at $\overrightarrow{a}$ if the tangent hyperplane approximation tends to zero significantly more rapidly than $||\overrightarrow{x} - \overrightarrow{a}||$ tends to 0 as $\overrightarrow{x} \to \overrightarrow{a}$. Specifically, f is differentiable if*

$$\lim_{\overrightarrow{x}\to\overrightarrow{a}} \frac{f(x,y) - f(\overrightarrow{a}) - (\nabla f)(\overrightarrow{a})\cdot(\overrightarrow{x} - \overrightarrow{a})}{||\overrightarrow{x} - \overrightarrow{a}||} \;=\; 0.$$

For example, if f is a function of two variables then f is differentiable at $(0,0)$ if

$$\lim_{(x,y)\to(0,0)} \frac{f(x,y) - f(0,0) - \frac{\partial f}{\partial x}(0,0)(x-0) - \frac{\partial f}{\partial y}(0,0)(y-0)}{||(x,y) - (0,0)||} \;=\; 0.$$

Definition 2.16 (Derivative notation (not really covered in 2011, included for completeness / future reference))**.** *Let $f : \mathbb{R}^n \to \mathbb{R}^m$; we may write*

$$f(\overrightarrow{x}) \;=\; (f_1(\overrightarrow{x}, \dots, f_m(\overrightarrow{x}))\,.$$

By $(Df)(\overrightarrow{x_0})$ we mean the matrix whose first row is $(\nabla f_1)(\overrightarrow{x})$, whose second row is $(\nabla f)(\overrightarrow{x})$, and so on until the last row, which is $(\nabla f_m)(\overrightarrow{x})$. In full glory, we have

$$(Df)(x_0) \;=\; \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(\overrightarrow{x}) & \cdots & \frac{\partial f_1}{\partial x_n}(\overrightarrow{x}) \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1}(\overrightarrow{x}) & \cdots & \frac{\partial f_m}{\partial x_n}(\overrightarrow{x}) \end{pmatrix}.$$

Note $(Df)(\overrightarrow{x})$ is a matrix with m rows and n columns.

Definition 2.17 (Differentiability: several variables and several outputs (not really covered in 2011, included for completeness / future reference))**.** *Let $f : \mathbb{R}^n \to \mathbb{R}^m$. We say f is differentiable at $\overrightarrow{a}$ if the tangent hyperplane approximation for each component tends to zero significantly more rapidly than $||\overrightarrow{x} - \overrightarrow{a}||$ tends to 0 as $\overrightarrow{x} \to \overrightarrow{a}$. Specifically, f is differentiable if*

$$\lim_{\overrightarrow{x}\to\overrightarrow{a}} \frac{f(\overrightarrow{x}) - f(\overrightarrow{a}) - (Df)(\overrightarrow{a})\cdot(\overrightarrow{x} - \overrightarrow{a})}{||\overrightarrow{x} - \overrightarrow{a}||} \;=\; \overrightarrow{0},$$

where we regard $\overrightarrow{x} - \overrightarrow{a}$ as a column vector being acted on by the matrix $(Df)(\overrightarrow{a})$.

Definition 2.18 (C^1 (not really covered in 2011, included for completeness / future reference)). *A function is said to be C^1 (or of class C^1) if all of its partial derivatives exist and if these partial derivatives are continuous.*

While all the partial derivatives of the function $f(x, y) = (xy)^{1/3}$ exist, this function is not C^1 as the partial derivatives are not continuous at the origin.

Definition 2.19 (Parametrization of paths / curves (not really covered in 2011, included for completeness / future reference)). *A map c from an interval to $\mathbb{R}^n$ traces out a path or curve in space. If $c : [a, b] \to \mathbb{R}^n$ then $c(a)$ is the initial point of the path and $c(b)$ is the endpoint. If $c(a) = c(b)$ then the path is closed. It is a path in the plane if $n = 2$ and a path in space if $n = 3$. When $n = 2$ we often write $c(t) = (x(t), y(t))$, and if $n = 3$ we write $c(t) = (x(t), y(t), z(t))$. The vector $c'(t)$ is the velocity vector, and the instantaneous speed at time t is given by $||c'(t)||$. The tangent line at time t_0 is given by $(x, y, z) = c(t_0) + c'(t_0)t$.*

Remark 2.20 ((Not really covered in 2011, included for completeness / future reference)). *Note that $(Dc)(t)$ is a column vector. The reason this is so is that $c : \mathbb{R}^n \to \mathbb{R}$ in general, so $c(t) = (c_1(t), \dots, c_n(t))$. The derivative matrix Dc has as its first row $Dc_1 = \nabla c_1$, ..., and its last row is $Dc_n = \nabla c_n$. As there is only one variable, $Dc_1 = \nabla c_1 = dc_1/dt$, and thus $(Dc)(t)$ is a column vector. This is needed so that the multiplication of matrices in the chain rule is well-defined.*

Definition 2.21 (Directional derivatives). *The directional derivative of f in the direction of $\overrightarrow{v}$ at $\overrightarrow{x}$ is defined by*

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

We typically take $\overrightarrow{v}$ to be a vector of unit length. One way to compute the directional derivative is $(\nabla f)(\overrightarrow{x}) \cdot \overrightarrow{v}$, which is much *better than having to take the limit.*

Definition 2.22 (Tangent plane from the gradient). *Let S be the level set of value k for the function f. The tangent plane at $\overrightarrow{x}_0 \in S$ is defined by*

$$(\nabla f)(\overrightarrow{x}_0) \cdot (\overrightarrow{x} - \overrightarrow{x}_0) \ = \ 0.$$

2.2. **Theorems.**

Theorem 2.23 (Limit Properties for sequences and functions). ***Provided all limits are finite,***

- *The limit of a constant times our sequence is that constant times our sequence: $\lim_{n\to\infty} cx_n = c\lim_{n\to\infty} x_n$.*
- *The limit of a sum is the sum of the limits.*
- *The limit of a difference is the difference of the limits.*
- *The limit of a product is the product of the limits.*
- *The limit of a quotient is the quotient of the limits, provided additionally that the limit of the denominator is non-zero.*
- *Sandwich Theorem: if $a_n \le b_n \le c_n$ and $\lim_{n\to\infty} a_n = \lim_{n\to\infty} c_n$, then $\lim_{n\to\infty} b_n$ also equals this common value.*

These limit laws imply corresponding results for continuous functions, namely the sum, difference or product of continuous functions is continuous, as well as the quotient (provided the denominator is non-zero). We also have the composition of continuous functions is continuous.

We have to avoid undefined expressions: $\infty-\infty$, $\infty\cdot 0$, $0/0$, ∞/∞; *other expressions, such as* $\infty+\infty$ *are okay.*

Remark 2.24. *To compute limits involving several variables, do not use L'Hopital's rule! That only works if you have a function of one variable. You should try simple paths first – if two give a different value, then the limit doesn't exist. Good paths to try are* $x=0$, $y=0$, $x=y$, $x=-y$ *or* $x=my$ *for some* $m\neq 0$. *Of course, just because the limit exists along all of these paths does not mean the limit exists. If the limits are the same, though, it becomes more likely that the limit does exist. One way to investigate is to switch to polar (in two variables) or spherical (in three). In polar, we set* $x=r\cos\theta$ *and* $y=r\sin\theta$. *If* $(x,y)\to(0,0)$ *then this is the same as* $r\to 0$ *and* θ *is free. One frequently tries this if we have terms such as* x^2+y^2 *in the denominator.*

It is essential that the limits are finite. For example, consider the three limits below:

$$\lim_{x\to 0}\left(x\cdot\frac{1}{x}\right),\quad \lim_{x\to 0}\left(x\cdot\frac{1}{x^2}\right),\quad \lim_{x\to 0}\left(x^3\cdot\frac{1}{x^2}\right).$$

All of these limits are of the form $0\cdot\infty$; the first is 1, the second is undefined and the third is zero. We can make $0\cdot\infty$, ∞/∞, $\infty-\infty$ and $0/0$ equal anything we want. More care is thus needed whenever one of these is encountered. As the fundamental limit of calculus involves $0/0$, the quotient rule is not applicable and we need more powerful arguments.

For example, consider $\lim_{(x,y)\to(0,0)}(x^4+y^6)/(x^2+y^2)^2$. If we take $x=0$ we have $\lim_{y\to 0}y^6/y^4=\lim_{y\to 0}y^2=0$; if we take $x=0$ then we have $\lim_{x\to 0}x^4/x=1$. As two paths give different values, the limit does not exist. If instead we had $\lim_{(x,y)\to(0,0)}(x^6+y^6)/(x^2+y^2)^2$, then both the path $x=0$ and the path $y=0$ give a limit of 0. More generally, if $y=mx$ we get

$$\lim_{x\to 0}\frac{x^6+m^6x^6}{(x^2+m^2x^2)^2} \;=\; \lim_{x\to 0}\frac{1+m^6}{(1+m^2)^2}\frac{x^6}{x^4} \;=\; \frac{1+m^6}{(1+m^2)^2}\lim_{x\to 0}x^2 \;=\; 0.$$

This suggests the limit might be zero, but is not a proof. Switching to polar coordinates gives

$$\lim_{\substack{r\to 0\\ \theta\ \text{free}}}\frac{r^6\cos^6\theta+r^6\sin^6\theta}{r^4} \;=\; \lim_{\substack{r\to 0\\ \theta\ \text{free}}} r^2(\cos^6\theta+\sin^6\theta).$$

As $|\cos t|,|\sin t|\leq 1$, the sum of the sixth powers always lies between -2 and 2 (actually, between 0 and 2). As $r\to 0$ the product tends to zero since it is sandwiched between 0 and $2r^2$, both of which tend to zero.

Theorem 2.25 (Main Theorem on Differentiation (not really covered in 2011, included for completeness / future reference))**.** *The following implications hold: (1) implies (2) implies (3), where*

(1) *The partial derivatives of* f *are continuous.*
(2) *The function* f *is differentiable.*

(3) *The partial derivatives of f exist.*

Each reverse implication may fail.

Note that (2) trivially implies (3) and (1) trivially implies (3); the meat of the argument is proving that (1) implies (2). The proof uses the Mean Value Theorem and adding zero.

Theorem 2.26 (Differentiation rules)**.** *Let $f, g : \mathbb{R}^n \to \mathbb{R}$ be differentiable functions, and let c be a constant. Then*

- *Constant rule: The derivative of $cf(\overrightarrow{x})$ is $c(Df)(\overrightarrow{x})$.*
- *Sum rule: The derivative of $f(\overrightarrow{x}) + g(\overrightarrow{x})$ is $(Df)(\overrightarrow{x}) + (Dg)(\overrightarrow{x})$.*
- *Difference rule: The derivative of $f(\overrightarrow{x}) - g(\overrightarrow{x})$ is $(Df)(\overrightarrow{x}) - (Dg)(\overrightarrow{x})$.*
- *Product rule: The derivative of $f(\overrightarrow{x})g(\overrightarrow{x})$ is $(Df)(\overrightarrow{x})g(\overrightarrow{x})+f(\overrightarrow{x})(Dg)(\overrightarrow{x})$.*
- *Quotient rule: The derivative of $f(\overrightarrow{x})/g(\overrightarrow{x})$ is*

$$\frac{(Df)(\overrightarrow{x})g(\overrightarrow{x}) - f(\overrightarrow{x})(Dg)(\overrightarrow{x})}{g(\overrightarrow{x})^2}.$$

Theorem 2.27 (Chain Rule (we only did one output in 2011; the rest is included for completeness))**.** *Let $g : \mathbb{R}^n \to \mathbb{R}^m$ and $f : \mathbb{R}^m \to \mathbb{R}^p$ be differentiable functions, and set $h = f \circ g$ (the composition). Then*

$$(Dh)(\overrightarrow{x}) \ = \ (Df)(g(\overrightarrow{x}))(Dg)(\overrightarrow{x}).$$

Important special cases are:

- *Let $c : \mathbb{R} \to \mathbb{R}^3$ and $f : \mathbb{R}^3 \to \mathbb{R}$, and set $h(t) = f(c(t))$. Then*

$$\frac{dh}{dt} \ = \ (\nabla f)(c(t)) \cdot c'(t) \ = \ \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt} + \frac{\partial f}{\partial z}\frac{dz}{dt}.$$

  Note that we could have written $\partial f/\partial x$ for df/dx.

- *Let $g(x_1, \dots, x_n) = (u_1(x_1, \dots, x_n), \dots, u_m(x_1, \dots, x_n)$ and set $h(x_1, \dots, x_n) = f(g(x_1, \dots, x_n))$, where $f : \mathbb{R}^m \to \mathbb{R}$. Then*

$$\frac{\partial h}{\partial x_i} \ = \ \frac{\partial f}{\partial u_1}\frac{\partial u_1}{\partial x_i} + \frac{\partial f}{\partial u_2}\frac{\partial u_2}{\partial x_i} + \cdots + \frac{\partial f}{\partial u_m}\frac{\partial u_m}{\partial x_i}.$$

The situation we saw the most is if $w(u,v) = f(x(u,v), y(u,v), z(u,v))$ then

$$\left.\frac{\partial w}{\partial u}\right|_{(u_0,v_0)} \ = \ \left.\frac{\partial f}{\partial x}\right|_{(x_0,y_0,z_0)}\left.\frac{\partial x}{\partial u}\right|_{(u_0,v_0)} + \left.\frac{\partial f}{\partial y}\right|_{(x_0,y_0,z_0)}\left.\frac{\partial y}{\partial u}\right|_{(u_0,v_0)} + \left.\frac{\partial f}{\partial z}\right|_{(x_0,y_0,z_0)}\left.\frac{\partial z}{\partial u}\right|_{(u_0,v_0)},$$

where

$$(x_0, y_0, z_0) \ = \ (x(u_0, v_0), y(u_0, v_0), z(u_0, v_0)).$$

Remark 2.28. *The key step in proving the chain rule is multiplying by 1. We need to do this to recognize terms as derivatives; in particular, to make sure we are dividing by the right quantity.*

Theorem 2.29 (Intermediate Value Theorem)**.** *Let f be a continuous function on an interval $[a, b]$. If C is any value between $f(a)$ and $f(b)$, then there is a c in $[a, b]$ such that $f(c) = C$.*

The key idea in the proof is using Divide and Conquer. For example, if $f(a) < 0$, $f(b) > 0$ and $C = 0$, look at $f(\frac{a+b}{2})$; if that is negative we look for the root in $[a, \frac{a+b}{2}]$ while if it is positive we look for the root in $[\frac{a+b}{2}, b]$.

Theorem 2.30 (Mean Value Theorem)**.** *Let f be a continuous, differentiable function. Then there is a c in $[a, b]$ such that*

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

or, equivalently, $f(b) - f(a) = f'(c)(b - a)$.

Informally this means that at some point in time the instantaneous speed equals the average speed. The key idea in the proof is to use the Intermediate Value Theorem. For example, if the average speed is 50mph and we are always traveling faster than 50mph, this is clearly impossible; similarly we cannot be traveling less than 50mph at every instant. We are thus either always traveling at 50mph or, at some point we exceed 50mph and at another point we are less than 50mph (and now use the Intermediate Value Theorem, applied to f').

Theorem 2.31 (**Very Important: Continuity of Partial Derivatives, the Tangent Plane and Differentiability**)**.** If *the partial derivatives are continuous* then *the tangent plane is an excellent approximation to the function (we also did the simpler case of the tangent line being an excellent approximation to the function).*

The key idea in the proof is to use the Mean Value Theorem to estimate the difference of the function at various arguments with partial derivatives at various arguments. The Mean Value Theorem is used to reduce the problem to the difference of partial derivatives at very close points, which is small by the assumed continuity of the partial derivatives.

Theorem 2.32 (Computing Directional Derivatives)**.** *The directional derivative of f at $\overrightarrow{x}$ in the direction $\overrightarrow{v}$ is $(\nabla f)(\overrightarrow{x}) \cdot \overrightarrow{v}$.*

Theorem 2.33 (Geometric interpretation of the gradient)**.** *If $(\nabla f)(\overrightarrow{x}) \neq \overrightarrow{0}$ then $(\nabla f)(\overrightarrow{x})$ points in the direction of fastest increase of f.*

Theorem 2.34 (Gradients and level sets)**.** *Let S be the level set of value k for a function $f : \mathbb{R}^n \to \mathbb{R}$. Then $(\nabla f)(\overrightarrow{x})$ is perpendicular to any tangent vector on S. More explicitly, if $c(t)$ is a path in S such that $c(0) = \overrightarrow{x}$ and $c'(0) = \overrightarrow{v}$ is the tangent vector at time 0, then $(\nabla f)(\overrightarrow{x}) \cdot \overrightarrow{v} = 0$.*

3. Part 3: Higher Order Derivatives: Maxima and Minima

Some of this material involves Taylor Series, and will be covered later in the semester.

3.1. **Definitions.**

Definition 3.1 (Higher partial derivatives)**.** *We write*

$$\frac{\partial^2 f}{\partial y \partial x}$$

for

$$\frac{\partial}{\partial y}\left(\frac{\partial f}{\partial x}\right);$$

note that we read the variables from right to left. In other words, the variable furthest to the right is what we differentiate with respect to first, while the one furthest to the left is the last one. This is similar to composition of functions, where in $f(g(x))$ we first apply g and then f, reading from right to left. If $y = x$ we write

$$\frac{\partial^2 f}{\partial x^2}.$$

We often write f_x for $\partial f/\partial x$ and f_{xy} for $\partial^2 f/\partial y \partial x$.

Definition 3.2 (Degree or order of a partial derivative)**.** *We say the number of partial derivatives taken in an expression is the order or degree of the term. For example, $\frac{\partial^3 f}{\partial x^2 \partial y}$ and $\frac{\partial^3 f}{\partial x \partial z \partial y}$ are all degree 3, while $\frac{\partial^2 f}{\partial y^2}$ and $\frac{\partial^2 f}{\partial x \partial y}$ are of degree 2.*

Definition 3.3 (Class $\mathcal{C}^2$)**.** *A function f is of class $\mathcal{C}^2$ if all of its partial derivatives up to order 2 exist and are continuous. For $f : \mathbb{R}^2 \to \mathbb{R}$, this implies that $\frac{\partial f}{\partial x}$, $\frac{\partial f}{\partial y}$, $\frac{\partial^2 f}{\partial x^2}$, $\frac{\partial^2 f}{\partial y^2}$, $\frac{\partial^2 f}{\partial x \partial y}$ and $\frac{\partial^2 f}{\partial y \partial x}$ all exist and are continuous.*

Definition 3.4 (Hessian matrix)**.** *Let f be a twice-differentiable function. The Hessian of f is the matrix of second partial derivatives:*

$$Hf \ = \ \begin{pmatrix} \frac{\partial^2 f}{\partial x_1 \partial x_1} & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_1 \partial x_n} & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_n} \end{pmatrix}.$$

The Hessian is very useful in Taylor series expansions and in determining if a critical point is a maximum or a minimum (the actual mechanism makes more sense after taking linear algebra).

One easy way to remember what the order of derivatives is for the Hessian is the following:

$$Df \ = \ \nabla f \ = \ \left(\frac{\partial f}{\partial x_1}, \dots, \frac{\partial f}{\partial x_n}\right).$$

The Hessian is the derivative of this, and so the first row of $Hf = D(Df)$ is $\nabla \frac{\partial f}{\partial x_1}$, and we continue in this manner until the last row, which is $\nabla \frac{\partial f}{\partial x_n}$.

Definition 3.5 (Local extrema)**.** *A function f has a local maximum at $\overrightarrow{x}_0$ if there is a ball B about $\overrightarrow{x}_0$ such that $f(\overrightarrow{x}_0) \geq f(\overrightarrow{x})$ for all $\overrightarrow{x} \in B$; the definition for minimum is similar. Equivalently, there is a local maximum at $\overrightarrow{x_0}$ if for all $\overrightarrow{x}$ sufficiently close to $\overrightarrow{x}_0$ we have $f(\overrightarrow{x}) \leq f(\overrightarrow{x}_0)$.*

Definition 3.6 (Critical point)**.** *A point $\overrightarrow{x}_0$ is a critical point of f if $(Df)(\overrightarrow{x}_0) = \overrightarrow{0}$.*

3.2. **Theorems.**

Theorem 3.7 (Equality of Mixed Partial Derivatives)**.** *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a function of class $\mathcal{C}^2$ (which means that all the partial derivatives of order at most 2 exist and are continuous). Then for any two variables x_i and x_j we have*

$$\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial^2 f}{\partial x_j \partial x_i}.$$

Theorem 3.8 (Taylor's Theorem (One Variable))**.** *The Taylor Series approximation to f of order n at the point x_0 is*

$$f(x_0) + f'(x_0)(x - x_0) + \frac{f''(x_0)}{2!}(x - x_0)^2 + \cdots + \frac{f^{(n)}(x_0)}{n!}(x - x_0)^n,$$

where $f^{(n)}(x_0)$ denotes the n^{th} derivative of f at x_0.

Common examples are

$$\begin{aligned}
e^x &= \sum_{n=0}^{\infty} \frac{x^n}{n!} = 1 + x + \frac{x^2}{2!} + \cdots \\
\cos x &= \sum_{n=0}^{\infty} \frac{(-1)^n x^{2n}}{(2n)!} = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \cdots \\
\sin x &= \sum_{n=0}^{\infty} \frac{(-1)^n x^{2n+1}}{(2n+1)!} = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \cdots \\
\log(1 - x) &= -\sum_{n=1}^{\infty} \frac{x^n}{n} = -\left(x + \frac{x^2}{2} + \frac{x^3}{3} + \cdots\right) \\
\log(1 + x) &= \sum_{n=1}^{\infty} \frac{(-1)^{n+1} x^n}{n} = x - \frac{x^2}{2} + \frac{x^3}{3} - \cdots \\
\frac{1}{1 - x} &= \sum_{n=0}^{\infty} x^n = 1 + x + x^2 + x^3 + \cdots .
\end{aligned}$$

Theorem 3.9 (Taylor's Theorem (Several Variables))**.** *We have*

$$f(\overrightarrow{x_0} + \overrightarrow{h}) = f(\overrightarrow{x_0}) + (\nabla f)(\overrightarrow{x_0}) \cdot \overrightarrow{h} + \frac{1}{2} \overrightarrow{h}^{\mathrm{T}} (Hf)(\overrightarrow{x_0}) \overrightarrow{h} + \cdots ,$$

where Hf is the Hessian of f and $\overrightarrow{h}^{\mathrm{T}}$ is a row vector and $\overrightarrow{h}$ is a column vector.

Theorem 3.10 (Tricks for Taylor Series Expansions)**.** *We give a few examples of some powerful tricks to find Taylor series expansions.*

(1) $\cos(x + y) = 1 - \frac{(x+y)^2}{2!} + \frac{(x+y)^4}{4!} - \cdots$.
(2) $\cos x \sin y = (1 - \frac{x^2}{2!} + \cdots)(y - \frac{y^3}{3!} + \cdots)$.
(3) $e^{x-y} \cos(x + y) = (1 + (x - y) + \frac{(x-y)^2}{2!} + \cdots)(1 - \frac{(x+y)^2}{2!} + \cdots)$.

To obtain a Taylor expansion of a given order, we just need to take enough terms above and expand.

Theorem 3.11 (First derivative test for local extremum)**.** *Let f be a differentiable function on an open set U. If $\overrightarrow{x}_0 \in U$ is a local extremum, then $(Df)(\overrightarrow{x}_0) = \overrightarrow{0}$.*

The following is an advanced theorem that is proved in an analysis class.

Theorem 3.12. *Let f be a continuous function on a closed and bounded set. Then f attains its maximum and minimum on this set.*

Remark 3.13. *The corresponding result fails for open sets. Consider, for example, $f(x) = \frac{1}{x} + \frac{1}{x-1}$. This function does not attain a maximum or minimum value; as $x \to 0$ the function diverges to ∞, while as $x \to 1$ the function diverges to $-\infty$. It may also fail for unbounded sets; the function $f(x) = x^3/(1+|x|^3)$ has no maximum or minimum if the input is all of $\mathbb{R}$; the reason is that as $x \to \infty$, $f(x) \to 1$ while as $x \to -\infty$, $f(x) \to -1$.*

Theorem 3.14 (Method of Lagrange Multipliers)**.** *Let $f, g : U \to \mathbb{R}$, where U is an open subset of $\mathbb{R}^n$. Let S be the level set of value c for the function g, and let $f|_S$ be the function f restricted to S (in other words, we only evaluate f at $\overrightarrow{x} \in U$). Assume $(\nabla g)(\overrightarrow{x}_0) \neq \overrightarrow{0}$. Then $f|_S$ has an extremum at $\overrightarrow{x}_0$ if and only if there is a λ such that $(\nabla f)(\overrightarrow{x}_0) = \lambda(\nabla g)(\overrightarrow{x}_0)$.*

General comments: These problems are all done the same way. Let's say we have functions of three variables, x, y, z. Find the function to maximize f, find the constraint function g, and then solve $\nabla f(x,y,z) = \lambda \nabla g(x,y,z)$ and $g(x,y,z) = c$. Explicitly, solve:

$$\begin{aligned}
\frac{\partial f}{\partial x}(x,y,z) &= \lambda \frac{\partial g}{\partial x}(x,y,z) \\
\frac{\partial f}{\partial y}(x,y,z) &= \lambda \frac{\partial g}{\partial y}(x,y,z) \\
\frac{\partial f}{\partial z}(x,y,z) &= \lambda \frac{\partial g}{\partial z}(x,y,z) \\
g(x,y,z) &= c.
\end{aligned}$$

For example, if we want to maximize xy^2z^3 subject to $x + y + z = 4$, then $f(x,y,z) = xy^2z^3$ and $g(x,y,z) = x+y+z = 4$. The hardest part is the algebra to solve the system of equations. *Remember to be on the lookout for dividing by zero. That is never allowed, and thus you need to deal with those cases separately. Specifically, if the quantity you want to divide by can be zero, you have to consider as a separate case what happens when it is zero, and as another case what happens when it is not zero.*

Remark 3.15. *There are lots of ways of doing the algebra. Remember that the goal is to find the point. It might help to first find λ, but if you can find the point without finding λ, that's fine. A common approach is to take ratios of equations, but this runs the risk of dividing by zero, which can lead to extra cases to check. It's still not a bad idea to do this, as in each case we now know a lot, and those extra assumptions frequently make it easy to handle a case.*

Theorem 3.16 (Finding extrema)**.** *To find the extrema (maximum and minima) of a function f, the candidates are the critical points (the points where Df vanishes) and* *then look for candidates on the boundary (which can often be found by Lagrange multipliers).*

IMPORTANT: Why does the Method of Lagrange Multipliers work? The key idea is directional derivatives. If we are looking at f and our input is constrained to satisfy say $g(x_1, \dots, x_n) = c$, we look at the directional derivative of f in any direction tangent to the level set $g(x_1, \dots, x_n) = c$. To be a max or a min, the directional derivative must be zero. As the directional derivative in the direction $\overrightarrow{v}$ at $(x_1, \dots, x_n)$ is $(\nabla f)(x_1, \dots, x_n) \cdot \overrightarrow{v}$, this forces the gradient of f, $(x_1, \dots, x_n)$ is $(\nabla f)(x_1, \dots, x_n)$, to be perpendicular to each possible tangent direction. The only direction left for the gradient of f is perpendicular to all the tangent vectors, whic is the direction of the normal to the surface; however, we know $(\nabla g)(x_1, \dots, x_n)$ is in the same direction as the normal to the surface. Thus the two gradients must be in the same direction. This forces these two vectors to be parallel, so one is a multiple of the other.

Theorem 3.17 (Method of Least Squares)**.** *Given a set of observations*

$$(x_1, y_1),\ (x_2, y_2),\ \dots,\ (x_N, y_N)$$

and a proposed linear relationship between x and y, namely

$$y \;=\; ax + b,$$

then the best fit values of a and b (according to the Method of Least Squares) are given by minimizing the error function given by

$$E(a, b) \;=\; \sum_{n=1}^{N} (y_n - (ax_n + b))^2 .$$

You do not need to know this for an exam, but the best fit values are

$$a \quad = \quad \frac{\sum_{n=1}^N 1 \sum_{n=1}^N x_n y_n - \sum_{n=1}^N x_n \sum_{n=1}^N y_n}{\sum_{n=1}^N 1 \sum_{n=1}^N x_n^2 - \sum_{n=1}^N x_n \sum_{n=1}^N x_n}$$

$$b \quad = \quad \frac{\sum_{n=1}^N x_n \sum_{n=1}^N x_n y_n - \sum_{n=1}^N x_n^2 \sum_{n=1}^N y_n}{\sum_{n=1}^N x_n \sum_{n=1}^N x_n - \sum_{n=1}^N x_n^2 \sum_{n=1}^N 1}. \tag{3.1}$$

Remark 3.18. *In the Method of Least Squares, we measure error by looking at the sum of the squares of the errors between the observed values and the predicted values. There are other measurements of error possible, such as summing the absolute values of the errors or just summing the signed errors. The advantage of measuring errors by squaring is that it is not a signed quantity and calculus is applicable; the disadvantage is that larger errors are given greater weight. Using absolute values weighs all errors equally, but as the absolute value function is not differentiable the tools of calculus are unaccessible. If we just summed signed errors, then positive errors could cancel with negative errors, which is quite bad.*

Remark 3.19. *The Method of Least Squares is applicable to far more than just linear relationships, and gives a great application of logarithms. If we believe the variables P and V are related by $P = CV^r$ for some constants C and r, we cannot use the*

Method of Least Squares to find the best fit values of C *and* r *as we do not have a linear relation. If, however, we take logarithms, we're in business. Let* $\mathcal{P} = \log P$, $c = \log C$, *and* $\mathcal{V} = \log V$. *Then* $P = CV^r$ *becomes* $\mathcal{P} = r\mathcal{V} + c$, *and now we can use the Method of Least Squares to find* r *and* c.

4. Part 4: Double and Triple Integrals

4.1. **Definitions.**

Definition 4.1 (Iterated Integral)**.** *The notation*

$$\int_a^b \int_c^d f(x,y) dy dx$$

means

$$\int_a^b \left[\int_c^d f(x,y) dy \right] dx.$$

Definition 4.2 (Bounded function)**.** *A real-valued function* f *is bounded by* B *if for any* $\overrightarrow{x}$ *in the domain of* f *we have*

$$-B \ \leq \ f(\overrightarrow{x}) \ \leq \ B;$$

equivalently,

$$|f(\overrightarrow{x})| \ \leq \ B.$$

Definition 4.3 (Rectangle)**.** *The rectangle* $[a,b] \times [c,d]$ *is the set of all* (x,y) *such that* $a \leq x \leq b$ *and* $c \leq y \leq d$.

Definition 4.4 (Integral over an interval)**.** *The integral of a continuous function* f *over an interval* $[a,b]$ *is the limit of the upper or lower sum as the partition becomes finer and finer (this means that the length of each subinterval used in partitioning the interval* $[a,b]$ *tends to 0).*

Definition 4.5 (Integral over a rectangle)**.** *The integral of a continuous function* f *over a rectangle* $R = [a,b] \times [c,d]$ *is the limit of the upper or lower sum as the partition becomes finer and finer (this means that the length and width of each sub-rectangle in the partition of the rectangle* $[a,b] \times [c,d]$ *tends to 0). We denote this by*

$$\int \int_R f(x,y) dA,$$

where we use dA *to denote area.*

Definition 4.6 (x-simple, y-simple, simple)**.** *A region* $D \subset \mathbb{R}^2$ *is* x*-simple if there are continuous* ψ_1 *and* ψ_2 *defined on* $[c,d]$ *such that*

$$\psi_1(y) \ \leq \ \psi_2(y)$$

and

$$D \ = \ \{(x,y) : \psi_1(y) \leq x \leq \psi_2(y) \ \text{ and } \ c \leq y \leq d\};$$

similarly, D *is* y*-simple if there are continuous functions* $\phi_1(x)$ *and* $\phi_2(x)$ *such that*

$$\phi_1(x) \ \leq \ \phi_2(x)$$

and

$$D = \{(x, y) : \phi_1(x) \leq y \leq \phi_2(x) \text{ and } a \leq x \leq b\}.$$

If D is both x-simple and y-simple then we say D is simple.

Definition 4.7 (Elementary region). *A region that is either x-simple, y-simple or simple is frequently called an elementary region.*

Definition 4.8 (Probability distribution). *A random variable X has a continuous probability distribution p if*

(1) $p(x) \geq 0$ *for all* x*;*
(2) $\int_{-\infty}^{\infty} p(x)dx = 1$*;*
(3) *the probability X takes on a value between a and b is $\int_a^b p(x)dx$.*

Definition 4.9 (Uniform distribution). *If $p(x) = \frac{1}{b-a}$ for $a \leq x \leq b$ and 0 otherwise, then p is the uniform distribution on $[a, b]$. We often consider the special case when $a = 0$ and $b = 1$. Note that for the uniform distribution on $[0, 1]$, the probability we take a value in an interval is just the length of the interval.*

Definition 4.10 (Mean, Variance). *The mean or expected value of a random variable is $\int_{-\infty}^{\infty} xp(x)dx$. We typically denote the mean by μ. The variance is defined by $\int_{-\infty}^{\infty}(x - \mu)^2 p(x)dx$, and measures how spread out a distribution is (the larger the variance, the more spread out it is). We typically denote the variance by σ^2 and the standard deviation by σ. Note that σ, μ and x all have the same units, while the variance hs units equal to the square of this.*

4.2. **Theorems.**

Theorem 4.11 (Fubini's Theorem). *Let f be a continuous, bounded function on a rectangle $[a, b] \times [c, d]$. Then*

$$\int\int_R f(x, y)dA = \int_a^b \int_c^d f(x, y)dydx = \int_c^d \int_a^b f(x, y)dxdy.$$

Theorem 4.12 (Integral over elementary regions). *Let $D \subset \mathbb{R}^2$ be an elementary region and $f : D \to \mathbb{R}$ be continuous on D. Let R be a rectangle containing D and extend f to a function f^* by*

$$f^*(x, y) = \begin{cases} f(x, y) & \text{if } (x, y) \in D \\ 0 & \text{otherwise.} \end{cases}$$

Then

$$\int\int_D f(x, y)dA = \int\int_R f^*(x, y)dA.$$

Theorem 4.13 (Reduction to iterated integrals). *Let D be a y-simple region given by continuous functions $\phi_1(x) \leq \phi_2(x)$ for $a \leq x \leq b$. Then*

$$\int\int_D f(x, y)dA = \int_a^b \int_{\phi_1(x)}^{\phi_2(x)} f(x, y)dydx.$$

4.3. **Special Topics. Mathematical modeling:** In mathematical modeling there are two competing factors. We want the model rich enough to capture the key features of the system, yet be mathematically tractable. In general the more complicated the system is, the more involved the model will be and the harder it will be to isolate nice properties of the solution. For example, in modeling baseball games we assumed runs scored and allowed were independent random variables. This clearly cannot be true (the simplest reason is that if a team scores r runs then they cannot allow r runs, as games do not end in ties). The hope is that simple models which clearly cannot be the entire story can nevertheless capture enough of the important properties of the system that the resulting solutions will provide some insight. This is somewhat similar to Taylor series, where we replace complicated functions with polynomials; as we've seen with the incredibly fast convergence of Newton's Method, it is possible to obtain very useful information from these approximations!

In Monte Carlo Integration we can use Chebyshev's Theorem to show convergence. We probably won't do this level of detail in the class; this is included below for future reference.

Theorem 4.14. *Let X be a random variable with finite mean μ and finite variance σ^2. Then for any $k > 0$ we have*

$$\mathrm{Prob}(|X-\mu| \geq k\sigma) \ \leq \ \frac{1}{k^2}.$$

Monte Carlo Integration: Let D be a nice region in $\mathbb{R}^n$, and assume for simplicity that it is contained in the n-dimensional unit hypercube $[0,1] \times [0,1] \times \cdots \times [0,1]$. Assume further that it is easy to verify if a given point $(x_1, \dots, x_n)$ is in D or not in D. Draw N points from the n-dimensional uniform distribution; in other words, each of the n coordinates of the N points is uniformly distributed on $[0,1]$. Then as $N \to \infty$ the n-dimensional volume of D is well approximated by the number of points inside D divided by the total number of points.

5. Part 5: Change of Variables Formula

5.1. Change of Variable Formula in the Plane.

Theorem 5.1 (Change of Variables Formula in the Plane)**.** *Let S be an elementary region in the xy-plane (such as a disk or parallelogram for example). Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be an invertible and differentiable mapping, and let $T(S)$ be the image of S under T. Then*

$$\int\int_S 1 \cdot dxdy \ = \ \int\int_{T(S)} 1 \cdot \left|\frac{\partial x}{\partial u}\frac{\partial y}{\partial v} - \frac{\partial x}{\partial v}\frac{\partial y}{\partial u}\right| dudv,$$

or more generally

$$\int\int_S f(x,y) \cdot dxdy \ = \ \int\int_{T(S)} f\left(T^{-1}(u,v)\right) \cdot \left|\frac{\partial x}{\partial u}\frac{\partial y}{\partial v} - \frac{\partial x}{\partial v}\frac{\partial y}{\partial u}\right| dudv.$$

Some notes on the above:

(1) We assume T has an inverse function, denoted T^{-1}. Thus $T(x,y) = (u,v)$ and $T^{-1}(u,v) = (x,y)$.

(2) We assume for each $(x, y) \in S$ there is one and only one (u, v) that it is mapped to, and conversely each (u, v) is mapped to one and only one (x, y).

(3) The derivative of $T^{-1}(u, v) = (x(u, v), y(u, v))$ is

$$(DT^{-1})(u, v) \ = \ \begin{pmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{pmatrix},$$

and the absolute value of the determinant of the derivative is

$$\left|\det\left(DT^{-1})(u, v)\right)\right| \ = \ \left|\frac{\partial x}{\partial u}\frac{\partial y}{\partial v} - \frac{\partial x}{\partial v}\frac{\partial y}{\partial u}\right|,$$

which implies the area element transforms as

$$dxdy \ = \ \left|\frac{\partial x}{\partial u}\frac{\partial y}{\partial v} - \frac{\partial x}{\partial v}\frac{\partial y}{\partial u}\right| dudv.$$

(4) Note that f takes as input x and y, but when we change variables our new inputs are u and v. The map T^{-1} takes u and v and gives x and y, and thus we need to evaluate f at $T^{-1}(u, v)$. Remember that we are now integrating over u and v, and thus the integrand must be a function of u and v.

(5) Note that the formula requires an absolute value of the determinant. The reason is that the determinant can be negative, and we want to see how a small area element transforms. Area is supposed to be positively counted. Note in one-variable calculus that $\int_a^b f(x)dx = -\int_b^a f(x)dx$; we need the absolute value to take care of issues such as this.

(6) While we stated T is a differentiable mapping, our assumptions imply T^{-1} is differentiable as well.

5.2. **Change of Variable Formula: Special Cases.**

Theorem 5.2 (Change of Variables Theorem: Polar Coordinates)**.** *Let*

$$x \ = \ r\cos\theta, \quad y \ = \ r\sin\theta$$

with $r \geq 0$ and $\theta \in [0, 2\pi)$; note the inverse functions are

$$r \ = \ \sqrt{x^2 + y^2}, \quad \theta \ = \ \arctan(y/x).$$

Let D be an elementary region in the xy-plane, and let D^* be the corresponding region in the $r\theta$-plane. Then

$$\int\int_D f(x, y)dxdy \ = \ \int\int_{D^*} f(r\cos\theta, r\sin\theta)rdrd\theta.$$

For example, if D is the region $x^2 + y^2 \leq 1$ in the xy-plane then D^* is the rectangle $[0, 1] \times [0, 2\pi]$ in the $r\theta$-plane.

Theorem 5.3 (Change of Variables Theorem: Cylindrical Coordinates)**.** *Let*

$$x \ = \ r\cos\theta, \quad y \ = \ r\sin\theta, \quad z \ = \ z$$

with $r \geq 0$, $\theta \in [0, 2\pi)$ and z arbitrary; note the inverse functions are

$$r \ = \ \sqrt{x^2 + y^2}, \quad \theta \ = \ \arctan(y/x), \quad z \ = \ z.$$

Let D be an elementary region in the xyz-plane, and let D^* be the corresponding region in the $r\theta z$-plane. Then

$$\int\int\int_D f(x,y,z)dxdydz \ = \ \int\int\int_{D^*} f(r\cos\theta, r\sin\theta, z) r dr d\theta dz.$$

Theorem 5.4 (Change of Variables Theorem: Spherical Coordinates)**.** *Let*

$$x \ = \ \rho\sin\phi\cos\theta, \quad y \ = \ \rho\sin\phi\sin\theta, \quad z \ = \ \rho\cos\phi$$

with $\rho \geq 0$, $\theta \in [0, 2\pi]$ and $\phi \in [0, \pi)$. Note that the angle ϕ is the angle made with the z-axis; many books (such as physics texts) interchange the role of ϕ and θ. Let D be an elementary region in the xyz-plane, and let D^* be the corresponding region in the $\rho\theta\phi$-plane. Then

$$\begin{aligned} &\int\int\int_D f(x,y,z)dxdydz \\ &= \ \int\int\int_{D^*} f(\rho\sin\phi\cos\theta, \rho\sin\phi\sin\theta, \rho\cos\phi)\rho^2\sin(\phi) d\rho d\theta d\phi. \end{aligned}$$

Note that the most common mistake is to have incorrect bounds of integration.

6. Part 6: Sequences and Series

6.1. Definitions.

Definition 6.1 (Sequence)**.** *A sequence $\{a_n\}_{n=1}^\infty$ is the collection $\{a_0, a_1, a_2, \dots\}$. Note sometimes the sequence starts with a_0 and not a_1.*

For example, if $a_n = 1/n^2$ then the sequence is $\{1, 1/4, 1/9, 1/16, \dots\}$.

Definition 6.2 (Series)**.** *A series is the sum of the terms in a sequence. If we have a sequence $\{a_n\}_{n=0}^\infty$ then the partial sum s_N is the sum of the first N terms in the sequence: $s_N = \sum_{n=1}^N a_n$. We often denote the infinite sum by s:*

$$s \ = \ \lim_{N\to\infty} s_N \ = \ \lim_{N\to\infty}\sum_{n=1}^N a_n.$$

Definition 6.3 (Alternating series)**.** *An alternating series is an infinite sum of a sequence where the terms alternate in sign.*

For example, $a_n = (-1)^n/2^n$ leads to an alternating series.

Definition 6.4 (Geometric Sequence / Series)**.** *A geometric sequence with common ratio r and initial value a is the sequence $\{a, ar, ar^2, ar^3, \dots\}$. The partial sums are*

$$s_N \ = \ \frac{a - ar^{N+1}}{1-r}$$

and the series sum (when $|r| < 1$) is

$$s \ = \ \frac{1}{1-r}.$$

Definition 6.5 (Absolutely convergent series)**.** *Consider the sequence* $\{a_n\}_{n=1}^{\infty}$*. The corresponding series is absolutely convergent (or converges absolutely) if the sum of the absolute values of the* a_n*'s converges; explicitly,*

$$\lim_{N\to\infty} \sum_{n=1}^{N} |a_n|$$

exists. If the sequence is just non-negative terms, we often say the series converges.

Definition 6.6 (Conditionally convergent series)**.** *Consider the sequence* $\{a_n\}_{n=1}^{\infty}$*. The corresponding series is conditionally convergent (or converges conditionally) if*

$$\lim_{N\to\infty} \sum_{n=1}^{N} a_n$$

exists.

Note a series may be conditionally convergent but not absolutely convergent. For example, consider $a_n = (-1)^n/n$; the series converges conditionally but not absolutely.

Definition 6.7 (Diverges)**.** *If* $\lim_{N\to\infty}\sum_{n=1}^{N} a_n$ *does not converge, then we say the series diverges.*

6.2. **Tests.**

Theorem 6.8 (End-term test)**.** *Let* $\{a_n\}_{n=1}^{\infty}$ *be a sequence. If* $\lim_{N\to\infty} a_n \neq 0$ *then the series diverges.*

Theorem 6.9 (Comparison Test)**.** *Let* $\{b_n\}_{n=1}^{\infty}$ *be a sequence of non-negative terms (so* $b_n \geq 0$*). Assume the series converges, and* $\{a_n\}_{n=1}^{\infty}$ *is another sequence such that* $|a_n| \leq b_n$ *for all* n*. Then the series attached to* $\{a_n\}_{n=1}^{\infty}$ *also converges.*

Theorem 6.10 (p-Test)**.** *Let* $\{a_n\}_{n=1}^{\infty}$ *be the sequence with* $a_n = 1/n^p$ *for some fixed* $p > 0$*; this is frequently called a* p*-series. If* $p > 1$ *then the series converges, while if* $p \leq 1$ *the series diverges.*

Theorem 6.11 (Ratio Test)**.** *Consider a sequence* $\{a_n\}_{n=1}^{\infty}$ *of positive terms. Let*

$$r \;=\; \lim_{n\to\infty} \frac{a_{n+1}}{a_n}.$$

If r *exists and* $r < 1$ *then the series converges, while if* $r > 1$ *then the series diverges; if* $r = 1$ *then this test provides no information on the convergence or divergence of the series.*

For example, applying this test to the geometric series with $a_n = r^n$ we find the series converges for $r < 1$ and diverges for $r > 1$. If we consider $b_n = 1/n$ and $c_n = 1/n^2$ then both of these have a corresponding value of r equal to 1, but the first diverges while the second converges. *In fact, the proof in general uses the comparison test applied to a related geometric series.*

Theorem 6.12 (Root Test)**.** *Consider a sequence* $\{a_n\}_{n=1}^{\infty}$ *of positive terms. Let*

$$\rho \;=\; \lim_{n\to\infty} a_n^{1/n},$$

the n^{th} *root of* a_n*. If* $\rho < 1$ *then the series converges, while if* $\rho > 1$ *then the series diverges; if* $\rho = 1$ *then the test does not provide any information.*

The key idea in the proof is to use the comparison test with a related geometric series.

Theorem 6.13 (Integral Test)**.** *Consider a sequence* $\{a_n\}_{n=1}^{\infty}$ *of non-negative terms. Assume there is some function* f *such that* $f(n) = a_n$ *and* f *is non-increasing. Then the series*

$$\sum_{n=1}^{\infty} a_n$$

converges if and only if the integral

$$\int_1^{\infty} f(x)dx$$

converges.

Theorem 6.14 (Alternating Test)**.** *If* $\{a_n\}_{n=1}^{\infty}$ *is an alternating sequence with* $\lim_{n\to\infty} |a_n| = 0$ *then the series converges.*

6.3. **Examples of divergent and convergent series.** We first list some convergent series. If $\{a_n\}_{n=1}^{\infty}$ and $\{b_n\}_{n=1}^{\infty}$ are convergent series and c_1, c_2 are any constants, then $\{c_1 a_n + c_2 b_n\}_{n=1}^{\infty}$ is a convergent series; i.e.,

$$\sum_{n=1}^{\infty} (c_1 a_n + c_2 b_n)$$

converges.

Convergent series

- $a_n = r^n$, $|r| < 1$.
- $a_n = 1/n^p$, $p > 1$.
- $a_n = x^n/n!$ for any x.

Divergent series

- $a_n = r^n$, $|r| > 1$.
- $a_n = 1/n^p$ for $p \leq 1$ (in particular, $a_n = 1/n$).
- If $\lim_{n\to\infty} a_n \neq 0$ then the series cannot converge.

We have many tests – Comparison Test, Ratio Test, Root Test, Integral Test. If possible, I like to try to use the Comparison Test first. It is very simple, but has the significant drawback that you need to be able to choose a good series to compare with. This is reminiscent of finding the roots of a quadratic. If you can 'see' how to factor it then the problem is easy; if not, you have to resort to the quadratic formula (which *will* solve the problem after some work).

So too it is here. We first try to see if we can be clever and choose the right series to compare with, but if we fail then we can resort to one of the other tests.