MATH 247 (Winter 2012 - 1121)

Advanced Calculus III

Prof. K. Morris
University of Waterloo

LaTeXer: W. Kong
http://wwkong.github.io
Last Revision: April 30, 2014

Table of Contents

These notes are currently a work in progress, and as such may be incomplete or contain errors. [Source: lambertw.com]

List of Figures

ACKNOWLEDGMENTS:

Special thanks to *Michael Baker* and his LaTeX formatted notes. They were the inspiration for the structure of these notes.
Also, thanks goes out to Sam Eisenstat who contributed to part of these notes.

Abstract

The purpose of these notes is to provide a guide to the advanced offering of Calculus III. The contents of this course include both differential calculus and of integral calculus in a multivariable framework, and the connections between them. The recommended prerequisite is Math 148 and readers should have a basic understanding of single-variable differential and integral calculus.

Some adjustments have been made in terms of the identity of several theorems as propositions and notation to better organize the material. Also some corrections were made due to errors that were not brought up in lectures.

These notes and other Math 247 notes contain the recommended content for those who wish to wish to pursue upper year analysis courses such as Real Analysis (Pmath 351) and are especially important to those in the mathematical physics and finance programs as well as those in the pure mathematics program.

1 Topology of $\mathbb{R}^n$

Definition 1.1. A **multivariable** function is a function that depends on a string of numbers $(x_1, x_2, ..., x_n)$ which we usually call a **vector**. If a vector x has n entries, we say that $x \in \mathbb{R}^n$.

Definition 1.2. A **vector space**[1] is a set $Y = \{y_1, y_2, ...\}$ such that for any $y_i, y_j, y_k \in Y$, and scalars $\alpha, \beta, 1 \in \mathbb{F}$, the following axioms hold:

1. $y_i + y_j \in Y$
2. $\alpha y_i \in Y$
3. $y_i + y_j = y_j + y_i$
4. $(y_i + y_j) + y_k = y_j + (y_i + y_k)$
5. $\exists \Theta$ s.t. (such that) $\Theta + y_i = y_i$
6. $\forall y_i \in Y, \exists \tilde{y_i} \in Y$ s.t. $y_i + \tilde{y_i} = \Theta$
7. $(\alpha\beta)y_i = \alpha(\beta y_i)$
8. $(\alpha + \beta)y_i = \alpha y_i + \beta y_i$
9. $1 \cdot y_i = y_i$
10. $\alpha(y_i + y_j) = \alpha y_i + \alpha y_j$

1.1 Norms

Definition 1.3. The **Euclidean norm** between two points $x, y \in \mathbb{R}^n$ is defined as

$$\|x - y\| = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}$$

Proposition 1.1. *The Euclidean norm satisfies the following* $\forall x, y \in \mathbb{R}^n$ *and* $\forall \alpha \in \mathbb{R}$*:*

(N1) $\|x\| > 0$ *and* $\|x\| = 0 \iff x = 0$
(N2) $\|\alpha x\| = |\alpha|\|x\|$
(N3) $\|x + y\| \leq \|x\| + \|y\|$

Proof. (N1) Trivial by observation.

(N2) $\|\alpha x\| = \left(\sum_{i=1}^{n} (\alpha x_i)^2\right)^{\frac{1}{2}} = \left(\alpha^2 \sum_{i=1}^{n} (x_i)^2\right)^{\frac{1}{2}} = |\alpha|\|x\|$

(N3) We first begin by proving the following theorem.

Theorem 1.1. *(Cauchy-Schwarz Inequality)*

For any $x, y \in \mathbb{R}^n$, $\left|\sum_{i=1}^{n} x_i y_i\right| \leq \sqrt{\sum_{i=1}^{n} x_i^2}\sqrt{\sum_{i=1}^{n} y_i^2}$

[1] See Appendix A for a more formal definition.

Proof. For any $x, y \in \mathbb{R}^n$, $\alpha \in \mathbb{R}$, we have

$$0 \leq \sum_{i=1}^{n}(x_i - \alpha y_i)^2 = \underbrace{\sum_{i=1}^{n} x_i^2}_{A} - 2\alpha \underbrace{\sum_{i=1}^{n} x_i y_i}_{B} + \alpha^2 \underbrace{\sum_{i=1}^{n} y_i^2}_{C}$$

Note that in the the equation $0 \leq A - 2\alpha B + \alpha^2 C$, the roots for α are $\alpha = B \pm \sqrt{B^2 - AC}$. Since $\alpha \in \mathbb{R}$, then we have

$$\begin{aligned} B^2 - AC \geq 0 \quad &\implies \quad B^2 \leq AC \\ &\implies \quad \left(\sum_{i=1}^{n} x_i y_i\right)^2 \leq \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right) \\ &\implies \quad \left|\sum_{i=1}^{n} x_i y_i\right| \leq \sqrt{\sum_{i=1}^{n} x_i}\sqrt{\sum_{i=1}^{n} y_i} \end{aligned}$$

□

Using this information, we manipulate $\|x + y\|^2$ to get

$$\begin{aligned} \|x+y\|^2 &= \sum_{i=1}^{n}(x_i + y_i)^2 \\ &= \sum_{i=1}^{n}(x_i^2 + 2x_i y_i + y_i^2) \\ &\leq \|x\|^2 + 2\|x\|\|y\| + \|y\|^2 \\ &= (\|x\| + \|y\|)^2 \end{aligned}$$

□

Definition 1.4. A **norm** is a function $\|\cdot\| : \mathbb{R}^n \to \mathbb{R}$ that satisfies (N1),(N2), and (N3) above. We call $(\mathbb{R}^n, \|\cdot\|)$ a **normed linear (vector) space**.

Example 1.1. $\|x\|_1 = \sum_{i=1}^{n} |x_i|$ is a norm on $\mathbb{R}^n$, also known as the **Taxicab norm**.

Proof. (N1) Trivial by observation.

(N2) $\|\alpha x\|_1 = \sum_{i=1}^{n} |\alpha x_i| = |\alpha| \sum_{i=1}^{n} |x_i| = |\alpha| \|x\|_1$

(N3) $\|x + y\|_1 = \sum_{i=1}^{n} |x_i + y_i| \leq \sum_{i=1}^{n} (|x_i| + |y_i|) = \sum_{i=1}^{n} |x_i| + \sum_{i=1}^{n} |y_i| = \|x\|_1 + \|y\|_1$ □

Other examples of norms include the **p-norms***:*

$$\|x\|_p = \left(\sum_{i=1}^{n} |x^p|\right)^{\frac{1}{p}} \quad \textit{for } p \in \mathbb{N}$$

and the **Chebyshev norm** also known as the **infinity norm***:*

$$\|x\|_\infty = \max_{1 \leq i \leq n} x_i$$

Proposition 1.2. $\|x\|_2 \leq \|x\|_1 \leq \sqrt{n}\|x\|_2$.

Proof. To show the first inequality, we have

$$\|x\|_1^2 = \left(\sum_{i=1}^{n} |x_i|\right)^2 = \sum_{i=1}^{n} |x_i|^2 + c \geq \sum_{i=1}^{n} |x_i|^2 = \|x\|_2^2 \implies \|x\|_1 \geq \|x\|_2$$

for $c > 0$. For the second inequality, we have

$$\|x\|_1 = \sum_{i=1}^{n} |x|_1 \cdot 1 \leq \left(\sum_{i=1}^{n} |x_i|^2\right)^{\frac{1}{2}} \left(\sum_{i=1}^{n} 1\right)^{\frac{1}{2}} = \sqrt{n}\|x\|_2$$

by Cauchy-Schwarz. □

1.2 Open and Closed Sets

In one-dimensional $\mathbb{R}$, the idea of an "open set" around a point a of "radius" r is like an open interval around a with length $2r$:

$$\left\{x \big| |x-a| < r\right\}.$$

In two dimensional $\mathbb{R}^2$, the idea is that we now think of these sets as **open balls** around a with radius r:

$$\mathcal{B}_r(a) = \left\{x \in \mathbb{R}^2 \big| \|x-a\| < r\right\}.$$

This can be generalized to $\mathbb{R}^n$ by intuition. Note that these balls are not necessarily circles or higher dimensional spheres and depend on the norm that is being used. For example, the Euclidean norm does produce a circle, the 1-norm produces a diamond and the infinity norm produces a square.

Definition 1.5. We define a ball of radius $r > 0$ and norm $\|\|_i$ around a point $a \in \mathbb{R}^n$ with the following notation:

$$\mathcal{B}_{r,i}(a) = \left\{x \in \mathbb{R}^n \big| \|x-a\|_i < r\right\}$$

otherwise if a norm is not given, then we use the notation:

$$\mathcal{B}_r(a) = \left\{x \in \mathbb{R}^n \big| \|x-a\| < r\right\}.$$

Definition 1.6. A set $\mathcal{V} \subseteq \mathbb{R}^n$ is **open** if for all $x \in \mathcal{V}$, there exists $\epsilon > 0$ such that $\mathcal{B}_\epsilon(x) \subset \mathcal{V}$.

Remark 1.1. Let $\|\|_a$, $\|\|_b$ be norms so that

$$m\|x\|_a \leq \|x\|_b \leq M\|x\|_a, \ \forall x \in \mathbb{R}^n$$

Suppose $\mathcal{B}_{\epsilon,a}(x_0) \subset \mathcal{V}$ such that $\|x - x_0\|_a < \epsilon$. Then $\|x - x_0\|_b < M\epsilon$ and so

$$\mathcal{B}_{\epsilon,a}(x_0) \quad \subset \quad \mathcal{B}_{M\epsilon,b}(x_0)$$

Similarly, suppose $\mathcal{B}_{\epsilon,b}(x_0) \subset \mathcal{V}$ such that $\|x - x_0\|_b < \epsilon$. Then $\|x - x_0\|_b < \frac{\epsilon}{m}$ and

$$\mathcal{B}_{\epsilon,b}(x_0) \subset \mathcal{B}_{\frac{\epsilon}{m},a}(x_0)$$

Thus, given any norms $\|\|_a$, $\|\|_b$ with the inequality above for any $\epsilon > 0$, we can always enclose a ball of radius ϵ of one norm by creating a ball of radius ϵ' of the other norm. ϵ' will just be defined as above depending on the norms used.

Proposition 1.3. *The set $\mathcal{B}_r(a)$ is open for $r > 0$, $a \in \mathbb{R}^n$.*

Proof. Choose any $x_0 \in \mathcal{B}_r(a)$, let $\delta = \|x_0 - a\|$. Choose $\epsilon < r - \delta$. For $x \in \mathcal{B}_\epsilon(x_0)$, we have

$$\begin{aligned} \|x-a\| &= \|x - x_0 + x_0 - a\| \\ &\leq \|x - x_0\| + \|x_0 - a\| \\ &\leq \epsilon + \delta \\ &< r - \delta + \delta \\ &= r \end{aligned}$$

□

Definition 1.7. A set $\mathcal{V}$ is **closed** if $\mathcal{V}^c$ is open.

Example 1.2. Define $\bar{\mathcal{B}}_r(a) = \{x \big| \|x-a\| \leq r\}$. The set $\bar{\mathcal{B}}_r(a)$ is closed for any $r > 0$, $a \in \mathbb{R}^n$.

Proof. We need to show $\bar{\mathcal{B}}_r(a)^c = \{x \big| \|x-a\| > r\} = S$ is open. Choose any $x_0 \in S$. Let $\delta = \|x_0 - a\| - r > 0$. Choose any $\epsilon < \delta$. For $x \in \mathcal{B}_\epsilon(x_0)$,

$$\begin{aligned} \|a - x_0\| &= \|a - x + x - x_0\| \\ &\leq \|a - x\| + \|x - x_0\| \end{aligned}$$

and

$$\begin{aligned} \|a - x\| &\geq \|a - x_0\| - \|x - x_0\| \\ &\geq r + \delta - \epsilon \\ &> r \end{aligned}$$

So for any $x \in \mathcal{B}_\epsilon(x_0)$, $x \in \bar{\mathcal{B}}_r(a)^c$. Thus, the set is closed. □

Proposition 1.4. $\mathbb{R}^n$ *and* $\emptyset$ *are both open and closed.*

Proof. Since the empty set contains no points, "every" point $x \in \emptyset$ satisfies $\mathcal{B}_\epsilon(x) \subset \emptyset$. So $\emptyset$ is open. Since $\mathcal{B}_\epsilon(x) \subset \mathbb{R}^n$ for all $\epsilon > 0$, $x \in \mathbb{R}^n$, $\mathbb{R}^n$ is open. Since $(\mathbb{R}^n)^c = \emptyset$, and $\emptyset^c = \mathbb{R}^n$, both sets are closed. □

Definition 1.8. A point $a \in \mathbb{R}^n$ is a **boundary point** of $\mathcal{V} \subset \mathbb{R}^n$ if $\forall \epsilon > 0$, $\mathcal{B}_\epsilon(a)$ contains points in $\mathcal{V}$ and points not in $\mathcal{V}$.

1.3 Other Set Concepts

Definition 1.9. Suppose $\alpha \subset \beta \subset \mathbb{R}^n$. If there is an open set $\mathcal{O}$ such that $\alpha = \mathcal{O} \cap \beta$ then α is **relatively open** in β. Similarly, if there's a closed set $\mathcal{C}$ such that $\alpha = \mathcal{C} \cap \beta$, α is **relatively closed** in β.

Definition 1.10. If there is $\alpha, \beta \subset \gamma$ such that $\alpha \neq \emptyset$, $\beta \neq \emptyset$, $\gamma = \alpha \cup \beta$, $\emptyset = \alpha \cap \beta$ with α and β relatively open in γ, we say that α and β *separate* γ.

If there are such α, and β, we say that γ is **disconnected**. Otherwise it is **connected**.

Example 1.3. Given $\gamma = \{x \in \mathbb{R}^2 \big| |x_2| \leq |x_1|, x \neq 0\}$, $\theta_1 = \{x \in \mathbb{R}^2 \big| x_1 < 0\}$, and $\theta_2 = \{x \in \mathbb{R}^2 \big| x_1 > 0\}$, define $\alpha = \theta_1 \cap \gamma$ and $\beta = \theta_2 \cap \gamma$. Note that $\alpha \cap \beta = \emptyset$ and $\alpha \cup \beta = \gamma$. Thus, α and β separate γ and γ is disconnected.

1.4 Sequences in $\mathbb{R}^n$

Before we continue, let us first define some notation:

- Let $x_{m,n}$ represent the the n^{th} component of the m^{th} vector in a sequence of vectors $\{x_i\}_{i\geq 1}$
- Given two statements A and B, let $(\Rightarrow)$ denote the beginning of a proof to show $A \Rightarrow B$ and likewise left $(\Leftarrow)$ denote the beginning of a proof to show $(B \Rightarrow A)$. Let $(\overline{\Rightarrow})$ denote the beginning of a proof to show $\overline{A} \Rightarrow \overline{B}$ and use a similar definition as above for $(\overline{\Leftarrow})$.

Definition 1.11. In $\mathbb{R}$, we consider a sequence $\{x_i\}_{i\geq 1},\ x_i \in \mathbb{R}$. The sequence is **convergent** if there is $a \in \mathbb{R}$ so for every $\epsilon > 0$, there is $N \in \mathbb{N}$ so

$$|x_i - a| < \epsilon,\ \forall i > N$$

We say that $\lim_{i\to\infty} a$.

Definition 1.12. In $\mathbb{R}^n$, we consider a sequence of vectors

$$\left\{ x_i \middle| x_i = \begin{bmatrix} x_{1,i} \\ x_{2,i} \\ \vdots \\ x_{n,i} \end{bmatrix} \right\}.$$

We say that this sequence converges if there is $a \in \mathbb{R}^n$ so for every $\epsilon > 0$, there is $N \in \mathbb{N}$ so

$$\|x_i - a\| < \epsilon,\ \forall i > N$$

for some norm $\|\cdot\|$. We can call this kind of convergence **norm convergence**.

Proposition 1.5. *For any two arbitrary norms $\|\|_a$ and $\|\|_b$ on $\mathbb{R}^n$, the following inequality will always hold:*

$$m\|x\|_a \leq \|x\|_b \leq M\|x\|_a,\ \forall x \in \mathbb{R}^n,\ m, M \in \mathbb{R}^n \tag{1.1}$$

(Proof in later chapters)

Proposition 1.6. *A sequence $\{x_i\}_{i\geq 1} \subset \mathbb{R}^n$ is convergent in one norm iff (if and only if) it is convergent in another norm.*

Proof. Using (1.1), suppose that $\{x_i\}$ converges in $\|\|_a$. There is $y \in \mathbb{R}^n$ so for any $\epsilon > 0$, there is an $N \in \mathbb{N}$ such that

$$\|x_i - y\|_a < \epsilon,\ \forall i > N.$$

Consider $\|\|_b$ and choose $\epsilon > 0$ so $\|x_i - y\|_a < \frac{\epsilon}{M},\ \forall i > N$. Then,

$$\|x_i - y\|_b < \epsilon,\ \forall i > N.$$

Thus, the sequence is also convergent in $\|\|_b$. A similar argument for the other direction can be made using the fact that $\|x\|_a \leq \frac{1}{m}\|x\|_b$. □

Proposition 1.7. *The sequence $\{x_i\}_{i\geq 1} \subset \mathbb{R}^n$ is convergent iff $\lim_{i\to\infty} x_{k,i} = a_k,\ 1 \leq k \leq n$ for some $a_k \in \mathbb{R}$.*

Proof. Use the max norm. Let $a \in \mathbb{R}^n$ be the limit of a convergent sequence. We have by observation

$$\|x_i - a\|_\infty < \epsilon \iff |x_{k,i} - a| < \epsilon,\ 1 \leq k \leq n$$

So norm convergence $\iff$ **component-wise convergence**. □

Definition 1.13. A sequence $\{x_i\} \subset \mathbb{R}^n$ is **Cauchy** if $\forall \epsilon > 0,\ \exists N \in \mathbb{N}$ such that

$$\|x_i - x_j\| < \epsilon,\ \forall i, j > N$$

over any arbitrary norm $\|\cdot\|$.

Proposition 1.8. *A sequence of vectors is convergent iff it is Cauchy.*

Proof. Suppose a sequence is convergent with a limit $a \in \mathbb{R}^n$. For any $\epsilon > 0$, $\exists N \in \mathbb{N}$ such that $\|x_i - a\| < \frac{\epsilon}{2}$, $\forall i > N$. Then,

$$\begin{aligned} \|x_i - x_j\| &= \|x_i - a + a - x_j\| \\ &\leq \|x_i - a\| + \|x_j - a\| \\ &< \epsilon \end{aligned}$$

and thus the sequence is Cauchy. Now suppose that the sequence is Cauchy. Use $\|\|_\infty$. Each component $\{x_{i,k}\}_{i\geq 1} \subset \mathbb{R}$ is Cauchy. So there are real numbers a_k, $k = 1, ..., n$ such that

$$\lim_{i\to\infty} x_{i,k} = a_k$$

and thus $\|x_i - a\|_\infty = 0$ where $a = (a_1, ..., a_n)$ for $i > N$ for some $N \in \mathbb{N}$. □

Proposition 1.9. *A set $A \subset \mathbb{R}^n$ is closed iff every convergent sequence $\{x_{i,k}\}_{i\geq 1}$ with $x_i \in A$, has its **limit point** in A.*

Proof. This is trivially true if $A = \emptyset$.

($\Rrightarrow$) Assume that $A \neq \emptyset$ and suppose that A^c is not open. That is, for some $x \in A^c$, no ball $\mathcal{B}_r(x) \subset A^c$. For $i = 1, 2, ...$ choose $x_i \in \mathcal{B}_{\frac{1}{i}}(x) \cap A$. Then $\{x_i\} \subset A$ and $\|x_i - x_j\| < \frac{1}{i}$ so $\lim_{i\to\infty} x_i = x$. So not every convergent sequence in A has a limit point in A.

($\Lleftarrow$) Let $\{x_i\} \subset A$, $\lim_{i\to\infty} x_i = a$, $a \in A^c$. By definition of limit, for all $\epsilon > 0$, there is $N \in \mathbb{N}$ such that

$$\|x_i - a\| < \epsilon,\ i > N$$

or $x_i \in \mathcal{B}_\epsilon(a)$, $i > N$. Since $x_i \in A$, this means every ball $\mathcal{B}_\epsilon(a)$ around a contains a point in A. Since $a \in A^c$, A^c is not open and A is not closed. □

Definition 1.14. For $A \subset \mathbb{R}^n$, the closure of A is defined to be:

$$\overline{A} = \left\{a \in \mathbb{R}^n \middle| \forall\epsilon > 0, \mathcal{B}_\epsilon(a) \cap A \neq 0\right\}$$

2 Functions in $\mathbb{R}^n$

Definition 2.1. Let $A \subset \mathbb{R}^n$ be non-empty, $a \in \mathbb{R}^n$. If there is $\{x_i\}_{i\geq 1} \subset A\backslash a$, we say that

$$\lim_{i\to\infty} x_i = a$$

where a is an **accumulation point** of A. The set of all accumulation points in A is denoted by A^a. If $a \in A\backslash A^a$, then we say that a is an **isolated point** of A.

2.1 Limits of Functions

Definition 2.2. Let $f : A \to \mathbb{R}^m$, $A \in \mathbb{R}^n$ non-empty. For $a \in A^a$ and $L \in \mathbb{R}^m$ we define the following:

- If $\forall\epsilon > 0,\ \exists\delta > 0$ such that $\|x - a\| < \delta$, $x \in A \implies \|f(x) - L\| < \epsilon$, we say that f has **limit** L. That is,

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

- If also, f is defined at a and $\lim_{x\to a} f(x) = f(a)$, f is **continuous** at a.

Proposition 2.1. *Let $A \subset \mathbb{R}^n$ be a non-empty set, $a \in A$, $f : A \to \mathbb{R}^m$. Then $\lim_{x\to a} f(x) = L$ iff $\lim_{i\to\infty} f(x_i) = L$ for every sequence $\{x_i\}_{i\geq 1} \subset A\backslash a$ with $\lim_{i\to\infty} x_i = a$. That is,*

$$\lim_{i\to\infty} f(x_i) = f\left(\lim_{i\to\infty} x_i\right)$$

Proof. ($\Rightarrow$) Suppose that $\lim_{x\to a} f(x) = L$ and so we have $\forall\epsilon > 0,\ \exists\delta > 0$ such that $\|x - a\| < \delta,\ x \in A \implies \|f(x) - L\| < \epsilon$. Let $\{x_i\}_{i\geq 1} \subset A\backslash a$ be convergent to a. then $\exists N$ such that $\|x_i - a\| < \delta$ for $i > N$. So $\|f(x_i) - L\| < \epsilon$ for $i > N$. Thus, $\lim_{i\to\infty} f(x_i) = L$.

($\not\Rightarrow$) Suppose $\exists\delta > 0$ such that $\|x - a\| < \delta$ but $\|f(x) - L\| \geq \epsilon$ for some $x \in A\backslash a$. For all $i \in \mathbb{N}$, $x_i \in A\backslash a$ and so

$$\begin{aligned} 0 < \|x_i - a\| < \frac{1}{i} \quad &\implies \quad \|f(x_i) - L\| \geq \epsilon \\ &\implies \quad \lim_{i\to\infty} x_i = a \\ &\implies \quad \lim_{i\to\infty} f(x_i) \neq L \end{aligned}$$

□

Example 2.1. Does the limit $\lim_{x\to 0} \underbrace{\frac{x_1^2 x_2}{x_1^4 + x_2^2}}_{f}$ exist? (f is defined on $\mathbb{R}^2\backslash 0$)

Let $x_1 = 0 \implies \lim_{x_2\to 0} f(0, x_2) = \lim_{x_2\to 0} \frac{0}{x_2^2} = 0$.

Now let $x_2 = x_1^2 \implies \lim_{x_1\to 0} f(x_1, x_1^2) = \lim_{x_1\to 0} \frac{x_1^4}{x_1^4 + x_1^4} = \frac{1}{2}$

Thus, since the limits are inconsistent, the limit does not exist.

Theorem 2.1. *(Limit Theorems)*

Let $a \in \mathbb{R}^n$, $\mathcal{V}$ an open set containing a, $f, g : \mathcal{V}\backslash a \to \mathbb{R}^n$. If $\lim_{x\to a} f(x) = L_f$, $\lim_{x\to a} fg(x) = L_g$, then the following hold.

- $\lim_{x\to a} [\alpha f(x) + g(x)] = \alpha L_f + L_g$, $\alpha \in \mathbb{R}$
- $\lim_{x\to a} f(x)g(x) = L_f L_g$
- If $L_g \neq 0$, $\lim_{x\to a} \frac{f(x)}{g(x)} = \frac{L_f}{L_g}$

Proof. Exercise for the reader. □

Theorem 2.2. *(Squeeze Theorem)*

Consider $f, gh : A \to \mathbb{R}$, with $A^a \neq \emptyset$ and let $a \in A^a$. Suppose,

$$f(x) \leq g(x) \leq h(x),\ \forall x \in A\backslash a \tag{2.1}$$

If $\lim_{x\to a} f(x) = b$, $\lim_{x\to a} h(x) = b$, then $\lim_{x\to a} g(x) = b$.

Proof. Choose any $\epsilon > 0$. There is $\delta > 0$ so if $\|x - a\| < \delta$, $x \in A$, then $|f(x) - b| < \epsilon$, $|g(x) - b| < \epsilon$. Since (2.1) holds, $f(x) - b \leq g(x) - b \leq h(x) - b$ and so $|g(x) - b| < \epsilon$. Thus, $\lim_{x\to a} g(x) = b$. □

Corollary 2.1. *If* $|g(x) - L| \le h(x)$ *for all* $x \in A \backslash a$ *and* $\lim_{x \to a} h(x) = 0$, *then* $\lim_{x \to a} g(x) = L$.

Example 2.2. Define $f(x_1, x_2) = \frac{x_1 x_2^4}{x_1^2 + x_2^6}$, $(x_1, x_2) \ne 0$, $A = \mathbb{R}^2 \backslash (0,0)$. Does $\lim_{x \to 0} f(x)$ exist?

For any cases that we come up with, this seems to be true. However, to be sure, we will apply the squeeze theorem and the following lemma:

Lemma 2.1. *(Young's inequality)*
$(|a| - |b|)^2 = a^2 + b^2 - 2|a||b| \ge 0 \implies 2|a||b| \le a^2 + b^2$

Back to our example, we have the following

$$0 \le \left| \frac{x_1 x_2^4}{x_1^2 + x_2^6} - 0 \right| \le \frac{|x_2|(x_1^2 + x_2^6)}{2(x_1^2 + x_2^6)} \le \frac{|x_2|}{2}$$

and thus $\lim_{x_2 \to 0} \frac{|x_2|}{2} = 0 \implies \lim_{x \to 0} f(x_1, x_2) = 0$.

2.2 Continuity

Definition 2.3. Let $a \in \mathbb{R}^n$, $\mathcal{V}$ an open set containing a and $f : \mathcal{V} \to \mathbb{R}^m$. The function is continuous at a if $\lim_{x \to a} f(x) = f(a)$.

Theorem 2.3. *(Continuity Theorems)*

Let $a \in \mathbb{R}^n$, $\mathcal{V}$ an open set containing a and $f, g : \mathcal{V} \to \mathbb{R}^m$. Assume f, g are continuous at a. Then the following hold true:

- $f + g$ is continuous at a
- αf is continuous at a, $\alpha \in \mathbb{R}$
- fg is continuous at a
- If $g \ne 0$, then $\frac{f}{g}$ is continuous at a

Proof. Exercise for the reader □

Theorem 2.4. *(Composition Continuity Theorem)*

Let $a \in \mathbb{R}^n$, $\mathcal{V}$ an open set containing a with $f : \mathcal{V} \to \mathbb{R}^m$ continuous at a, and let $g : W \subset \mathbb{R}^m \to \mathbb{R}^p$ be continuous on an open set W containing $f(a)$. Then the composite function $h = g \circ f$, defined by $h(x) = g(f(x))$ is continuous at a.

Proof. Exercise for the reader. □

A visualization of Theorem 2.4 can be seen below:

Example 2.3. The function $f(x_1, x_2) = \begin{cases} \frac{\sin(x_1^2 + x_2^2)}{x_1^2 + x_2^2} & (x_1, x_2) \ne (0,0) \\ 1 & (x_1, x_2) = (0,0) \end{cases}$ is continuous on $\mathbb{R}^2$.

Proof. We note that the following functions are continuous: $x_1^2, x_2^2, x_1^2 + x_2^2$ and $g(z) = \text{sinc}(z) = \begin{cases} \frac{\sin z}{z} & z \ne 0 \\ 1 & z = 0 \end{cases}$. By Theorem 2.4, $g(x_1^2 + x_2^2)$ is continuous. □

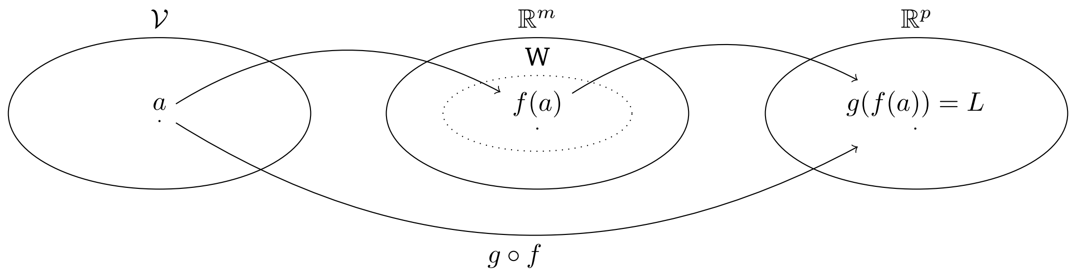

Figure 2.1: Visualization of CCT

2.3 Continuity and Sets

Recall that if $A \subset S \subset \mathbb{R}^n$ and $\exists \theta$ such that $S \cap \theta = A$ where θ is some open set, then A is relatively open in S.

Proposition 2.2. *Consider $A \subset S \subset \mathbb{R}^n$.*

1. A is relatively open in S iff $\forall a \in A$, $\exists r > 0$ such that $\mathcal{B}_r(a) \cap S \subset A$.
2. If S is open, A is relatively open in S iff A is open.

Proof. See Wade 8.2.7., p. 292 □

Example 2.4. Take $A = [1, 5)$, $S = [1, \infty)$. Pick $\theta = (-5, 5)$, and so $A = S \cap \theta$. Then

$$\mathcal{B}_r(1) = (1 - r, 1 + r) \implies \mathcal{B}_r(1) \cap S = [1, 1 + r) \subset A.$$

Define $S = \{x \in \mathbb{R}^2 \big| |x_2| \leq |x_1|, x \neq 0\}$, $\theta = \{x \in \mathbb{R}^2 \big| x_1 < 0\}$. Take $A = S \cap \theta = \{x \in \mathbb{R}^2 \big| |x_2| \leq |x_1|, x < 0\}$. Then,

$$\mathcal{B}_r(a) \cap S = \{x \big| \|x - a\| < r, |x_1| \leq |x_2|\} \subset A$$

Remark 2.1. What if $A = S$? Well, $\mathbb{R}^n$ is open, so $A = \mathbb{R}^n \cap S = A = S$. Thus, A is relatively open in itself.

Next, recall that a set $\gamma \in \mathbb{R}^n$ is disconnected if $\exists \alpha, \beta \subset \gamma$ such that $\alpha \neq \emptyset$, $\beta \neq \emptyset$, $\gamma = \alpha \cup \beta$, $\alpha \cap \beta = \emptyset$ with α, β relatively open in γ. We say that if such α, β exist, then α, β separate γ with γ being disconnected. Otherwise, γ is connected.

Proposition 2.3. *If $A \subset \mathbb{R}^n$ connected and $f : A \to \mathbb{R}^m$ is continuous on A, then $f(A)$ is connected (in $\mathbb{R}^m$).*

Proof. Suppose that $f(A)$ is disconnected. So there $\exists U, V \subset f(A)$ non-empty and relatively open in $f(A)$ such that $U \cap V = \emptyset$, $f(A) = U \cup V$. Set $\mathcal{U}_1 = f^{-1}(U)$ and $\mathcal{V}_1 = f^{-1}(V)$[2]. Then $\mathcal{U}_1$ and $\mathcal{V}_1$ are relatively open in A.[3]Since $f(A) = U \cup V$ and $f^{-1}(U), f^{-1}(V) \subset A$, then $A = f^{-1}(U) \cup f^{-1}(V)$. Also note that $U \cap V = \emptyset \implies f^{-1}(U) \cap f^{-1}(V) = \emptyset$.[4] Thus,

$$A = \mathcal{U}_1 \cap \mathcal{V}_1$$

and so A is disconnected. □

Theorem 2.5. *(Intermediate Value Theorem)*

Let $f : A \to \mathbb{R}$ be continuous. If A is connected, then $\forall a, b \in A$, with $f(a) < f(b)$, and $\forall v \in (f(a), f(b))$, $\exists c \in A$ such that $f(c) = v$.

[2]Note here that we define $f(A) = \{y \in \mathbb{R}^m | y = f(x), x \in A\}$ and $f^{-1}(U) = \{x \in A | f(x) \in U\}$.
[3]See Assignment 2.
[4]$f : X \to Y, E_\alpha \in Y \implies f^{-1}(\bigcup_\alpha E_\alpha) = \bigcup_\alpha f^{-1}(E_\alpha)$ (See Wade 137(iii) for proof).

Proof. Suppose the contrary, that is, $v \notin f(A)$. Let

$$\begin{aligned} S_1 &= (-\infty, v) \cap f(A) \\ S_2 &= (v, \infty) \cap f(A) \end{aligned}$$

Both are relatively open in $f(A)$ with $f(a) \in S_1$, $f(b) \in S_2$. So they are both non-empty with $S_1 \cap S_2 = \emptyset$ and $S_1 \cup S_2 = f(A)$. Thus, $f(A)$ is disconnected so A is disconnected. □

Proposition 2.4. *For $f : \mathbb{R}^n \to \mathbb{R}^m$, the following are equivalent:*

1. f is continuous on $\mathbb{R}^n$
2. $\forall V \in \mathbb{R}^m$ where V is open, $f^{-1}(V)$ is open (in $\mathbb{R}^n$)
3. $\forall V \in \mathbb{R}^m$ where V is closed, $f^{-1}(V)$ is closed (in $\mathbb{R}^n$)

Proof. (1)⇒(2): Let V be open and choose any $x \in f^{-1}(V)$ so $f(x) \in V$. Since V is open, there is $\mathcal{B}_\epsilon(f(x)) \subset V$ for some $\epsilon > 0$. Because f is continuous, $\exists \delta_x > 0$ so $\|y - x\| < \delta_x$ implies $\|f(y) - f(x)\| < \epsilon$. Then $\mathcal{B}_{\delta_x}(x) \subset f^{-1}(V)$. Therefore $f^{-1}(V)$ is open.

(2)⇒(1): Choose any $\epsilon > 0$, $x_0 \in \mathbb{R}^n$, $y_0 = f(x_0)$. $f^{-1}(\mathcal{B}_\epsilon(y_0))$ is open . Since $x_0 \in f^{-1}(\mathcal{B}_\epsilon(y_0))$ there is $\delta > 0$ such that $\mathcal{B}_\delta(x_0) \subset f^{-1}(\mathcal{B}_\epsilon(y_0))$. In other words, for any $\epsilon > 0$, there is $\delta > 0$ so $\|x - x_0\| < \delta$ implies $\|f(x) - f(x_0)\| < \epsilon$. Thus, f is continuous at x_0.

(2) ⇒(3): We first note that

$$\mathbb{R}^n = f^{-1}(V) \cup f^{-1}(V^c) \implies f^{-1}(V) = [f^{-1}(V^c)]^c \implies [f^{-1}(V)]^c = f^{-1}(V^c)$$

If V is closed, V^c is open, $f^{-1}(V^c)$ is open and so by the above, $f^{-1}(V)$ is closed.

(3) ⇒(2): Can be proven using similar logic as the above.[5] □

Example 2.5. $f(x) = x^2$, $V = (-1, 1)$, $f(V) = [0, 1)$ (counter-example)

Example 2.6. $f(x) = \frac{1}{x}$, $V = [1, \infty)$, $f(V) = (0, 1]$

Proposition 2.5. *Suppose that $A \subset \mathbb{R}^n$, $f : A \to \mathbb{R}^n$. Then f is continuous on A iff for every open set $V \in \mathbb{R}^m$, $f^{-1}(V)$ is relatively open on A.*

Proof. See Wade Thm. 9.2.6 □

Definition 2.4. A set $A \in \mathbb{R}^n$ is **compact** if every sequence $\{x_i\}_{i \geq 1} \subset A$ has a subsequence convergent to some element of A.

Definition 2.5. A sequence $\{x_i\}_{i \geq 1}$ is bounded if there is $M > 0$ such that

$$\|x_i\| \leq M, \forall i$$

Theorem 2.6. *(Bolzano-Weierstrauss Theorem)*

Every bounded sequence of vectors in $\mathbb{R}^n$ has a convergent subsequence.

Proof. Sequence $\{x_{i1}\}$,the first component, has a convergent subsequence with indices i_1. 2^{nd} component $\{x_{i_1 2}\}$ has a convergent subsequence i_2. Continue in this fashion for all n components. Note that we are using the result in the one dimensional case. □

[5] An alternative proof of Prop. 2.4. can be found in Harrier and Wanner (H+W) IV, 2.8.

Proposition 2.6. *A set $A \subset \mathbb{R}$ is compact iff it is closed and bounded.*

Proof. Suppose A is closed and bounded, $\|x\| \leq M$, for all $x \in A$, some M. Choose any sequence $\{x_i\}_{i\geq 1} \subset A$. By B-W (Bolzano-Weierstrauss Theorem), it has a convergent subsequence. Since A is closed, the limit is in A.

Now suppose A is compact. Then A is closed because every subsequence of a convergent sequence has the same limit. Now suppose A is not bounded. Then there is a sequence $\{x_i\}_{i\geq 1} \subset A$, $\|x_i\| \to \infty$. This has no convergent subsequence and so A is not compact. □

Definition 2.6. Let $A \subset \mathbb{R}^n$. An **open covering** is a family of open sets $\{U_\lambda\}_{\lambda \in L}$ with

$$\bigcup_{\lambda \in L} U_\lambda \supset A.$$

If there exists a finite covering,

$$U_{\lambda_1} \cup U_{\lambda_2} \cup \ldots \cup U_{\lambda_m} \supset A$$

this is said to be a **finite subcovering**.

Theorem 2.7. *(Heine-Borel Theorem)*

A set A is compact iff every open covering has a finite subcovering.

Proof. H+W, I.21 (P. 283) □

Proposition 2.7. *Let $A \subset \mathbb{R}^n$ be non-empty and compact. If $f \in \mathcal{C}(A, \mathbb{R}^n)$ then $f(A)$ is compact.*

Proof. See Wade. Or this link. □

Example 2.7. $f(x) = e^{-x}$, $A = [1, \infty)$, $f(V) = (0, e^{-1}]$. Choose any sequence $\{y_i\} \subset f(A)$, $y_i = e^{-x_i}$. As $y_i \to 0$, $x_i \to \infty$. For a different interval, try $A = [1, \ln 10]$, $f(V) = [\frac{1}{10}, \frac{1}{e}]$.

Theorem 2.8. *(Extreme Value Theorem (EVT))*

Let $A \subset \mathbb{R}^n$ be a non-empty compact set, $f \in \mathcal{C}(A, \mathbb{R})$. Then there is $x_0 \in A$, $x_1 \in A$ such that

$$f(x_0) \leq f(x) \leq f(x_1), \forall x \in A$$

Proof. Write $S = f(A)$ which is closed and bounded. Define $\alpha = \inf(S)$ and $\beta = \sup(S)$. For every $i \in \mathbb{N}$, there is $x_i \in A$,

$$\alpha \leq f(x_i) \leq \alpha + \frac{1}{i}.$$

Thus, $\lim_{i\to\infty} f(x_i) = \alpha$. Since S is closed, $\alpha \in S$ and there is $x_0 \in A$ so $f(x_0) = \alpha$. Existence of x_1 is shown similarly. □

Example 2.8. Approximations and recognition: Let $A \subset \mathbb{R}^n$ be a non-empty compact set. We wish to find an element $x \in A$ that's closest to a given $x_0 \in \mathbb{R}^n$

$$\|x - x_0\| = \inf_{x \in A} \|x - x_0\|$$

by showing $f(x) = \|x - x_0\|$:

$$\begin{aligned} \|x - y\| &= \|x - x_0\| - \|y - x_0\| \\ &\geq |\|x - x_0\| - \|y - x_0\|| \\ &= |f(x) - f(y)| \end{aligned}$$

So for any $\epsilon > 0$, if $\|x-y\| < \epsilon$, ,$|f(x) - f(y)| > \epsilon$ and so f is continuous on $\mathbb{R}^n$. By EVT (Extreme Value Theorem), $\exists \bar{x} \in A$ such that

$$\|\bar{x} - x_0\| = \inf_{x\in A} \|x - x_0\|$$

Proposition 2.8. *All norms on $\mathbb{R}^n$ are equivalent.*

Proof. We will show that for any $\|\|$, $\exists A > 0$ and $B > 0$ such that

$$A\|x\|_2 \leq \|x\| \leq B\|x\|_2, \forall x \in \mathbb{R}^n.$$

Consider $S = \left\{x \in \mathbb{R}^n \middle| x_1^2 + x_2^2 + ... + x_n^2 = 1\right\}$ which is closed and bounded, hence compact. There is A and B such that

$$A \leq \|y\| \leq B, \forall y \in S$$

and $A > 0$ since $\|\|$ is a norm, $0 \neq S$. Consider now, any $x \in \mathbb{R}^n$, $x \neq 0$,

$$x = \underbrace{\|x\|_2}_{\alpha} \underbrace{\frac{x}{\|x\|_2}}_{y\in S}$$

Since $\|\|$ is a norm, then

$$\|x\| = \|x\|_2 \underbrace{\left\| \frac{x}{\|x\|_2} \right\|}_{y} \leq \|x\|_2 B$$

and similarly

$$\|x\| \geq \|x\|_2 A.$$

Thus,

$$A\|x\|_2 \leq \|x\| \leq B\|x\|_2, \forall x \in \mathbb{R}^n.$$

□

Definition 2.7. A function is $f : A \subset \mathbb{R}^n \to \mathbb{R}^m$ is **continuous** at $x_0 \in A$ if for any $\epsilon > 0$, $\exists \delta > 0$ so $\|x - x_0\| < \delta, x \in A \implies \|f(x) - f(x_0)\| < \epsilon$. We say that it is continuous on A it is continuous at all $x_0 \in A$. It is said to be **uniformly continuous** on A if the same δ can be used for all $x_0 \in A$.

Proposition 2.9. *Let $f : A \subset \mathbb{R}^n \to \mathbb{R}^m$ be continuous on A. If A is compact, then f is uniformly continuous on A.*

Proof. See H+W, IV.2 □

3 Differential Multivariate Calculus

In this section, we will try to define differentiability in a multivariable framework.

Consider $f : \mathbb{R}^2 \to \mathbb{R}$. We define the **level curves** as

$$\{(x_1, x_2) | c = f(x_1, x_2), \forall c\}$$

and similarly, we define the two define **cross sections** as

$$\{(x_1, x_2) | x_1 = c, x_2 = f(c, x_2), \forall c\}$$
$$OR$$
$$\{(x_1, x_2) | x_2 = c, x_2 = f(x_1, c), \forall c\}$$

and these will be quite useful for visualizing some of the later proofs in this section.

3.1 Partial Derivatives

Now, we define the rate of change in the x_1 direction at (a_1, a_2) as

$$\lim_{h\to 0} \frac{f(a_1+h, a_2) - f(a_1, a_2)}{h} = \frac{\partial f}{\partial x_1}(a) = D_1 f(a) = f_{x_1}(a).$$

We call this a **partial derivative**.

Definition 3.1. A point a is an **interior point** of $U \subset \mathbb{R}^n$ if there is $\mathcal{B}_\epsilon(a) \subset U$ for some $\epsilon > 0$.

Definition 3.2. Assume a is an interior point of U. Let $f : U \subset \mathbb{R}^n \to \mathbb{R}$. The partial derivatives are

$$\begin{aligned}
\tfrac{\partial f}{\partial x_1}(a) &= \lim_{h\to 0} \tfrac{f(a_1+h, a_2, \ldots, a_n) - f(a)}{h} \\
\tfrac{\partial f}{\partial x_2}(a) &= \lim_{h\to 0} \tfrac{f(a_1, a_2+h, \ldots, a_n) - f(a)}{h} \\
&\vdots \\
\tfrac{\partial f}{\partial x_n}(a) &= \lim_{h\to 0} \tfrac{f(a_1, a_2, \ldots, a_n+h) - f(a)}{h}
\end{aligned}$$

Note that if all the partial derivatives exist for a function, it does not mean that it is continuous.

Definition 3.3. The **directional derivative** of $f : U \subset \mathbb{R}^n \to \mathbb{R}$ at $a \in U$ in the direction u, $\|u\| = 1$ is defined as

$$D_u f(a) = \lim_{h\to 0} \frac{f(a+hu) - f(a)}{h} = \frac{d}{dh} f(a+hu)\bigg|_{h=0}$$

if the limit exists.

Example 3.1. Let

$$f(x_1, x_2) = \begin{cases} \frac{x_1^2 x_2}{x_1^4 + x_2^2} & (x_1, x_2) \neq 0 \\ 0 & (x_1, x_2) = 0 \end{cases}.$$

If $u = (u_1, u_2)$, then

$$\begin{aligned}
u_2 = 0 \quad &\implies \quad D_{(1,0)} f(0) = f_{x_1}(0) = 0 \\
u_2 \neq 0 \quad &\implies \quad \lim_{h\to 0} \frac{f(0 + hu_1, hu_2) - f(0,0)}{h} \\
&\qquad = \ldots \\
&\qquad = \lim_{h\to 0} \frac{u_1^2 u_2^2}{h^2 u_1^4 + u_2^2} \\
&\qquad = \frac{u_1^2}{u_2}
\end{aligned}$$

3.2 Linear Approximations and Differentiability

Definition 3.4. The **linear approximation** for a function f at an interior point $a \in U$ is defined as $L_a(x) = f(a) + f'(a)(x-a)$ where $f'(a) \in \mathbb{R}^{m\times n}$.

Proposition 3.1. *A function $f : U \subset \mathbb{R}^n \to \mathbb{R}$ is said to be **differentiable** at an interior point $a \in U$ if the following is satisfied*

$$\lim_{x\to a} \frac{\|f(x) - L_a(x)\|}{\|x - a\|} = 0$$

where $L_a(x)$ is the linear approximation of f at a. An alternative definition is that there exists a linear map $f'(a) : \mathbb{R}^m \to \mathbb{R}^n$ and $r(x) : U \to \mathbb{R}$, with $r(a) = 0$, such that

$$f(x) = f(a) + f'(a)(x-a) + r(a)\|x-a\|$$

Proof. For $f:\mathbb{R}\to\mathbb{R}$, $g'(a)=\lim_{x\to a}\frac{g(x)-g(a)}{x-a}$ and

$$\begin{aligned} 0 &= \lim_{x\to a}\left[\left(\frac{f(x)-f(a)}{x-a}\right)-\left(f'(a)\frac{x-a}{x-a}\right)\right] \\ &= \lim_{x\to a}\left[\frac{f(x)-f(a)-f'(a)(x-a)}{x-a}\right] \\ &= \lim_{x\to a}\frac{|f(x)-L_a(x)|}{|x-a|} \end{aligned}$$

□

Proposition 3.2. *If $f:U\subset\mathbb{R}^n\to\mathbb{R}$ is differentiable at a, all partial derivatives exists at a and*

$$f'(a)=\nabla f(a)=\left[\begin{array}{cccc} \frac{\partial f}{\partial x_1}(a) & \frac{\partial f}{\partial x_2}(a) & \cdots & \frac{\partial f}{\partial x_n}(a)\end{array}\right]$$

which we call the ***gradient*** *of f.*

Proof. Consider $x_1\to a$, with $x_i=a_i$, $i\neq 1$. Then,

$$\begin{aligned} 0 &= \lim_{x_1\to a_1}\frac{|f(x_1,a_2,...,a_n)-L_a(x_1,a_2,...,a_n)|}{|x_1-a_1|} \\ &= \lim_{x_1\to a_1}\frac{|f(x_1,a_2,...,a_n)-f(a_1,a_2,...,a_n)-m_1(x_1-a_1)-m_2(a_2-a_2)-...-m_n(a_n-a_n)|}{|x_1-a_1|} \\ &= \lim_{h\to 0}\left|\frac{f(a_1+h,a_2,...,a_n)-f(a)}{h}-m_1\right| \end{aligned}$$

and so $\frac{\partial f}{\partial x_1}(a)$ exists and is equal to m_1. The same argument can be applied to the other terms to obtain their partials. □

Example 3.2. $f(x_1,x_2)=\sqrt{|x_1x_2|}$, $f(0,0)=0$, $f_{x_1}(0,0)=0$, $f_{x_2}(0,0)=0$. If f differentiable at 0? We know that

$$L_0(x)=0+0(x_1-0)+0(x_2-0)=0$$

and so

$$f'(0)=\frac{|f(x_1,x_2)-0|}{\|(x_1,x_2)-0\|_2}=\frac{\sqrt{|x_1x_2|}}{\sqrt{x_1^2+x_2^2}}.$$

Try the line $x_2=x_1 \implies \lim_{x_1\to 0}\frac{\sqrt{x_1^2}}{\sqrt{2x_1^2}}=\frac{1}{\sqrt{2}}\neq 0$ and hence f is not differentiable at 0.

Exercise 3.1. Is $f(x_1,x_2)$ differentiable at 0? (hint: try the Squeeze Theorem)

Proposition 3.3. *A vector valued function f is differentiable iff each component function is differentiable.*

Proof. Exercise. □

Later on, we will prove that $f'(a)$ is in the form of

$$f'(a)=\begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \ddots & & \vdots \\ \vdots & & & \\ \frac{\partial f_m}{\partial x_1} & \cdots & & \frac{\partial f_m}{\partial x_n} \end{bmatrix}=Df(a)$$

where $f'(a)$ is called the **Jacobian** of f.

Remark 3.1. An alternate way of defining differentiability is the following. Let $f(x) - L_a(x) = R(x) = r(x)\|x-a\|$ which implies that

$$\|r(x)\| = \frac{\|f(x) - L_a(x)\|}{\|x-a\|}.$$

We say that f is differentiable if $\lim\limits_{x\to a} \|r(x)\| = 0$.

Proposition 3.4. *Let $A \in \mathbb{R}^{m\times n}$. Then $\|Ax\|_\infty \le M\|x\|_\infty$, $\forall x \in \mathbb{R}^n$ where $M = \max\limits_i \sum\limits_{j=1}^{n} |a_{ij}|$ and $a_{ij} = [A]_{ij}$.*

Proof. $\|Ax\|_\infty = \max\limits_i |\sum\limits_{j=1}^{n} a_{ij}x_j| \le \max\limits_i \sum\limits_{j=1}^{n} |a_{ij}x_j| \le \max\limits_i \sum\limits_{j=1}^{n} |a_{ij}|\|x_j\|_\infty$.[6] □

Proposition 3.5. *Any mapping $x \to Ax$ where A is a matrix is uniformly continuous.*

Proof. Using Prop. 3.2, $\forall\epsilon > 0$ use $\delta = \frac{\epsilon}{M}$. Then when $\|x-y\| < \delta = \frac{\epsilon}{M}$ we have $\|Ax-Ay\| = \|A(x-y)\| \le M\|x-y\| < \epsilon$. □

Proposition 3.6. *If f is differentiable at a then it is continuous at a.*

Proof. It can be shown through a rearranging of $\|f(x) - f(a)\|$:

$$\begin{aligned} 0 \le \|f(x) - f(a)\| &= \|f(x) - f(a) - f'(a)(x-a) + f'(a)(x-a)\| \\ &\le \|f(x) - f(a) - f'(a)(x-a)\| + \|f'(a)(x-a)\| \\ &= \underbrace{\frac{\|f(x) - f(a) - f'(a)(x-a)\|}{\|x-a\|} \cdot \|x-a\|}_{\to 0 \text{ by definition}} + \|f'(a)(x-a)\| \end{aligned}$$

Since the limit of R.H.S (right-hand side) is 0, $\lim\limits_{x\to a} f(x) = f(a)$ by squeeze theorem. □

Proposition 3.7. *Consider $f : U \subset \mathbb{R}^n \to \mathbb{R}^m$. If all partial derivatives $\frac{\partial f_i}{\partial x_j}$ are continuous at a, then f is differentiable at a.*

Proof. Consider $f : \mathbb{R}^2 \to \mathbb{R}$.

$$\begin{aligned} R(x) &= f(x) - L_a(x) \\ &= f(x) - f(a) - \frac{\partial f}{\partial x_1}(a)(x_1 - a_1) - \frac{\partial f}{\partial x_2}(a)(x_2 - a_2) \\ &= [f(x_1, x_2) - f(a_1, x_2)] - \frac{\partial f}{\partial x_1}(a)(x_1 - a_1) + [f(x_1, a_2) - f(a_1, a_2)] - \frac{\partial f}{\partial x_2}(a)(x_2 - a_2) \end{aligned}$$

which implies that

$$f(x_1, x_2) - f(a_1, x_2) = \frac{\partial f}{\partial x_1}(c_1, x_2)(x_1 - a_1)$$

for some c_1 (assuming w.l.g (without loss of generality) $a_1 \le x_1$ so $a_1 \le c_1 \le x_1$) by the mean value theorem for single variable calculus. As $x_1 \to a_1$, $c_1 \to a_1$ and since $\frac{\partial f}{\partial x_1}$ is continuous

$$\lim_{x\to a} \frac{\partial f}{\partial x_1}(c_1, x_2) = \frac{\partial f}{\partial x_1}(a_1, a_2).$$

Similarly for c_2,

$$f(a_1, x_2) - f(a_1, a_2) = \frac{\partial f}{\partial x_2}(a_1, c_2)(x_2 - a_2) \implies \lim_{x\to a} \frac{\partial f}{\partial x_2}(a_1, c_2) = \frac{\partial f}{\partial x_2}(a_1, a_2).$$

[6] The proof using the 2-norm can be found in H+W (P. 293).

So

$$\begin{aligned} R(x) &= [f(x_1,x_2) - f(a_1,x_2)] - \frac{\partial f}{\partial x_1}(a)(x_1-a_1) + [f(x_1,a_2) - f(a_1,a_2)] - \frac{\partial f}{\partial x_2}(a)(x_2-a_2) \\ &= \left[\frac{\partial f}{\partial x_1}(c_1,x_2) - \frac{\partial f}{\partial x_1}(a_1,a_2)\right](x_1-a_1) + \left[\frac{\partial f}{\partial x_2}(a_1,c_2) - \frac{\partial f}{\partial x_2}(a_1,a_2)\right](x_2-a_2) \end{aligned}$$

So

$$\frac{|R(x)|}{\|x-a\|} \le \left|\frac{\partial f}{\partial x_1}(c_1,x_2) - \frac{\partial f}{\partial x_1}(a_1,a_2)\right| \frac{|x_1-a_1|}{\|x-a\|} + \left|\frac{\partial f}{\partial x_2}(a_1,c_2) - \frac{\partial f}{\partial x_2}(a_1,a_2)\right| \frac{|x_2-a_2|}{\|x-a\|}$$

and since the R.H.S$= 0$, by continuity of the partials, $\lim\limits_{x\to a} \frac{|R(x)|}{\|x-a\|} = 0$ and f is differentiable at a. □

You can see a summary of what we know about differentiability below:

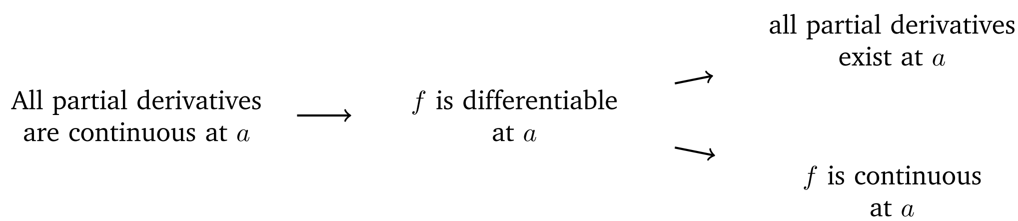

Figure 3.1: Differentiability Theorems

3.3 Geometry

Proposition 3.8. *Let $U \subset \mathbb{R}^n$, $a \in intU$ and $f : U \to \mathbb{R}$ be differentiable at a. Then the following hold true.*

1. The vector $(\nabla f(a), -1)$ is orthogonal at the tangent hyperplane of the graph $x_{n+1} = f(x)$ at $(a, f(a))$.
2. $D_u f(a) = \nabla f(a) \cdot u$.
3. If $\nabla f(a) \neq 0$ then $D_u f(a)$ has a maximum at $u = \frac{\nabla f(a)}{\|\nabla f(a)\|}$.

Proof. (1) By the linear approximation of the tangent hyperplane,

$$x_{n+1} = f(a) + \nabla f(a) \cdot (x-a)$$

which implies

$$(\nabla f(a), -1) \cdot (x-a, x_{n+1} - f(a)) = 0.$$

If a point $(x, x_{n+1}) \in$ hyperplane, then $(x-a, x_{n+1} - f(a))$ is a vector in the hyperplane. Thus, the vector $(\nabla f(a), -1)$ is orthogonal to the tangent hyperplane.

(2) Choose any u, $\|u\| = 1$. By the definition of differentiability,

$$\begin{aligned} 0 &= \lim_{t\to 0} \frac{|f(a+tu) - f(a) - \nabla f(a)\cdot(tu)|}{t} \\ &= \lim_{t\to 0} \left|\frac{f(a+tu) - f(a)}{t} - \nabla f(a)\cdot(u)\right|. \end{aligned}$$

But $D_u f(a) = \lim\limits_{t\to 0} \frac{f(a+tu)-f(a)}{t}$. So $D_u f(a) = \nabla f(a) \cdot (u)$.

(3) $D_u f(a) = \nabla f(a) \cdot (u) = A \cdot B$ for some matrices A and B. This is largest if $u = \frac{1}{\|\nabla f(a)\|}\nabla f(a)$ because $|D_u f(a)| \le \|\nabla f(a)\| \|u\|$. □

3.4 Rules of Differentiation

(In lectures, an example similar to Example 3.2 was done here, so I will exclude it from these notes. The function in question was $f(x_1, x_2) = x_1^3 x_2^{\frac{1}{3}}$ so determining differentiability will be left as an exercise)

Theorem 3.1. *(Chain Rule)*

Let $A \subset \mathbb{R}^n$,$B \subset \mathbb{R}^m$, and $g : A \to B$, $f : B \to \mathbb{R}^l$. If g is differentiable at $a \in$ int$A$ and f is differentiable at $b \in$ int$B$, then $h = f(g(x)) = (f \circ g)(x)$ is differentiable at a with

$$h'(x) = f'(g(x))g'(x)$$

Proof. For $p \in \mathbb{R}^n$ and $q \in \mathbb{R}^m$, choose p, q such that $\|p\|$ and $\|q\|$ are sufficiently small. We know that

$$g(a+p) = b + g'(a) \cdot p + R_g(p)$$

and

$$f(a+q) = f(b) + f'(b) \cdot q + R_f(q)$$

where $R(x)$ is some error. By the differentiability of g and f, we know that

$$\lim_{\|p\| \to 0} \frac{\|R_g(p)\|}{\|p\|} = 0 \implies R_g(0) = 0$$

and similarly

$$\lim_{\|q\| \to 0} \frac{\|R_f(q)\|}{\|q\|} = 0 \implies R_f(0) = 0.$$

So

$$\begin{aligned} h(a+p) &= f(g(a+p)) \\ &= f(b + \underbrace{g'(a) \cdot p + R_g(p)}_{q = g(a+p) - g(a)}) \\ &= f(b+q) \\ &= f(b) + f'(b) \cdot q + R_f(q) \end{aligned}$$

and this implies that

$$h(a+p) = \underbrace{f(b) + \overbrace{f'(g(a))g'(a)}^{h'(a)}}_{h(a)} \cdot p + \underbrace{f'(b)R_g(p) + R_f(q)}_{R_h(p)}.$$

Thus, all we need to show is that

$$\lim_{\|p\| \to 0} \frac{\|R_h(p)\|}{\|p\|} = 0$$

by showing

$$(1) \lim_{\|p\| \to 0} \frac{\|f'(b)R_g(p)\|}{\|p\|} = 0 \text{ and } (2) \lim_{\|p\| \to 0} \frac{\|R_f(q)\|}{\|p\|} = 0.$$

The proofs for (1) and (2) can be found in Wade. □

Remark 3.2. Note that in the chain rule proof, we are generalizing differentiability in the directional derivative sense,

$$(1) \lim_{h \to 0} \frac{\|f(a+hu) - f(a) - f'(a)hu\|}{|h|}$$

into a stronger statement,

$$(2)\ \lim_{\|p\|\to 0} \frac{\|f(a+hp)-f(a)-f'(a)p\|}{\|p\|}.$$

So, in other words, $(2) \implies (1)$.

Example 3.3. In this example, we will investigate the case when $D_u f(a) = v \cdot u$, for all u and some v. So suppose that

$$f(x_1,x_2) = \begin{cases} \frac{x_1^3 x_2}{x_1^6+x_2^6} & (x_1,x_2)\neq 0 \\ 0 & (x_1,x_2)=0 \end{cases}$$

(we will check for continuity and differentiability) and so

$$\begin{aligned} \lim_{h\to 0} \frac{f(hu_1,hu_2)-f(0,0)}{h} &= \lim_{h\to 0} \frac{h^3u_1^3(hu_2)}{h(h^6u_1^6+h^2u_2^6)} \\ &= \lim_{h\to 0} \frac{hu_1^3u_2}{h^4u_1^6+u_2^2}. \end{aligned}$$

Now if $u_2 = 0 \implies D_{(1,0)}f(0,0) = 0$ and if $u_2 \neq 0 \implies D_u f(0,0) = \frac{0}{u_2} = 0$. We have $D_u f(0,0) = 0, \forall u$ and $f_x = 0$, $f_y = 0 \implies \nabla f(0) = 0$. Thus,

$$D_u f(0,0) = \begin{bmatrix} 0 & 0 \end{bmatrix} u$$

but

$$\lim_{x_1\to 0} f(x_1,x_1^3) = \lim_{x_1\to 0} \frac{x_1^6}{x_1^6+x_1^6} = \frac{1}{2}$$

and

$$\lim_{x_1\to 0} f(x_1,0) = \lim_{x_1\to 0} \frac{0}{x_1^3} = 0.$$

Thus, f is not continuous at 0 and hence not differentiable.

Suppose that $f(x_1,x_2) = (x_1^2+1, x_2^2+1)$, $g(u_1,u_2) = u_1+u_2$. What is the linear approximation of $g\circ f$ at $(1,1)$? We know that $g(f(1,1)) = 3$. Next, by the chain rule,

$$\begin{aligned} \left.\frac{\partial(g\circ f)}{\partial x_1}\right|_{(1,1)} &= \left.\left(\frac{\partial g}{\partial u_1}\frac{\partial u_1}{\partial x_1} + \frac{\partial g}{\partial u_2}\frac{\partial u_2}{\partial x_1}\right)\right|_{(1,1)} = \left.(2x_1+x_2^2)\right|_{(1,1)} = 3 \\ \left.\frac{\partial(g\circ f)}{\partial x_2}\right|_{(1,1)} &= \left.\left(\frac{\partial g}{\partial u_1}\frac{\partial u_1}{\partial x_2} + \frac{\partial g}{\partial u_2}\frac{\partial u_2}{\partial x_2}\right)\right|_{(1,1)} = \left.(2x_1x_2)\right|_{(1,1)} = 2 \end{aligned}$$

and so

$$L(x_1,x_2) = 3 + 3(x_1-1) + 2(x_2-1).$$

Note that

$$\begin{aligned} \left.\frac{\partial g}{\partial u_1}\right|_{u_2} &= 1 \quad \text{(holding } u_2 \text{constant)} \\ \left.\frac{\partial f}{\partial x_1}\right|_{x_2} &= 2x_1+x_2^2 \quad \text{(holding } x_2 \text{constant)} \end{aligned}$$

Example 3.4. Let $f(x,y,z) = e^x yz^2$, $x = r\cos z$ and $y = r\sin z$. Find $\left.\frac{\partial f}{\partial z}\right|_{x,y}$ and $\left.\frac{\partial f}{\partial z}\right|_{r}$. Using the chain rule on the second,

$$\begin{aligned} \left.\frac{\partial f}{\partial z}\right|_{r} &= \frac{\partial f}{\partial x}\frac{\partial x}{\partial z} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial z} + \frac{\partial f}{\partial z}\frac{\partial z}{\partial z} \\ &= (e^x yz^2)(-r\sin z) + (e^x z^2)(r\cos z) + 2ze^x y \end{aligned}$$

and the first is just a simple evaluation

$$\left.\frac{\partial f}{\partial z}\right|_{x,y} = 2e^x yz.$$

Theorem 3.2. *(Mean Value Theorem (MVT))*

Let $f : A \subset \mathbb{R}^n \to \mathbb{R}$ be differentiable on $S \subset \text{int}A$ where $S = \{a + t(b - a), t \in (0, 1)\}$, where $a, b \in A$ and f continuous on $\bar{S}$. Then, there is $c \in S$ such that $f(b) - f(a) = \underbrace{f'(c)}_{\nabla f(c)}(b - a)$.

Proof. Define $h(t) = f(g(t))$, $g(t) = a + (b - a)t$. By the chain rule, h is differentiable with $h'(t) = f'(g(t))(b - a)$. By the mean value theorem in one dimension, for $h : \mathbb{R} \to \mathbb{R}$, there is $t_0 \in (0, 1)$ so that $h'(t_0) = \frac{h(1)-h(0)}{1-0}$. Writing $c = a + (b - a)t_0$, $h(t_0) = f'(c)(b - a) = f(b) - f(a)$ through straight substitution. □

Definition 3.5. A set is **convex** if for any $x, y \in \theta$, $x + t(y - x) \in \theta$, $\forall t \in [0, 1]$.

Corollary 3.1. *Let $\theta \subset \mathbb{R}^n$ be non-empty, open and convex. If $f : \theta \to \mathbb{R}$ is differentiable on θ with $f'(x) = 0$, $\forall x \in \theta$, then f is constant on θ.*

Proof. Choose any $x, y \in \theta$, $x \neq y$. Since θ is convex, there is a line, in θ, connecting x to y. By the MVT (mean value theorem),

$$\begin{aligned} f(x) - f(y) &= f'(c)(x - y) \\ &= 0, \forall c \end{aligned}$$

and so $f(x) = f(y)$. Since x, y were arbitrary, then f is constant. □

Example 3.5. In this example, we check to see if the MVT can be extended into vector valued functions. Let $f(x_1, x_2) = \begin{bmatrix} x_1(x_2 - 1) \\ x_2^2(x_1 - 1) \end{bmatrix}$, $a = (0, 0)$, $b(1, 1) \implies f(a) = 0, f(b) = 0$. So

$$f'(x_1, x_2) = \begin{bmatrix} x_2 - 1 & x_1 \\ x_2^2 & 2x_2(x_1 - 1) \end{bmatrix}.$$

Consider the line a to b:

$$\begin{bmatrix} 0 \\ 0 \end{bmatrix} + t\begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} t \\ t \end{bmatrix}$$

and note that $f(b) - f(a) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$. Next,

$$f'\left(\begin{bmatrix} t \\ t \end{bmatrix}\right)(b - a) = \begin{bmatrix} t - 1 & t \\ t^2 & 2t(t - 1) \end{bmatrix}\begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 2t - 1 \\ 3t^2 - 2t \end{bmatrix}$$

and so we want t such that $2t - 1 = 0$ and $3t^2 - 2t = 0$. However, there is not solution to this system of equations and so we cannot generalize the MVT this way.

Theorem 3.3. *(Generalized Mean Value Theorem)*[a]
Let $f : U \subset \mathbb{R}^n \to \mathbb{R}^m$ be differentiable on $S \subset int\ U$ where $S = \{a + t(b-a), t \in (0,1)\}$, where $a, b \in U$ and f continuous on $\bar{S}$ and suppose that there is M such that $\|f'(x)\|_{2,2} \leq M$.[b] *Then,*

$$\|f(b) - f(a)\|_2 \leq M\|b - a\|_2$$

Proof. For any v_i, $i = 1, ..., m$ define

$$\begin{aligned} g(t) &= \sum_{i=1}^{m} v_i f_i(a + (b-a)t) = v^t f(a + (b-a)t) \\ g'(x) &= v^t f'(a + (b-a)t)(b-a). \end{aligned}$$

Apply MVT. There is $t_0 \in (0,1)$ such that $g(1) - g(0) = g'(t_0)(1-0)$. So,

$$v^t(f(b) - f(a)) = v^t f'(c)(b-a), c = a + (b-a)t_0.$$

Choose $v = f(a) - f(b)$. Then, using the dot product,

$$\begin{aligned} \|f(b) - f(a)\|_2^2 &= (f(b) - f(a))^t f'(c)(b-a) \\ &\leq \|f(b) - f(a)\|_2 \|f'(c)(b-a)\|_2 \\ &\leq \|f(b) - f(a)\|_2 M\|b-a\|_2 \end{aligned}$$

and so $\|f(b) - f(a)\|_2 \leq M\|(b-a)\|_2$. □

[a] See also H+W, IV 3.7
[b] $\|f'(a)\|_{2,2} \leq M$ means $\|f'(a)y\|_2 \leq M\|y\|_2, \forall y$

3.5 Implicit Functions

In this section, we are interested in: given a differentiable function f, when is $f(x, y) = 0$ the graph of the a differentiable function $y = g(x)$?

Theorem 3.4. *(Implicit Function Theorem)*

Consider a point (a, b) and $f : \mathbb{R}^2 \to \mathbb{R}$. If $f(a, b) = 0$, $f_y(a, b) \neq 0$ and f has continuous partial derivatives in a neighbourhood of (a, b), then there is a neighbourhood of (a, b) in which $f(x, y) = 0$ has a unique solution for y in terms of x :$y = g(x)$. Moreover, g has a continuous partial derivative at a.

Proof. (From H+W) Assume w.l.g. that $f_y(a, b) > 0$. Since $f_y$is continuous at (a, b), there is a ball B of radius ϵ around (a, b) so

$$f_y(x, y) > 0, (x, y) \in B.$$

This means that $f(a, y)$ is an increasing function of y for small enough ϵ. Since f is continuous, then there is a $\delta > 0$ so $|x - a| < \delta$ implies

$$f(x, b - \epsilon) < 0 < f(x, b + \epsilon).$$

For each $x \in [a - \delta, a + \delta]$ apply IVT (Intermediate Value Theorem) to $f(x, y)$, considered as a function of x. This yields $g(x) = y$ with $f(x, g(x)) = 0$, since f is a monotonic function of of y. This therefore defines a function $g(x)$ with $f(x, g(x)) = 0$. Choose some x near a. For small enough $|h|$,

$$\begin{aligned} 0 &= f(x + h, g(x + h)) - f(x, g(x)) \\ &= [f((x + h), g(x + h)) - f(x + h, g(x))] + [f((x + h), g(x)) - f(x, g(x))]. \end{aligned}$$

By 1-variable MVT, $\exists c_1$ between $g(x)$ and $g(x + h)$ and $\exists c_2$ between x and $x + h$ such that

$$0 = f_y(x + h, c_1)(g(x + h) - g(x)) + f_x(c_2, g(x))h.$$

Rearrange.

$$\frac{g(x+h)-g(x)}{h} = \frac{-f_x(c_2,g(x))}{f_y(x+h,c_1)}, h \neq 0.$$

As $h \to 0$, $c_2 \to x$ and since f_x is continuous, $f_x(c_2,g(x)) \to f_x(x,g(x))$. Similarly, as $h \to 0$, $c_2 \to x$ and $f_y(x+h,c_1) \to f_y(x,g(x))$. So

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

□

Exercise 3.2. Show that g is continuous. (hint: use the fact that both $|f_x|$ and $|f_y|$ are bounded)

Exercise 3.3. Generalize the above theorem for a vector valued function f.

3.6 Locally Invertible Functions

Definition 3.6. We define the set of all functions with continuous partial derivatives as

$$\mathcal{C}^1(U,\mathbb{R}^m) = \{f : U \subset \mathbb{R}^n \to \mathbb{R}^m | U \neq 0\}$$

Definition 3.7. Let $f \in \mathcal{C}^1(U,\mathbb{R}^m)$. The function f is said to be **locally injective** at $x_0 \in U$ if there is a ball $\mathcal{B}_r(x_0)$, $r > 0$ such that f is **injective** (one-to-one) on $\mathcal{B}_r(x_0) \cap U$. [7]

Lemma 3.1. *Let $f \in \mathcal{C}^1(U,\mathbb{R}^m)$ where $U \subset \mathbb{R}^n$, and U is open such that $\det(f'(\underline{a})) \neq 0$ at $\underline{a} \in U$[8]. Then, the following hold true:*

(1) There is a neighbourhood $\mathcal{B}$ of $\underline{a}$ so that $\det(f'(c)) \neq 0$ for all $\underline{c} \in \mathcal{B}$.

(2) f is locally injective at $\underline{a}$.

Proof. (1) Define $h : U^n \to \mathbb{R}$ as

$$h(\underline{c}_1,\underline{c}_2,...,\underline{c}_n) = \det \begin{bmatrix} \frac{\partial f_1}{x_1}(\underline{c}_1) & \cdots & \frac{\partial f_1}{x_n}(\underline{c}_1) \\ \vdots & \ddots & \vdots \\ \frac{\partial f_n}{x_1}(\underline{c}_n) & \cdots & \frac{\partial f_n}{x_n}(\underline{c}_n) \end{bmatrix}$$

with $U^n = \{(\underline{x}_1,\underline{x}_2,...,\underline{x}_n) | x_i \in U\}$. Now since h is continuous on U^n,

$$h(\underline{a},\underline{a},...,\underline{a}) = \det(f'(\underline{a})).$$

If $\det(f'(\underline{a})) \neq 0$, then there is $r > 0$ so,

$$h(\underline{x}_1,\underline{x}_2,...,\underline{x}_n) \neq 0, \underline{x}_i \in \mathcal{B}_r(\underline{a})$$

Since $h(\underline{c}_1,\underline{c}_2,...,\underline{c}_n) = \det(f'(\underline{c}))$ then $\det(f'(\underline{c})) \neq 0$ for $\underline{c} \in \mathcal{B}_r(\underline{a})$.

(2) Suppose f is not locally injective at $\underline{a}$ for all $r > 0$. Then $\exists x, y \in \mathcal{B}_r(\underline{a})$, $\underline{x} \neq \underline{y}$, $f(\underline{x}) = f(\underline{y})$. By the MVT, $\exists \underline{c}_i \in \mathcal{B}_r(\underline{a})$ such that

$$0 = f_i(\underline{x}) - f_i(\underline{y}) = \begin{bmatrix} \frac{\partial f_i}{\partial x_1}(\underline{c}_i) & \cdots & \frac{\partial f_i}{\partial x_n}(\underline{c}_i) \end{bmatrix} (\underline{y}-\underline{x}).$$

Doing this for each component of f gives

$$\begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix} = \underbrace{\begin{bmatrix} \frac{\partial f_1}{\partial x_1}(\underline{c}_1) & \cdots & \frac{\partial f_1}{\partial x_n}(\underline{c}_1) \\ \vdots & \ddots & \vdots \\ \frac{\partial f_n}{\partial x_1}(\underline{c}_n) & \cdots & \frac{\partial f_n}{\partial x_n}(\underline{c}_n) \end{bmatrix}}_{M} (\underline{y}-\underline{x})$$

[7] That is, $a \neq b$ implies $f(a) \neq f(b)$.
[8] Note that $\underline{a}$ is an $n-$dimensional vector.

and note that $\det(M) = h(\underline{c}_1, \underline{c}_2, ..., \underline{c}_n)$. The existence of a non-trivial solution by fact that f is not injective implies that $h(\underline{c}_1, \underline{c}_2, ..., \underline{c}_n) = 0$, $\forall r > 0$, $\underline{c}_i \in \mathcal{B}_r(\underline{a})$. Thus, using (1), $\det(f'(\underline{a})) = 0$. □

Example 3.6. $f'(p,\theta) = \begin{bmatrix} p\cos\theta \\ p\sin\theta \end{bmatrix}$, $U = \{(p,\theta), p > 0\}$. Note that $f(p,\theta) = f(p, \theta + 2\pi)$. Thus,

$$f'(p,\theta) = \begin{bmatrix} \cos\theta & -p\sin\theta \\ \sin\theta & p\cos\theta \end{bmatrix} \implies \det(f'(p,\theta)) = p \neq 0$$

Proposition 3.9. *Let $f \in \mathcal{C}^1(U, \mathbb{R}^m)$, $U \subset \mathbb{R}^n$, U open and $\det(f'(\underline{x})) \neq 0$ for $\underline{x} \in U$. Then $f(U)$ is open.*

Proof. Choose $y_0 \in f(U)$ and let $f(x_0) = y_0$, $x_0 \in U$. From the above lemma, $\exists \delta > 0$ so $\overline{\mathcal{B}}_\delta(x_0) \subset U$ where f is injective on $\overline{\mathcal{B}}_\delta(x_0)$. Since $f(\partial\overline{\mathcal{B}}_\delta(x_0))$ is compact where

$$\partial\overline{\mathcal{B}}_\delta(x_0) = \{x, \|x - x_0\| = \delta\}$$

it doesn't include y_0. Next, define

$$\epsilon = \frac{1}{3}\inf\{\|y_0 - f(x)\|_2, x \in \partial\overline{\mathcal{B}}_\delta(x_0)\} > 0.$$

We will show that $\mathcal{B}_\epsilon(y_0) \subset U$. Pick $y \in \mathcal{B}_\epsilon(y_0)$ and define

$$g : \overline{\mathcal{B}}_\delta(x_0) \to \mathbb{R}, g(x) = \|f(x) - y\|^2.$$

g is continuous and contains a minimum $\tilde{x}$, by the EVT. Suppose that $\tilde{x} \in \partial\overline{\mathcal{B}}_\delta(x_0)$ which implies

$$\begin{aligned} \sqrt{g(\tilde{x})} &= \|f(\tilde{x}) - y\| \\ &\geq \|f(\tilde{x}) - y_0\| - \|y_0 - y\| \\ &\geq 3\epsilon - \epsilon \\ &= 2\epsilon \\ &> \epsilon \\ &\geq \|f(x_0) - y\| = \sqrt{g(x_0)} \end{aligned}$$

and so we have a contradiction ($g(\tilde{x})$ is not the minimum of g). Thus, $\tilde{x} \in \mathcal{B}_\delta(x_0)$ and $g(\tilde{x})$ is minimum of g on $\mathcal{B}_\delta(x_0)$. Using some one-dimensional calculus rules (critical points),

$$\frac{\partial g}{\partial x_k}(\tilde{x}) = 0, k = 0, 1, ..., n$$

and so

$$\begin{aligned} g(x) = \sum_{j=1}^{n}(f_j(x) - y_j)^2 \quad &\implies \quad g'(x) = 2\sum_{j=1}^{n}(f_j(x) - y_j) \cdot \frac{\partial g}{\partial x_k}(\tilde{x}) = 0 \\ &\implies \begin{bmatrix} f_1(\tilde{x}) - y_1 & \cdots & f_n(\tilde{x}) - y_n \end{bmatrix} \begin{bmatrix} \frac{\partial f_1}{\partial x_1}(\tilde{x}) & \cdots & \frac{\partial f_1}{\partial x_n}(\tilde{x}) \\ \cdots & \ddots & \vdots \\ \frac{\partial f_n}{\partial x_1}(\tilde{x}) & \cdots & \frac{\partial f_n}{\partial x_n}(\tilde{x}) \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}. \end{aligned}$$

Since $\det(f'(\tilde{x})) \neq 0$, we obtain only the trivial solution $f(\tilde{x}) = y$. Thus, $y \in \mathcal{B}_\delta(x_0) \subset f(U)$. □

From here we deduce a few interesting propositions about the locally invertible function f^{-1}.

Proposition 3.10. *Let $K \subset \mathbb{R}^n$ be compact, non-empty and $f : K \to \mathbb{R}^m$ be injective and continuous. Then, $f^{-1} : f(K) \to K$ is continuous.*

Proof. Suppose f^{-1} is not continuous. Then, there is $\{y_n\} \in f(K)$, so $\lim_{n\to\infty} y_n = y_0$ but $\lim_{n\to\infty} x_n \neq x_0$ where $f^{-1}(y_n) = x_n$ and $f^{-1}(y_0) = x_0$. This means for $\epsilon_0 > 0$, the subsequence $\{x_{n_k}\}$ is such that $\|x_{n_k} - x_0\| \geq \epsilon_0$ for all k. Since K is compact, there is also a subsequence, also called $\{x_{nk}\}$ such that $\lim_{n\to\infty} x_{n_k} = \tilde{x}$. Thus

$$y_0 = \lim_{n\to\infty} y_{n_k} = \lim_{n\to\infty} f(x_{n_k}) = f(\lim_{n\to\infty} x_{n_k}) = f(\tilde{x})$$

but $y_0 = f(x_0)$ so $f(x_0) = f(\tilde{x})$ and f is not injective. □

Theorem 3.5. *(Inverse Function Theorem)*
Let $f \in \mathcal{C}^1(U, \mathbb{R}^m)$ where $U \subset \mathbb{R}^n$ is open. If for $a \in U$, $\det f'(a) \neq 0$, then there is an open set B containing a so that

- f is injective on B
- f^{-1} is $\mathcal{C}^1$ on $f(B)$
- For each $f \in f(B)$, $(f^{-1})'(y) = [f'(x)]^{-1}$

Proof. By Lemma 3.5. we already have show that there is an open ball B centered at a so that f is injective on B, $\det f'(x) \neq 0$, $\forall x \in B$, and f^{-1} is defined on B. Next, we show that f is differentiable on B.

Pick $x_0 \in B$ and define

$$g : B \to \mathbb{R}^n, g(x) = \begin{cases} \frac{f(x)-f(x_0)-f'(x_0)(x-x_0)}{\|x-x_0\|} & x \neq x_0 \\ 0 & x = x_0 \end{cases}$$

so g is continuous on B. Since $\det f'(x_0) \neq 0$,

$$\|x - x_0\| f'(x_0)^{-1} g(x) \quad = \quad f'(x_0)^{-1}(f(x) - f(x_0)) - (x - x_0), \forall x \in \mathbb{R}^n \tag{3.1}$$

So considering $\|x\|$ with the above we get

$$\begin{aligned} \|x\| = \|f'(x_0)^{-1} f'(x_0)x\| \leq \|f'(x_0)^{-1}\|_{2,2} \|f'(x_0)x\| \quad &\implies \quad \frac{1}{\|f'(x_0)^{-1}\|_{2,2}} \|x\| \leq \|f'(x_0)\| \\ &\implies \quad 2C\|x\| \leq \|f'(x_0)x\| \end{aligned}$$

where $C = \frac{1}{2\|f'(x_0)^{-1}\|_{2,2}}$. Choose $\epsilon > 0$, $\mathcal{B}_\epsilon(x_0) \subset B$ so

$$\|f(x) - f(x_0) - f'(x_0)(x - x_0)\| = \|x - x_0\| \|g(x)\| \leq C\|x - x_0\|$$

and note that we can find such an ϵ since g is continuous ($\|x - x_0\| < \epsilon \implies \|g(x)\| \leq C$).

For $x \in \mathcal{B}_\epsilon(x_0)$,

$$\begin{aligned} C\|x - x_0\| \quad &\geq \quad \|f(x) - f(x_0) - f'(x_0)(x - x_0)\| \\ &\geq \quad \|f'(x_0)(x - x_0)\| - \|f(x) - f(x_0)\| \\ &\geq \quad 2C\|x - x_0\| - \|f(x) - f(x_0)\| \end{aligned}$$

and so

$$\|f(x) - f(x_0)\| \geq C\|x - x_0\|. \tag{3.2}$$

Next, consider $\frac{1}{C}\|y - y_0\| \|f'(x_0)^{-1} g(x)\|$, taking note that $y_0 = f(x_0)$ and $y = f(x)$

$$\begin{aligned} \frac{1}{C}\|y - y_0\| \|f'(x_0)^{-1} g(x)\| \quad &= \quad \frac{1}{C}\|f(x) - f(x_0)\| \|f'(x_0)^{-1} g(x)\| \\ &\geq \quad \underbrace{\|x - x_0\|}_{(3.1)} \|f^{-1}(x_0) g(x)\| \\ &= \quad \| \underbrace{f^{-1}(x_0)(f(x) - f(x_0)) - (x - x_0))}_{(3.1)} \|. \end{aligned}$$

Next, taking note that $x = f^{-1}(y)$ and $x_0 = f^{-1}(y_0)$, we can rearrange the above to get the following

$$\begin{aligned}\frac{\|f^{-1}(x_0)\left(f(f^{-1}(y)) - f(f^{-1}(y_0))\right) - (f^{-1}(y) - f^{-1}(y_0))\|}{\|y - y_0\|} &= \frac{\|f^{-1}(y) - f^{-1}(y_0) - f^{-1}(x_0)(y - y_0)\|}{\|y - y_0\|} \\ &\leq \frac{1}{C}\|f'(x_0)^{-1}g(x)\|\end{aligned}$$

and since f^{-1} is continuous at x_0, then as $x \to x_0$, $y \to y_0$ and so $\lim_{y \to y_0} g(x) = 0$ by the composition continuity theorem. By the squeeze theorem, the limit of the left hand side is 0 and f^{-1} is differentiable. That is

$$(f^{-1})'(y_0) = [f'(x_0)]^{-1}, x_0 = f^{-1}(y_0)$$

or

$$(Df^{-1})(y_0) = [Df(x_0)]^{-1}$$

□

Remark 3.3. If f^{-1} is differentiable at $f(a) = b$, then

$$\begin{aligned} I = (f^{-1} \circ f)(a) &\implies I = (f^{-1})'(f(a))f(a) \\ &\implies 1 = \det\left[(f^{-1})(b)\right] \det\left[f'(a)\right]\end{aligned}$$

meaning that $\det f'(a) \neq 0$. The converse of the above, under a couple of other conditions is the inverse function theorem.

Exercise 3.4. Show that the inverse function theorem is true if and only if the implicit function theorem is true.

3.7 Non-Linear Approximations

Recall that Taylor's theorem in one variable calculus states that if $f \in \mathcal{C}^2(I)$ [9], for every $a, x \in I$, there is a c between x and a so

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

We will try to generalize this theorem in a multivariable setting.

Proposition 3.11. *If $f \in \mathcal{C}^2(U)$, then $f \in \mathcal{C}^1(U)$.*

Proof. Since $g_i = \frac{\partial f}{\partial x_i}$ continuous partials $\frac{\partial g_i}{\partial x_j}$ for all combinations i and j, then g_i is differentiable. Since g_i is differentiable, it is also continuous for all i. □

Proposition 3.12. *Consider $f : U \subset \mathbb{R}^2 \to \mathbb{R}$ where U is open. If $\frac{\partial^2 f}{\partial x \partial y}$ and $\frac{\partial^2 f}{\partial y \partial x}$ exist in a neighbourhood of $a \in U$ and are continuous at a, then*

$$\frac{\partial^2 f}{\partial x \partial y}(a) = \frac{\partial^2 f}{\partial y \partial x}(a)$$

Proof. (See H+W, Thm. IV.4.3) □

Definition 3.8. We define the second degree **Taylor polynomial** of a function $f : \mathbb{R}^2 \to \mathbb{R}$ as the following

$$P_2(x) = f(a) + f'(a)(x - a) + A(x_1 - a_1) + B(x_1 - a_1)(x_2 - a_2) + C(x_2 - a_2)^2$$

where

$$P_2(a) = f(a), \frac{\partial P_2}{\partial x_1}(a) = \frac{\partial f}{\partial x_1}(a), \frac{\partial P_2}{\partial x_2}(a) = \frac{\partial f}{\partial x_2}(a)$$
$$\frac{\partial^2 P_2}{\partial x_1^2}(a) = 2A = \frac{\partial^2 f}{\partial x_1^2}(a), \frac{\partial^2 P_2}{\partial x_2^2}(a) = 2C = \frac{\partial^2 f}{\partial x_2^2}(a)$$

[9] $\mathcal{C}^n(U) = \mathcal{C}^n(U, \mathbb{R})$

$$\frac{\partial^2 P_2}{\partial x_1 \partial x_2}(a) = \frac{\partial^2 P_2}{\partial x_2 \partial x_1}(a) = B = \frac{\partial^2 f}{\partial x_2 \partial x_1}(a) = \frac{\partial^2 f}{\partial x_1 \partial x_2}(a)$$

Definition 3.9. We define the **Hessian** of $f : V \subset \mathbb{R}^n \to \mathbb{R}$ at a point $a \in \mathbb{R}^n$ to be

$$H_f(a) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1 \partial x_1}(a) & \frac{\partial^2 f}{\partial x_1 \partial x_2}(a) & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n}(a) \\ \frac{\partial^2 f}{\partial x_2 \partial x_1}(a) & \frac{\partial^2 f}{\partial x_2 \partial x_2}(a) & & \frac{\partial^2 f}{\partial x_2 \partial x_n}(a) \\ \vdots & & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1}(a) & \frac{\partial^2 f}{\partial x_n \partial x_2}(a) & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_n}(a) \end{bmatrix}$$

Thus, another way to write our second degree Taylor polynomial is

$$P_2(x) = f(a) + f'(a)(x-a) + \frac{1}{2}(x-a)^t (H_f(a))(x-a)$$

Theorem 3.6. *(Generalized Taylor's Theorem)*

Consider $f : V \subset \mathbb{R}^n \to \mathbb{R}$ where V is open and convex. If $f \in \mathcal{C}^2(V)$, then for any $a, x \in V$, there is c on the line joining x to a so that

$$f(x) = \underbrace{f(a) + f'(a)(x-a)}_{L(x)} + \frac{1}{2}(x-a)^t (H_f(c))(x-a)$$

Proof. Define $\phi(t) = a + t(x-a)$, $0 \le t \le 1$. Next, define $g(t) = f(\phi(t))$. So,

$$\begin{aligned} g'(t) &= f'(\phi(t)) \cdot \phi'(t) \\ &= f'(\phi(t))(x-a) \\ &= f_{x_1}(\phi(t))(x_1 - a_1) + \ldots + f_{x_n}(\phi(t))(x_n - a_n) \end{aligned}$$

and

$$\begin{aligned} g''(t) &= \sum_{1 \le i,j \le n} f_{x_i x_j}(\phi(t))(x_i - a_i)(x_j - a_j) \\ &= (x-a)^t H_f(\phi(t))(x-a). \end{aligned}$$

Using Taylor's theorem on g, we get

$$g(t) = g(t_0) + g'(t_0)(t - t_0) + \frac{1}{2} f''(\alpha)(t - t_0)^2$$

for some α between t and t_0. Setting $t = 1$ and $t_0 = 0$ we have

$$g(1) = g(0) + g'(0) + \frac{1}{2} g''(\alpha).$$

So $\phi(1) = x$, $\phi(0) = a$ and $c = \phi(\alpha)$. Thus

$$f(x) = f(a) + f'(a)(x-a) + \frac{1}{2}(x-a)^t H_f(c)(x-a).$$

□

Example 3.7. Compute the linear and second order Taylor approximations of $f(x, y) = e^x \cos y$ at $(0, 0)$. Show

$$|f(x) - L(x)| \le e\|x\|_2^2, \ \forall x, \|x\| \le 1$$

Solution. By observation, we may note that $f(0,0)=1$, $f_x=e^x\cos y \implies f_x(0,0)=1$, $f_y=-e^x\sin y \implies f_y(0,0)=0$. So

$$L(x)=f(0,0)+f'(0,0)\begin{bmatrix} x-0 \\ y-0 \end{bmatrix}=1+x$$

Next, $f_{xx}=e^x\cos y \implies f_{xx}(0,0)=1$, $f_{xy}=f_{yx}=-e^x\sin y = \implies f_{xy}(0,0)=0$, $f_{yy}=-e^x\cos y \implies f_{yy}(0,0)=-1$. So,

$$P_2(x,y)=L(x,y)+\frac{1}{2}\begin{pmatrix} x & y \end{pmatrix}\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}\begin{pmatrix} x \\ y \end{pmatrix}=1+x+\frac{1}{2}x^2-\frac{1}{2}y^2.$$

From Taylor's theorem,

$$\begin{aligned} f(x,y)-L(x,y) &= \frac{1}{2}\begin{pmatrix} x & y \end{pmatrix}\begin{bmatrix} f_{xx}(c) & f_{xy}(c) \\ f_{yx}(c) & f_{yy}(c) \end{bmatrix}\begin{pmatrix} x \\ y \end{pmatrix} \\ &= \frac{1}{2}\left[f_{xx}(c)(x^2)+2f_{xy}(c)(xy)+f_{yy}(c)(y^2)\right] \end{aligned}$$

for some $c\in\mathbb{R}^2$.

By the triangle inequality,

$$\begin{aligned} |f(x,y)-L(x,y)| &\le \frac{1}{2}\left|f_{xx}(c)(x^2)+2f_{xy}(c)(xy)+f_{yy}(c)(y^2)\right| \\ &\le \frac{1}{2}\left|ex^2+2e|x||y|+ey^2\right| \\ &\le \frac{1}{2}\left|ex^2+e|x|^2+e|y|^2+ey^2\right| \\ &= e(x^2+y^2) \end{aligned}$$

and so

$$|f(x,y)-L(x,y)|\le e\|(x,y)\|_2^2$$

Higher Order Taylor Polynomials

An extension into $\mathbb{R}^3$ of the second order Taylor polynomials can be proven in a similar fashion as above. However the notation does get messy and we will just leave it as an exercise to the reader. The general form should be as follows.

Suppose $f\in\mathcal{C}^3(V)$, V convex and open, $V\subset\mathbb{R}^2$. Then,

$$f(x,y)=f(a,b)+f'(a,b)h+\frac{1}{2}h^tH_f(a,b)h+R_3,\, h=\begin{bmatrix} x-a \\ y-b \end{bmatrix}$$

and

$$R_3=\frac{1}{3!}\left[\sum_{i,j,k\in\{x,y\}}\left[f_{ijk}(c)\right](i-q(i))(j-q(j))(k-q(k))\right]$$

where

$$q(z)=\begin{cases} a & if\ z=x \\ b & if\ z=y \end{cases}$$

4 Optimization in $\mathbb{R}^n$

Notes have been provided in class and a copy can be found here.

5 Integral Multivariate Calculus

We first begin by review some basic concepts from Math 148.

5.1 Integration in $\mathbb{R}$

Definition 5.1. We define a **partition** or a **division** over an interval $[a, b]$ as $D = \{a = x_0, x_1, ..., x_{n-1}, x_n = b\}$ with $a = x_0 < x_1 < ... < x_{n-1} < x_n = b$. We say D' is a **refinement** of D if $D' \supset D$ and $D' \neq D$.

Definition 5.2. We define the upper and lower **Darboux Sums**, $S(D)$ and $s(D)$ respectively, of a bounded function $f : [a, b] \to \mathbb{R}$ on a division

$$D = \{a = x_0, x_1, ..., x_{n-1}, x_n = b\}$$

as

$$S(D) = \sum_{i=1}^{n} F_i\delta_i,\ s(D) = \sum_{i=1}^{n} f_i\delta_i$$

where $f_i = \inf_{x_{i-1}\le x\le x_i} f(x)$, $F_i = \sup_{x_{i-1}\le x\le x_i} f(x)$ and $\delta_i = x_i - x_{i-1}$. When f_i and F_i are chosen arbitrarily on the interval $[x_{i-1}, x_i]$, we call $S(D)$ and $s(D)$ the upper and lower **Riemann Sums**, respectively.

Lemma 5.1. *Let D, D' be divisions of $[a, b]$ and $f : [a, b] \to \mathbb{R}$ a bounded function. Then*

1. $s(D) \le S(D)$
2. If D' is a refinement of D, then $s(D) \le s(D') \le S(D') \le S(D)$
3. $s(D) \le S(D')$ where D' need not be a refinement of D

Proof. Claim (1) and (2) are obvious so we move on to proving (3). Take $\tilde{D} = D \cup D'$, where we are counting parts found in both D and D' only once. $\tilde{D}$ is a refinement of D and D' so $s(D) \le s(\tilde{D}) \le S(\tilde{D}) \le S(D')$. □

We can see, from Lemma 5.1., that the definitions for $\inf_D (S(D))$ and $\sup_D (s(D))$ are well defined.

Definition 5.3. We say that a bounded function $f[a, b] \to \mathbb{R}$ is **integrable** if the upper and lower quantities, $\inf_D (S(D))$ and $\sup_D (s(D))$, are equal. If so, we write:

$$\int_a^b f(x)\, dx = \inf_D (S(D)) = \sup_D (s(D))$$

Proposition 5.1. *A bounded function $f : [a, b] \to \mathbb{R}$ is integrable iff for $\epsilon > 0$, there exists some partition D such that $S(D) - s(D) < \epsilon$.*

Proof. By definition of the integral for a given $\epsilon > 0$, $\exists D_1$, D_2 such that

$$S(D_2) - s(D_1) < \epsilon.$$

Taking $D = D_1 \cup D_2$, we have

$$s(D_1) \le s(D) \le S(D) \le S(D_2) \quad \Longrightarrow \quad S(D) - s(D) > \epsilon$$

□

Definition 5.4. We define the **norm of a division** $D = \{a = x_0, x_1, ..., x_{n-1}, x_n = b\}$ as

$$\|D\| = \max_{1\le i\le n} |x_i - x_{i-1}|$$

Theorem 5.1. *(Darboux-Reymond-Du Bois)*

An equivalent definition for intergrability is the following. Given a bounded function, $f : [a, b] \to \mathbb{R}$, f is said to be integrable iff for all $\epsilon > 0$, there exists a $\delta > 0$ such that every division D with $\|D\| < \delta$ has the property $S(D) - s(D) < \epsilon$.

Proof. Suppose f is integrable. Then, we can find a D such that $S(D) - s(D)$ from Theorem 5.1.. We can then refine D such that $\|D\| < \delta$. Suppose the converse. Given some $\epsilon > 0$, we pick a $\tilde{D}$ such that $S(\tilde{D}) - s(\tilde{D}) < \frac{\epsilon}{2}$ with $\tilde{n}$ points in the division. We pick D such that $\|D\| < \delta$. Let $D = D' \cup \tilde{D}$ and set $\delta = \frac{\epsilon}{4(\tilde{n}-1)M}$, where M is the upper bound on f. Now

$$S(D) \leq S(\tilde{D}) + (\tilde{n} - 1)\delta M$$

and

$$s(D) \geq s(\tilde{D}) - (\tilde{n} - 1)\delta M$$

because of how we have chosen $\tilde{D}$. Note that $\|D\| < \delta$ as well. [10] So evaluating $S(D) - s(D)$ directly, we get

$$\begin{aligned} S(D) - s(D) &\leq S(D') + (\tilde{n} - 1)\delta M - s(D') + (\tilde{n} - 1)\delta M \\ &\leq S(D') - s(D') + \frac{\epsilon}{2} \\ &< \epsilon. \end{aligned}$$

□

Proposition 5.2. *If f is continuous except at a finite number of points in $[a, b]$, it is integrable on $[a, b]$.*

Proof. Left as an exercise for the reader. (Hint: Use continuity) □

Proposition 5.3. *A function $f : [a, b] \to \mathbb{R}$ is also integrable on $[a, b]$ iff a sequence of divisions D_i exists such that $\|D_i\| \to 0$ and*

$$I(f) = \lim_{\|D_i\| \to 0} \sum_{i=1}^{n} f(t_i)(x_i - x_{i-1})$$

exists, where $x_{i-1} \leq t_i \leq x_i$. If so, we say that

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

Proof. See Wade Thm 5.18. □

5.2 Integration in $\mathbb{R}^n$

We now extend these concepts into $\mathbb{R}^n$ and use examples and cases in $\mathbb{R}^2$ to simplify the proofs and avoid tedious notation.

Definition 5.5. We define the **boundary** of a set A, denoted as bdy(A), as the closure of A subtract the interior of A.

Definition 5.6. We define a **rectangle** in $\mathbb{R}^2$ as $I = [a, b] \times [a, b]$. A partition $D = D_x \times D_y$ of the rectangle I is defined by $D_x = \{a = x_0, x_1, ..., x_n = b\}$ and $D_y = \{a = y_0, y_1, ..., y_n = b\}$. We denote the sub-rectangle I_{ij} as $I_{ij} = [x_{i-1}, x_i] \times [y_{j-1}, y_j]$ and its area as

$$\mu(I_{ij}) = (x_i, x_{i-1})(y_j, y_{j-1}).$$

Generalizing this notion into $\mathbb{R}^n$ is fairly easy.

Definition 5.7. In $\mathbb{R}^2$, we define the upper and lower Darboux/Riemann Sums in a similar way from Definition 5.2.. For a bounded function $f : I \to \mathbb{R}$ and partitions D (using the definition from Definition 5.6), the upper sum $S(D)$ is given by

$$S(D) = \sum_{i=1}^{n} \sum_{i=1}^{m} F_{ij} \cdot \mu(I_{ij})$$

and the lower sum $s(D)$ is given by

$$s(D) = \sum_{i=1}^{n} \sum_{i=1}^{m} f_{ij} \cdot \mu(I_{ij})$$

where $F_{ij} = \sup_{(x,y)\in I_{ij}} f(x,y)$ and $f_{ij} = \inf_{(x,y)\in I_{ij}} f(x,y)$. Again, one can easily generalize this notion into $\mathbb{R}^n$.

Definition 5.8. Similar to $\mathbb{R}$, we say that a bounded function $f : I \subset \mathbb{R}^n \to \mathbb{R}$, where I is a rectangle, is integrable on I if

$$\sup_D (s(D)) = \inf_D (S(D))$$

and we denote this value by

$$\int_I f(\mathbf{x}) d\mathbf{x}$$

Proposition 5.4. *Let $f : I \subset \mathbb{R}^n \to \mathbb{R}$ be a bounded function. Then f is integrable iff for all $\epsilon > 0$, there is a division D so that*

$$S(D) - s(D) < \epsilon.$$

Proof. See Assignment 8 Question 2. □

Definition 5.9. In $\mathbb{R}^2$, we define the norm of a division D as

$$\|D\| = \max \left(\max_{1\leq i\leq n} |x_i - x_{i-1}| \, , \max_{1\leq i\leq m} |y_i - y_{i-1}| \right)$$

which is easily generalized into $\mathbb{R}^n$.

Proposition 5.5. *A bounded function $f : I \subset \mathbb{R}^n \to \mathbb{R}$, where I is a rectangle, is integrable iff for $\epsilon > 0$, there exists $\delta > 0$ such that for all D with $\|D\| < \delta$, $S(D) - s(D) < \epsilon$.*

Proof. The proof is very similar to Theorem 5.2. and will be left as an exercise for the reader. □

Proposition 5.6. *A function $f : I \subset \mathbb{R}^n \to \mathbb{R}$ is integrable on I iff for all sequences of divisions D_i, $t_i \in I_i$, $\|D_i\| \to 0$,*

$$I(f) = \lim_{\|D_i\|\to 0} \sum_{i=1}^{n} f(t_j)\mu(I_j) = \lim_{i\to\infty} \sum_{I_k \in D_i} f(x)\mu(I_k), \ x \in I_k$$

exists, where we are indexing our rectangles for a particular D_i by I_i, $i = 1, ..., n$. If this is the case, we say

$$I(f) = \int_I f(\mathbf{x}) \, d\mathbf{x}$$

Proof. See A8Q3. □

Non-Rectangular (General) Domains

In this section, we discuss the possibility of integrating functions over domains that are not rectangles.

Definition 5.10. A set $X \subset \mathbb{R}^n$ is called a **null set** if

- There is a rectangle I such that $X \subset I$
- For all $\epsilon > 0$, there exists a finite set of rectangles I_k, $k = 1, ..., n$ such that $X \subset \bigcup_{i=1}^{n} I_k$ and $\sum_{i=1}^{n} \mu(I_k) < \epsilon$.

Proposition 5.7. *Let $\phi : [0,1] \to \mathbb{R}^n$ be a curve such that for all $s,t \in [0,1]$*

$$\|\phi(s) - \phi(t)\|_\infty \leq M|s-t|. \tag{5.1}$$

Then the image $\phi([0,1])$ is a null set.

Proof. ($\mathbb{R}^2$) Divide $[0,1]$ into n even intervals, I_1,..., I_n, each of length $\frac{1}{n}$. For $s,t \in I_k$, $\|\phi(s) - \phi(t)\| \leq \frac{M}{n}$. So $\phi(I_k)$ is contained within a square of sides $\frac{2M}{n}$ and we will denote these squares as J_k. So

$$\phi([0,1]) \subset \bigcup_{k=1}^{n} J_k$$

where

$$\sum_{k=1}^{n} \mu(J_k) = \sum_{k=1}^{n} \left(\frac{2M}{n}\right)^2 = \frac{4M^2}{n}$$

and $\frac{4M^2}{n}$ can be made arbitrarily small by increasing n. □

Proposition 5.8. *If $\phi : [0,1] \to \mathbb{R}^n$ is $\mathcal{C}^1([0,1], \mathbb{R}^n)$, $\exists M$ such that (5.1) holds.*

Proof. Since ϕ' is continuous on $[0,1]$, $\exists M \geq 0$ such that $\sup_{0\leq t\leq 1} \|\phi'(t)\| \leq M$. The result follows from the generalized mean-value theorem. □

Proposition 5.9. *If $f : I \subset \mathbb{R}^n \to \mathbb{R}$ is bounded on I and continuous on $I \backslash X$ where X is a null set, f is integrable on I.*

Proof. Suppose $|f| \leq M$. Let $\epsilon > 0$ be given and choose I_k, $k = 1, ..., n$, such that $X \subset \bigcup_{k=1}^{n} I_k$ and $\sum_{k=1}^{n} \mu(I_k) < \epsilon$. Enlarge the I_k's slightly into open $J_k \supset I_k$ with $\sum_{k=1}^{n} \mu(J_k) < 2\epsilon$. Define $H = I \backslash \bigcup_{k=1}^{n} J_k$ which is closed and therefore compact. f is continuous on H so there is a $\delta > 0$ such that $|f(x) - f(y)| < \epsilon$ when $\|x - y\| < \delta$. Create a division $D = \{J_1, ..., J_m\}$, $m > n$, for I that contains the vertices of J_k and refine so that $\|D\| < \delta$. Evaluating $S(D) - s(D)$ directly, we get

$$\begin{aligned}
S(D) - s(D) &= \sum_{J_k} (F_k - f_k)\mu(J_k) \\
&= \sum_{J_k \in H} (F_k - f_k)\mu(J_k) + \sum_{J_k \notin H} (F_k - f_k)\mu(J_k) \\
&\leq \epsilon \sum_{J_k \in H} \mu(J_k) + 2M \sum_{J_k \notin H} \mu(J_k) \\
&< \epsilon \cdot \mu(I) + (2M)(2\epsilon) \\
&= \epsilon(\mu(I) + 4M)
\end{aligned}$$

□

Remark 5.1. We can put a more general region, D, inside a rectangle, since we already know how to integrate over rectangles. Then, in order to integrate $f(x) : D \to \mathbb{R}$, over D, we can integrate $F(x) = \begin{cases} f(x) & x \in D \\ 0 & x \notin D \end{cases}$ over our rectangle $I \supset D$.

Definition 5.11. Let $f \colon D \to \mathbb{R}$ where $D \subset I$ for some rectangle I. Define F as above. Then, if F is integrable on I, we say f is integrable on D.

$$\int_A f(x)\, dx = \int_I F(x)\, dx$$

Definition 5.12. A point $x \in \mathbb{R}^n$ is a **boundary point** of $A \subset \mathbb{R}^n$ if for every $r > 0$, $B_r(x)$ contains a point in A and a point not in A. The set of all boundary points is written ∂A.

Definition 5.13. The set $A \subset \mathbb{R}^n$ is a **Jordan region** if (1) $A \subset I$ for some rectangle I, and (2) ∂A is a null set.

Proposition 5.10. *If $f\colon A \to \mathbb{R}$ is continuous and A is a Jordan region, then f is integrable on A.*

Theorem 5.2. *(Jordan Region Properties)*
Assume f, g are integrable on a Jordan region $A \subset \mathbb{R}^n$, α a scalar. Then we have the following properties (proofs left as an exercise):

- Linearity

$$\int_A f(x) + \alpha g(x)\, dx = \int_A f(x)\, dx + \alpha \int_A g(x)\, dx$$

- Equality: If $f(x) \le g(x) \quad \forall x \in A$, then $\int_A f(x)\, dx \le \int_A g(x)\, dx$.
- Decomposition: If $A = A_1 \cup A_2$ and $A_1 \cap A_2 = \emptyset$ for Jordan regions A_1, A_2

$$\int_A f(x)\, dx = \int_{A_1} f(x)\, dx + \int_{A_2} f(x)\, dx$$

Note. We can define the volume of a Jordan region A as $\text{Vol}(A) = \int_A dx$. This corresponds to area in $\mathbb{R}^2$ and volume in $\mathbb{R}^3$.

Proposition 5.11. *If f and g are integrable on a Jordan region $A \subset \mathbb{R}^n$, fg is integrable on A.*

Proof. First show f^2 is integrable on A. For a division D of a rectangle containing A, the difference between the upper and lower sums of f^2 on each subrectangle of the division is, letting M be the bound of f over A (we are implicitly assuming that all functions that are being integrated are bounded in this course), $F_i^2 - f_i^2 = (F_i + f_i)(F_i - f_i) \le 2M(F_i - f_i)$. Therefore, $S_{f^2}(D) - s_{f^2}(D) \le 2M(S_f(D) - s_f(D))$. Integrability of f^2 follows from integrability of f. Also. g^2 and $(f+g)^2$ are integrable by the same argument. But, $fg = \frac{1}{2}\left((f+g)^2 - f^2 - g^2\right)$, so fg is integrable. □

Theorem 5.3. *(Stolz' Theorem)*
Let $f: I \to \mathbb{R}$ be integrable on $I = [a,b] \times [c,d]$. If for each $x \in [a,b]$, $y \mapsto f(x,y)$ is integrable on $[c,d]$, then $x \mapsto \int_c^d f(x,y)\, dy$ is integrable on $[a,b]$ and

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

Proof. See Wade. □

Example 5.1. $f(x,y) = y^3 e^{xy^2}$ on $[0,1] \times [0,2]$ and f is continuous on I. We can evaluate this integral as an iterated integral in any order:

$$\int_0^1 \int_0^2 y^3 e^{xy^2}\, dy\, dx = \ldots = \frac{1}{2}(e^4 - 4 - 1)$$

Example 5.2. Consider the following function.

$$f(x,y) = \begin{cases} 1 & (x,y) = \left(\frac{p}{2^n}, \frac{q}{2^n}\right), 0 < p, q < 2^n \\ 0 & \text{otherwise} \end{cases}$$

If we fix x to be $x_0 = \frac{p}{2^n}$ for some value p, then $f(x,y) = 1$ only if $y = \frac{q}{2^n}$ where $q \in \{1, \ldots, 2^{n-1}\}$. Note that once we fix x, we are also fixing n. Hence for each $x_0 \in [0,1]$, $f(x,y) = 1$ for finite number of $y's$. So $\forall x \in [0,1]$, $\int_0^1 f(x,y)\, dy = 0$ and similarly $\forall y \in [0,1]$, $\int_0^1 f(x,y)\, dx = 0$. Thus,

$$\int_0^1 f(x,y)\, dy\, dx = \int_0^1 f(x,y)\, dx\, dy = 0.$$

So what about the integrability of f? On any division of $[0,1]\times[0,1]$, any subrectangle contains both irrational points and points of the form $\left(\frac{p}{2^n},\frac{q}{2^n}\right)$. Thus, $s(f)=0$ and $S(f)=1$ and f is not integrable.

Example 5.3. Consider the function

$$f(x,y)=\begin{cases}1 & x=0,1, y\in\mathbb{Q}\\ 1 & y=0,1, x\in\mathbb{Q}\\ 0 & \text{otherwise}\end{cases}.$$

Note that $x_0=0,1$, $y\mapsto f(x_0,y)$ is the Dirichlet (characteristic) function which is not integrable. Hence the iterated integrals do not exist. However,

$$\iint\limits_I f(x,y)\,d(x,y)$$

where $I=[-1,2]\times[1,-2]$ exists and is integrable because the set of discontinuities is a null set.

Theorem 5.4. *(Fubini's Theorem)*

Let f be continuous on A.

- If $A=\{(x,y),a\le x\le b,y_l(x)\le y\le y_h(x)\}$ where $y_l,y_u\in\mathcal{C}[a,b]$, then

$$\int\limits_A f(x,y)\,d(x,y)=\int\limits_a^b\int\limits_{y_l(x)}^{y_u(x)} f(x,y)\,dy\,dx$$

- If $A=\{(x,y),c\le y\le d,x_l(y)\le x\le x_h(y)\}$ where $x_l,x_u\in\mathcal{C}[c,d]$, then

$$\int\limits_A f(x,y)\,d(x,y)=\int\limits_c^d\int\limits_{x_l(y)}^{x_u(y)} f(x,y)\,dx\,dy$$

Example 5.4. Use Fubini's Theorem to evaluate $\int\limits_A \left(x^2y+\cos x\right)dx$, $A=\{(x,y),x\in[0,\frac{\pi}{2}],y\in[0,x]\}$.

$$\int\limits_A f(x,y)\,dx=\int\limits_0^{\frac{\pi}{2}}\int\limits_0^x (x^3y+\cos x)dy=\ldots=\frac{\pi^6}{768}+\frac{\pi}{2}-1$$

Example 5.5. Evaluate $\iint\limits_D y^2\sqrt{x}\,d(x,y)$, $D=\{(x,y),x>0,y>x^2,y<10-x^2\}$. Starting with x first:

$$\int\limits_0^5\int\limits_0^{\sqrt{y}} y^2\sqrt{x}\,dx\,dy+\int\limits_5^{10}\int\limits_0^{\sqrt{10-y}} y^2\sqrt{x}\,dx\,dy$$

which is very difficult to evaluate. Going with y first, we get:

$$\int\limits_0^{\sqrt{5}}\int\limits_{x^2}^{10-x^2} y^2\sqrt{x}\,dx\,dy=\ldots=380.2$$

Note. There are couple more examples that I left out, but the above should be enough for practice.

Example 5.6. Let $D\subset\mathbb{R}^3$ be the region bounded by $x=0,x=2,y=0,z=0,y+z=1$. Evaluate $\iiint\limits_D y\,dV$. We start with z

first, noting that $0 \le z \le 1-y$ and $D_{xy} = \{0 \le x \le 2, 0 \le y \le 1\}$.

$$\iint_{D_{xy}} \int_0^{1-y} y\,dz\,d(x,y) = \int_0^1 \int_0^2 \int_0^{1-y} y\,dz\,dx\,dy = \frac{1}{3}$$

Example 5.7. Determine the volume of the region bounded by $z = x^2 + 3y^2$ and $z = 9 - x^2$. Write the answer as a triple integral. We first note that $x^2 + 3y^2 \le x \le 9 - x^2$ and take $D_{xy} = \{(x,y), 2x^2 + 3y^2 \le 9\}$. Thus the volume is

$$Vol = \iint_{D_{xy}} \int_{x^2+3y^2}^{9-x^2} dz\,d(x,y) = \int_{-\sqrt{3}}^{\sqrt{3}} \int_{-\sqrt{\frac{9-3y^2}{2}}}^{\sqrt{\frac{9-3y^2}{2}}} \int_{x^2+3y^2}^{9-x^2} dz\,d(x,y)$$

which the answer will be left as an exercise.

5.3 Change of Variables in $\mathbb{R}^n$

Notation. We denote the determinant of the Jacobian of a function ϕ at x as $\triangle_\phi(x)$.

Notation. We denote the set of first Riemann integrable functions $I \mapsto \mathbb{R}$ as $\mathcal{L}^1(I)$.

In the simple one dimensional case, the formula for a change of variable on a function f from a domain $\phi([a,b])$ to $[a,b]$, where $\phi'(x) \neq 0$ is bijective and $\mathcal{C}^1[a,b]$, is

$$\int_{\phi([a,b])} f(t)\,dt = \int_a^b f(\phi(x))\,|\phi'(x)|\,dx.$$

We generalize this into $\mathbb{R}^n$ by making the following claim.

Claim 5.1. Given a function f that is integrable on E, where $\phi \in \mathcal{C}^1(E)$, bijective and $\triangle_\phi(x) \neq 0$, then

$$\int_{\phi(E)} f(u)\,du = \int_E f(\phi(x))\,|\triangle_\phi(x)|\,dx.$$

In order for this to be true, we need the following to be true as well.

1. E is a Jordan region
2. f in integrable on $\phi(E)$
3. $\phi(E)$ is a Jordan region
4. $f \circ \phi \cdot |\triangle_\phi(x)|$ is integrable on E

From here on out, the proof of the theorem will have to be found in Wade. We will only create a sketch of the lemmas and propositions needed (without proof).

Lemma 5.2. *Let $V \subset \mathbb{R}^n$ be a bounded open set and $\phi \in \mathcal{C}(V, \mathbb{R}^n)$. If K is a null set, $\phi(K)$ is a compact null set. If moreover,* $\det \phi'(u) \neq 0$, $\forall u \in V$*, then*

$$\{u \in K \mid \phi(u) \in \partial\phi(K)\} \subset \partial K \implies \partial\phi(K) \subset \phi(\partial K)$$

Proposition 5.12. *Let $V \subset \mathbb{R}^n$ be a bounded open set and $\phi \in \mathcal{C}^1(V, \mathbb{R}^n)$ be bijective on V with* $\det \phi'(u) \neq 0$, $\forall u \in V$*. If $E \subset V$ is a Jordan region, $\phi(E)$ is a Jordan region.*

Proposition 5.13. *Suppose $\phi : \mathbb{R}^n \to \mathbb{R}^n$ is a linear function defined by $\phi(u) = Mu$ for some matrix M. Let $I \subset \mathbb{R}^n$ be a rectangle. Then Vol$(\phi(I)) = |\det M| \cdot Vol(I)$.*

Lemma 5.3. *Let $V \subset \mathbb{R}^n$ be a bounded set and $\phi \in \mathcal{C}^1(V, \mathbb{R}^n)$ be bijective. If $\det \phi'(a) \neq 0$ then there exists a rectangle $I \subset V$, $a \in I$, and $\phi^{-1} \in \mathcal{C}^1$ with a non-zero Jacobian on $\phi(I)$. Therefore, if $J \subset \phi(I)$ is a rectangle, then $\phi^{-1}(J)$ is a Jordan region and*

$$Vol(J) = \int\limits_{\phi^{-1}(J)} |\triangle_\phi(u)| \, du$$

An interesting application of the above lemma is Mercator's Projection which uses loxodromes, which are lines that cut the meridians of the 2-sphere at a constant angle.

Theorem 5.5. *(Change of Variables)*
Let $\phi : V \to \mathbb{R}^n$ where V is a an open set and $\phi \in \mathcal{C}^1(V, \mathbb{R}^n)$ and let $E \subset V$ be a closed Jordan region. Suppose ϕ is one-to-one and $\triangle_\phi(x) \neq 0$ on $E \backslash Z$ where Z is a null set. Then $\phi(E)$ is a closed Jordan region and

$$\int\limits_{\phi(E)} f(u) \, du = \int\limits_{E} f(\phi(x)) \, |\triangle_\phi(x)| \, dx$$

holds for all continuous functions $f : \phi(E) \to \mathbb{R}^n$.

Example 5.8. Integrate $\int\limits_A (x+y)(2x-y) \, dX$, $A = \{(x,y), y \leq 2x, y \leq 3-x, y \geq 2x-3, y \geq -x\}$. Using the above, we set $u = x+y$, $v = 2x-y$ with $g(x,y) = (u,v)$, $\phi = g^{-1}$ and

$$|\triangle_\phi(x)| = \begin{vmatrix} 1 & 1 \\ 2 & -1 \end{vmatrix} = -3.$$

Thus, we get

$$\int\limits_A (x+y)(2x-y) \, dX = \int\limits_0^3 \int\limits_0^3 uv \cdot \left| -\frac{1}{3} \right| \, dX.$$

Example 5.9. Evaluate $I = \int\limits_A \sqrt{x^2+y^2} dX$ where $A = \{1 \leq x^2+y^2 \leq 4, y \leq x, y \geq 0\}$. We first note that we can change this into a polar coordinate system with $\sqrt{x^2+y^2} = r$ and $|\triangle_\phi(x)| = r$ as follows.

$$I = \int\limits_{r=1}^{r=2} \int\limits_{\theta=0}^{\theta=\frac{\pi}{4}} rr \, dr \, d\theta = \frac{7\pi}{12}$$

Remark 5.2. Note that the change of variables does not work for a change from Cartesian to polar coordinates if we do not restrict $r > 0$. Otherwise the map $(r,\theta) \mapsto (x,y)$ is zero everywhere for $r = 0$ and arbitrary θ.

Definition 5.14. A useful change of variables is the **cylindrical coordinate system**. The map $(r,\theta,z) \mapsto (x,y,z)$ and determinant of the map is given by

$$\begin{cases} x = r\cos\theta \\ y = r\sin\theta \\ z = z \end{cases}, |\triangle_\phi| = \begin{vmatrix} \cos\theta & -r\sin\theta & 0 \\ \sin\theta & r\cos\theta & 0 \\ 0 & 0 & 1 \end{vmatrix} = r$$

where we have to restrict $r > 0$.

Example 5.10. Use a change of variables (cylindrical) to find the volume bounded by the region $\{z \geq x^2+y^2, (z-2)^2 \leq x^2+y^2, z \leq 2\}$. We first determine the range for r and z. First, $x^2+y^2 = z \implies z = r^2$ and $(z-2)^2 = x^2+y^2 \implies z = 2-r$ since $z \geq 2$. So $r^2 = 2-r \implies r = 1$ and our integral is

$$V = \int\limits_0^1 \int\limits_0^{2\pi} \int\limits_{r^2}^{2-r} r \, dz \, d\theta \, dr = \frac{5\pi}{6}$$

Definition 5.15. Another useful change of variables is the **spherical coordinate system**. The map $(\rho, \phi, \theta) \mapsto (x, y, z)$ and determinant of the map is given by

$$\begin{cases} x = \rho\sin\phi\cos\theta \\ y = \rho\sin\phi\sin\theta \\ z = \rho\cos\phi \end{cases}, |\triangle_\phi| = \begin{vmatrix} \sin\phi\cos\theta & \rho\cos\phi\cos\theta & -\rho\sin\phi\sin\theta \\ \sin\phi\sin\theta & \rho\cos\phi\sin\theta & \rho\sin\phi\cos\theta \\ \cos\phi & -\rho\sin\phi & 0 \end{vmatrix} = \rho^2\sin\phi$$

where we have to restrict $\rho > 0$.

Example 5.11. Find the volume enclosed by the surfaces

$$x^2 + y^2 + z^2 = 4, z^2 = x^2 + y^2, z^2 = 3(x^2 + y^2), z = 0.$$

With a change of variables (spherical), we get the iterated integral:

$$V = \int_{\theta=0}^{\theta=2\pi} \int_{\phi=\frac{\pi}{6}}^{\phi=\frac{\pi}{4}} \int_{\rho=0}^{\rho=2} \rho^2 \sin\phi \, d\rho \, d\phi \, d\theta$$

which can be more easily understood if one draws a diagram.

We will end off this section with some exercises. This answers will be provided in footnotes

Exercise 5.1. Let $A \subset \mathbb{R}^3$ be the region bounded by $x = 0, x = 2, y = 0, z = 0, y + z = 1$. What is $\int_A y \, dX$?[11]

Exercise 5.2. Let $W \subset \mathbb{R}^3$ be the region bounded by $x = 0, y = 0, z = 0, z = \pi, x + y = 1$. What is $\int_W x^2 \cos z \, dX$?[12]

Exercise 5.3. What is the volume of a sphere of radius b (no cheating, now)?[13]

Exercise 5.4. Let $A \subset \mathbb{R}^3$ be the region bounded by $y = 1 - x^2 - z^2, z^2 + y^2 + z^2 = 3, y < 1 - x^2 - z^2$. Find $vol(A)$[14]

Exercise 5.5. Convert the following into an iterated integral over Cartesian, cylindrical, and spherical coordinates. Evaluate in spherical coordinates.[15]

$$I = \iiint_D \frac{z}{(x^2 + y^2 + z^2)^{\frac{3}{2}}} e^{-\sqrt{x^2+y^2+z^2}} dV$$

where $D = \{(x, y, z), x^2 + y^2 + z^2 \leq 4, z^2 \geq x^2 + y^2, z \geq 0\}$.

Exercise 5.6. Let $W = \{(x, y, z), x = 0, y = 0, z = 0, x^2 + y^2 + z^2 = 4\}$. Evaluate the following integral.[16]

$$I = \iiint_D \frac{e^{\sqrt{x^2+y^2+z^2}}}{x^2 + y^2 + z^2} dV$$

Exercise 5.7. Compute the volume of the solid bounded by $x^2 + 2y^2 = 2$, $z = 0$ and $x + y + 2z = 2$ as an iterated integral.[17]

Exercise 5.8. Evaluate

$$\iint_D \frac{y}{x^2 + y^2} \, dA$$

where $D = \{(x, y), x^2 + (y + 1)^2 \leq 1, y \leq -1\}$.[18]

[11] Ans: 1
[12] Ans: 0
[13] Ans: $\frac{4\pi r^3}{3}$
[14] Ans: $\frac{2\pi}{3}(\sqrt{27} - 1)$
[15] Ans: $\frac{\pi}{2}(1 - e^{-2})$
[16] Ans: $\frac{\pi^2}{4}(e^2 - 1)$
[17] Ans: $\sqrt{2}\pi$
[18] Ans: -1

Appendix A

Vector Spaces

More formally, we can define a vector space V over a field $\mathbb{F}$ as the set of all finite linear combinations (elements are allowed to be scaled through elements in $\mathbb{F}$) of a set S, called the basis, together with two bilinear operators, $+ : V \times V \to V$ and $* : V \times \mathbb{F} \to V$ called addition and scalar multiplication respectively. The axioms defined in **Definition 1.2.** must also hold for V to be a vector space.

Note that depending on how we choose our basis, V can contain either finitely many, countably many, or uncountably many elements.

Index