MATH247 Final Exam Review

LaTeXer: W. Kong

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

Theorem 1.1. *(Cauchy-Schwarz Inequaity)*

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

Definition 1.1. A **norm** is a function $\|\cdot\| : \mathbb{R}^n \to \mathbb{R}$ that satisfies (N1),(N2), and (N3) below. We call $(\mathbb{R}^n, \|\cdot\|)$ a **normed linear (vector) space**.

(N1) $\|x\| > 0$ and $\|x\| = 0 \iff x = 0$

(N2) $\|\alpha x\| = |\alpha|\|x\|$

(N3) $\|x + y\| \leq \|x\| + \|y\|$

Notation 1. We define a ball of radius $r > 0$ and norm $\|\|_i$ around a point $a \in \mathbb{R}^n$ with the following notation:

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

Definition 1.2. otherwise if a norm is not given, then we use the notation:

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

Definition 1.3. 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 1.1. *The set $\mathcal{B}_r(a)$ is open for $r > 0$, $a \in \mathbb{R}^n$.*

Definition 1.4. A set $\mathcal{V}$ is **closed** if $\mathcal{V}^c$ is open.

Definition 1.5. 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 $\mathcal{V}$. Suppose $\alpha \subset \beta \subset \mathbb{R}^n$. If there is an open set O such that $\alpha = O \cap \beta$ then α is **relatively open** in β. Similarly, if there's a closed set C such that $\alpha = C \cap \beta$, α is **relatively closed** in β.

Definition 1.6. 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**.

Notation 2. Let $x_{m,n}$ represent the the n^{th} component of the m^{th} vector in a sequence of vectors $\{x_i\}_{i\geq 1}$.

Definition 1.7. 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 $\lim\limits_{i\to\infty} a$.

Definition 1.8. In $\mathbb{R}^n$, we consider a sequence of vectors

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

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

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

Proposition 1.3. *A sequence $\{x_i\}_{i\geq 1} \subset \mathbb{R}^n$ is convergent in one norm iff it is convergent in another norm.*

Proposition 1.4. *The sequence $\{x_i\}_{i\geq 1} \subset \mathbb{R}^n$ is convergent iff $\lim\limits_{i\to\infty} x_{k,i} = a_k$, $1 \leq k \leq n$ for some $a_k \in \mathbb{R}$.*

Definition 1.9. 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 1.5. *A sequence of vectors is convergent iff it is Cauchy.*

Proposition 1.6. *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.*

Definition 1.10. 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.

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,*

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

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 $L_g \neq 0$, $\lim_{x \to a} \frac{f(x)}{g(x)} = \frac{L_f}{L_g}$

Theorem 2.2. *(Squeeze Theorem)*

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

$$f(x) \leq g(x) \leq h(x),\ \forall x \in A \backslash a \tag{1}$$

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

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

Lemma 2.1. *(Young's inequality)*

$(|a| - |b|)^2 = a^2 + b^2 - 2|a||b| \geq 0 \implies 2|a||b| \leq a^2 + b^2$

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 $g \neq 0$, then $\frac{f}{g}$ is continuous at a

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.

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.

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.

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 $\mathbb{R}^m$).*

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 $f(c) = v$.

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 $\mathbb{R}^n$)

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.*

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.

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

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,

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

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.

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.*

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

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

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

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 $x_0 \in A$.

3 Differential Multivariate Calculus

Definition 3.1. 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 3.2. 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 3.3. Assume a is an interior point of U. Let $f : U \subset \mathbb{R}^n \to \mathbb{R}$. The partial derivatives are

$$\begin{array}{c} \frac{\partial f}{\partial x_1}(a) = \lim\limits_{h \to 0} \frac{f(a_1+h, a_2, \ldots, a_n) - f(a)}{h} \\ \frac{\partial f}{\partial x_2}(a) = \lim\limits_{h \to 0} \frac{f(a_1, a_2+h, \ldots, a_n) - f(a)}{h} \\ \vdots \\ \frac{\partial f}{\partial x_n}(a) = \lim\limits_{h \to 0} \frac{f(a_1, a_2, \ldots, a_n+h) - f(a)}{h} \end{array}$$

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

Definition 3.4. 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.

Definition 3.5. 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

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

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

$$f'(a) = \nabla f(a) = \begin{bmatrix} \frac{\partial f}{\partial x_1}(a) & \frac{\partial f}{\partial x_2}(a) & \cdots & \frac{\partial f}{\partial x_n}(a) \end{bmatrix}$$

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

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

Definition 3.6. The **Jacobian** of f is

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

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 $\lim_{x \to a} \|r(x)\| = 0$.

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

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

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

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.*

Summary 1.

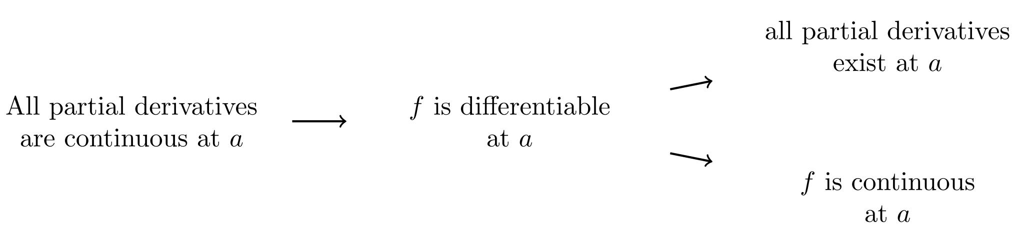

Figure 1: Differentiability Theorems

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 $u = \frac{\nabla f(a)}{\|\nabla f(a)\|}$.

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 $a \in intA$ and f is differentiable at $b \in intB$, then $h = f(g(x)) = (f \circ g)(x)$ is differentiable at a with

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

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).

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

Let $f : A \subset \mathbb{R}^n \to \mathbb{R}$ be differentiable on $S \subset int A$ 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 $f(b) - f(a) = \underbrace{f'(c)}_{\nabla f(c)}(b-a)$.

Definition 3.7. 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.*

Theorem 3.3. *(Generalized Mean Value Theorem)*[1]

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 $\|f'(x)\|_{2,2} \leq M$.[2] *Then,*

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

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.

Definition 3.8. 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 3.9. 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 $\mathcal{B}_r(x_0) \cap U$. [3]

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 $\underline{a} \in U$*[4]*. 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 $\underline{a}$.

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.*

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.*

---

[1] See also H+W, IV 3.7

[2] $\|f'(a)\|_{2,2} \leq M$ means $\|f'(a)y\|_2 \leq M\|y\|_2$,$\forall y$

[3] That is, $a \neq b$ implies $f(a) \neq f(b)$.

[4] Note that $\underline{a}$ is an $n-$dimensional vector.

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 $y \in f(B)$, $(f^{-1})'(y) = [f'(x)]^{-1}$

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.

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

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*

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

Definition 3.10. 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

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

Definition 3.11. 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

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

4 Optimization

Definition 4.1. Let $f : V \subset \mathbb{R}^n \to \mathbb{R}$. The point x^o minimizes f over V if $f(x^o) \leq f(x)$, $\forall x \in V$.

The point x^o is a **local minimum** if there is $\epsilon > 0$ such that $f(x^o) \leq f(x)$, $\forall x \in \mathcal{B}_\epsilon(x^o) \cap V$ and $x \in \text{int}V$. The definition for the **local maximum** is similar to the previous definition except with the change that $f(x^o) \geq f(x)$.

An **extreme point** or an **extremum** is a local maximum (max) or minimum (min).

Proposition 4.1. *Assume $f : V \subset \mathbb{R}^n \to \mathbb{R}$ is differentiable on V. If $x^o \in V$ is a local extremum, $f'(x^o) = 0$.*

Definition 4.2. A point x^o at which a differentiable function f has $f'(x^o) = 0$ is called a **stationary** or **critical point**.

Remark 4.1. Not every critical point is an extreme point. (e.g. the classical example in $\mathbb{R}$ is $y = x^3$ (standard cubic) and in $\mathbb{R}^2$ it is $f(x_1, x_2) = x_1^2 - x_2^2$ (standard saddle))

Definition 4.3. A set $\mathcal{D} \subset \mathbb{R}^n$ is **convex** if $\forall x, w \in \mathcal{D}$, we have $\alpha x + (1-\alpha)w \in \mathcal{D}, 0 \leq \alpha \leq 1$.

Definition 4.4. A function $f : \mathcal{D} \subset \mathbb{R}^n \to \mathbb{R}$ is a **convex** function if for all $x, w \in \mathcal{D}$ and $\alpha \in (0, 1)$, we have

$$f(\alpha x + (1-\alpha)w) \leq \alpha f(x) + (1-\alpha)f(w)$$

where $\mathcal{D}$ is convex. If we have $<$ holding instead of $\leq$, we say that the function is **strictly convex**.

Notation. We define the **epigraph** of a function $f : \mathcal{D} \subset \mathbb{R}^n \to \mathbb{R}$ to be $epi(f) = \{(x, y) \in \mathcal{D} \times \mathbb{R}, y \geq f(x)\}$.

Remark 4.2. Two equivalent definitions to Definition 5.4 are

- Secant lines with points in $\mathcal{D}$ will always lie above the graph of f
- The epigraph of f is a convex set

Proposition 4.2. *Let $f : \mathcal{D} \subset \mathbb{R}^n \to \mathbb{R}$ be differentiable on $\mathcal{D}$.Then f is convex on $\mathcal{D}$* ***if and only if*** *$f(w + v) \geq f(w) + f'(w) \cdot v$ where $w, v \in \mathcal{D}$.*

Proposition 4.3. *Let $f : \mathcal{D} \subset \mathbb{R}^n \to \mathbb{R}$ be convex on $\mathcal{D}$. Then, it has a one sided directional derivative*

$$D_+ f(x, v) = \lim_{t \to 0^+} \frac{f(x + tv) - f(x)}{t}$$

for all $x \in int(\mathcal{D})$ and arbitrary unit vector $v \in \mathbb{R}^n$.

Proposition 4.4. *If $f : \mathcal{D} \subset \mathbb{R}^n \to \mathbb{R}$ is differentiable and convex, then every critical point minimizes f on $\mathcal{D}$.*

Corollary 4.1. *If f is differentiable strictly convex, then a critical point is a unique minimizer of f on $\mathcal{D}$.*

Proposition 4.5. *If $f \in \mathcal{C}^2([a, b])$, then f is convex on (a, b)* ***if and only if*** *$f''(x) \geq 0$, $\forall x \in (a, b)$.*

Definition 4.5. A symmetric matrix $M \in \mathbb{R}^{m \times n}$ is

- positive semi-definite if $x^t M x \geq 0$, for all $x \in \mathbb{R}^n$ (denoted as $M \geq 0$) and positive definite if the previous holds and $x^t M x = 0 \implies x = 0$ (denoted as $M > 0$)
- negative semi-definite if $x^t M x \leq 0$, for all $x \in \mathbb{R}^n$ (denoted as $M \leq 0$) and negative definite if the previous holds and $x^t M x = 0 \implies x = 0$ (denoted as $M < 0$)
- indefinite if $x^t M x > 0$, $y^t M y > 0$ for some $x, y \in \mathbb{R}^n$ and $x \neq y$.

Remark 4.3. Alternatively, a matrix is positive (negative) semi-definite if all the eigenvalues are greater than (less than) or equal to 0, positive (negative) definite if the eigenvalues are all positive (negative), and indefinite if there are both positive and negative eigenvalues present.

Proposition 4.6. *If $f \in \mathcal{C}^2(\mathcal{D})$, where $\mathcal{D}$ is a convex set, f is convex on $\mathcal{D}$ **if and only if** the Hessian is positive semi-definite at each point on $\mathcal{D}$.*

Proposition 4.7. *Consider $A = \begin{bmatrix} A & B \\ B & C \end{bmatrix}$ and define $D = AC - B^2$. If $D = 0$ then M is semi-definite, $D < 0$ then M is indefinite, $D > 0$, $A > 0$ then M is positive definite and $D > 0$, $A < 0$ then M is negative definite.*

Proposition 4.8. *If for an open set $\mathcal{D} \subset \mathbb{R}^n$, $f \in \mathcal{C}^2(\mathcal{D})$ and $f'(a) = 0$ for some $a \in \mathcal{D}$, then if:*

- $H_f(x) \geq 0$ for all x on a neighbourhood $\mathcal{B}_r(a)$ of a, then a is a strict local minimum of f
- $H_f(x) \leq 0$ for all x on a neighbourhood $\mathcal{B}_r(a)$ of a, then a is a strict local maximum of f

Lemma 4.1. *Consider a symmetric matrix $M \in \mathbb{R}^n$. If $M > 0$ there is a constant $m > 0$ such that*

$$x^t M x > m\|x\|^2$$

for all $x \in \mathbb{R}^n$, $x \neq 0$.

Proposition 4.9. *Consider $f \in \mathcal{C}^2(\mathcal{D})$, $D \subset \mathbb{R}^n$ is open. Let $a \in \mathcal{D}$ be such that $f'(a) = 0$. Then if:*

- $H_f(a) > 0$, then a is a strict local minimum of f
- $H_f(a) < 0$,then a is a strict local maximum of f
- $H_f(a)$ is indefinite,then a is a saddle point of f

Theorem 4.1. *(Extended Extreme Value Theorem)*

Suppose f is differentiable on a compact set A. Then by the extreme value theorem, f achieves its minimum/maximum at some $x^o \in A$. If $x^o \in int(A)$ or $x \in bdy(A)$, then x^o is a critical point ($f'(x_0) = 0$) or $x \in bdy(A)$.

Note. We build up motivation for the Lagrange multiplier theorem in the following way. Suppose we are given some differentiable function $f : \mathbb{R}^2 \to \mathbb{R}$ and we restrict the domain through the condition $g(x, y) = 0$ for some function g. Using the implicit function theorem, we parametrize f with $(x, y) \mapsto (x(t), y(t))$ where $h(t) = f(x(t), y(t)) \implies h'(t) = f'(x(t), y(t)) \cdot \begin{pmatrix} x'(t) \\ y'(t) \end{pmatrix}$ and $x(t), y(t) \in \mathcal{C}^1$. Suppose h has an extremum at t^o. Then

$$\underbrace{f'(x(t^o), y(t^o))}_{\text{gradient to } f} \cdot \left[\underbrace{\begin{pmatrix} x'(t^o) \\ y'(t^o) \end{pmatrix}}_{\text{tangent vector to the curve}} \right] = 0 \implies f'(x(t_0), y(t_0)) \| g'(x(t^o), y(t^o))$$

Theorem 4.2. *(Lagrange Multiplier Theorem)*

Let $f, g \in \mathcal{C}^1(V)$ where $V \subset \mathbb{R}^n$ is open. If $x^o \in V$ is a local extremum of f subject to $g(x^o) = 0$ then either

- $g'(x^o) = 0$ OR $\exists \lambda \in \mathbb{R}$ such that $f'(x^o) = \lambda g'(x^o)$

5 Integral Multivariate Calculus

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 $f_i = \inf\limits_{x_{i-1} \leq x \leq x_i} f(x)$, $F_i = \sup\limits_{x_{i-1} \leq x \leq x_i} f(x)$ 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. $s(D) \leq S(D)$
2. If D' is a refinement of D, then $s(D) \leq s(D') \leq S(D') \leq S(D)$
3. $s(D) \leq S(D')$ where D' need not be a refinement of D

Definition 5.3. We say that a bounded function $f[a, b] \to \mathbb{R}$ is **integrable** if the upper and lower quantities, $\inf\limits_{D} (S(D))$ and $\sup\limits_{D} (s(D))$, 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 $S(D) - s(D) < \epsilon$.*

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

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

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

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

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

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

Definition 5.5. We define the **boundary** of a set A, denoted as $\text{bdy}(A)$, 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 $F_{ij} = \sup\limits_{(x,y) \in I_{ij}} f(x,y)$ and $f_{ij} = \inf\limits_{(x,y) \in I_{ij}} f(x,y)$. 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

$$\int\limits_I f(\mathbf{x}) d\mathbf{x}$$

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*

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

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

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

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 $\|D\| < \delta$, $S(D) - s(D) < \epsilon$.*

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

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

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 $s, t \in [0,1]$*

$$\|\phi(s) - \phi(t)\|_\infty \leq M|s - t|. \tag{2}$$

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

Proposition 5.8. *If $\phi : [0,1] \to \mathbb{R}^n$ is $\mathcal{C}^1([0,1], \mathbb{R}^n)$, $\exists M$ such that (2) holds.*

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.*

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 $f : D \to \mathbb{R}$ 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.

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

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 $f : A \to \mathbb{R}$ 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

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

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

$$\int\limits_A f(x)\, dx = \int\limits_{A_1} f(x)\, dx + \int\limits_{A_2} f(x)\, dx$$

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.*

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 $x \mapsto \int\limits_c^d f(x,y)\,dy$ is integrable on $[a,b]$ and

$$\int\limits_I f(x,y)\,d(x,y) = \int\limits_a^b \int\limits_c^d f(x,y)\,dy\,dx$$

Theorem 5.4. *(Fubini's Theorem)*

Let f be continuous on A.

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

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

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

Notation. 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

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

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

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

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 $\phi : V \to \mathbb{R}^n$ whereV 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 $f : \phi(E) \to \mathbb{R}^n$.

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 $(r, \theta, z) \mapsto (x, yz)$ 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 $r > 0$.

Definition 5.15. Another useful change of variables is the **spherical coordinate system**. The map $(\rho, \phi, \theta) \mapsto (x, yz)$ 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 $\rho > 0$.