MATH 247: Calculus III (Advanced Level)

Felix Zhou*

Winter 2019
University of Waterloo

*from Professor Henry Shum's lectures

Contents

Introduction

From the University of Waterloo's website: topics convered include Topology of real n-dimensional space: completeness, closed and open sets, connectivity, compact sets, continuity, uniform continuity. Differential calculus on multivariable functions: partial differentiability, differentiability, chain rule, Taylor polynomials, extreme value problems. Riemann integration: Jordan content, integrability criteria, Fubini's theorem, change of variables. Local properties of continuously differentiable functions: open mapping theorem, inverse function theorem, implicit function theorem.

1 Euclidean Space

HW IV.1 up to Definition 1.4
GF 1.1

1.1 Definitions

Definition 1.1.1 (distance)
For two vectors in two-dimensional space.
$d(\vec{x}, \vec{y}) = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$

Note that we work in $\mathbb{R}^n$, or "Euclidean Space".

Definition 1.1.2 (Euclidean Inner Product (dot/scalar product))
$\langle \vec{x}, \vec{y} \rangle = \sum x_i y_i$

Reminder of the Law of Cosines

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

Definition 1.1.3 (Euclidean Norm)
$\|\vec{x}\| = \sqrt{\langle \vec{x}, \vec{x} \rangle}$

Proposition 1.1.1 (Properties of the Inner Product)

1. $\langle \vec{x}, \vec{y} \rangle = \langle \vec{y}, \vec{x} \rangle$ (Symmetry)
2. $\langle \vec{x}, \vec{x} \rangle \geq 0$ and $\langle \vec{x}, \vec{x} \rangle = 0 \iff \vec{x} = 0$ (Positive Definite)
3. $\langle \alpha\vec{x} + \beta\vec{y}, \vec{z} \rangle = \alpha\langle \vec{x}, \vec{z} \rangle + \beta\langle \vec{y}, \vec{z} \rangle$ (bilinearity)

Proposition 1.1.2 (Properties of the Euclidean Norm)

1. $\|\vec{x}\| \geq 0$ and $\|\vec{x}\| = 0 \iff \vec{x} = \vec{0}$ (Positive Definite)
2. $\|\alpha\vec{x}\| = |\alpha|\|\vec{x}\|$ (Homogeneity)
3. $\|\vec{x} + \vec{y}\| \leq \|\vec{x}\| + \|\vec{y}\|$ (Triangle Inequality)

Note that we also have the reverse triangle inequality:

$$|\|\vec{x}\| - \|\vec{y}\|| \leq \|\vec{x} - \vec{y}\|$$

.

1.2 Inequalities

Theorem 1.2.1 (Cauchy-Schwartz-Inequality)
$\forall \vec{x}, \vec{y} \in \mathbb{R}^n \quad |\langle \vec{x}, \vec{y} \rangle| \leq \|\vec{x}\| \cdot \|\vec{y}\|$

Proof
TBD

Theorem 1.2.2 (Triangle Inequality for Euclidean Norm)
$\|\vec{x} + \vec{y}\| \leq \|\vec{x}\| + \|\vec{y}\|$

Proof
TBD

2 Sequences

H-W IV.1 Definition 1.5 to Theorem 1.8
G-F 1.4-1.5

2.1 Definition of Sequences and Convergence in $\mathbb{R}^n$

Definition 2.1.1
infinite enumerated list of vectors or points $(\vec{x_k})_{k=1}^{\infty}$, where each $\vec{x_k} \in \mathbb{R}^n$.

Definition 2.1.2 (Sequential Convergence)
A sequence of points $(\vec{k_k})$ **converges** to a point $\vec{a}$ if

$$\forall \epsilon > 0, \exists N \in \mathbb{N}, k \geq N \implies \|\vec{x} - \vec{a}\| < \epsilon$$

If such a point exists, we say the sequence is **convergent** and that $\vec{a}$ is the **limit** of the sequence.

$$\lim_{k \to \infty} \vec{x}_k = \vec{a}$$

Lemma 2.1.1
For sequences in $\mathbb{R}^n$
$\lim_{k\to\infty} \vec{x_k} = \vec{a} \iff \lim_{k\to\infty} \|\vec{x_k} - \vec{a}\| = 0$

Lemma 2.1.2
For sequences in $\mathbb{R}^n$, each $\vec{x_k} = (x_{k,1}, x_{k,2}, \dots, x_{k,n})$
$\lim_{k\to\infty} \vec{x_k} = (a_1, a_2, \dots, a_n) \iff \lim_{k\to\infty} x_{k,i} = a_i \quad \forall 1 \leq i \leq n$

Proof
$\implies$
Since the terms converge, then, the entry-wise difference must go to zero as well
$\impliedby$
We simply choose N sufficiently large to make the entry-wise differences a fraction of the total, giving an upper bound that way.

2.2 Cauchy Sequences

Definition 2.2.1
A sequence $(\vec{x}_k)_{k=1}^{\infty}$ is **Cauchy** if there is some interger N such that

$$\|\vec{x}_k - \vec{x}_l\| < \epsilon$$

for all $k, l \geq N$.

Lemma 2.2.1
A sequence in $\mathbb{R}^n$ is Cauchy if and only its entry-wise sequences are cauchy in $\mathbb{R}$.

Proof
TBD

2.3 Completeness

Definition 2.3.1
$S \subseteq \mathbb{R}^n$ is **complete** if every Cauchy sequence of points in S converges to a point in S.

Theorem 2.3.1 (Completeness of $\mathbb{R}^n$)
A sequence $(\vec{x}_k)$ in $\mathbb{R}^n$ converges if and only if it is Cauchy.

Proof
By Lemma 2.1.2, the sequence converges if and only if each sequece of components converges.
We know that $\mathbb{R}$ is complete, so the sequence of components converges if and only if the component-wise sequences are Cauchy.
By Lemma 2.2.1, each sequence of components is Cauchy if and only if the original sequence is Cauchy.

3 Bounded, Closed, and Open

H-W IV.1 Theorem 1.9 to Figure 1.10
GF 1.5, 1.2

3.1 Bounded Sequences and Sets

Definition 3.1.1 (Bounded Sequence)
A sequence $(\vec{x}_k)_{k=1}^{\infty}$ in $\mathbb{R}^n$ is **bounded** if there is come real number R such that

$$\|\vec{x}_k\| < R$$

for all k.

Definition 3.1.2 (Bounded Set)
A set $X \subseteq \mathbb{R}^n$ is **bounded** if there is a real number R such that

$$\|\vec{x}\| < R$$

for all $\vec{x} \in X$

Theorem 3.1.1 (Bolzano-Weierstrass)
Every bounded sequence $(\vec{x}_k)_{k=1}^{\infty}$ in $\mathbb{R}^n$ has a convergent subsequence.

Proof
Definition of Upper / Lower Bound for $\mathbb{R}^n$ and Monotone Convergence Theoerem.

3.2 Closed Sets in $\mathbb{R}^n$

Definition 3.2.1 (limit point)
$X \subseteq \mathbb{R}^n$.
$\vec{a} \in \mathbb{R}^n$ is a **limit point** of X if there is a sequence $(\vec{x}_k)_{k=1}^{\infty}$ of points in X that converges to $\vec{a}$.

Definition 3.2.2 (closed)
$X \subseteq \mathbb{R}^n$ is **closed** if it contains all of its limit points.

Definition 3.2.3 (closure)
the set of all limit points of $X \subseteq \mathbb{R}^n$.
Denoted $\bar{X}$.

Proposition 3.2.1
For any $X \subseteq \mathbb{R}^n$, the closure of X is closed.
Moreover, it is the smallest closed set that contains X.

Proof
By definition

3.3 Examples

Example 3.3.1
$\emptyset, \mathbb{R}^n$ are both closed.

Example 3.3.2
$(0, 1] \times [0, 5]$ is not closed.

Example 3.3.3
Every closed interval $[a, b] \subseteq \mathbb{R}$ is closed.

Example 3.3.4
The closure of $(0, 1)$ is $[0, 1]$.

3.4 Open Sets in $\mathbb{R}^n$

Definition 3.4.1 (open ball)
define the **open ball** of radius r about a point $\vec{a} \in \mathbb{R}^n$ as the set

$$B_r(\vec{a}) := \{\vec{x} \in \mathbb{R}^n : \|\vec{x} - \vec{a}\| < r\}$$

Definition 3.4.2 (open)
$U \subseteq \mathbb{R}^n$ is **open** if for all $\vec{a} \in U$, there is come $r > 0$ such that $B_r(\vec{a}) \subseteq U$.

Definition 3.4.3 (open neighbourhood)
if U is an open set containing a point $\vec{a}$, then U is an **open neighbourhood** of $\vec{a}$.

Definition 3.4.4 (interior point)
If $S \subseteq \mathbb{R}^n$ and $\vec{a} \in \mathbb{R}^n$ is such that $B_r(\vec{a}) \subseteq S$ for some $r > 0$, then $\vec{a}$ is an **interior point** of S.

Definition 3.4.5 (interior)
set of all interior points of $S \subseteq \mathbb{R}^n$ denoted $\text{int}(S)$.
Note it is the largest open subset of S (by defintion).
If $\text{int}(S)$ is empty, then we say that S has an **empty interior**
Else, it has a nonempty interior

3.5 Examples

Example 3.5.1
$\emptyset, \mathbb{R}^n$ are open

Example 3.5.2
Every open interval (a, b) is open.

Example 3.5.3
$B_r(\vec{a})$ is open for every $\vec{a} \in \mathbb{R}^n$, every $r > 0$.

Example 3.5.4
The interior of the closed interval $[a, b]$ is the open interval (a, b).

Example 3.5.5
The set $X := \{s \in \mathbb{Q} : |s| < 1\}$ is not open.
It has an empty interior.

4 More Open and Closed

H-W IV.1 Theorem 1.9 to Remark 1.17
GF 1.2

4.1 Open and Closed Sets

Proposition 4.1.1
The ony subsets of $\mathbb{R}^n$ that are both open and closed are $\emptyset, \mathbb{R}^n$.

Proof
Take $\vec{x} \in X \subseteq \mathbb{R}^n$ both open and closed as well as $\vec{y} \in Y := \mathbb{R}^n \setminus X$.
Take the maximize sized ball around a $\vec{x}$, $B_R(\vec{x})$.
By the closedness of X, the closure of the ball is in X.
Also, any $B_{R+\epsilon}(\vec{x})$ would not be in X by the definition of the supremum.
Take $z_k \in B_{R+\frac{1}{k}}(\vec{x}) \cap Y$ and note that the closed and boundedness of $B_{R+1}(\vec{x})$ implies that there is convergent subsequence $(z_{k_j})_{j=1}^\infty \to \vec{z} \in B_{R+1}(\vec{x})$.
Note that by construction, $\vec{z} \in B_R(\vec{x})$. But then X cannot be open since the entire sequence $(\vec{z}_k)_{k=1}^\infty$ is in Y.

Theorem 4.1.2
A set $X \subseteq \mathbb{R}^n$ is open if and only if its complement,

$$X' := \{\vec{x} \in \mathbb{R}^n : \vec{x} \notin X\}$$

is closed.

Proof
$\implies$
Suppose X is open. By the openess of X no point in its complement can converge to X.
$\impliedby$
Suppose X not open. If some point $\vec{x} \in X$ is not open, we can construct a sequence in X' converging to $\vec{x} \notin X'$ so X' is not closed.

4.2 Properties of Closed Subsets

Proposition 4.2.1
(finite) union of closed subsets of $\mathbb{R}^n$ is closed

Proof
By definition

Proposition 4.2.2
(uncountably infinite) intersection between closed subsets of $\mathbb{R}^n$ is closed.

Proof
By definition

4.3 Properties of Open Subsets

Proposition 4.3.1
(finite) intersection of open subsets of $\mathbb{R}^n$ is open.

Proof
Theorem 4.1.2

Proposition 4.3.2
(uncountably infinite) unions of open subsets of $\mathbb{R}^n$ is open.

Proof
Theorem 4.1.2

5 Compact Sets

H-W IV.1 Definition 1.18 onwards
GF 1.6

5.1 Definitions

Definition 5.1.1 (sequential compactness)
$K \subseteq \mathbb{R}^n$ is **compact** if every sequence of points in K has a subsequence that converges to a point in K.

5.2 Nonexamples

Example 5.2.1
$\mathbb{R}^n$ is not compact.

Example 5.2.2
The interval $(0, 1]$ is not compact.

5.3 Example

Lemma 5.3.1
The cube $[a, b]^n$ is a compact subset of $\mathbb{R}^n$ for any real numbers a and b with $a \leq b$.

Proof
let $(\vec{x}_k)_{k=1}^{\infty}$ be any sequence of points in the cube X.
Write

$$\vec{x}_k = (x_{k,1}, \ldots, x_{k,n})$$

for each $k \geq 1$.
Note that X is bounded so by the Bolzano Weirstrauss Theorem, we can find a subsequence $\left(\vec{x}_{k_j}\right)_{j=1}^{\infty}$ that converges to a limit

$$\vec{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$$

By Lemma 2.1.2 (Component-Wise Convergence),

$$\lim_{j \to \infty} x_{k_j,i} = x_i$$

for $i = 1, \ldots, n$.
Since each i and for all $k_j \geq 1$, $x_{k_j,i} \in [a, b]$ and $[a, b]$ is a closed subset of $\mathbb{R}$, we conclude that $x_i \in [a, b]$ for each i and that $\vec{x} \in [a, b]$ as desired.
So X is compact.

Corollary 5.3.1.1
The cube $[a, b]^n$ is a closed subset of $\mathbb{R}^n$ for any real numbers $a \leq b$

Proof
Trivial

5.4 Heine-Borel Theorem

Lemma 5.4.1
Every compact subset of $\mathbb{R}^n$ is closed and bounded

Proof
To see closed, any subsequence must converge to the same limit as the actual sequence.
To see boundedness, we can construct a sequence $(\vec{x}_k)_{k=1}\infty$ such that $\|\vec{x}_k\| \geq k$ for each $k \geq 1$, so no subsequence can converge.

Lemma 5.4.2
If $K \subseteq \mathbb{R}^n$ is compact, and C is a closed subset of K, then C is compact.

Proof
By definition of compact and closed

Theorem 5.4.3 (Heine-Borel)
A subset of $\mathbb{R}^n$ is compact if and only if it is closed and bounded.

Proof
$\Longleftarrow$
Lemma 5.4.1
$\Longrightarrow$
Supose a subset C of $\mathbb{R}^n$ is closed and bounded.
We can contain it in cube.
Then C is a closed subset of a compact subset of $\mathbb{R}^n$, so Lemma 5.4.2 tells us that C is compact.

5.5 Other Definitions of Compactness

Definition 5.5.1 (compactness)
$K \subseteq \mathbb{R}^n$ is compact if it is closed and bounded.

Definition 5.5.2 (open cover)
Suppose U_i is an open subset of $\mathbb{R}^n$ for each i in a possibly infinite indexing set I. If X is any subset of $\mathbb{R}^n$ and $X \subseteq \bigcup_{i \in I} U_i$, then $\{U_i : i \in I\}$ is an **open cover** of X.

Definition 5.5.3 (finite subcover)
If there is a finite subset $\{i_1, i_2, \ldots, i_l\}$ of I such that $X \subseteq \bigcup_{k=1}^{l} U_{i_k}$, then $\{U_{i_k} : 1 \leq k \leq l\}$ is a **finite (open) cover** of X.

Definition 5.5.4 (topological compactness)
$K \subseteq \mathbb{R}^n$ is compact if every open cover of K has a finite subcover.

6 Compact and Connected

H-W IV.1 Theorem 1.21 onwards
GF 1.7

6.1 Examples of Open Covers, Subcovers, and (Topological) Compactness

Example 6.1.1
every finite set

$$X := \{\vec{x}_k : 1 \leq k \leq N\}$$

is compact.

Example 6.1.2
The open ball $B_r(\vec{a})$ is not compact.

Example 6.1.3
R^n is not compact.

Example 6.1.4
The cube $[a, b]^n$ is compact.

6.2 Connectedness

Definition 6.2.1 (seperation)
A **seperation** of $X \subseteq \mathbb{R}^n$ is a pair (U, V) of open sets such that

1. $X \cap U \neq \emptyset$
2. $X \cap V \neq \emptyset$
3. $X \subseteq U \cap V$
4. $X \cap U \cap V = \emptyset$

Definition 6.2.2 (disconnected)
A set is **disconnected** if there exists a seperation of it

Definition 6.2.3 (connected)
A set is **connected** if there are no seperations of it

6.3 Examples

Example 6.3.1
$X_1 := B_1(-1, 0), X_2 := B_2(1, 0)$.
$X = X_2 \cup X_2$ is disconnected since (X_1, X_2) is a separatino of X.

Proposition 6.3.2
$\mathbb{R}^n$ is connected.

Proof
Suppose (U, V) is a seperation of R^n.
Then U is the complement of V in R^n by the definition of a seperation.
But then, U, V are both open and closed, meaning one of them is the emptyset by Proposition 4.1.1.

Proposition 6.3.3
The interval $X = [0, 1]$ is connected.

Proof
By assignment.

7 Limits of Functions

H-W III.3, IV.2 up to THeorem 2.2
G-F 1.3

7.1 The Limit of a Function

Definition 7.1.1 (accumulation point)
$S \subseteq \mathbb{R}^n$.
$\vec{a}$ is an **accumulation** point of S if it it a limit point of $S \setminus \{\vec{a}\}$.
The set of all accumulation points of S is denoted S^a.

Definition 7.1.2 (isolated point)
If $\vec{a} \in S \setminus S^a$, it is an **isolated point** of S.

Definition 7.1.3 (limit)
$f : A \subseteq \mathbb{R}^n \to \mathbb{R}^m, \vec{a} \in A^a$.
$\vec{v} \in R^m$ is the **limit** of f at $\vec{a}$ if for all $\epsilon > 0$, there is some $\delta > 0$ such that

$$\|f(\vec{x}) - \vec{v}\| < \epsilon$$

for all $\vec{x} \in A$ such that $0 < \|\vec{x} - \vec{a}\| < \delta$.
We write $\lim_{\vec{x} \to \vec{a}} f(\vec{a}) = \vec{v}$.

7.2 Continuity

Definition 7.2.1 (continuity at a point)
$f : A \subseteq \mathbb{R}^n \to \mathbb{R}^m$ is **continuous at the point** $\vec{a} \in A$ if

$$\forall \epsilon > 0, \exists \delta > 0, \|\vec{x} - \vec{a}\| < \delta \implies \|f(\vec{x}) - \vec{a}\| < \epsilon$$

Definition 7.2.2 (point-wise continuity)
If f is continuous at every point $\vec{a} \in A$, the f is **continuous on** A

Definition 7.2.3 (discontinuity)
If f is not continuous at some point $\vec{a}$, the it is **discontinuous**.

7.3 Example of a Continuous Function

Example 7.3.1
$f : (0, \infty) \to \mathbb{R}, x \mapsto \frac{1}{x}$ is continuous on its domain.

Proof
By definition.

7.4 Properties of Continuous Functions

Proposition 7.4.1
$f : A \subseteq \mathbb{R}^n \to \mathbb{R}^m$.
f is continuous at $\vec{a} \in A \cap A^a$ if and only if $\lim_{\vec{x}\to\vec{a}} = f(\vec{a})$.

Proof
By definition

Proposition 7.4.2
$f : A \subseteq \mathbb{R}^n \to \mathbb{R}^m$.
If $\vec{a}$ is an isolated point of A, the f is continuous at $\vec{a}$.

Proof
By definition

Theorem 7.4.3 (Sequential Characterization of Limits)
Let $A \subseteq \mathbb{R}^n, f : A \to \mathbb{R}^m$. For all $\vec{a} \in A$, the following are equivalent:

1. $\lim_{\vec{x}\to\vec{a}} f(\vec{x}) = \vec{b}$
2. $\lim_{k\to\infty} f(\vec{x_k}) = \vec{b}$ for all sequences in $A \setminus \{\vec{a}\}$ such that $(\vec{x_k})_{k=1}^{\infty} \to \vec{a}$.

Proof
By definition.

Theorem 7.4.4 (Sequential Characterization of Continuity)
Let $A \subseteq \mathbb{R}^n, f : A \to \mathbb{R}^m$. For all $\vec{a} \in A$, the following are equivalent:

1. f is continuous at $\vec{a}$
2. $\lim_{k\to\infty} f(\vec{x_k}) = f(\vec{a})$ for all sequences in A such that $(\vec{x_k})_{k=1}^{\infty} \to \vec{a}$.

Proof
By Theorem 7.4.3.

Theorem 7.4.5
$f_j : A \subseteq \mathbb{R}^n \to \mathbb{R}$ for $1 \leq j \leq m$.
$f := (f_1, \dots f_m) : A \to \mathbb{R}^m$ is continuous at $\vec{x} \in A$ if and only if f_j is continuous at $\vec{x}$ for all j.

Proof
By Sequential Characterization of Continuity with Lemma 2.1.2 (Component-Wise Convergence of a Sequence).

8 More on Limits and Continuity

H-W III.3, IV.2 up to Theorem 2.2
GF 1.3

8.1 Some Properties of Limits and Continuous Functions

Theorem 8.1.1 (Squeeze Theorem)
$A \subseteq \mathbb{R}^n, \vec{a} \in A^a$.
Suppose $f, g, h : A \to \mathbb{R}$ are such that $f(\vec{x}) \leq g(\vec{x}) \leq h(\vec{x})$ for all $\vec{x} \in A \setminus \{\vec{a}\}$.
If $\lim_{\vec{x} \to \vec{a}} f(\vec{x}) = \lim \vec{x} \to \vec{a} h(\vec{x}) = L$ then $\lim_{\vec{x} \to \vec{a}} f(\vec{x}) = L$.

Proof
By definition

Theorem 8.1.2 (Combining Limits)
let $f, g : A \subseteq \mathbb{R}^n \to \mathbb{R}^m$.
Suppose $\vec{a} \in A$ such that

$$\lim_{\vec{x} \to \vec{a}} f(\vec{x}) = \vec{u} \in \mathbb{R}^m, \lim_{\vec{x} \to \vec{a}} f(\vec{x}) = \vec{v}$$

Then

1. $\lim_{\vec{x} \to \vec{a}} f(\vec{x}) + g(\vec{x}) = \vec{u} + \vec{v}$
2. $\lim_{\vec{x} \to \vec{a}} \alpha f(\vec{x}) = \alpha \vec{u}, \forall \alpha \in \mathbb{R}$
3. $\lim_{\vec{x} \to \vec{a}} f(\vec{x}) g(\vec{x}) = uv$, if $m = 1$
4. $\lim_{\vec{x} \to \vec{a}} \frac{f(\vec{x})}{g(\vec{x})} = u/v$ for $m = 1$ provided $v \neq 0$

Proof
Sequential Characterization of Limits plus results for sequences

8.2 Combining Continuous Functions

Theorem 8.2.1
$f, g : A \subseteq \mathbb{R}^n \to \mathbb{R}^m$.
Suppose there is a point $\vec{a} \in A$ where f, g are continuous.
Then

1. $f + g$ is continuous at $\vec{a}$
2. αf is continuous at $\vec{a}$ for any $\alpha \in \mathbb{R}$
3. fg is continuous at $\vec{a}$ if $m = 1$
4. f/g is continuous at $\vec{a}$ for $m = 1$ provided $f(\vec{a}) \neq 0$

Proof
Similar for Combining Limits

8.3 Example: Polynomials

Definition 8.3.1 (monomial)
Let $\vec{\alpha} = (\alpha_1, \dots, \alpha_n)$ be a **multi-index**. For this case, $\alpha_i \in \mathbb{N}$.
for any $\vec{x} \in \mathbb{R}^n$ define the **monomial**

$$\vec{x}^{\vec{\alpha}} := x_1^{\alpha_1} \dots x_n^{\alpha_n}$$

Definition 8.3.2 (polynomial)

$$p(\vec{x}) : \mathbb{R}^n \to \mathbb{R}, p(\vec{x}) = \sum_{\vec{a} \in A} a_{\vec{\alpha}} \vec{x}^{\vec{\alpha}}$$

where A must be a finite set of multi-indices and $a_{\vec{\alpha}} \in \mathbb{R}$.

Corollary 8.3.0.1
Every polynomial $p : \mathbb{R}^n \to \mathbb{R}$ is continuous on $\mathbb{R}^n$

Proof
$p(x) = x$ is continuous.
So x^k is continuous for all $k \in \mathbb{N}$.
so all $\vec{x}^{\vec{\alpha}}$ is continuous.
Then, any linar combination is continuous (ie all polynomials.)

Corollary 8.3.0.2
Any rational function

$$f(\vec{x}) = p(\vec{x})/q(\vec{x})$$

where p, q are polynomials are continuou at every point $\vec{s} \in \mathbb{R}^n$ which $q(\vec{a}) \neq 0$.

Proof
Combining polynomials

8.4 Compositions of Continuous Functions

Theorem 8.4.1
$f : A \subseteq \mathbb{R}^n \to T \subseteq \mathbb{R}^m, g : T \to \mathbb{R}^l$.
If f is continuous at a point $\vec{a} \in A$, g is continuous at the point $f(\vec{a}) \in T$, then $g \circ f$ is continuous at $\vec{a}$.

Proof
Sequential continuous applied twice

8.5 Example Euclidean Norm

Proposition 8.5.1
The Euclidean Norm function $N : \mathbb{R}^n \to \mathbb{R}_{\geq 0}$ defined by

$$N(\vec{x}) = \|\vec{x}\|$$

is continuous

Proof
By Reverse Triangle Inequality and definition.

8.6 More Properties of Continuous Functions

Definition 8.6.1 (Image)
$\forall f : A \subseteq \mathbb{R}^n \to Y \subseteq \mathbb{R}^m$, and $X \subseteq A$.
The **image** of X, $f(X) = \{\vec{y} \in \mathbb{R}^m : f(\vec{x}) = \vec{y} \text{ for some } \vec{x} \in X\}$.

Definition 8.6.2 (Preimage)
$\forall f : A \subseteq \mathbb{R}^n \to Y \subseteq \mathbb{R}^m$, and $X \subseteq A$.
The **preimage** of Y, $f^{-1}(Y) = \{\vec{x} \in \mathbb{R}^n : f(\vec{x}) = \vec{y} \text{ for some } \vec{y} \in Y\}$.

Theorem 8.6.1
Suppose $f : \mathbb{R}^n \to \mathbb{R}^m$ is continuous, $Y \subseteq \mathbb{R}^m$.

1. $f^{-1}(Y)$ is open if Y is open
2. $f^{-1}(Y)$ is closed if Y is closed

Proof

1. By continuity of f and the definition of pre-image
2. Show that $f^{-1}(Y')$ is open so

   $$f^{-1}(Y) = f^{-1}(Y')'$$

   is closed

9 Continuous Functions and Compactness

H-W III.3 THeorem 3.6, IV.2 Theorem 2.3
GF 1.6

9.1 The Extreme Value Theorem

Theorem 9.1.1
$K \subseteq \mathbb{R}^n$ compact, $f : K \to \mathbb{R}^m$ is a continuous function on K.
Then $f(K)$ is compact

Proof
We show that an arbitrary sequence $(\vec{y}_k)_{k=1}^{\infty}$ in $f(K)$ has a subsequence that converges to a point in $f(K)$.
By the definition of the image set, there is a pre-image sequence $(\vec{x}_k)_{k=1}^{\infty}$.
By the compactness of K we have a convergent subsequence of $(\vec{x}_k)$ whose image sequence is also convergent by the continuity of f.
We are done

Theorem 9.1.2 (Extreme Value Theorem)
$\emptyset \neq K \subseteq \mathbb{R}^n$ compact and $f : K \to \mathbb{R}$ is continuous on K.
Then there are $\vec{a}, \vec{b} \in K$ such that

$$f(\vec{a}) \leq f(\vec{x}) \leq f(\vec{b})$$

for all $\vec{x} \in K$

Proof
By Theorem 9.1.1, $f(K)$ is compact and therefore closed and bounded by the Heine-Borel Theorem.
Since the image is non-empty, the Least Upper / Greatest Lower Bound Principle says that the Supremem / Infimum both exist.
By the definition of the Supremum and Infinimum, we can find a sequence that converges to both in $f(K)$.
Then, by the closedness of $f(K)$, we know that the Supremum and Infimum are in $f(K)$ and then, we are done.

10 Continuous Functions and Connectedness

10.1 Continuous Functions on Connected Domains

Theorem 10.1.1
$A \subseteq \mathbb{R}^n, f : A \to \mathbb{R}^m$ continuous
If A is non-empty and connected, then $f(A)$ is connected

Proof (contradiction)
Take the proposed seperation and consider the respective open pre-images (pre-image is open only if set is open, as seperations are opens).
It can be verified that it forms a seperation of A.

Theorem 10.1.2 (Intermediate Value Theorem)
$A \subseteq \mathbb{R}^n$ is non-empty and connected, $f : A \to \mathbb{R}$ continuous
For $\vec{a}, \vec{b} \in A$ distinct and $f(\vec{a}) < f(\vec{b})$
Then for every $y \in \mathbb{R}$ satisfying $f(\vec{a}) < y < f(\vec{b})$, there is some $\vec{c} \in A$ such that $f(\vec{c}) = y$.

Proof (contrapositive)
Suppose $y \notin f(A)$.
Let

$$U := (-\infty, y) \subset \mathbb{R}, V := (y, \infty)$$

It can be shown that (U, V) is a seperation of A.

Corollary 10.1.2.1
Continuous function map closed interval to closed intervals

Proof
EVT determines bounds are included while IVT says all values in between are reached

10.2 Path-Connectedness

Definition 10.2.1
$A \subseteq \mathbb{R}^n$, and for any two distinct points $\vec{x}, \vec{y} \in A$
We say there is a path from $\vec{x}$ to $\vec{y}$ in A if there exists $\phi : [0,1] \to A$ continuous such that

$$\phi(0) = \vec{x}, \phi(1) = \vec{y}$$

The **path** is the image of the function $\phi([0,1])$.
If there is a path from between two distinct points in A, then we say A is **path-connected**

Corollary 10.2.0.1
If $A \subseteq \mathbb{R}^n$, $f : A \to \mathbb{R}$ continuous
If $\vec{a}, \vec{b} \in A$ are two distinct points connected by a path in A such that $f(\vec{a}) < f(\vec{b})$, then for every $y \in (f(\vec{a}), f(\vec{b})), \exists \vec{c} \in A$ with $f(\vec{c}) = y$ on the path!

Proof
This is a direct result from the composition of continuous functions, intermediate value theorem, and the definition above.

Theorem 10.2.1
Every path-connected set is connected.
Note the inverse might not be true!

Proof (example)
$A := \{\{(0,0)\} \cup \{(x, f(x)) : 0 < x \leq 1\}\}$, called "The Topologist's Sine Curve" is connected but not path-connected

Definition 10.2.2
$f : [a,b] \to \mathbb{R}$
the **graph** of f is defined as $F := \{(x, f(x)) : x \in [a,b]\}$

Theorem 10.2.2
$f : [a,b] \to \mathbb{R}$
f is continuous $\iff$ its graph is a path connected subset of $\mathbb{R}^2$

Proof (contradiction)
This is the highschool curve drawing, handwavy proof rigorously shown. But the result is quite obvious.
If there is a point of discontinuity, we cannot have any paths from that point to other

points in the graph.

Theorem 10.2.3
If $A, B \neq \emptyset$ are path-connected sets with $A \cap B \neq \emptyset$, then $A \cup B$ is path-connected

Proof
join the path from A to common point and B to same common point

Theorem 10.2.4
The continuous image of a path-connected set is path-connected

Proof (continuity is presersed thorugh composition)
take $\phi[0,1] \to A$ continuous, with $\phi(0) = \vec{a_1}, \phi(1) = \vec{a_2}$, and the continuous map $f : A \to B$.
Since the composition $f \circ \phi$ is continuous, its image is the path we desire.

11 Convex Sets and Uniform Continuity

HW IV.2 Theorem 2.5-2.6
GF 1.8

11.1 Convex Sets

Definition 11.1.1
$\emptyset \neq X \subseteq \mathbb{R}^n$ is convex if for any two points $\vec{x}, \vec{y} \in X \wedge t \in [0,1]$, the point $\vec{x} + t(\vec{y} - \vec{x})$ is in X.
Note the ambiguity of definition from different sources for the emptyset as it is not that interesting.

Note the remark above also holds for connectedness, path-connected, disconnectedness.
Also note that all convex sets must be path-connected and therefore connected (trivial).

Proposition 11.1.1
any $S \subseteq \mathbb{R}^n$ convex is path-connected, so connected

11.2 Uniformly Continuous Functions

Definition 11.2.1
$f : A \subseteq \mathbb{R}^n \to \mathbb{R}^m$ is **uniformly continuous** if

$$\forall \epsilon > 0, \exists \delta > 0 \quad \forall \vec{x}, \forall \vec{y} \in A, \|\vec{x} - \vec{y}\| < \delta \implies \|f(\vec{x}) - f(\vec{y})\| < \epsilon$$

11.3 Examples

Example 11.3.1
For any matrix $M \in \mathbb{R}^{m \times n}$ and vector $\vec{b} \in \mathbb{R}^m$
The function $f : \mathbb{R}^n \to \mathbb{R}^m$ defined by $f(\vec{x}) = M\vec{x} + \vec{b}$ is uniformly continuous

Example 11.3.2
$f(x) = x^2$ on $[a, b]$ is uniformly continuous but NOT on $\mathbb{R}$

Example 11.3.3
$f(x) = \frac{1}{x}$ on $(0, 1]$ is not uniformly continuous

Example 11.3.4
$f(x) = \sin\left(\frac{1}{x}\right)$ on $(0, 1]$ is NOT uniformly continuous

Example 11.3.5
$f(x) = x\sin\left(\frac{1}{x}\right)$ on $(0, 1]$ is uniformly continuous

11.4 Compactness and Uniform Continuity

Theorem 11.4.1
Let $f : K \subseteq \mathbb{R}^n \to \mathbb{R}^m$ be continuous.
If K is compact, then f is uniformly continuous on K.

Proof (Contradiction)
Suppose f is not uniformly continuous.
There is some $\epsilon > 0$ such that for all $\delta > 0$ there are points $0 < \|\vec{x} - \vec{y}\| < \delta$ but $\|f(\vec{x}) - f(\vec{y})\| \geq \epsilon$.
Let us construct a sequence $(\vec{x_k}), (\vec{y_k})$ such that for each $k \geq 1$, define $\delta_k = 1/k$.
Let $\vec{x_k}, \vec{y_k}$ to satisfy the above condition
There are converging subsequences by compactness which converge to the same thing by continuity, thus contradicting the negation of uniform continuity.

11.5 Lipschitz Functions

Definition 11.5.1
$f : A \subseteq \mathbb{R}^n \to \mathbb{R}^m$ is **Lipschitz** if there is a constant $c \in \mathbb{R}$ such that

$$\|f(\vec{x}) - f(\vec{y})\| \leq c\|\vec{x} - \vec{y}\|$$

for all $\vec{x}, \vec{y} \in A$
Note if there is such a constant, there are infinitely many.
We are interested in the smallest such function (**Lipschitz Constant**)

Proposition 11.5.1
Every Lipschitz function is uniformly continuous

Proof
Trivial

Definition 11.5.2
n A function $T : \mathbb{R}^n \to \mathbb{R}^m$ is a **linear map** if for any $\alpha, \beta \in \mathbb{R}$ and points $\vec{x}, \vec{y} \in \mathbb{R}^n$

$$T(\alpha\vec{x} + \beta\vec{y}) = \alpha T(\vec{x}) + \beta T(\vec{y})$$

Proposition 11.5.2
Every linear map from $\mathbb{R}^n$ to $\mathbb{R}^m$ is uniformly continuous

Proof
Show that T is Lipschitz.

12 Derivatives

H-W IV.3 up to definition 3.3
GF 2.10

12.1 Single Variable Differentiation

Definition 12.1.1
Let $f : \mathbb{R} \to \mathbb{R}$ be a function.
f is differentiable at $a \in \mathbb{R}$ if the limit

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

exists
Denote it $f'(a)$ or $\frac{df(a)}{dx}$.

Proposition 12.1.1
If f is differentiable at $a \in \mathbb{R}$, then it is continuous at a.

Proof
Note that for any $h \neq 0$,

$$f(a+h) - f(a) = \frac{f(a+h) - f(a)}{h} \cdot h$$

Taking limits as $h \to 0$ on both sides, then

$$\lim_{h \to 0} [f(a+h) - f(a)] = \lim_{h \to 0} \frac{f(a+h) - f(a)}{h} \cdot \lim_{h \to 0} h = f'(a) \cdot 0 = 0$$

So, $\lim_{h \to 0} f(a+h) = f(a) \implies \lim_{x \to a} f(x) = f(a)$
By proposition 7.4.1, f is continuous at a.

12.2 Directional Derivatives

Definition 12.2.1
Let $\emptyset \neq A \subseteq \mathbb{R}^n$ with non-empty interior.
Let $f : A \to \mathbb{R}^n$.
Given an interior point $\vec{a} \in \text{int}(A)$ and a unit vector $\vec{u} \in \mathbb{R}^n$,
the **directional derivative** of f at $\vec{a}$ in the direction $\vec{u}$ is defined as

$$D_{\vec{u}} f(\vec{a}) := \lim_{h \to 0} \frac{f(\vec{a} + h\vec{u}) - f(\vec{a})}{h}$$

if it exists.

Example 12.2.1
$f : \mathbb{R}^2 \to \mathbb{R}$

$$f(x, y) = \begin{cases} \frac{xy^2}{x^2+y^4}, & (x, y) \neq (0, 0) \\ 0, & (x, y) = (0, 0) \end{cases}$$

The directional derivatives of f at $(0, 0)$:
Consider arbitrary unit vector $\vec{u} = (u, v), u^2 + v^2 = 1$

$$D_{\vec{u}} f(\vec{0}) = \lim_{h \to 0} \frac{f(hu, hv) - f(0, 0)}{h} = \lim_{h \to 0} \frac{hu(hv)^2}{(hu)^2 + (hv)^4} \cdot \frac{1}{h} = \lim_{h \to 0} \frac{uv^2}{u^2 + h^2 v^4}$$

This is 0 if $u = 0$ else $v^2/u = (1 - u^2)/u$ if $u \neq 0$.
So all directional derivatives exist at $\vec{x} = \vec{0}$
Note that the limit of f does not exist at the origin so f is NOT continuous!

12.3 Partial Derivatives

Definition 12.3.1
Let $\{\vec{e_j} : 1 \leq j \leq n\}$ be the standard basis of $\mathbb{R}^n$
Let $A \subseteq \mathbb{R}^n$, $f : A \to \mathbb{R}^m$
Given a point $\vec{a} \in \text{int}(A)$, define **partial derivative** of f with respect to x_j at $\vec{a}$ as

$$\frac{\partial f}{\partial x_j}(\vec{a}) := D_{\vec{e_j}} f(\vec{a})$$

This is commonly denoted $\partial_{x_j} f(\vec{a}), \partial_j f(\vec{a}), f_{\vec{x_j}}(\vec{a})$.
An equivalent definition is

$$\frac{\partial f}{\partial x_j}(\vec{a}) = \lim_{h \to 0} \frac{f(a1, \ldots, a_j + h, \ldots, a_n) - f(a_1, \ldots, a_j, \ldots, a_n)}{h}$$

12.4 Some Results

Proposition 12.4.1
Suppose $\vec{a}$ is in the interior of some subset $A \subseteq \mathbb{R}^n$ and $f : A \to \mathbb{R}^m$ with components $f_j : A \to \mathbb{R}$, for $1 \leq j \leq m$.
Let $\vec{u} \in \mathbb{R}^n$ be a unit vector, then $D_{\vec{u}} f(\vec{a})$ exists if and only if $D_{\vec{u}} f_j(\vec{a})$ exists for each $j \in \{1, 2, \ldots, n\}$.

Proof
Consider the function $g : (-\epsilon, \epsilon) \to \mathbb{R}^m$ defined by $g(h) := \frac{f(\vec{a}+h\vec{u})-f(\vec{a})}{h}$ for ϵ sufficiently small so that $\vec{a} + h\vec{u} \in A$.
By the Sequential Characterization of Limits and Lemma 2.1.2 (Component-Wise Convergence of Sequences),

$$D_{\vec{u}} f(\vec{a}) = \lim_{h \to 0} g(h)$$

exists if and only if each $D_{\vec{u}_i} f_i(\vec{a}) = \lim_{h\to 0} g_i(h)$ exists for $1 \leq i \leq m$.

Proposition 12.4.2
Let $A \subseteq \mathbb{R}^n, \vec{a} \in \operatorname{int}(A), f : A \to \mathbb{R}^m$.
If $\frac{\partial f}{\partial x_j}(\vec{a})$ exists for some $j \in \{1, 2, \ldots, n\}$.
Then $\frac{\partial f_i}{\partial x_j}(\vec{a})$ exists all $i \in \{1, 2, \ldots, m\}$.
And $\frac{\partial f}{\partial x_j}(\vec{a}) = \left(\frac{\partial f_1}{\partial x_j}(\vec{a}), \ldots, \frac{\partial f_m}{\partial x_j}(\vec{a})\right)$.

Proof
Trivial

13 Differentiability

13.1 Differentiability

Recall if f is differentiable

$$\lim_{h \to 0} \left| \frac{f(a+h) - f(a) - f'(a)h}{h} \right| = 0$$

Definition 13.1.1
$A \subseteq \mathbb{R}^n, \vec{a} \in \text{int}(A), f : A \to \mathbb{R}^m$.
We say f is differentiable at $\vec{a}$ if there exists a linear map $T : \mathbb{R}^n \to \mathbb{R}^m$ such that

$$\lim_{\vec{h} \to \vec{0}} \frac{\left\| f(\vec{a} + \vec{h}) - f(\vec{a}) - T(\vec{h}) \right\|}{\left\| \vec{h} \right\|} = 0$$

We call T the **derivative** of f at $\vec{a}$.
Do not confuse $T(h)$ with $f'(a)$

Note that in the one-variable case, we can express the limit definition of the derivative as

$$\lim_{h \to 0} \frac{f(a+h) - g(h)}{h} = 0$$

where $g(h) = f(a) + f'(a)(a+h)$.
We can interpret this as that $g(h)$ is a good approximation of f around the value a.
In other words, f is differentiable if there exists a **linear function** L(h) such that

$$\lim_{h \to 0} \frac{f(a+h) - f(a) - L(h)}{h} = 0$$

So our generalization for the derivative in n-dimensional Real vector space is that f is differentiable at $\vec{a}$ if there exists a linear map $T : \mathbb{R}^n \to \mathbb{R}^m$ such that

$$\lim_{\vec{h} \to 0} \frac{\left\| f(\vec{a} + \vec{h}) - f(\vec{a}) - T(\vec{h}) \right\|}{\vec{h}} = 0$$

In other words,

$$f(\vec{a}) + T(\vec{h})$$

is the best linear approximation of f near $\vec{a}$.
Note T is a function! The approximation happens like

$$f(\vec{a}) + T(\vec{h}) = f(\vec{a}) + T \begin{bmatrix} h_1 \\ h_2 \\ \dots \\ h_n \end{bmatrix}$$

We can think of $\vec{h}$ as a indication of the direction as we approach the point $\vec{a}$.

Example 13.1.1
$f(x) = x^2, f'(x) = 2x$
$T(h) = 0$ NOT $2x$!!

Theorem 13.1.2 (Uniqueness of the Derivative)
Derivatives are unique if they exist

Proof
Suppose T_1, T_2 linear maps that satisfy the definition of a derivative.
For $\left\|\vec{h}\right\|$ sufficiently small (but non-zero), we can define the functions
$r_k := f(\vec{a}+\vec{h}) - f(\vec{a}) - T_k(\vec{h}) \quad k = 1, 2$
Note that by the definition of the derivative, $\lim_{\vec{h}\to\vec{0}} \frac{\left\|r_1(\vec{h})\right\|}{\left\|\vec{h}\right\|} = \lim_{\vec{h}\to\vec{0}} \frac{\left\|r_2(\vec{h})\right\|}{\left\|\vec{h}\right\|} = 0$
By calculation, $\left\|T_1(\vec{h}) - T_2(\vec{h})\right\| = \left\|r_2(\vec{h}) - r_1(\vec{h})\right\| \leq \left\|r_2(\vec{h})\right\| + \left\|r_1(\vec{h})\right\|$
Then, $0 \leq \frac{\left\|T_1(\vec{h})-T_2(\vec{h})\right\|}{\left\|\vec{h}\right\|} \leq \frac{\left\|r_2(\vec{h})\right\|}{\left\|\vec{h}\right\|} + \frac{\left\|r_1(\vec{h})\right\|}{\left\|\vec{h}\right\|}$
By the Squeeze Theorem(8.1.1), $\lim_{\vec{h}\to\vec{0}} \frac{\left\|T_1(\vec{h})-T_2(\vec{h})\right\|}{\vec{h}} = 0$
Let $\vec{u} \in \mathbb{R}^n \setminus \{\vec{0}\}$ and write $\vec{h} = h\vec{u}$
By linearity of T_1, T_2

$$0 = \lim_{h\to 0} \frac{\|T_2(h\vec{u}) - T_1(h\vec{u})\|}{\|h\vec{u}\|} = \frac{\|T_2(\vec{u}) - T_1(\vec{u})\|}{\|\vec{u}\|}$$

So they are equal on all but the zero vector.
By linearity, we also have $T_1(\vec{0}) = T_2(\vec{0}) = \vec{0}$
So T_1, T_2 act equally on all of $\vec{x} \in \mathbb{R}^n$ so they must be equal

Since derivatives are unique, we call T as **the derivative** of f at $\vec{a}$, denoted $Df(\vec{a})$
if $f : A \to \mathbb{R}^m$ has component functions $f_1 : A \to \mathbb{R}$ for $1 \leq i \leq n$
we can relate the derivatives of f to f_i.
Note that $Df(\vec{a}) = Df(\vec{a})(\sum \vec{e_i})$ is an m by n-dimensional matrix.

Proposition 13.1.3
Let $A \subseteq \mathbb{R}^n, \vec{a} \in \text{int}(A), f : A \to \mathbb{R}^m$
$Df(\vec{a}) = T$ is equivalent to saying $Df_i(\vec{a}) = T_i$ for each component from $1, \ldots, m$

So the rows of the matrix correspond to the component-wise derivatives.

Proof (Exercise)
Note that $Df(\vec{a}) = T$ means

$$\lim_{h\to 0} \left\| \frac{f(\vec{a}+\vec{h}) - f(\vec{a}) - T(\vec{h})}{\left\|\vec{h}\right\|} \right\| = 0$$

while $Df_i(\vec{a}) = T_i$ means that

$$\lim_{h\to 0} \left| \frac{f_i(\vec{a}+\vec{h}) - f_i(\vec{a}) - T_i(\vec{h})}{\left\|\vec{h}\right\|} \right| = 0$$

Theorem 13.1.4
Let $A \subseteq \mathbb{R}^n, \vec{a} \in \text{int}(A), f : A \to \mathbb{R}^m$
Suppose f is differentiable at $\vec{a}$.
Let $T := Df(\vec{a})$ be the derivative of f at $\vec{a}$.
Then

1. For every unit vector $\vec{u} \in \mathbb{R}^n$, the directional derivative of f at $\vec{a}$ in the direction $\vec{u}$ exists and is equal to $D_{\vec{u}}f(\vec{a}) = T(\vec{u})$
2. All partial derivatives $\frac{\partial f_i}{\partial x_j}(\vec{a}), 1 \le i \le m, 1 \le j \le n$ exist
3. The $m \times n$ matrix representing T in the standard basis is the Jacobian Matrix

$$J := \sum \frac{\partial f_i}{\partial x_j}(\vec{a}) e_{ij}$$

Note that the partial derivative $\frac{\partial f_i}{\partial x_j}(\vec{a})$ can be understood as the slope of the i-th component-wise function when we vary the j-th component of the input.

Proof
1. From the definition of the derivative

$$\lim_{h\to 0} \frac{\|f(\vec{a}+h\vec{u}) - f(\vec{a}) - T(h\vec{u})\|}{|h|\|\vec{u}\|} = 0$$

So

$$\lim_{h\to 0} \left\| \frac{f(\vec{a}+h\vec{u}) - f(\vec{a})}{|h|} - T(\vec{u}) \right\| = 0$$

And

$$\lim_{h\to 0} \frac{\|f(\vec{a}+h\vec{u}) - f(\vec{a})\|}{\|h\vec{u}\|} = T(\vec{u})$$

Note the subtlety at the end uses the fact that the Euclidean Norm is positive definite.

2. This literally follows from 1

3. By our work in Math 146, the matrix describing a linear matrix T in the standard basis are defined by

$$\begin{aligned} J_{ij} &= T_i(\vec{e_j}) \\ &= Df_i(\vec{a})(\vec{e_j}) \\ &= \frac{\partial f_i}{\partial x_j}(\vec{a}) \end{aligned}$$

Example 13.1.5
$f : \mathbb{R}^3 \to \mathbb{R}^2, (x, y, z) \mapsto (x^2 + y^2, e^{x+z})$
The Jacobian at $\vec{a} = (1, 2, 3)$ can be computed as follows:
The component-wise functions are

$$\begin{aligned} f_1(x, y, z) &= x^2 + y^2 \\ f_2(x, y, z) &= e^{x+z} \end{aligned}$$

So

$$Df(1,2,3) = \begin{bmatrix} \frac{\partial f_1}{\partial x}(1,2,3) & \frac{\partial f_1}{\partial y}(1,2,3) & \frac{\partial f_1}{\partial z}(1,2,3) \\ \frac{\partial f_2}{\partial x}(1,2,3) & \frac{\partial f_2}{\partial y}(1,2,3) & \frac{\partial f_2}{\partial z}(1,2,3) \end{bmatrix} = \begin{bmatrix} 2 & 4 & 0 \\ e^4 & 0 & e^4 \end{bmatrix}$$

14 Conditions for Differentiability

H-W IV.3 - Theorem 3.6
GF 2.10

14.1 An Alternative View of Differentiable Functions

Theorem 14.1.1
$A \subseteq \mathbb{R}^n, \vec{a} \in \text{int}(A), f : A \to \mathbb{R}^m$
f is differentiable at $\vec{a}$ if and only if there exists a linear map $l : \mathbb{R}^n \to \mathbb{R}^m$ and a function $r : A \to \mathbb{R}^m$ that is continuous at $\vec{a}$ and satisfies $r(\vec{a}) = \vec{0}$, such that

$$f(\vec{x}) = f(\vec{a}) + l(\vec{x} - \vec{a}) + r(\vec{x})\|\vec{x} - \vec{a}\|$$

Intuitively, this says that f can be approximated at a certain point by a linear function with a remainder that infinitely approaches zero as we approach this point.

Proof
exercise
Note that the proof does not refer to limits

14.2 Continuity of Differentiable Functions

Theorem 14.2.1
$A \subseteq \mathbb{R}^n, \vec{a} \in \text{int}(A), f : A \to \mathbb{R}^m$.
If f is differentiable at $\vec{a}$, then it is continuous at $\vec{a}$.

Proof
This follows directly from Theorem 14.1.1, since we see that at point $\vec{a}$, f can be expressed as a sum of continuous functions.

14.3 Sufficient Conditions for Differentiability

Theorem 14.3.1
$A \subseteq \mathbb{R}^n, \vec{a} \in \text{int}(A), f : A \to \mathbb{R}^m$.
Let $r > 0$ be such that $B_r(\vec{a}) \subseteq A$. If all partial derivatives $\frac{\partial f_i}{\partial x_j}(\vec{a})$ exist on $B_r\vec{a}$ and are continuous at $\vec{a}$, then f is differentiable at $\vec{a}$.

Proof
Note that differentiability of f is equivalent to differentiability of all component functions f_i so it suffices to consider the case $m = 1$.

We will show that the difference $f(\vec{a}+\vec{h}) - f(\vec{a}) - J\vec{h})$ can be expressed in terms of the component-wise difference between values of the partial derivatives. The result then follows by continuity.
Let $\epsilon > 0$. We need a linear map $T : \mathbb{R}^n \to \mathbb{R}^m$ and $\delta > 0$ such that

$$\frac{\left| f(\vec{a}-\vec{h}) - f(\vec{a}) - T(\vec{h}) \right|}{\left\| \vec{h} \right\|} < \epsilon$$

for all $\vec{h} \in \mathbb{R}^n$, $\left\| \vec{h} \right\| < \delta$.
For each $1 \leq j \leq n$, the partial derivative exists and is continuous at $\vec{a}$ so there are $\delta_j \in (0, r)$ such that

$$\left| \frac{\partial f}{\partial x_j}(\vec{x}) - \frac{\partial f}{\partial x_j}(\vec{a}) \right| < \frac{\epsilon}{\sqrt{n}}$$

for all $\vec{x}$ satisfying $\|\vec{x} - \vec{a}\| < \delta_j$
Let $\delta := \min\{\delta_j\}$ and fix $\vec{h} \in B_\delta(\vec{0})$ and define vectors $\vec{v_k}$, $0 \leq k \leq n$ by

$$\vec{v_0} := \vec{a}, \vec{v_k} := \vec{a} + \sum_{j=1}^{k} h_j \vec{e_j} = v_{\vec{k-1}} + h_k \vec{e_j}$$

Note that $\vec{v_k} \in B_r(\vec{a})$ for each $k \in \{0, 1, \dots, n\}$ because

$$\|\vec{v}_k - \vec{a}\| = \sqrt{\sum_{j=1}^{k} h_j^2} \leq \left\| \vec{h} \right\| < r$$

Also, $\vec{v}_n = \vec{a} + \vec{h}$ so

$$f(\vec{a}+\vec{h}) - f(\vec{a}) = f(\vec{v}_n) - f(\vec{v}_0) = \sum_{k=1}^{n} [f(\vec{v}_k) - f(\vec{v}_{k-1})]$$

by the telescoping sum
For each $k \in \{1, \dots, n\}$, consider the line segment $S_k := \{\vec{v}_{k-1} + t\vec{e}_k : 0 \leq t \leq h_k\}$ from $\vec{v}_{k-1}$ to $\vec{v}_k$.
Suppose $h_k > 0$, else we simply take $h_k \leq t \leq 0$ instead.
Since $B_r(\vec{a})$ is convex, each line segment is a subset of the ball.
Thus, we can define a function $g_k : [0, h_k] \to \mathbb{R}$ by

$$g_k(t) := f(\vec{v}_{k-1} + t\vec{e}_k) = f(a_1 + h_1, \dots, a_{k-1} + h_{k-1}, a_k + t, a_{k+1}, \dots, a_n)$$

Note that the derivative of g_k exists and is precisely the partial derivative of f with respect to x_k:

$$g_k'(t) = \frac{\partial f}{\partial x_k}(\vec{v}_{k-1} + t\vec{e}_k)$$

By assumption, the partial derivatives of f exist on $B_r(\vec{a})$ so each g_k is differentiable on $(0, h_k)$ and continuous on $[0, h_k]$.
Hence, the Mean Value Theorem tells us that there exists $t_k \in (0, h_k)$ such that

$$g_k'(t_k)h_k = g_k(h_k) - g_k(0) \iff \frac{\partial f}{\partial x_k}(\vec{c}_k)h_k = f(\vec{v}_k) - f(\vec{v}_{k-1})$$

where $\vec{c}_k := \vec{v}_{k-1} + t_k\vec{e}_k$
Then, by the accumulation of our work above

$$f(\vec{a} + \vec{h}) - f(\vec{a}) = \sum_{k=1}^{n} h_k \frac{\partial f}{\partial x_k}(\vec{c}_k)$$

Recall that the Jacobian is the derivative $Df(\vec{a})$ (if it exists) with respect to the standard basis.
So $Df(\vec{a})(\vec{h}) = J\vec{h}$.
We claim that this is the linear map that satisfies the definition of differentiability.
We have

$$f(\vec{a} + \vec{h}) - f(\vec{a}) - J\vec{h} = \sum_{k=1}^{n} \left[\frac{\partial f}{\partial x_k}(\vec{c}_k) - \frac{\partial f}{\partial x_k}(\vec{a})\right] h_k = \langle \vec{w}, \vec{h}\rangle$$

Where $\vec{w}$ is the vector with components $w_k = \left[\frac{\partial f}{\partial x_k}(\vec{c_k}) - \frac{\partial f}{\partial x_k}(\vec{a})\right]$.
By the Cauchy-Schwartz Inequality, we can write

$$\left\| f(\vec{a} + \vec{h}) - f(\vec{a}) - J\vec{h} \right\| \leq \|\vec{w}\| \left\|\vec{h}\right\|$$
$$\frac{\left| f(\vec{a} + \vec{h}) - f(\vec{a}) - J\vec{h} \right|}{\left\|\vec{h}\right\|} \leq \sqrt{\sum_{j=1}^{n} w_j^2} < \sqrt{\sum_{j=1}^{n} \frac{\epsilon^2}{n}} = \epsilon$$

Note that the above relies heavily on the assumptions that the partial derivatives exist and are continuous!

15 Examples and Combinations of Functions

15.1 Conditions in Theorem 14.3 are not necessary for differentiability

Example 15.1.1
Consider $f : \mathbb{R}^2 \to \mathbb{R}$

$$f(x,y) = \begin{cases} 0, x = y = 0 \\ (x^2+y^2)\sin\left(\frac{1}{\sqrt{x^2+y^2}}\right) \end{cases}$$

It can be shown that partial derivatives exist everywhere but is not continuous at the origin, yet the derivative exists at the origin.
The derivative at the origin is given by

$$Df(0,0) = (0,0)$$

Proof
Write

$$\frac{\left|f(\vec{h}) - f(\vec{0}) - J\vec{h}\right|}{\left\|\vec{h}\right\|} = \frac{\left\|\vec{h}\right\| \left|\sin\left(\frac{1}{\left\|\vec{h}\right\|}\right)\right|}{\left\|\vec{h}\right\|}$$

15.2 Examples of Computing Derivatives

Example 15.2.1
for $f(x,y,z) = x^2y^4 + z$, find

$$\frac{\partial}{\partial y}(2,1,-1)$$

Example 15.2.2
Find the Jacobian Matric for $(x,y) = (x+2y, xy, e^{-y}\sin y)$ at $(\pi/2, 1)$.

16 Combinations of Differentiable Functions

16.1 Rules for Differentiating Combinations of Functions

Theorem 16.1.1 (Chain Rule)
Let $A \subseteq \mathbb{R}^n, B \subseteq \mathbb{R}^m, f : A \to B, g : B \to \mathbb{R}^l$.
If f is differentiable at $\vec{a}$ and g is differentiable at $f(\vec{a})$,
then the composition $h = g \circ f$ if differentiable at $\vec{a}$ and the derivative is

$$Dh(\vec{a}) = Dg(f(\vec{a})) \circ Df(\vec{a})$$

Proof
write

$$r_f(\vec{h}) = f(\vec{a}+\vec{h}) - f(\vec{a}) - Df(\vec{a})(\vec{h})$$

and

$$r_g(\vec{t}) = g(\vec{b}+\vec{t}) - g(\vec{b}) - Dg(\vec{b})(\vec{t})$$

By differentiability,

$$\lim_{\vec{h}\to\vec{0}} \frac{\left\|r_f(\vec{h})\right\|}{\|h\|} = 0, \lim_{\vec{t}\to\vec{0}} \frac{\|r_g(\vec{t})\|}{\|t\|} = 0$$

We wish to show that

$$\lim_{\vec{h}\to\vec{0}} \frac{\left\|r_h(\vec{h})\right\|}{\|h\|} = 0$$

where

$$r_h(\vec{h}) = h(\vec{a}+\vec{h}) - h(\vec{a}) - Dh(\vec{a})(\vec{h}) = g(f(\vec{a}+\vec{h})) - g(f(\vec{a})) - Dg(f(\vec{a}))Df(\vec{a})(\vec{h})$$

Define $\vec{t} = f(\vec{a}+\vec{h}) - f(\vec{a}) = f(\vec{a}+\vec{h}) - \vec{b}$.
By the continuity (differentiability) of f at $\vec{a}$, $\left\|\vec{t}\right\|$ will be small if $\left\|\vec{h}\right\|$ is small.
Then,

$$\begin{aligned} g(f(\vec{a}+\vec{h})) &= g(\vec{b}+\vec{t}) \\ &= g(\vec{b}) + Dg(\vec{b})(\vec{t}) + r_g(\vec{t}) \\ &= g(\vec{b}) + Dg(\vec{b})(f(\vec{a}+\vec{h}) - f(\vec{a})) + r_g(\vec{t}) \\ &= g(\vec{b}) + Dg(\vec{b})(Df(\vec{a})(\vec{h})) - Dg(\vec{b})(r_f(\vec{h})) + r_g(\vec{t}) \qquad \text{linearity} \end{aligned}$$

From here,

$$\frac{\left\|r_h(\vec{h})\right\|}{\left\|\vec{h}\right\|} = \frac{\left\|Dg(\vec{b})(r_f(\vec{h})) + r_g(\vec{t})\right\|}{\left\|\vec{h}\right\|} \qquad \text{cancellation}$$

$$\leq \frac{\left\|Dg(\vec{b})(r_f(\vec{h}))\right\|}{\left\|\vec{h}\right\|} + \frac{\left\|r_g(\vec{t})\right\|}{\left\|\vec{h}\right\|}$$

We can write

$$\left\|D_g(\vec{b})(r_f(\vec{h}))\right\| \leq M\left\|r_f(\vec{h})\right\|$$

Where M denotes the Frobenius Norm of the Matrix respresenting $D_g(\vec{b})$.
So

$$\frac{\left\|D_g(\vec{b})(r_f(\vec{h}))\right\|}{\left\|\vec{h}\right\|} \leq \frac{\left\|r_f(\vec{h})\right\|}{\left\|\vec{h}\right\|}$$

And hence, $\lim_{\vec{h}\to\vec{0}} = 0$ for the above.

Theorem 16.1.2
Let $A \subseteq \mathbb{R}^n, \vec{a} \in \text{int}(A), f, g : A \to \mathbb{R}$ differentiable at $\vec{a}$.
Then

1. $D(f+g)(\vec{a}) = Df(\vec{a}) + Dg(\vec{a})$
2. $D(\alpha f)(\vec{a}) = \alpha Df(\vec{a})$ for any scalar
3. $D(fg)(\vec{a}) = f(\vec{a})Dg(\vec{a}) + g(\vec{a})Df(\vec{a})$
4. $D(f/g)(\vec{a}) = \frac{g \cdot Df(\vec{a}) - f \cdot Dg(\vec{a})}{g(\vec{a})^2}$

Proof (1)
$h(x) = (f(g), g(x))$.
Consider function $s : \mathbb{R}^2 \to \mathbb{R}$ defined by $s(x, y) = x + y$.
Note that $f + g = s \circ h$.
Then we simply apply the Chain Rule.

16.2 Examples

Proposition 16.2.1
Every polynomial in the Reals is differentiable on $\mathbb{R}^n$.

Proposition 16.2.2
Let p, q be two polynomials with $A := \{\vec{a} : q \neq \vec{0}\} \neq \emptyset$.
Then $\frac{p}{q} : A \to \mathbb{R}$ is differentiable on A by the quotient rule.
Note that we should check that A is open

17 Mean Value Theorem and Gradients

17.1 Mean Value Theorem

Theorem 17.1.1 (Mean Value Theorem)
Let $A \subseteq \mathbb{R}^n, \vec{a}, \vec{b} \in A$.
define the set

$$S := \{\vec{a} + t(\vec{b} - \vec{a}) : 0 < t < 1\}$$

Suppose $S \subseteq \text{int}(A)$ and $\bar{S} \subseteq A$.
If $f : A \to \mathbb{R}$ is continuous on $\bar{S}$ and differentiable on S, then there exists $\vec{c} \in S$ such that

$$f(\vec{b}) - f(\vec{a}) = Df(\vec{c})(\vec{b} - \vec{a})$$

Proof
let $\phi : [0, 1] \to S$ be a function defined by $\phi(t) = \vec{a} + t(\vec{b} - \vec{a})$.
Note that ϕ is continuous on $[0, 1]$ and differentiable on $(0, 1)$.
Now, define $g : [0, 1] \to \mathbb{R}$ by $g(t) = f(\phi(t))$.
Since f and ϕ are continuous g is also continuous.
Also, g is differentiable by similar logic (Theorem 16.1.1).
In particular, $D\phi(t) = \vec{b} - \vec{a}$.
So

$$\begin{aligned} Dg(t) &= Df(\phi(t)) \circ D\phi(t) \\ &= Df(\phi(t))(\vec{b} - \vec{a}) \end{aligned}$$

By construction, $\phi(0) = \vec{a}, \phi(1) = \vec{b}$.
Applying Mean Value Theorem for scalar valued functions of a single variable, there exists some $t_0 \in (0, 1)$ such that

$$f(\vec{b}) - f(\vec{a}) = g(1) - g(0) = Dg(t_0)(1 - 0) = Dg(t_0) = Df(\vec{c})(\vec{b} - \vec{a})$$

where $\vec{c} = \phi(t_0) \in S$.

17.2 Linear Approximation

Definition 17.2.1
Consider $A \subseteq \mathbb{R}^n$ and $\vec{a} \in \text{int}(A)$.
Recall that if a function $f : A \to \mathbb{R}$ is differentiable at $\vec{a}$, then

$$f(\vec{x}) = f(\vec{a}) + Df(\vec{a})(\vec{x} - \vec{a}) + r(\vec{x} - \vec{a})$$

where the function r is continuous at $r(\vec{0}) = 0$.
So if $\|\vec{x} - \vec{a}\|$ is small, then $|r|$ should be small.

$$f(\vec{x}) \approx f(\vec{a}) + Df(\vec{a})(\vec{x} - \vec{a})$$

we define the function

$$l^f_{\vec{a}} = f(\vec{a}) + Df(\vec{a})(\vec{x} - \vec{a})$$

and we call this the **linear approximation** to f at $\vec{a}$.

Definition 17.2.2 (gradient)
the **gradient** of a function $f : A \to \mathbb{R}$ at $\vec{a}$ is the vector

$$\nabla f(\vec{a}) := \left(\frac{\partial f}{\partial x_i}(\vec{a})\right) = J^T$$

Note that is it common to extend this definition to vector valued functions.
The differential operator

$$\nabla := \left(\frac{\partial}{\partial x_i}\right)$$

is called "nabla" or "del" or "grad".

$$\nabla \cdot f = \sum \frac{\partial f_i}{\partial x_i}$$

$$\nabla \times \vec{v}$$

Definition 17.2.3
Consider a scalar-valued function $f : A \to \mathbb{R}$, where $A \subseteq \mathbb{R}^n$.
The graph of f is the set

$$S_f := \{(\vec{x}, f(\vec{x})) : \vec{x} \in A\}$$

In general, if f is continuous, its graph will be an n-dimensional hypersurface define by

$$x_{n+1} = f(\vec{x})$$

embedded in $\mathbb{R}^{n+1}$.
The graph of the linear approximation to f at $\vec{a}$ is the n-dimensional tangent hyperplane of S_f at $(\vec{a}, f(\vec{a}))$.

Theorem 17.2.1
$A \subseteq \mathbb{R}^n, \vec{a} \in \operatorname{int}(A), f : A \to \mathbb{R}$.
If f is differentiable at $\vec{a}$, then

1. $\vec{v} := (\nabla f(\vec{a}), -1) \in \mathbb{R}^{n+1}$ is orthogonal to the tangent hyperplane of the hyperspace $x_{n+1} = f(\vec{x})$ at the point $(\vec{a}, f(\vec{a}))$.

2. If $\nabla f(\vec{a}) \neq \vec{0}$, then the directional derivative $D_{\vec{u}} f(\vec{a})$ is maximised over all unit vectors $\vec{u}$ when
$$\vec{u} = \frac{\nabla f(\vec{a})}{\|\nabla f(\vec{a})\|}$$
and minimised when
$$\vec{u} = -\frac{\nabla f(\vec{a})}{\|\nabla f(\vec{a})\|}$$

Proof (Part 2: From Theorem 13.1.3)

$$D_{\vec{u}} f(\vec{a})(\vec{u}) = \langle \nabla f(\vec{a}), \vec{u} \rangle$$

By Cauchy-Schwartz,

$$|\langle \nabla f(\vec{a}), \vec{u} \rangle| \leq \|\nabla f(\vec{a})\| \|\vec{u}\| = \|\nabla f(\vec{a})\|$$

with equality if and only if they are linearly dependent.
By assumption, the gradient is non-zero so $\vec{u} = \alpha \nabla f(\vec{a})$.
Since $\vec{u}$ is a unit vector
$$\alpha = \pm \frac{1}{\|\nabla f(\vec{a})\|}$$
So $|\langle \nabla f(\vec{a}), \vec{u} \rangle| = \|\nabla f(\vec{a})\|$ if and only if

$$\alpha = \pm \frac{\nabla f(\vec{a})}{\|\nabla f(\vec{a})\|}$$

$D_{\vec{u}} f(\vec{a}) = \langle \nabla f(\vec{a}), \vec{u} \rangle > 0$ when $\vec{u} = \vec{u}_+$ and vice versa.

Proof (Part 1 Sketch for $n = 1$ case)
$\vec{v} = (\nabla f(\vec{a}), -1)$

18 Higher Order Derivatives

18.1 Higher Order Derivatives

Let $A \subseteq \mathbb{R}^n, f : A \to \mathbb{R}$.
We have already defined the partial first order derivatives for $\vec{a} \in \text{int}(A)$.

Definition 18.1.1 (first order partial derivative function)
$A \subseteq \mathbb{R}^n, f : A \to \mathbb{R}$
The function

$$\frac{\partial f}{\partial x_j} : \text{int}(A) \to \mathbb{R}$$

is called a **first order partial derivative function**.

Definition 18.1.2 (second order partial derivative function)
For $j = 1, \dots, n$

$$\frac{\partial^2 f}{\partial x_j \partial x_i} := \frac{\partial}{\partial x_j}\left(\frac{\partial f}{\partial x_i}\right)$$

We the (inductively) define higher order partial derivatives.

$$\frac{\partial^m f}{\partial x_{i_m} \partial x_{i_{m-1}} \dots \partial x_{i_1}} := \frac{\partial}{\partial x_{i_m}} \frac{\partial^{m-1} f}{\partial x_{i_{m-1}} \dots \partial x_{i_1}}$$

It is common to write

1. $\frac{\partial f}{\partial x_i} = f_{x_i}$
2. $\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}$
3. $\frac{\partial^2 f}{\partial x_j \partial x_i} = f_{x_i x_j}$
4. $\frac{\partial^2 f}{\partial x \partial y}$
5. $\frac{\partial^2 f}{\partial x^2}$

18.2 Example

Example 18.2.1
second order partial of $f(x, y) = x^2 y^3$

Example 18.2.2
first and second order partial of $f(x, y) = x^2 y^2$

18.3 Mixed Partial Derivatives

Theorem 18.3.1 (Equality of mixed Partial Derivative)
$A \subseteq \mathbb{R}^n, \vec{a} \in \text{int}(A), f : A \to \mathbb{R}$.
$B_\delta(\vec{a}) \subseteq A$ for some $\delta > 0$.
Suppose that for some $i, j \in \{1, \dots n\}$, the partial derivatives $f_{x_i}, f_{x_j}, f_{x_i x_j}, f_{x_j x_i}$ exist on the ball and are continuous at $\vec{a}$, then

$$f_{x_i x_j} = f_{x_j x_i}$$

We begin by nothing several things.
Consider $n = 2, x_i = 1, x_j = 2, f_{x_i x_j}$.

$$\begin{aligned}
f_{xy}(a,b) &= \lim_{h \to 0} \frac{f_y(a+h,b) - f_y(a,b)}{h} \\
&= \lim_{h \to 0} \left[\lim_{k \to 0} \frac{f(a+h,b+k) - f(a,b)}{k} - \lim_{k \to 0} \frac{f(a,b+h) - f(a,b)}{h} \right] \\
&= \lim_{h \to 0} \lim_{k \to 0} \frac{1}{hk} \left[f(a+h,b+k) - f(a+h,b) - f(a,b+k) + f(a,b) \right]
\end{aligned}$$

But we cannot proceed as we **need** to be able to exchange limits, which is not always valid.
Consider

$$f(x,y) = \frac{x^2 - y^2}{x^2 + y^2}$$
$$\lim_{x \to 0} \lim_{y \to 0} f(x,y) = 1$$
$$\lim_{y \to 0} \lim_{x \to 0} f(x,y) = -1$$

We will note that the above applies to derivatives.
Consider

$$f(x,y) = xyf(x,y)$$

$$
\begin{aligned}
g_x(x,y) &= yf(x,y) + xyf_x(x,y) \\
g_y(x,y) &= xf(x,y) + xyf_y(x,y) \\
g_{xy}(x,y) &= f(x,y) + yf_y(x,y) + xf_x(x,y) + yxf_{xy}(x,y) \\
g_{yx}(x,y) &= f(x,y) + xf_x(x,y) + yf_y(x,y) + xyf_{yx}(x,y) \\
g_{xy}(0,0) &= \lim_{x\to 0}\lim_{y\to 0} f(x,y) \\
g_{yx}(0,0) &= \lim_{y\to 0}\lim_{x\to 0} f(x,y)
\end{aligned}
$$

Proof

The actual proof will be motivative by differences of limits.
We must however be very careful HOW we take these limits.
Consider $n = 2$.
$x_i = 1, x_j = 2$. In other words $f_{xy}(a,b) = f_{yx}(a,b)$.
Let $p, q \in \mathbb{R}$ satisfy $0 < p, q < \frac{\delta}{\sqrt{2}}$.
Note that $(a,b), (a+q,b), (a+q,b+p), (a,b+p)$ are all in $B_\delta(a,b)$.
Define the functions

$$
\begin{aligned}
g(s) &:= f(a+s,b+p) - f(a+s,b) \\
h(t) &:= f(a+q,b+t) - f(a,b+t)
\end{aligned}
$$

The above are well defined functions if $0 < t < p, 0 < s < q$.
Note that g, h are differentiable (and hence continuous) by assumption since we are given that first order derivatives exist inside the ball.

$$g(q) - g(0) = f(a+q,b+p) - f(a+q,b) - f(a,b+p) + f(a,b)$$

$$h(p) - h(0) = f(a+q,b+p) - f(a,b+p) - f(a+q,b) + f(a,b)$$

So $g(q) - g(0) = h(p) - h(0)$.
By the Mean Value Theorem

$$g(q) - g(0) = g'(s_0)q = h'(t_0)p = h(p) - h(0)$$

for some s_0, t_0.
By definitions of g, h

$$g'(s_0) = f_x(a+s_0,b+p) - f_x(a+s_0,b) = \varphi(p) - \varphi(0)$$

where $\varphi(u) := f_x(a+s_0,b+u)$ and similarly for $h'(t_0)$.
By the Mean Value Theorem again,

$$g'(s_0)q = [\varphi(p) - \varphi(0)]\, q = \varphi'(u_0)qp = f_{xy}(a+s_0,b+u_0)pq$$

for some $u_0 \in (0,p)$.

Do this for $h'(t_0)$.

$$f_{xy}(a+s_0, b+u_0) = f_{yx}(a+v_0, b+t_0)$$

for some $v_0 \in (0, q)$.
The final step requires the use of sequential continuity of $f_{xy}(a, b)$ and $f_{yx}(x, y)$ to show the result.

Corollary 18.3.1.1
Let $A \subseteq \mathbb{R}^n$ be nonempty and open.
If $f \in C^k(A, \mathbb{R})$ for some $k \geq 2$,
For any partial derivative of order $2 \leq l \leq k$, the order in which partial derivatives are taken do not matter.

19 Taylor's Theorem

19.1 Continuously differentiable functions

Notation:
$C(A, \mathbb{R})$ is the set of continuous functions from $A \subseteq \mathbb{R}^n$ to $\mathbb{R}$.
If $A \subseteq \mathbb{R}^n$ is open, then $C^k(A, \mathbb{R})$ denotes the set of functions from $A \to \mathbb{R}$ for which all partial derivatives up to and including order k exist and are continuous on A.
If A is not open and has a nonempty interior, then $C^k(A, \mathbb{R})$ denotes the set of functions for which all partial derivatives up to and including order k exist and are continuous on $\text{int}(A)$, and are extendable to continuous functions on A.
We can shorten the notation to $C^k(A)$
We say that a function $f \in C^k(A, \mathbb{R})$ is **of class** C^k (on A).
If $A \subseteq \mathbb{R}$ then $C^k(A, \mathbb{R})$ is the set of k-times continuously differentiable (real-valued) functions.

19.2 Taylor's Theorem

Theorem 19.2.1 (Taylor's Theorem for one variable)
$I \subseteq \mathbb{R}$ be an interval, $a \in I$.
If $f : I \to \mathbb{R}$ is $p+1$ times differentiable on I for some integer $p \geq 0$, then for any $x \in I$ with $x \neq a$, there is $\xi \in (a, x)$ such that

$$f(x) = \sum_{k=0}^{p} \frac{f^{(k)}(a)}{k!}(x-a)^k + \frac{f^{(p+1)}(\xi)}{(p+1)!}(x-a)^{p+1}$$

Proof
Textbook

Recall the gradient we will use this notationally like a vector

1. $$[(\vec{h} \cdot \nabla) f](\vec{a}) = [(h_1 \frac{\partial}{\partial x_1} + \cdots + h_n \frac{\partial}{\partial x_n}) f](\vec{a})$$

2. for $n = 2$
$$[(\vec{h} \cdot \nabla)^2 f](\vec{a}) = [(h_1 \frac{\partial}{\partial x_1} + h_2 \frac{\partial}{\partial x_2})^2 f](\vec{a})$$

Theorem 19.2.2 (Taylor's Theorem)
Let $U \subseteq \mathbb{R}^n$ be open and convex.
$f \in C^{p+1}(U, \mathbb{R})$ for some integer $p \geq 0$.
Suppose $\vec{a}, \vec{x} \in U$ and define $\vec{h} := \vec{x} - \vec{a}$.
Then there is some $\xi \in (0, 1)$ such that

$$f(\vec{x}) = f(\vec{a} + \vec{h}) = f(\vec{a}) + \sum_{k=1}^{p} \frac{1}{k!} \left[(\vec{h} \cdot \nabla)^k f\right](\vec{a}) + R_p(\xi)$$

where $R_p(t) := \frac{1}{p+1} \left[(\vec{h} \cdot \nabla)^{p+1} f\right](\vec{a} + t\vec{h})$.

Proof
By convexity of U, the line segment $\{\vec{a} + t\vec{h} : \vec{0} < t < 1\}$ is contained in U for any $\vec{a}, \vec{x} = \vec{a} + \vec{h} \in U$.
Define $g : [0, 1] \to \mathbb{R}$ by $g(t) = f(\vec{a} + t\vec{h})$.
By the Chain Rule, g is differentiable and

$$g'(t) = Df(\vec{a} + t\vec{h})(\vec{h}) = \nabla f\left(\vec{a} + t\vec{h}\right) \cdot \vec{h} = \left[\left(\vec{h} \cdot \nabla\right) f\right]\left(\vec{a} + t\vec{h}\right)$$

We proceed by induction to show that

$$g^{(k)}(t) = \left[\left(\vec{h} \cdot \nabla\right)^k f\right]\left(\vec{a} + t\vec{h}\right)$$

for each $0 \leq k \leq p + 1$.
For $k = 0$, $g(t) = f(\vec{a} + t\vec{h})$ by definition.
We have already shown the case $k = 1$.
For induction, suppose that the result holds for $k = m$ for some $m \in \{1, \ldots, p\}$.
So $g^{(m)}(t)$ is differentiable and by the Chain Rule:

$$\begin{aligned}
g^{(m+1)}(t) &= \left[\nabla\left(\vec{h} \cdot \nabla\right)^m f\right]\left(\vec{a} + t\vec{h}\right) \cdot \nabla(\vec{a} + t\vec{h}) \\
&= \left[\nabla\left(\vec{h} \cdot \nabla\right)^m f\right]\left(\vec{a} + t\vec{h}\right) \cdot \vec{h} \\
&= \sum_{i=1}^{n} \frac{\partial}{\partial x_i}\left\{\left(\vec{h} \cdot \nabla\right)^m f\right\}\left(\vec{a} + t\vec{h}\right) h_i \\
&= \sum_{i=1}^{n} h_i \frac{\partial}{\partial x_i}\left\{\left(\vec{h} \cdot \nabla\right)^m f\right\}\left(\vec{a} + t\vec{h}\right) \\
&= \left(\vec{h} \cdot \nabla\right)\left\{\left(\vec{h} \cdot \nabla\right)^m f\right\}\left(\vec{a} + t\vec{h}\right) \\
&= \left[\left(\vec{h} \cdot \nabla\right)^{m+1} f\right]\left(\vec{a} + t\vec{h}\right)
\end{aligned}$$

Note that g is of class C^{p+1} thus by applying Taylor's Theorem for one variable to g with endpoints $0, 1$ shows that

$$g(1) = g(0) + \sum_{k=1}^{p} \frac{g^{(k)}(0)}{k!}(1-0)^k + \frac{g^{(p+1)}(\xi)}{(p+1)!}(1-0)^{p+1}$$

$$f(\vec{a}+\vec{h}) = f(\vec{a}) + \sum_{k=1}^{p} \frac{\left[\left(\vec{h}\cdot\nabla\right)^k f\right](\vec{a})}{k!} + \frac{\left[\left(\vec{h}\cdot\nabla\right)^{p+1} f\right](\vec{a}+\xi\vec{h})}{(p+1)!}$$

We can write Taylor's Theorem in a different way using the notation

$$\left(D^{(k)}f\right)_{\vec{a}}(\vec{h}) := \begin{cases} f(\vec{a}) & , k = 0 \\ \sum_{j_1,\ldots,j_k=1}^{n} f_{x_{j_k}\ldots x_{j_1}}(\vec{a})h_{j_1}\ldots h_{j_k} \end{cases}$$

with this notation, Taylor's Theorem becomes

$$f(\vec{a}+\vec{h}) = \sum_{k=0}^{p} \frac{\left(D^{(k)}f\right)_{\vec{a}}(\vec{h})}{k!} + \frac{1}{(p+1)!}\left(D^{(p+1)}f\right)_{\vec{a}+\xi\vec{h}}(\vec{h})$$

Corollary 19.2.2.1
$U \subseteq \mathbb{R}^n$ open, $f \in C^1(U, \mathbb{R})$.
If $K \subseteq U$ is compact, then there is some $M > 0$ such that for any convex subset $C \subseteq K$

$$\|f(\vec{x}) - f(\vec{y})\| \leq M\|\vec{x} - \vec{y}\|$$

So f becomes Lipschitz on the convex set.

20 Taylor Polynomials and Critical Points

20.1 Taylor Poynomials

Definition 20.1.1
$U \subseteq \mathbb{R}^n$ open (not necessarily convex).
$f \in C^p(U, \mathbb{R})$ for some integer $p \geq 0$.
Suppose $\vec{a}, \vec{x} \in U$, the **Taylor Polynomial** of order p for f at the points $\vec{a}$ is

$$p^f_{p,\vec{a}}(\vec{x}) := f(\vec{a}) + \sum_{k=1}^{p} \frac{1}{k!} \left[\left(\vec{h} \cdot \nabla \right)^k f \right] (\vec{a})$$

Definition 20.1.2
Taylor remainder of order p is

$$R^f_{p,\vec{a}}(\vec{x}) := f(\vec{x}) - P^f_{p,\vec{a}}(\vec{x})$$

Theorem 20.1.1 (Alternate Taylor's Theorem)
$U \subseteq \mathbb{R}^n$ open, $f \in C^p(U, \mathbb{R})$ for some integer $p \geq 0$.
For any $\vec{a} \in U$ the Taylor remainder of order p satisfies

$$\lim_{\vec{x} \to \vec{a}} \frac{R^f_{p,\vec{a}}(\vec{x})}{\|\vec{x} - \vec{a}\|^p} = 0$$

20.2 Critical Points

Definition 20.2.1
$A \subseteq \mathbb{R}^n, \vec{a} \in \text{int}(A), f : A \to \mathbb{R}$.
$\vec{a}$ is a

1. **critical point** (stationary point) of f is $\nabla f(\vec{a}) = 0$.
2. **local maximum** of f is there is some $\delta > 0$ such that $f(\vec{x}) \leq f(\vec{a})$ for all $\vec{x} \in B_\delta(\vec{a})$.
3. **local minimum** of f if ...
4. **saddle point** of f is for any $\delta > 0$ there is some $\vec{x}, \vec{y} \in B_\delta(\vec{a})$ such that $f(\vec{x}) < f(\vec{a}) < f(\vec{y})$

Theorem 20.2.1
$A \subseteq \mathbb{R}^n, \vec{a} \in \text{int}(A), f : A \to \mathbb{R}$.
If $\vec{a}$ is a local minimum/maximum and the gradient exists at $\vec{a}$, then it must be zero.

Proof
Exercise

21 Second Derivative Test

21.1 The Hessian Matrix

Let $U \subseteq \mathbb{R}^n$ be open and let $f \in C^2(U, \mathbb{R})$.
Suppose $\vec{a} \in U$ is a critical point of f.
By Taylor's Theorem, for sufficiently small $\left\|\vec{h}\right\|$, there is some

$$\vec{c} \in \{\vec{a} + t\vec{h} : 0 < t < 1\}$$

such that

$$f(\vec{a} + \vec{h}) = f(\vec{a}) + \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} f_{x_j x_i}(\vec{c}) h_i h_j$$

(see explicit example from lecture 19)
The remainder term could tell us $\vec{a}$ is a local minimum of maximum or a saddle.
We can use the Alternative Taylor Theorem

$$f(\vec{a} + \vec{h}) = f(\vec{a}) + \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} f_{x_i x_j}(\vec{c}) h_i h_j + R^f_{2,\vec{a}}(\vec{a} + \vec{h})$$

Definition 21.1.1 (Hessian Matrix)
$H \in \mathbb{R}^{n \times m}$ of f at $\vec{a}$ by

$$H = [H_{ij}]$$

where

$$H_{ij} := f_{x_i x_j}(\vec{a})$$

We will also use the notation $D^2 f(\vec{a})$ for H.

Note that the Hessian is symmetrix by the Equality of Mixed Partial Derivatives.

Definition 21.1.2
$Q : \mathbb{R}^n \to \mathbb{R}$

$$Q(\vec{u}) := \vec{u}^T H \vec{u} = \sum_{i=1}^{n} \sum_{j=1}^{n} u_i H_{ij} u_j$$

So at any critical point

$$f(\vec{a} + \vec{h}) = f(\vec{a}) + \frac{1}{2} Q(\vec{h}) + R^f_{2,\vec{a}}(\vec{a} + \vec{h})$$

21.2 Quadratic Forms

Definition 21.2.1
function $A : \mathbb{R}^n \to R$ is a **quadratic form** if there is a real symmytric matrix $A \in \mathbb{R}^{n\times m}$ such that

$$Q(\vec{u}) = \vec{u}^T A\vec{u}$$

for all $\vec{u} \in \mathbb{R}^n$.

Definition 21.2.2
$Q : \mathbb{R}^n \to \mathbb{R}$ is a quadratic form. Then Q is

1. **positive definite** if $Q(\vec{u}) > 0$ for all $\vec{u} \neq 0$
2. **negative definite** if $Q(\vec{u}) < 0$ for all $\vec{u} \neq 0$
3. **positive semi-definite** if $Q(\vec{u}) \geq 0$ for all $\vec{u}$
4. **negative semi-definite** if $Q(\vec{u}) \leq 0$ for all $\vec{u}$
5. **indefinite** if there exists $\vec{x}, \vec{y} \in \mathbb{R}^n$ such that $Q(\vec{x}) > 0, Q(\vec{y}) < 0$

Note that we can also write equivalent definitions for the matrix A.
Recall from linear algebra

Theorem 21.2.1 (Spectral Theorem for real, symmetrix matrices)
If $A \in \mathbb{R}^{n\times m}$ is symmetrix, then the eigenvalues

$$\{\lambda_i : 1 \leq i \leq n\}$$

are all real and there is an orthonormal matrix $P \in \mathbb{R}^{n\times n}$ such that

$$P^T AP = D$$

Where D is the diagonal matrix of eigenvalues

Define $\vec{v} := P^T\vec{u}$

$$\begin{aligned} Q(\vec{u}) &= (P\vec{v})^T A(P\vec{v}) \\ &= \vec{v}^T D\vec{v} \\ &= \sum_{i=1}^{n} \lambda_i v_i^2 \end{aligned}$$

Proposition 21.2.2
Let $Q : R^n \to \mathbb{R}$ be a quadratic form with associated matrix $A \in \mathbb{R}^{n\times m}$

1. Q is positive definite if all eigenvalues are positive
2. Q is negative definite if all eigenvalues are negative
3. Q is positive semi-definite if all eigenvalues are non-negative
4. Q is negative semi-definite if all eigenvalues are non-positive
5. Q is indefinite if there is one positive and one negative eigenvalue

21.3 Second Derivative Test

Lemma 21.3.1
Let $Q : \mathbb{R}^n \to \mathbb{R}$ be a quadratic form

1. if Q is positive definite, then there is some $M > 0$ such that $Q(\vec{u}) \geq M\|\vec{u}\|$ for all $\vec{u} \in \mathbb{R}^n$
2. if Q is negative definite, then there is some $M > 0$ such that $Q(\vec{u}) \leq -M\|\vec{u}\|$ for all $\vec{u} \in \mathbb{R}^n$

Proof
exercise

Theorem 21.3.2 (Second Derivative Test)
$U \subseteq \mathbb{R}^n$ be open and $f \in C^2(U, \mathbb{R})$.
Suppose $\vec{a} \in U$ be a critical point of f and let $A : \mathbb{R}^n \to \mathbb{R}$ be the quadratic form associated with the Hessian matrix of f at $\vec{a}$.

1. $\vec{a}$ is a local maximum of f is Q is negative definite
2. $\vec{a}$ is a local minimum of f is Q is positive definite
3. $\vec{a}$ is a saddle of f is Q is indefinite

Proof
By the alternative Taylor

$$f(\vec{a}+\vec{h}) = f(\vec{a}) + \frac{1}{2}Q(\vec{h}) = R^f_{2,\vec{a}}(\vec{a}+\vec{h})$$

where $R^f_{2,\vec{a}}$ satisfies

$$\lim_{\vec{h}\to\vec{0}} \frac{R^f_{2,\vec{a}}(\vec{a}+\vec{h})}{\left\|\vec{h}\right\|} = 0$$

Part I
If Q is negative definite ....
By the definition of limits

$$\frac{\left|R^f_{2,\vec{a}}(\vec{a}+\vec{h})\right|}{\left\|\vec{h}\right\|} < \frac{M}{2}$$

whenever $0 < \left\|\vec{h}\right\| < \delta$.
Hence

$$-\frac{M}{2}\left\|\vec{h}\right\|^2 < R^f_{2,\vec{a}}(\vec{a}+\vec{h}) < \frac{M}{2}\left\|\vec{h}\right\|^2$$

whenever $0 < \left\|\vec{h}\right\| < \delta$.
Therefore,

$$\begin{aligned} f(\vec{a}+\vec{h}) &= f(\vec{a}) + \frac{1}{2}Q(\vec{h}) + R^f_{2,\vec{a}}(\vec{a}+\vec{h}) \\ &< f(\vec{a}) - \frac{1}{2}M\left\|\vec{h}\right\|^2 + \frac{M}{2}\left\|\vec{h}\right\|^2 \\ &< f(\vec{a}) && \text{for } \vec{h} \text{ satisfying } 0 < \|h\| < \delta \end{aligned}$$

21.4 Examples

Example 21.4.1
$f(x, y) = x^2 + y^2$ on domain $A = \{(x, y) : x^2 + y^2 \leq 1\}$.
f has a critical point at the origin.

$$H = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$$

The eigenvalues are $2, 2$ so H is positive definite, the origin is a local minimum

Example 21.4.2
$f(x,y) = x^2 + y^4$

$$\nabla f(x,y) = (2x, 4y^3)$$

So the critical point is the origin with the Hessian given by

$$H = \begin{pmatrix} 2 & 0 \\ 0 & 12y^2 \end{pmatrix}$$

so the Hessian does not tell us anything

We say that the critical point is **degenerate** if the Hessian exists but $\det(H) = 0$.

22 Inverse Function Theorem I

22.1 Invertibility

Suppose $A \subseteq \mathbb{R}^n, B \subseteq \mathbb{R}^m, f : A \to B$.
For any input $\vec{x} \in A$, can we find $\vec{y} \in B$ such that $\vec{y} = f(\vec{x})$ ($\vec{y}$ is unique)?
Is there a function $g : B \to A$ such that for any $\vec{x} \in A, \vec{y} \in B$,

$$g(\vec{y}) = \vec{x} \iff f(\vec{x}) = \vec{y}$$

If $f \in C^k$ and is invertible, is its inverse also in C^k?

Definition 22.1.1 (inverse function)
$f^{-1} : B \to A$ such that

$$f^{-1}(\vec{y}) = \vec{x} \iff f(\vec{x}) = \vec{y}$$

Note that the inverse function is unique
Also note that if f is invertible then f^{-1} is invertible and f is the inverse of f^{-1}.

Proposition 22.1.1
If $f : A \to B$ is inveritble then

1. $f^{-1} \circ f(\vec{x}) = \vec{x}$
2. $f \circ f^{-1}(\vec{y}) = \vec{y}$

22.2 Inverse Function Theorem

We now focus on propeties of function of class $C^k, k \geq 1$.
For any point $\vec{a} \in A$, we can define the linear approximation of f near $\vec{a}$

$$l^f_{\vec{a}}(\vec{x}) := \vec{a} + Df(\vec{a})(\vec{x} - \vec{a})$$

Provided that the Jacobian Matrix is invertible, we can write

$$\vec{x} - \vec{a} = [Df(\vec{a})]^{-1} \left(l^f_{\vec{a}}(\vec{x}) - l^f_{\vec{a}}(\vec{a}) \right)$$
$$[l^f_{\vec{a}}]^{-1}(\vec{y}) = \vec{a} + [Df(\vec{a})]^{-1}[\vec{y} - f(\vec{a})]$$

So for $\vec{x}$ near $\vec{a}$ we expect f to also be invertible with

$$f^{-1}(\vec{y}) \approx \vec{a} + [Df(\vec{a})]^{-1}[\vec{y} - f(\vec{a})]$$

Recall, the invertibility of a matrix $Df(\vec{a})$ means

$$\det Df(\vec{a}) \neq 0$$

This suggests that the condition for the determinant not to be zero is sufficient for local invertibility of f, while also suggests a formula for the derivative of the inverse.

Definition 22.2.1 (locally invertible)
If the restriction of f to some nonempty subset of the domain is invertible

Theorem 22.2.1 (Global Inverse Function Theorem)
Let $U \subseteq \mathbb{R}^n$ be nonempty and open, suppose $V \subseteq \mathbb{R}^n$, and $f \in C^k(U, V)$ be a bijective, for some $k \geq 1$.
If $\det Df(\vec{a}) \neq 0$ for all $\vec{x} \in U$, then

1. $V = f(U)$ is open
2. $f^{-1} \in C^k(V, U)$
3. $Df^{-1}(f(\vec{x})) = [Df(\vec{x})]^{-1}$ for every $\vec{x} \in U$

Proof (Part I)
Since f is a bijection, $f(U) = V$.
By Lemma 23.1.1, V is open

Proof (Part II)
By Lemma 23.1.4

Proof (Part III)
Let $\vec{a} \in U$ and $\vec{b} := f(\vec{a}) \in V$.
Define the Jacobian Matrix $J := Df(\vec{a})$.
We will show that f^{-1} is differentiable at $\vec{b}$ and that $Df^{-1}(\vec{b}) = J^{-1}$.
Define a function $u : U \to \mathbb{R}^n$ by

$$u(\vec{x}) = \begin{cases} \frac{f(\vec{x}) - \vec{b} - J(\vec{x} - \vec{a})}{\|\vec{x} - \vec{a}\|}, & \vec{x} \neq \vec{a} \\ 0, & \vec{x} = \vec{a} \end{cases}$$

Note that u is continuous at $\vec{a}$ (and all other points) as it is a product / quotient of continuous functions.
Since $\det J \neq 0$ we have (*)

$$\|\vec{x} - \vec{a}\| J^{-1} u(\vec{a}) = J^{-1}[f(\vec{x}) - \vec{b}] - (\vec{x} - \vec{a})$$

Note that this holds for all $\vec{a}$ including $\vec{a}$.
By Lemma 23.1.2, there exists $\delta, m > 0$ such that $B_\delta(\vec{a}) \subseteq U$ and (**)

$$\left\| f(\vec{x}) - \vec{b} \right\| \geq m \|\vec{x} - \vec{a}\|$$

for all $\vec{x} \in B_\delta(\vec{a})$.
WLOG pick δ such that $\overline{B_\delta(\vec{a})} \subseteq U$.
Given any $\vec{y} \in f(B_\delta(\vec{a}))$, let $\vec{x} := f^{-1}(\vec{y}) \in B_\delta(\vec{a})$. Then

$$\begin{aligned}
\frac{\left\| f^{-1}(\vec{y}) - f^{-1}(\vec{b}) - J^{-1}(\vec{y} - \vec{b}) \right\|}{\left\| \vec{y} - \vec{b} \right\|} &= \frac{\left\| \vec{x} - \vec{a} - J^{-1}\left[f(\vec{x}) - \vec{b} \right] \right\|}{\left\| \vec{y} - \vec{b} \right\|} \\
&= \frac{\|\vec{x} - \vec{a}\| \|J^{-1}u(\vec{x})\|}{\left\| \vec{y} - \vec{b} \right\|} && \text{from (*)} \\
&\le \frac{\frac{1}{m}\left\| \vec{y} - \vec{b} \right\| \|J^{-1}u(\vec{x})\|}{\left\| \vec{y} - \vec{b} \right\|} && \text{from (**)} \\
&= \frac{1}{m}\left\| J^{-1}u(f^{-1}(\vec{y})) \right\| && \text{for any } \vec{y} \neq \vec{b}
\end{aligned}$$

We want the RHS to tend to 0 as $\vec{y} \to \vec{b}$. Specifically, we want to write

$$\begin{aligned}
\lim_{\vec{y} \to \vec{b}} \frac{1}{m}\left\| J^{-1}u(f^{-1}(\vec{y})) \right\| &= \frac{1}{m}\| J^{-1}u\left(f^{-1}\left(\lim_{\vec{y} \to \vec{b}} \vec{y} \right) \right) \| \\
&= \frac{1}{m}\left\| J^{-1}u(f^{-1}(\vec{b})) \right\| \\
&= \frac{1}{m}\left\| J^{-1}u(\vec{a}) \right\| \\
&= 0
\end{aligned}$$

The first equality needs to be verified.
We want to use Proposition 7.4.1 (passing limits to argument of continuous functions) but first need to check the hypothesis.
Define $g : f(\overline{B_\delta(\vec{a})}) \to \mathbb{R}$ by

$$g(\vec{y}) := \left\| J^{-1}u(f^{-1}(\vec{y})) \right\|$$

We must show that

1. $\vec{b} \in f(\overline{B_\delta(\vec{a})}) \cap [f(\overline{B_\delta(\vec{a})})]^a$
2. g is continuous at $\vec{b}$

Since $\vec{b} \in f(\vec{a}), \vec{b} \in f(B_\delta(\vec{a})) \subseteq f(\overline{B_\delta(\vec{a})})$
By Lemma 23.1.1, we know that the $f(B_\delta(\vec{a}))$ is open, so $\vec{b} = f(\vec{a})$ is an interior point of $f(\overline{B_\delta(\vec{a})}))$ so $\vec{b} \in \left[f(\overline{B_\delta(\vec{a})}) \right]^a$
Lemma 23.1.3 says that if $f \in C\left(K, \mathbb{R}^n\right)$ is a continuous injection on a compact set K, then $f^{-1} \in C\left(f(K), K\right)$ So f^{-1} is continuous at $\vec{b}$ and we have previously shown that g is continuous at $f^{-1}(\vec{b}) = \vec{a}$, so g is continuous at $\vec{b}$.
Finally, by the Squeeze Theorem, f^{-1} is differentiable at $f(\vec{a})$ and $Df^{-1}(f(\vec{a})) = J^{-1}$, as required.

23 Inverse Function Theorem II

23.1 Lemmas

Lemma 23.1.1
Let $U \subseteq \mathbb{R}^n$ be nonempty and open, $f \in C(U, \mathbb{R}^n)$.
If $\det(Df(\vec{x})) \neq 0$ for all $\vec{x} \in U$, $f(U)$ is open.

Proof
Given any $\vec{b} \in f(U)$, there exists $\vec{a} \in U$ such that $\vec{b} = f(\vec{a})$.
Since U is open, we can apply Lemma 23.1.5 to show that there exists $r > 0$ such that $B_r(\vec{a}) \subseteq U$ and f is an injection on $B_r(\vec{a})$
Choose any $\rho > 0$ with $\rho < r$, we have $A := \overline{B_\rho(\vec{a})} \subseteq U$ and f is an injection on A.
Define the boundary of A, which is the sphere

$$S := \{\vec{x} \in \mathbb{R}^n : \|\vec{x} - \vec{a}\| = \rho\} \subseteq U$$

Note that S is compact.
The image of a continuous function on compact domains are compact (exercise), so $f(S)$ is compact.
Since f is injective on A, $f(\vec{x}) \neq f(\vec{a})$ for any $\vec{x} \in S$.
Define

$$\delta := \frac{1}{2} \inf_{\vec{x} \in S} \left\| f(\vec{x}) - \vec{b} \right\|$$

We must have $\delta > 0$ because EVT says that the infimum is obtained by f on A.
By the definition of δ, $\left\| f(\vec{x}) - \vec{b} \right\| \geq 2\delta$ for any $\vec{x} \in S$.
We want to show that $B_\delta(\vec{b}) \subseteq f(U)$.
Let $\vec{v} \in B_\delta(\vec{b})$ and define the function $\phi : A \to \mathbb{R}$ by

$$\phi(\vec{x}) := \|f(\vec{x}) - \vec{v}\|^2$$

We seek a vector $\vec{u}$ such that $\phi(\vec{u}) = 0$.
This would mean that $f(\vec{u}) = \vec{v}$, showing that $\vec{v} \in f(A) \subset f(U)$
By EVT, there is come $\vec{u} \in A$ such that $\phi(\vec{x}) \geq \phi(\vec{u})$ for all $\vec{x} \in A$.
We claim that $\vec{u} \in \text{int}(A)$ and prove it by contradiction.
Suppose $\vec{u} \in S$.

$$\begin{aligned}
\sqrt{\phi(\vec{u})} &= \|f(\vec{u}) - \vec{v}\| \\
&\geq \left\| f(\vec{u}) - \vec{b} \right\| - \left\| \vec{b} - \vec{v} \right\| \\
&> 2\delta - \delta \\
&= \delta
\end{aligned}$$

Since

$$\vec{v} \in B_\delta(\vec{b}), \sqrt{\phi(\vec{a})} = \left\|\vec{b} - \vec{v}\right\| < \delta < \sqrt{\phi(\vec{u})}$$

Hence $\vec{u}$ cannot be the minimum.
By contradiction, $\vec{u}$ must be in the interior of A.
By the Chain Rule, $D\phi(\vec{x}) = 2\left[f(\vec{x}) - \vec{v}\right]^T Df(\vec{x})$ for all $\vec{x} \in A$.
Since $\vec{u} \in \text{int}(A)$ and the gradient vector $\nabla\phi(\vec{u}) = [D\phi(\vec{u})]^T$ exists, we deduce that

$$D\phi(\vec{u}) = \vec{0}^T$$

by Theorem 20.2.1.

$$\begin{aligned} D\phi(\vec{u}) = 2\left[f(\vec{u}) - \vec{v}\right]^T Df(\vec{u}) = \vec{0}^T \\ D\phi(\vec{u})[Df(u)]^{-1} = 2[f(\vec{u}) - \vec{v}]^T = \vec{0}^T \\ f(\vec{u}) = \vec{v} \end{aligned}$$

Lemma 23.1.2
Let $\vec{a} \in \mathbb{R}^n, r > 0$.
If $f \in C^1(B_r(\vec{a}), \mathbb{R}^n)$ and $\det Df(\vec{a}) \neq 0$, then there is some $\delta \in (0, r]$, $m > 0$ such that

$$\|f(\vec{x}) - f(\vec{a})\| \geq m\|\vec{x} - \vec{a}\|, \forall \vec{x} \in B_\delta(\vec{a})$$

Proof
Define the Jacobian Matrix $J := Df(\vec{a})$.
Since $\det J \neq 0$, $\|J\vec{u}\| \geq 0$ for every nonzero vector in $\mathbb{R}^n$.
Let

$$S := \{\vec{x} \in \mathbb{R}^n : \|\vec{x} - \vec{a}\| = 1\}$$

be the uniqe sphere, which is clearly compact.
Define $m := \frac{1}{2}\inf_{\vec{u}\in S}\{\|J\vec{u}\|\}$.
EVT guarantees that the infimum is attained so $m > 0$.
For all $\vec{u} \in \mathbb{R}^n$, $\|J\vec{u}\| \geq 2m\|\vec{u}\|$.
By the definition of the derivative, there is $\delta \in (0, r]$ such that for all $\vec{x} \in B_\delta(\vec{a})$

$$\|f(\vec{x}) - f(\vec{a}) - J(\vec{x} - \vec{a})\| \leq m\|\vec{x} - \vec{a}\|$$

By the Reverse Triangle Inquality

$$\begin{aligned} m\|\vec{x} - \vec{a}\| &\geq \|J(\vec{x} - \vec{a}) - f(\vec{x}) + f(\vec{a})\| \\ &\geq \|J(\vec{x} - \vec{a})\| - \|f(\vec{x}) - f(\vec{a})\| \\ &\geq 2m\|\vec{x} - \vec{a}\| - \|f(\vec{x}) - f(\vec{a})\| \end{aligned}$$

for all $\vec{x} \in B_\delta(\vec{a})$. Hence $\|f(\vec{x}) - f(\vec{a})\| \geq m\|\vec{x} - \vec{a}\|$ for all $\vec{x} \in B_\delta(\vec{a})$.

Lemma 23.1.3
Let $K \subseteq \mathbb{R}^n$ be nonempty and compact, and let $f \in C(K, \mathbb{R}^n)$ be an injection.
Then $f^{-1} \in C(f(K), K)$.

Proof
Since f is an injection, $f : K \to f(K)$ is a bijection.
Hence the inverse $f^{-1} \to K$ makes sense.
Let $b \in f(K)$.
We want to show that the inverse is continuous on at $\vec{b}$.
Suppose for a contradiction that there is a sequence $(\vec{y}_k)_{k=1}^{\infty}$ in $f(K)$ that converges to $\vec{b}$ but

$$\exists \epsilon > 0, \|f^{-1}(\vec{y}_k) - f^{-1}(\vec{b})\| \geq \epsilon, \forall k \geq 1$$

By the compactness of K, there must be a subsequence $(f^{-1}(\vec{y}_{k_j}))$ that converges to $\vec{a} \in K$.
By the continuity of f,

$$\lim_{j\to\infty} \vec{y}_{k_j} = \lim_{k\to\infty} = \vec{b} \implies \vec{a} = f^{-1}(\vec{b})$$

But the convergence of the image of the subsequence in the inverse of f contradicts our assumptions. So f^{-1} must be continuous.

Lemma 23.1.4
Let $U \subseteq \mathbb{R}^n$ be nonempty and open, $V \subseteq \mathbb{R}^n$, and $f \in C(U, V)$ be a bijection, for $k \geq 1$.
If $\det Df(\vec{x}) \neq 0$ for all $\vec{x} \in U$, then

$$f^{-1} : V \to U \in C^k(V, U)$$

Proof
We showed in the proof of the Gloval Inverse Theorem that

$$Df^{-1}(Df(\vec{a})) = [Df(\vec{a})]^{-1}$$

In other words, $Df^{-1} : V \to GL_n$ is the composition

$$Df^{-1} = i \circ Df \circ f^{-1}$$

where $i : GL_n \to GL_n$ defined by $i(M) := M^{-1}$ is the matrix inverse function.
We now proceed by induction.
For the base case, suppose $f \in C^1(U, V)$. By the hypothesis, each component of the matrix function $Df(\vec{x})$ is continuous at $\vec{x}$ and $\det Df(\vec{x}) \neq 0$ for all $\vec{x} \in U$.
Recall

$$M \in \mathbb{R}^{n\times m} \implies M^{-1} = \frac{1}{\det M} C^T$$

with C being the cofactor matrix of M.
The inverse is then a rational function of the components of M and the denominator i nonzero. Hence i is of class C^∞.
Sine f^{-1}, Df, i are continuous, $Df^{-1} = i \circ Df \circ f^{-1}$ is continuous, so $f^{-1} \in C^1(f(U), U)$.
For the inductive step, we suppose $f \in C^m(U,V) \implies f^{-1} \in C^m(V,U)$, for some $1 \le m < k$.
If f is of class C^{m+1}, then f is of class C^m. Hence f^{-1}, Df, i are all of class C^m. SO Df^{-1} is of class C^m. This means that f is of class C^{m+1}.

Lemma 23.1.5
Let $\vec{a} \in \mathbb{R}^n, r > 0$.
If $f \in C^1(B_r(\vec{a}), \mathbb{R}^n)$ and let $\det Df(\vec{a}) \neq 0$.
Then there is some $\delta \in (0, r]$ such that f is injective on $B_\delta(\vec{a})$.

Proof
Since $f \in C^1(b_r(\vec{a}), \mathbb{R}^n)$ and $\det Df(\vec{a}) \neq 0$ and the determinant of a matrix is a multinomial in the components a_{ij}, there is some $\delta \in (0, r]$ such that

$$\det \begin{pmatrix} Df_1(\vec{p}_1) \\ \dots \\ Df_n(\vec{p}_n) \end{pmatrix} = \det \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(\vec{p}_1) & \dots & \frac{\partial f_1}{\partial x_n}(\vec{p}_n) \\ \dots & \dots & \dots \\ \frac{\partial f_n}{\partial x_1}(\vec{p}_1) & \dots & \frac{\partial f_n}{\partial x_n}(\vec{p}_n) \end{pmatrix} \neq 0$$

for all $\vec{p}_1, \dots, \vec{p}_n \in B_\delta(\vec{a})$ by the continuity of multinomials.
Suppose $f(\vec{x}) = f(\vec{y}), \vec{x}, \vec{y} \in B_\delta(\vec{a})$.
We wish to show that $\vec{x} = \vec{y}$.
By the Mean Value Theorem, for each $i \in \{1, 2, \dots, n\}$ there exists $\vec{c}_i$ on the line segment from $\vec{x}$ to $\vec{y}$ such that

$$f_i(\vec{x}) - f_i(\vec{y}) = Df_i(\vec{c}_i)(\vec{x} - \vec{y})$$

Hence

$$\vec{x} - \vec{y} = \begin{pmatrix} Df_1(\vec{c}_1) \\ \dots \\ Df_n(\vec{c}_n) \end{pmatrix}^{-1} [f(\vec{x}) - f(\vec{y})] = \vec{0}$$

24 Inverse Function Theorem III

24.1 Local Invertibility

Theorem 24.1.1 (Local Inverse Function Theorem)
Let $U \in\subseteq \mathbb{R}^n$ be nonempty and open and suppose $f \in C(U, \mathbb{R}^n)$ satisfies

$$\det Df(\vec{a}) \neq 0$$

for some $\vec{a} \in U$.
Then, f is locally invertible.
Moreover, there is $\delta > 0$ such that

1. $A := B_\delta(\vec{a}) \subseteq U$
2. f is injective on A
3. $f(A)$ is open
4. $f^{-1} \in C\left(f(A), A\right)$
5. $Df^{-1}\left(f(\vec{x})\right) = \left[Df(\vec{x})\right]^{-1}$ for every $\vec{x} \in A$.

Proof
Since U is open, there is some $f > 0$ such that $B_r(\vec{a}) \subseteq U$ with $f \in C\left(B_r(\vec{a}), \mathbb{R}^n\right)$.
So by Lemma 23.1.5, there is $\delta \in (0, r]$ such that Parts 1-2 hold.
From the proof of Lemma 23.1.3, $\det Df(\vec{x}) \neq 0$ for all $\vec{x} \in B_\delta(\vec{a}) =: A$.
We can apply the Global Inverse Function Theorem, to the restricted function

$$f : A \to f(A)$$

which yields Parts 3-5.

25 Implicit Function Theorem

25.1 Implicit Function Theorem

$U \subseteq \mathbb{R}^{n+m}$ open and $\phi : U \to \mathbb{R}^p$.
Suppose $\vec{x} \in \mathbb{R}^n, \vec{y} \in \mathbb{R}^m$ satisfies $\phi(\vec{x}, \vec{y}) = \vec{0}$.
Then the equation

$$\phi(\vec{x}, \vec{y}) = \vec{0}$$

defines an implicit relationship between $\vec{x}$ and $\vec{y}$.
Can we find an explicit relaship?
In other words, we week to write $\vec{y}$ as a function of $\vec{x}$.

$$\phi(\vec{x}, \vec{y}(\vec{x})) = 0$$

(for a subset of the domain at least)
If so, $\phi(\vec{x} + \vec{y}) = \vec{0}$ implicitly defines $\vec{y}$ as a function of $\vec{x}$.

25.2 Linear Examples

Example 25.2.1

$$\phi(\vec{x}, \vec{y}) = A \begin{pmatrix} \vec{x} \\ \vec{y} \end{pmatrix} = (A_n A_m) \begin{pmatrix} \vec{x} \\ \vec{y} \end{pmatrix}$$

Where $A_n \in \mathbb{R}^{p \times n}, A_m \in \mathbb{R}^{p \times m}$.
Suppose $\phi(\vec{x}, \vec{y}) = \vec{0}$, meaning

$$A_m \vec{y} = -A_n \vec{x}$$

If A_m is invertible, then we could write

$$\vec{y} = -A_m^{-1} A_n \vec{x}$$

expresssing $\vec{y}$ as a function fo $\vec{x}$.
If ϕ is instead not linear but a smooth function that satisfies $\phi(\vec{x}_0, \vec{y}_0) = \vec{0}$ for some $\vec{x}_0, \vec{y}_0 \in \mathbb{R}^m$, then we could linearize near $(\vec{x}_0, \vec{y}_0$ and hope that the linearization gives an approximate solution for $\vec{y}(\vec{x})$.

25.3 Implicit Function Theorem

Theorem 25.3.1 (Implicit Function Theorem)
Let $U \subseteq \mathbb{R}^{n+m}$ be nonempty and open, and let $\phi \in C^1(U, \mathbb{R}^m)$.
Suppose there exists points $\vec{x}_0 \in \mathbb{R}^n, \vec{y}_0 \in \mathbb{R}^m$ such that $\phi(\vec{x}_0, \vec{y}_0) = \vec{0}$, and

$$\det\left(D\vec{y}\phi(\vec{x}_0, \vec{y}_0)\right) \neq 0$$

Then there is $a, b > 0$ such that $B_a(\vec{x}_0) \times B_b(\vec{y}_0) \subseteq U$ and there exists a function $f \in C^1(B_a(\vec{x}_0), B_b(\vec{y}_0))$ such that $\phi(\vec{x}, f(\vec{x})) = \vec{0}$ for all $\vec{x} \in B_a(\vec{x}_0)$.
Moreover, the function f is unique.

Note that $D_{\vec{y}} f$ is the Jacbian matrix only including partial derivatives with respect to $\vec{y}$.

Proof
Define an auxiliary function $g : U \to \mathbb{R}^{n+m}$ by

$$g(\vec{x}, \vec{y}) := (\vec{x}, \phi(\vec{x}, \vec{y}))$$

Observe that g is of class C^1

$$Dg = \begin{pmatrix} D_{\vec{x}}\vec{x} & D_{\vec{y}}\vec{x} \\ D_{\vec{x}}\phi & D_{\vec{y}}\phi \end{pmatrix} = \begin{pmatrix} I_n & 0_{n\times m} \\ D_{\vec{x}}\phi & D_{\vec{y}}\phi \end{pmatrix}$$

By our results in Linear Algebra

$$\det D_g = \det \begin{pmatrix} I_n & 0_{n\times m} \\ D_{\vec{x}}\phi & D_{\vec{y}}\phi \end{pmatrix} = \det D_{\vec{y}}\phi$$

by Cofactor Expansion.
But by assumption $\det D_{\vec{y}}\phi \neq 0$ so we know our entire matrix is invertible.
We now apply the Local Inverse Function Theorem

1. there is $\delta > 0$ such that $A := B_\delta(\vec{x}_0, \vec{y}_0) \subseteq U$
2. g is injective on A
3. $g(A)$ is open
4. $g^{-1} \in C^1(g(A), A)$
5. $Dg^{-1}(g(\vec{x}, \vec{y})) = [Dg(\vec{x}, \vec{y})]^{-1}$ for every $(\vec{x}, \vec{y}) \in A$.

By the continuity of Dg, we can assume $\delta > 0$ is such that

$$\det D_g(\vec{x}, \vec{y}) \neq 0$$

for all $(\vec{x}, \vec{y}) \in A$ (proof of Lemma 23.1.5)
Choose any $\tilde{a}, b$ satistfying $\tilde{a}^2 + b^2 < \delta^2$.
Define

$$X := B_{\tilde{a}}(\vec{x}_0) \times B_b(\vec{y}_0) \subseteq \mathbb{R}^{n+m}$$

Note that $X \subseteq A \subseteq U$ and g is a bijection from X from $g(X)$.
Recall the definition $g(\vec{x}, \vec{y}) = (\vec{x}, \phi(\vec{x}, \vec{y}))$

$$g^{-1}(\vec{x}, \vec{z}) = (\vec{x}, h(\vec{x}, \vec{z}))$$

Applying the Cancellation Property of Inverses

$$g(g^{-1}(\vec{x}, \vec{z})) = (\vec{x}, \phi(\vec{x}, h(\vec{x}, \vec{z}))) = (\vec{x}, \vec{z})$$

So $\phi(\vec{x}, h(\vec{x}, \vec{z})) = \vec{z}$ (for all $(\vec{x}, \vec{z}) \in g(X)$, ie inverse is defined).
The rest is trivial.

26 Constrained Optimization

26.1 Constrained Optimization

Suppose we wish to optimize a function $f : \mathbb{R}^2 \to \mathbb{R}$, which we assume to be differentiable.
By Theorem 20.2.1, if $(\vec{x}_0, \vec{y}_0)$ is a local maximum or a local minimum of f, the $\nabla f(\vec{x}_0, \vec{y}_0) = (0, 0)$.
So we could look for extrema be finding the solutions of

$$f_x(\vec{x}_0, \vec{y}_0) = f_y(\vec{x}_0, \vec{y}_0) = 0$$

Now, suppose we want to optimize f subject to constraints (ie $x > 0, x + y = 1$).
In other words, optimize over a domain defined by these constraints.
Theorem 20.2.1 only applies in $(\vec{x}_0, \vec{y}_0) \in \text{int}(S)$.
We could use 20.2.1 to look for extrema in the interior, but then we must look for potential extrema on the boundaries.
In the example above, with constraint $x + y = 1$, S has an empty interior, so Theorem 20.2.1 is of little use.
We will focus on constraints of the form

$$\phi(\vec{x}) = \vec{0}$$

26.2 Lagrange Multipliers

Theorem 26.2.1 (Lagrange Multiplier Theorem)
Suppose we have $\emptyset \neq U \subseteq \mathbb{R}^n$ open with $n \geq 2$
Let $f \in C^1(U, \mathbb{R})$ and $\phi \in C^1(U, \mathbb{R}^m)$ with $1 \leq m < n$, and define the set

$$S := \{\vec{x} \in U : \phi(\vec{x}) = \vec{0}\}$$

If the points $\vec{a} \in S$ is a local minimum or a local maximum of the restriction of f to S and

$$\text{rank}\, D\phi(\vec{a}) = m$$

then there exists $\vec{\lambda} \in \mathbb{R}^m$ such that $Df(\vec{a}) = \vec{\lambda}^T D\phi(\vec{a})$.

Proof
Suppose $\vec{a} \in S$ is a local minimum or maximum of the $f\big|_S$.
Also suppose rank $D\phi(\vec{a}) = m$. WLOG the last m columns are linearly independent.
Write variable

$$\vec{x} = (x_1, \ldots, x_n) = (\vec{\xi}, \vec{\nu})$$

If the last m columns of
$$D\phi(\vec{a}) = \left(D_{\vec{\xi}}\phi(\vec{a}), D_\nu\phi(\vec{a})\right)$$
are linearly independent, then $\det D_{\vec{\nu}}\phi(\vec{a}) \neq 0$.
Applying the Implicit Function Theorem 25.3.1 gives us a function $g \in C^1$ such that in a neighbourhood of $\vec{a}$
$$\phi(\vec{x}) = \vec{0} \iff \vec{x} = (\vec{\xi}, g(\vec{\xi}))$$
So, if $\vec{x}$ is close to $\vec{a}$, then $f(\vec{x}) = f(\vec{\xi}, g(\vec{\xi})) =: h(\vec{\xi})$.
By the hypothesis, $f\big|_S$ has a local extremum at $\vec{a}$ so h has a local extremum at
$$\vec{\xi_0} = (a_1, \ldots, a_{n-m})$$
Since h is of class C^1, Theorem 20.2.1 tells us that
$$\nabla h(\vec{\xi_0}) = \vec{0} \iff Df(\vec{\xi_0}) = \vec{0}^T$$
By the Chain Rule
$$Dh(\vec{\xi_0}) = D_{\vec{\xi}}f(\vec{a}) + D_\nu f(\vec{a})Dg(\vec{\xi_0}) = \vec{0}^T$$
Evaluating at $\vec{a}$
$$\begin{aligned} D_{\vec{\xi_0}}\phi(\vec{a}) + D_{\vec{\nu}}\phi(\vec{a})Dg(\vec{\xi_0}) &= \vec{0}^T \\ Dg(\vec{\xi_0}) &= -\left[D_{\vec{\nu}}\phi(\vec{a})\right]^{-1} D_{\vec{\xi}}\phi(\vec{a}) \end{aligned}$$
$$D_{\vec{\xi}}f(\vec{a}) - D_{\vec{\nu}}f(\vec{a})\left[D_{\vec{\nu}}\phi(\vec{a})\right]^{-1} D_{\vec{\xi}}\phi(\vec{a}) = \vec{0}^T$$
$$\begin{aligned} D_{\vec{\xi}}f(\vec{a}) &= D_{\vec{\nu}}f(\vec{a})\left[D_{\vec{\nu}}\phi(\vec{a})\right]^{-1} D_{\vec{\xi}}\phi(\vec{a}) \\ &= \vec{\lambda}^T D_\xi\phi(\vec{a}) \\ &\ldots \\ Df(\vec{a}) &= \vec{\lambda}^T D\phi(\vec{a}) \end{aligned}$$
Note by definition
$$D_{\vec{\nu}}f(\vec{a}) = \vec{\lambda}^T D_\nu\phi(\vec{a})$$

27 Riemann Integerals

27.1 Boxes and Partitions

Definition 27.1.1 (Box)

$$I = [a_1, b_1] \times \cdots \times [a_n, b_n] \subseteq \mathbb{R}^n$$

where $[a_k, b_k]$ is a closed interval for $k \in \{1, 2, \ldots, n\}$.

The volume of a box is

$$\mu(I) = \prod_{k=1}^{n} (b_k - a_k)$$

Definition 27.1.2 (partition)

$$[a_k, b_k] = \bigcup_{i=0}^{l-1} [x_k^{(i)}, x_k^{(i+1)}]$$

with $a_k = x^{(0)} < x^{(1)} < \cdots < x^l = b_k$.
Let

$$P_k := \{x_k^i : 0 \leq i \leq k\}$$

be a partition of $[a_k, b_k]$.
Then

$$P := \{P_k : 1 \leq k \leq n\}$$

is a partition of I.

Definition 27.1.3 (norm of one-dimensional partition)

$$\|P_k\| := \max_{1 \leq j \leq l_k} \{x_k^{(j)} - x_k^{(j-1)}\}$$

Definition 27.1.4 (norm of a general partition)

$$\|P\| := \max_{1 \le k \le n} \|P_k\|$$

Denote the set of all possible partitions of I by $\mathbb{P}$.
For a given partition P of I, the associated indexing set is

$$J := \{1, \ldots, l_1\} \times \cdots \times \{1, \ldots, l_n\}$$

Elements of J, $\vec{\alpha}$, are multi-indices.

Definition 27.1.5 (sub-box)
For each $\vec{\alpha} \in J$, we define the sub-box

$$I^{(\vec{\alpha})} := [x_1^{(\alpha_1 - 1)}, x_1^{(\alpha_1)}] \times [x_n^{(\alpha_n - 1)}, x_n^{(\alpha_n)}]$$

The box I is partitioned by the set of all those sub-indices

$$I := \bigcup_{\vec{\alpha} \in J} I^{(\vec{\alpha})}$$

28 Riemann Integrability

28.1 Riemann Sums and Riemann Integrals

Definition 28.1.1 (Riemann Sum)
$I \subseteq \mathbb{R}^n$ a box.
P a partition of I.
$f : I \to \mathbb{R}$.
For each $\alpha \in J$, choose some point $\vec{x}^{(\vec{\alpha})} \in I^{(\vec{\alpha})}$.
Then

$$S(f, P) := \sum_{\vec{\alpha} \in J} f(\vec{x}^{(\vec{\alpha})})\mu(I^{(\vec{\alpha})})$$

is a Riemann Sum of f with respect to the partition P.

Not that the Riemann Sum depends on our choice of $\{\vec{x}^{(\vec{\alpha})}\}$. But note that each sub-box is compact so if f is continuous, then by EVT, there is a supremum and infimum attained by f in each sub-box.
If we relax the restriction of continuity to bounded functions.
We can still define

$$M^{(\vec{\alpha})} := \sup_{\vec{x} \in I^{(\vec{\alpha})}} \{f(\vec{x})\}$$

Note that the image of f on a compact domain (box) is guaranteed to be uniformly continuous (bounded).

Definition 28.1.2 (Upper Riemann Sum)

$$U(f, P) := \sum_{\vec{\alpha}} M^{(\vec{\alpha})}\mu(I^{(\vec{\alpha})})$$

and similarly for the Lower Riemann Sum.
It follows that

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

for any choice of $\vec{x}^{(\vec{\alpha})}$ for S.
then we have

Definition 28.1.3 (Lower Riemann Integral)

$$\underline{\int_I} f(\vec{a})d\vec{x} := \sup_{P\in\mathbb{P}_I} \{L(f,P)\}$$

and

Definition 28.1.4 (Upper Riemann Integral)

$$\overline{\int_I} f(\vec{a})d\vec{x} := \inf_{P\in\mathbb{P}_I} \{U(f,P)\}$$

Definition 28.1.5 (Riemann Integrable)
If the lower and upper Riemann Integrals are equivalent we say the Riemann Integral is

$$\int_I f(\vec{x})d\vec{x} := \overline{\int_I} f(\vec{x})d\vec{x} = \underline{\int_I} f(\vec{x})d\vec{x}$$

Note $d\vec{x}$ can be intepreted as the "volume".

28.1.1 Notation

For $n = 1$ we write

$$\int_{I=[a,b]} f(x)dx = \int_a^b f(x)dx$$

For $n = 2$

$$\int_I f(\vec{x})d\vec{x} = \int_I f(\vec{x})d^2d\vec{x} = \iint_I f(\vec{x})d\vec{x} = \int_I f(x,y)d(x,y)$$

28.2 Integrability

Lemma 28.2.1
Let $I \subseteq \mathbb{R}^n$ be a box and let P be a partition of I with indexing set J and sub-boxes $\{I^{(\vec{\alpha})} : \vec{\alpha} \in J\}$.
Then

$$\mu(I) = \sum_{\vec{\alpha}\in J} \mu(I^{(\vec{\alpha})})$$

Proof
Exercise.

Definition 28.2.1 (refinement)
P, Q two partitions of a box I.
We say Q is a refinement of P if

$$P_k \subseteq Q_k, \forall k \in 1, \ldots, n$$

Proposition 28.2.2
P, Q two partitions, there is a partition R of I that is a refinement of both.
Simple Common Partition.

Proposition 28.2.3
$I \subseteq \mathbb{R}^n$ a box.
$f : I \to \mathbb{R}$ a bounded function.

1. for any $P \in \mathbb{P}_I$, $L(f, P) \leq U(f, P)$.
2. for any refinement Q of $P \in \mathbb{P}_I$,

$$L(f, P) \leq L(f, Q) \wedge U(f, P) \geq U(f, Q)$$

3. For any two partitions P, Q of I,

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

Proof
Exercise

Theorem 28.2.4
$I \subseteq \mathbb{R}^n$ a box.
$f : I \to \mathbb{R}$ a bounded function.
f is Riemann integrable if and only if
for all $\epsilon > 0$ there is $P \in \mathbb{P}_I$ such that

$$0 \leq U(f, P) - L(f, Q) < \epsilon$$

28.3 Riemann Integerals over Arbitrary Domains

Definition 28.3.1 (Integrability)
$f : S \to \mathbb{R}$ is a bounded function on a domain $S \subseteq \mathbb{R}^n$ that is not necessarily a box.
If S is nonempty and bounded, Let $I \subseteq \mathbb{R}^n$ be a box which contains S.
Define a function $g : I \to \mathbb{R}$ by

$$g(\vec{x}) := \begin{cases} f(\vec{x}), & \vec{x} \in S \\ 0, & \vec{x} \in I \setminus S \end{cases}$$

If f is Riemann Integrable on S if g is Riemann Integrable on I.
We the define

$$\int_S f(\vec{x})d\vec{x} := \int g(\vec{x})d\vec{x}$$

Proposition 28.3.1
Let $S \subseteq \mathbb{R}^n$ be nonempty and bounded, $f : \mathbb{R}^n \to \mathbb{R}$ be bounded and satisfy $f(\vec{x}) \neq 0$ for all $\vec{x} \in S$.
If I_1, I_2 and two boxes in $\mathbb{R}^n$ that each contain S, then if f is Riemann Integrable on I_1, then f is Riemann integrable on I_2 and

$$\int_{I_1} f = \int_{I_2} f$$

Proof
Exercise

29 Jordan Content

29.1 Content

Definition 29.1.1 (characteristic function)
of a set $S \subseteq \mathbb{R}^n$ is a function $\chi_S : \mathbb{R}^n \to \mathbb{R}$ defined by

$$\chi_S(\vec{x}) := \begin{cases} 1, & \vec{x} \in S \\ 0, & \text{else} \end{cases}$$

Definition 29.1.2 (Jordan Content)
If the characteristic function of a nonempty, bounded set $S \subseteq \mathbb{R}^n$ is integrable on S, then we say that S has **(Jordan) Content**.

If S has content, the its volume is

$$\mu(S) := \int_S \chi_S(\vec{x}) d\vec{x}$$

If S has content and $\mu(S) = 0$, then we say S has content zero.

Proposition 29.1.1
Let $S \subseteq \mathbb{R}^n$ be nonempty and bounded. S has content zero if and only if for all $\epsilon > 0$, there is a finite set of boxes

$$\{I_i \subseteq \mathbb{R}^n : 1 \leq i \leq m\}$$

such that

$$S \subseteq \bigcup_{i=1}^{m} I_i$$

and

$$\sum_{i=1}^{m} \mu(I_i) < \epsilon$$

Proof
Exercise
$0 \leq U(f, p) - L(p, P) < \epsilon$
We wish to show that $0 \leq U(\chi_S, P) < \epsilon$ for some P.

Corollary 29.1.1.1
Let $S, R \subseteq \mathbb{R}^n$ be nonempty and bounded

1. If T has content zero and $S \subseteq T$, then S has content zero

2. If S and T both have content zero, then $S \cup U$ has content zero

29.2 Examples

Example 29.2.1
Singleton $\{\vec{x}\}$ has content zero

Example 29.2.2
$I := [0, 1] \subseteq \mathbb{R}$ has content (not zero)

Example 29.2.3
$S := [0, 1] \cap \mathbb{Q}$ has no content

Example 29.2.4
$T := [0, 1] \setminus \mathbb{Q}$ has no content

Example 29.2.5
note $[0, 1] = S \cup T$ which has a nonzero content!

29.3 More Results

Proposition 29.3.1
Let $f \in C([a, b], \mathbb{R})$ with $a < b$. Then the graph

$$G := \{(x, f(\vec{x})) : \vec{x} \in [a, b]\} \subseteq \mathbb{R}^2$$

has content zero

Proof
By Theorem 11.4.1, f is uniformly continuous on $[a, b]$.
Fix $\epsilon > 0$, there exists $\delta > 0$ such that for any $x, y \in [a, b]$ with $|x - y| < \delta$, then

$$|f(x) - f(y)| < \epsilon$$

Partition $[a, b]$ into boxes of length at most δ, then the box around the graph would be bounded by ϵ. So each box has content zero. We take finite unions, leading something of content zero.

30 More about Integrability

30.1 Content and Integrablity

Theorem 30.1.1 (Lebesgue's Criterion)
Let $I \subseteq \mathbb{R}^n$ be a box and $S \subseteq I$ be nonempty.
Suppose $f : I \to \mathbb{R}$ is bounded and $f(\vec{x}) = 0$ for all $\vec{x} \in I \setminus S$.
Let $D \subseteq I$ be the set of points at which f is discontinuous.
If D has content zero, then f is integrable on S.

Proof
By the boundedness of f, there exists M such that

$$|f(\vec{x})| \leq M, \forall \vec{x} \in I$$

If D has content zero, then the characteristic function χ_D is integrable on D and hence on I (it is constant on outside of D), and

$$\int_I \chi_D(\vec{x}) d\vec{x} = 0$$

Let $\epsilon > 0$.
By integrability of χ_D,

$$\inf_{P \in \mathbb{P}_I} \{U(\chi_D, P)\} = 0$$

.
So there is a partition P such that

$$U(\chi_D, P) < \frac{\epsilon}{4M}$$

Let J_p denote the indexing set of P and let

$$\{I_p^{(\vec{\alpha})} : \vec{\alpha} \in J_p\}$$

be the set of sub-boxes
Let

$$J_p^1 = \{\vec{\alpha} \in J_p : I_p^{(\vec{\alpha})} \cap D \neq \emptyset\}$$

and define $J_p^2 = J_p \setminus J_p^1$.
Also define

$$K := \{I_p^{(\vec{\alpha})} : \vec{\alpha} \in J_p^2\}$$

By construction

$$\sum_{\vec{\alpha} \in J_p^1} \mu(I_p^{(\vec{\alpha})}) = U(\chi_D, P) < \frac{\epsilon}{4M}$$

Since f is continuous on K, which is a compact set, f is uniformly continuous on K. So there is $\delta > 0$ such that

$$|f(\vec{x}) - f(\vec{y})| < \frac{\epsilon}{2\mu(I)}$$

for all $\vec{x}, \vec{y} \in K$ satisfying $\|\vec{x} - \vec{y}\| < \delta$.
Suppose Q is a refinement of P with indexing set J_Q, sub-boxes

$$\{I_Q^{(\vec{\beta})} : \vec{\beta} \in J_Q\}$$

and norm $\|Q\| < \frac{\delta}{\sqrt{n}}$
For any $\vec{x}, \vec{y}$ in the same sub-box of Q,

$$\begin{aligned}\|\vec{x} - \vec{y}\| &= \left(\sum_{i=1}^{n}(x_i - y_i)^2\right)^{\frac{1}{2}} \\ &\leq \left[n\left(\frac{\delta}{\sqrt{n}}\right)^2\right]^{\frac{1}{2}}\end{aligned}$$

Let

$$J_Q^1 := \{\vec{\beta} \in J_Q : I_Q^{(\vec{\beta})} \subseteq I_P^{(\vec{\alpha})} \text{ for some } \vec{\alpha} \in J_p^1\}$$

and define $J_Q^2 := J_Q \setminus J_Q^1$
Furthermore, define

$$\Delta(\vec{\beta}) := \sup_{\vec{x} \in I_Q^{(\vec{\beta})}} \{f(\vec{x})\} - \inf_{\vec{x} \in I_Q^{(\vec{\beta})}\{f(\vec{x})\}} \{f(\vec{x})\}$$

note that $\Delta^{(\vec{\beta})} \leq 2M$ for any $\vec{\beta} \in J_Q$, and $\Delta \leq \frac{\epsilon}{2\mu(I)}$ for any $\vec{\beta} \in J_Q^2$.
Hence,

$$\begin{aligned}0 &\leq U(f,Q) - L(f,Q) \\ &= \sum_{\vec{\beta} \in J_Q} \Delta^{(\vec{\beta})} \mu(I_Q^{(\vec{\beta})}) \\ &= \sum_{\vec{\beta} \in J_Q^1} + \sum_{\vec{\beta} \in J_Q^2} \\ &\leq 2M \sum_{\beta \in J_Q^1} \mu(I_Q^{(\vec{\beta})}) + \frac{\epsilon}{2\mu(I)} \sum_{\vec{\beta} \in J_Q^2} \mu(I_Q^{(\vec{\beta})}) \\ &\leq 2M\frac{\epsilon}{4M} + \frac{\epsilon}{2\mu(I)}\mu(I) \\ &= \epsilon\end{aligned}$$

Corollary 30.1.1.1
Let $S \subseteq \mathbb{R}^n$ be nonempty and bounded, and have a boundary

$$\partial S := \overline{S} \setminus \text{int}(S)$$

with content zero.
Then, every function $f : S \to \mathbb{R}$ that is bounded and continuous is integrable on S.

Proof
We extend f to a box $S \subseteq I$ (ie $f(\vec{x}) = 0, \vec{x} \in I \setminus S$).
Discontinuities must be confined to the ∂S, so Lebesgue's Criterion applies.

Proposition 30.1.2
Let $S \subseteq \mathbb{R}^n$ be nonempty and bounded, and has content zero
Then every function $f : S \to \mathbb{R}$ that is bounded is integrable and

$$\int_S f(\vec{x}) d\vec{x} = 0$$

Proposition 30.1.3
Let S be as in the previous proposition.
Then S has content if and only if ∂S has content zero.

Proof
$\Longrightarrow$
If S has content zero, then by Proposition 29.1.1, for any $\epsilon > 0$, we can find a finite set of boxes which cover S such that the total volume of the boxes are less than ϵ. We can simply extend the box to also cover the boundary, so the boundary necessarily has content zero.
Elsewise, suppose S has non-zero content. In addition, the supremum of the Lower Riemann Sums is non-zero. Then for any $\epsilon > 0$, there is some partition such that the difference between Upper Riemann Sum and Lower Riemann Sum is less than ϵ. This means in particular means that there are very few sub-boxes in the select partition (each having a very small volume) which intersect a point $\vec{x} \in I, \chi_S(\vec{x}) = 0$. We can then contain the points of the boundary in those boxes and thus cover the boundary with boxes of arbitrarily small volume. Again by Proposition 29.1.1, the boundary necessarily has content zero.
$\Longleftarrow$
The reverse is a corollary of Corollary 30.1.1.1 since the characteristic function of S is continuous on S.

31 Properties of Riemann Integrals

31.1 Properties of the Riemann Integral

Theorem 31.1.1
Let $S \subseteq \mathbb{R}^n$ be nonempty and bounded, let $f, g : S \to \mathbb{R}$ be integrable on S.

1. For any $\alpha, \beta \in \mathbb{R}$, $h := \alpha f + \beta g$ is integrable on S and

$$\int_S \alpha f(\vec{x}) + \beta g(\vec{x}) d\vec{x} = \alpha \int_S f(\vec{x}) d\vec{x} + \beta \int_S g(\vec{x}) d\vec{x}$$

2. If $f(\vec{x}) \leq g(\vec{x})$ for all $\vec{x} \in S$, then

$$\int_S f(\vec{x}) d\vec{x} \leq \int_S g(\vec{x}) d\vec{x}$$

3. $|f|$ is integrable and

$$\int_S |f(\vec{x})| \geq |\int_S f(\vec{x}) d\vec{x}|$$

4. If S has content, then

$$m\mu(S) \leq \int_S f(\vec{x}) d\vec{x} \leq M\mu(S)$$

where m, M are respectively the lower and upper bounds of f on S.

Proposition 31.1.2
Suppose we have nonempty sets S, T that are bounded and satisfty $\mu(S \cap T) = 0$. If $f : S \cup T \to \mathbb{R}$ is bounded and integrable on S, T, then f is integrable on $S \cup T$ and

$$\int_{S \cup T} f(\vec{x}) d\vec{x} = \int_S f(\vec{x}) d\vec{x} + \int_T f(\vec{x}) d\vec{x}$$

Proof
Since S, T are bounded, $S \cup T$ is bounded and we can contain in within a box $I \subseteq \mathbb{R}^n$. Define functions

$$\begin{aligned} u, v, w &: I \to \mathbb{R} \\ u &= f\chi_S, && \vec{x} \in S \cup T \\ v &= f\chi_T, && \vec{x} \in S \cup T \\ w &= f\chi_{S \cap T}, && \vec{x} \in S \cap T \end{aligned}$$

Then $f(\vec{x}) = u(\vec{x}) + v(\vec{x}) - w(\vec{x})$ for all $\vec{x} \in S \cup T$.
By the hypothesis, u, v are integrable on S, T repsectively and therefore on I. By proposition further down, w is integrable on I and

$$\int_I w(\vec{x})d\vec{x} = 0$$

By theorem 31.1.1 Part 1, f is integrable on I, so on $S \cup T$. We also get the required formula.

Theorem 31.1.3
Let $S \subseteq \mathbb{R}^n$ be nonempty and bounded, and connected. Suppose that S has content. For any continuous function $f : S \to \mathbb{R}$, there is some $\vec{c} \in S$ such that

$$\int_S f(\vec{x})d\vec{x} = f(\vec{x})\mu(S)$$

Proof
Exercise.

If $S \subseteq \mathbb{R}^n$ is bounded and has nonzero content and $f : S \to \mathbb{R}$ is integrable on S, then we define the average or mean value of f over S by

$$\text{mean}(f, S) := \frac{\int_S f(\vec{x})d\vec{x}}{\mu(S)}$$

31.2 Alternative Characterization of Integrability

Theorem 31.2.1
Let $I \subseteq \mathbb{R}^n$ be a box and let $f : I \to \mathbb{R}$ be bounded. f is Riemann Integrable on I and

$$\int_I f(\vec{x})d\vec{x} = V$$

if and only if:
for all $\epsilon > 0$, there is $\delta > 0$ such that for any partition P of I with $\|P\| < \delta$, every Riemann Sum $S(f, P)$ satistisfies

$$|S(f, P) - V| < \epsilon$$

Proof (sketch)
From Theorem 28.2.4, for any $\epsilon > 0$, we can find P such that

$$0 \leq U(f, P) - L(f, P) < \epsilon$$

...

$$U(f, P) - V < \epsilon \wedge V - L(f, p) < \epsilon$$

So, for any Riemann Sum,

$$S(f, P) - V < \epsilon \wedge V - S(f, P) < \epsilon \implies |S(f, p) - V| < \epsilon$$

Suppose Q is a partition of I (not necessarily a refinement of P). Take their common refinement R

32 Fubini's Theorem I

32.1 Fubini with Two Variables

Theorem 32.1.1 (Fubini's Theorem with Two Variables)
Let $I := [a, b] \times [c, d] \subseteq \mathbb{R}^n$ be a box with $a < b$ and $c < d$.
Let $f : I \to \mathbb{R}$ be bounded and integrable on I.
If for each $x \in [a, b)$, the function $f(x, \cdot)$ is integrable on $[c, d]$, then

$$\int_c^d f(\cdot, y)dy$$

is integrable on the inverval $[a, b]$ and

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

Proof
Let $\epsilon > 0$. We will show that for partitions P of $[a, b]$ with small enough norm, the Riemann Sums satisfy

$$\left| S\left(\int_c^d f(\cdot, y)dy, P\right) - \int_I f(x, y)d(x, y)\right| < \epsilon$$

.
By Theorem 31.2.1, there is $\eta > 0$ such that for any partition T of I with $\|T\| < \eta$, any Riemann Sum $S(f, P)$ satisfies

$$\left| S(f, T) - \int_I f(x, y)d(x, y)\right| < \frac{\epsilon}{2}$$

.
Let $P := \{x_0, x_1, \ldots, x_l\}$ be a parititon of $[a, b]$ with $\|P\| < \eta$ and let $Q := \{y_0, y_1, \ldots, y_m\}$ be a partition of $[c, d]$ with $\|Q\| < \eta$.
Define $T := \{P, Q\}$ to be a partition of I.
Note that $\|T\| < \eta$.
For each $i \in \{1, \ldots, l\}$, choose $\overline{x}^{(i)} \in [x^{(i-1)}, x^{(i)}]$, and choose $\overline{y}^{(j)} \in [y^{(j-1)}, y^{(j)}]$ for each $j \in \{1, 2, \ldots, m\}$.
Then,

$$\left| S(f, T) - \int_I f(x, y)d(x, y)\right| < \frac{\epsilon}{2}$$

where

$$S(f, T) := \sum_{i=1}^{l} \sum_{j=1}^{m} f(\overline{x}^{(i)}, \overline{y}^{(j)})[x^{(i)} - x^{(i-1)}][y^{(j)} - y^{(j-1)}]$$

For any $\in \{1, 2, \ldots, l\}$, we know (by hypothesis) that $f(\overline{x}^{(i)}, \cdot)$ is integrable on $[c, d]$. By Theorem 31.2.1, there is $\delta^{(i)} > 0$ such that

$$\left|\sum_{i=1}^{m} f(\overline{x}^{(i)}, \overline{y}^{(j)})[y^{(j)} - y^{(j-1)}] - \int_c^d f(\overline{x}^{(i)}, y)dy\right| < \frac{\epsilon}{2(b-a)}$$

provided $\|Q\| < \delta^{(i)}$.
Define $\delta := \min\{\eta, \delta^{(1)}, \ldots, \}$. If $\|Q\| < \delta$, then

$$\left|S(f,T) - \sum_{i=1}^{l} \int_c^d f(\overline{x}^{(i)}, y)dy[x^{(i)} - x^{(i-1)}]\right| < \frac{\epsilon}{2(b-a)} \sum_{i=1}^{l}[x^{(i)} - x^{(i-1)}] = \frac{\epsilon}{2}$$

By the Triangle Inequality,

$$\begin{aligned}
&\left|\sum_{i=1}^{l} \int_c^d f(\overline{x}^{(i)}, y)dy[x^{(i)} - x^{(i-1)}] - \int_I f(x,y)d(x,y)\right| \\
&\leq \left|S(f,T) - \sum_{i=1}^{l} \int_c^d f(\overline{x}^{(i)}, y)dy[x^{(i)} - x^{(i-1)}]\right| + \left|S(f,T) - \int_I f(x,y)d(x,y)\right| \\
&< \epsilon
\end{aligned}$$

Corollary 32.1.1.1
If in addition to the hypotheses of the theorem, the function $f(\cdot, y)$ is integrable on $[ab]$ for each $y \in [c, d]$. Then $\int_a^b f(\cdot, y)dy$ is integrable on $[c, d]$ and

$$\int_I f = \int_a^b \int_c^d f = \int_c^d \int_a^b f$$

Corollary 32.1.1.2
Let $S := \{(x, y) \in \mathbb{R}^2 : a \leq x \leq b, l(x) \leq y \leq u(x)\}$ where $a < b$ and $l(x) \leq u(x)$ for all $x \in [a, b]$, and $l, u \in C([a, b], \mathbb{R})$. Then for all $f \in C(S, \mathbb{R})$,

$$\int_S f(x,y)d(x,y) = \int_a^b \int_{l(x)}^{u(x)} f(x,y)dydx$$

Integrals of the form $\int_{S_x} \int_{S_y} f(x,y)dydx$ are called **iterated integrals**.

33 Fubini's Theorem II and Change of Variables

33.1 Fubini with Possibly More than Two Variable

Theorem 33.1.1 (Fubini's Theorem)
Let $A \subseteq \mathbb{R}^n$ and $B \subseteq \mathbb{R}^m$ be two boxes and let $f : A \times B \to \mathbb{R}$ be a bounded and integrable function on $A \times B$.
If for each $\vec{x} \in A$, the function $f(\vec{x}, \cdot)$ is integrable on B, then the $\int_B f(\cdot, \vec{y}) d\vec{y}$ is integrable on A, and

$$\int_{A\times B} f(\vec{x}, \vec{y}) d(\vec{x}, d\vec{y}) = \int_A \int_B f(\vec{x}, \vec{y}) d\vec{y} d\vec{x}$$

Corollary 33.1.1.1
Let $A \subseteq \mathbb{R}^2$ be a compact set that has content and let

$$S := \{(x, y, z) \in \mathbb{R}^3 : (x, y) \in A, 0 \leq z \leq f(x, y)\}$$

where f is continuous and non-negative.
Then

$$\mu(S) = \int_A f(x, y) d(x, y)$$

Proof
A is compact and hence bounded.
Suppose $I_A \subseteq \mathbb{R}^2$ is a box that contains A.
Since f is continuous on a compact set, there is an upper bound b. So $S \subseteq I_A \times [0, b] =: I$.
We want to apply Fubini's Theorem to say that

$$\begin{aligned} \mu(S) &= \int_I \chi_S(x, y, z) d(x, y, z) \\ &= \int_{I_A} \left[\int_0^b \chi_S(x, y, z) dz \right] d(x, y) \\ &= \int_{I_A} \left[\int_0^{f(x,y)} 1 dz \right] d(x, y) \\ &= \int_{I_A} f(x, y) d(x, y) \end{aligned}$$

We must first check that hypothesis of Fubini are satisfied.

i) χ_S is integrable on I (ie S has content). This is true by the Corollary to Lebesgue's Criterion, which says that χ_S is integrable if the boundary ∂S has content zero.

Note that $\partial S \subseteq U \cup V \cup W$, where

$$U := \partial A \times [0,b], V := \{(x,y,f(x,y)) : (x,y) \in A\}, W := \{(x,y,0) : (x,y) \in A\}$$

It is easy to see that W has content zero. Also, V has content zero by Proposition 29.3.1. A has content hypothsiss, so this implies U has content zero.

$$\begin{aligned}
&\implies U \cup V \cup W \text{ has content zero} \\
&\implies \partial S \text{ has content zero} \\
&\implies S \text{ has content}
\end{aligned}$$

ii) $\chi_S(x,y,\cdot)$ is integrable on $[0,b]$ for all

$$(x,y) \in A := \begin{cases} 1, & z \leq f(x,y) \\ 0 \end{cases}$$

34 Change of Variables

Theorem 34.0.1
Let $U \subseteq \mathbb{R}^n$ be nonempty and open and let $S \subseteq U$ be nonempty, compact, and have content.
Let $\psi \in C^1(U, \mathbb{R}^n)$ be a transformation that is an injection on $S \setminus T$, where T is either empty or has content zero.
If $\det(D\psi(\vec{x})) \neq 0$ for all $\vec{x} \in S \setminus T$, then $\psi(S)$ has content.
Futhermore, if $f : \psi(S) \to \mathbb{R}$ is bounded and integrable on $\psi(S)$, then

$$\int_{\psi(S)} f(\vec{u})d\vec{u} = \int_S f(\psi(\vec{x})) \cdot |\det(D\psi(\vec{x}))|d\vec{x}$$

To see that the above aligns with the one-dimensional case, suppose $\psi \in C^1, \psi'(x) \neq 0$ on $[a, b]$.
Let $f : [a, b] \to \mathbb{R}$

$$\int_{\psi([a,b])} f(u)du = \int_a^b f(\psi(x))|\psi'(x)|dx$$

Two Cases:

1. $\psi'(x) > 0$ on $[a, b]$. So $\psi(a) < \psi(b)$ and $|\psi'(x)| = \psi'(x)$.

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

2. $\psi'(x) < 0$ on $[a, b]$ so $\psi(a) > \psi(b)$.

$$\int_{\psi([a,b])} f(u)du = \int_b^a f(\psi(x))\psi'(x)dx$$

34.1 Polar Coordinates

How to we convert $\int_X f(x, y)d(x, y)$ into an integral over (r, θ)?
Suppose $X \subseteq \mathbb{R}^2$ is closed, bounded, and has content.
Then, ther exists $R > 0$ such that $\|\vec{x}\| \leq R$ for all $\vec{x} \in X$.
If $f : X \to \mathbb{R}$ is bounded and integrable on X, then we can apply Change of Variables.
Define $\psi \in C^1(\mathbb{R}^2, \mathbb{R}^2)$ by

$$\psi(r, \theta) := (r \cos\theta, r \sin\theta)$$

By previous computation,

$$\det(D\psi(r, \theta)) = r$$

Let $S := \psi^{-1}(X) \cap \{(r, \theta) : r \geq 0, 0 \leq \theta \leq 2\pi\}$
Since $\|\vec{x}\| = r \leq \mathbb{R}$ for all $\vec{x} \in X$, so S is bounded.
Also, it is closed since we take the intersection of two closed sets.
It follows that S is compact.
We must assume S has content and is nonempty. By our previous work, ∂S has content zero.
From our previous definition,

$$T := S \cap \{(r, \theta) : r = 0 \vee \theta = 2\pi\} \subseteq \partial S$$

has content zero.
Then, we may apply the Change of Variables Formula!

$$\int_X f(x, y) d(x, y) = \int_S f(r \cos \theta, r \sin \theta) r d(r, \theta)$$

. Note that we may be able to use Fubini to decompose the integral.

35 Final Exam Format

Section 1-3 less weight, 4-7 more weight.