

# Multivariable Calculus

Labix

December 12, 2024

## Abstract

The aim of these notes is to develop the notion of continuity, differentiability and some parts of integrability in  $\mathbb{R}^n$  by attempting to generalize the same ideas in  $\mathbb{R}$ . Another major aim is to investigate non linear transformations through the lens of linear algebra.

Assumes knowledge of Linear Algebra 1, Real analysis.

Note:  $M_{m \times n}(\mathbb{R})$  means the set of all matrices with  $m$  rows and  $n$  columns.

## Contents

<b>1</b>	<b>Analysis on <math>\mathbb{R}^n</math> and <math>M_{m \times n}(\mathbb{R})</math></b>	<b>3</b>
1.1	Sequences on $\mathbb{R}^n$	3
1.2	Continuity on $\mathbb{R}^n$	4
1.3	Topological Properties	6
1.4	The Space of Matrices	8
<b>2</b>	<b>Differentiation</b>	<b>11</b>
2.1	Frechet Derivative	11
2.2	Jacobian Matrix and Directional Derivatives	13
2.3	Properties of the Derivative	15
2.4	Mean Value Inequality	17
2.5	Conditions for Differentiability	17
<b>3</b>	<b>More Properties of Differentiability</b>	<b>20</b>
3.1	Inverse Function Theorem	20
3.2	Implicit Function Theorem	23
3.3	Higher Order Derivatives	23
3.4	Second Order Derivative Test	25
<b>4</b>	<b>Vector Calculus</b>	<b>26</b>
4.1	Regions	26
4.2	Differentiation and Integration of Curves	26
4.3	Vector Fields	28
<b>5</b>	<b>2 Dimensional Calculus</b>	<b>30</b>
5.1	Double Integrals	30
5.2	Green's Theorem for a Planar Region	31
5.3	Divergence Theorem for a Planar Region	32
<b>6</b>	<b>3 Dimensional Calculus</b>	<b>33</b>
6.1	Triple Integrals	33
6.2	Surface Integrals	34
6.3	Divergence Theorem	34
6.4	Stoke's Theorem	35

# 1 Analysis on $\mathbb{R}^n$ and $M_{m \times n}(\mathbb{R})$

## 1.1 Sequences on $\mathbb{R}^n$

### Definition 1.1.1: Euclidean Metric

The Euclidean Metric of  $x, y \in \mathbb{R}^n$  is defined to be

$$d(x, y) = \left( \sum_{k=1}^n (x_k - y_k)^2 \right)^{\frac{1}{2}}$$

We denote it as  $|x - y|$ .

### Proposition 1.1.2

$\mathbb{R}^n$  and the Euclidean Metric forms a metric space.

*Proof.* We show that the Euclidean Metric is indeed a metric on  $\mathbb{R}^n$ . Let  $x, y, z \in \mathbb{R}^n$ .

- $(x_k - y_k)^2 \geq 0$  for all  $k \in \{1, \dots, n\}$  with equality if and only if  $x_k = y_k$
- $(x_k - y_k)^2 = (y_k - x_k)^2$  for all  $k \in \{1, \dots, n\}$
- $d(x, y) \leq d(x, z) + d(y, z)$  for any  $z \in \mathbb{R}^n$

□

### Definition 1.1.3: Convergence in $\mathbb{R}^n$

A sequence  $\{x_n\} \subseteq \mathbb{R}^n$  is said to converge to  $x \in \mathbb{R}^n$  if for every  $\epsilon > 0$  there exists  $N \in \mathbb{N}$  such that  $d(x_n, x) < \epsilon$  for all  $n > N$ .

### Proposition 1.1.4: Componentwise Convergence

$x_k \in \mathbb{R}^n$  converges if and only if  $x_{i,k}$  converges for every  $i \in \{1, \dots, n\}$ .

*Proof.* Suppose that  $|x_k - x| < \epsilon$ . Then in particular  $|x_{i,k} - x_i| < |x_k - x| < \epsilon$  and thus is convergent componentwise. Suppose that  $\{x_k\}$  is convergent componentwise. Then there exists  $N_1, \dots, N_n$  such that  $n_j > N_j$  implies  $|x_{j,k} - x_j| < \epsilon$  for  $j \in \{1, \dots, n\}$ . Take the max of all  $N_1, \dots, N_n$ . Then

$$\begin{aligned} |x_k - x| &= \left( \sum_{j=1}^n (x_j - x_{j,k})^2 \right)^{\frac{1}{2}} \\ &< \sqrt{n}\epsilon \end{aligned}$$

□

### Theorem 1.1.5: Bolzano-Weierstrass Theorem

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

*Proof.* Let  $\{x_n\} \subset \mathbb{R}^n$  be bounded. By Bolzano-Weierstrass Theorem on  $\mathbb{R}$ , the sequence  $\{x_{1,m}\} \subset \mathbb{R}$  has a convergent subsequence  $\{x_{1,m_k}\}$ . Apply Bolzano-Weierstrass Theorem to  $\{x_{2,m_k}\}$  and keep going until you reach the  $n$ th component. We will end up with a subsequence that converges for all components and thus is convergent in  $\mathbb{R}^n$ .

□

**Proposition 1.1.6: Sum Rule of Sequences**

Let  $\{x_n\} \subset \mathbb{R}^n$  converge to  $x \in \mathbb{R}^n$  and  $\{y_n\} \subset \mathbb{R}^n$  converge to  $y \in \mathbb{R}^n$ . Let  $a, b \in \mathbb{R}$ . Then

$$ax_n + by_n \rightarrow ax + by$$

*Proof.* Trivial using the characterization of componentwise convergence and real analysis.  $\square$

**1.2 Continuity on  $\mathbb{R}^n$** **Definition 1.2.1: Limits**

Let  $f : U \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$ . Let  $a \in \mathbb{R}^n$  be a limit point of  $U$ . Let  $b \in \mathbb{R}^m$ . We say that

$$\lim_{x \rightarrow a} f(x) = b$$

if for every  $\epsilon > 0$  there is some  $\delta > 0$  such that for every  $x \in U$  such that  $\|x - a\| < \delta$  implies  $\|f(x) - b\| < \epsilon$

$$\|x - a\| < \delta \implies \|f(x) - b\| < \epsilon$$

**Definition 1.2.2: Continuity**

Let  $f : U \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^m$ . We say that  $f$  is continuous at  $a \in U$  if

$$\lim_{x \rightarrow a} f(x) = f(a)$$

**Theorem 1.2.3: Componentwise Continuity**

Let  $f : U \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^m$  be a function. Then  $f$  is continuous if and only if each of its components  $f_1, \dots, f_n : U \rightarrow \mathbb{R}$  is continuous.

*Proof.* Suppose that  $f$  is continuous at  $a \in \mathbb{R}^n$ . Let  $\epsilon > 0$ . Then there exists  $\delta > 0$  from continuity such that  $\|x - a\| < \delta$  implies  $\|f(x) - f(a)\| < \epsilon$ . Then  $\|x - a\| < \delta$  implies

$$\|f_i(x) - f_i(a)\| \leq \|f(x) - f(a)\| < \epsilon$$

Thus each component is continuous.

Now let  $f$  be component wise continuous at  $a \in \mathbb{R}^n$ . Let  $\epsilon > 0$ . For each component  $f_i : \mathbb{R}^n \rightarrow \mathbb{R}$ , using continuity with  $\frac{\epsilon}{\sqrt{n}} > 0$  there exists  $\delta_i > 0$  such that  $\|x - a\| < \delta$  implies  $\|f_i(x) - f_i(a)\| < \frac{\epsilon}{\sqrt{n}}$ . Choose  $\delta = \max\{\delta_1, \dots, \delta_n\}$ . Then  $\|x - a\| < \delta$  implies that

$$\begin{aligned} \|f(x) - f(a)\|^2 &= \sum_{k=1}^n \|f_k(x) - f_k(a)\|^2 \\ &\leq n \max_{k \in \{1, \dots, n\}} \|f_k(x) - f_k(a)\|^2 \\ \|f(x) - f(a)\| &\leq \sqrt{n} \max_{k \in \{1, \dots, n\}} \|f_k(x) - f_k(a)\| \\ &\leq \sqrt{n} \frac{\epsilon}{\sqrt{n}} \\ &= \epsilon \end{aligned}$$

Thus  $f$  is continuous.  $\square$

**Theorem 1.2.4: Sequential Continuity**

Let  $f : U \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^m$ . Then  $f$  is continuous at  $c \in U$  if and only if for every sequence  $\{x_n | n \in \mathbb{N}\} \subset U$  that  $x_n \rightarrow c \in U$  it has the property that

$$f(x_n) \rightarrow f(c)$$

*Proof.* Exactly the same proof as that of real analysis. □

**Proposition 1.2.5: Sum Rule**

Let  $f, g : U \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^m$  be continuous at  $p \in \mathbb{R}^n$ . Then for any  $a, b \in \mathbb{R}^m$ ,  $af(x) + bg(x)$  is continuous at  $p$ .

*Proof.* Simple proof involving componentwise continuity and real analysis. □

**Proposition 1.2.6: Product and Quotient Rule**

Let  $f, g : U \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$  be continuous at  $p \in \mathbb{R}^n$ . Then

- $f(x)g(x)$  is continuous at  $p$
- $\frac{f(x)}{g(x)}$  is continuous at  $p$  given that  $g(x) \neq 0$  for all  $x \in U$

*Proof.* Also a simple proof involving componentwise continuity and real analysis. □

**Proposition 1.2.7: Composition Rule**

Let  $f : U \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^m$  and  $g : V \subseteq \mathbb{R}^m \rightarrow \mathbb{R}^k$  be continuous at  $p \in \mathbb{R}^n$  and  $f(p) \in V$  respectively. Then  $g(f(x))$  is continuous at  $p$ .

*Proof.* Another simple proof involving componentwise continuity and real analysis. □

**Definition 1.2.8: Linear Continuity**

A function  $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$  is linearly continuous if given any line  $g(t) = x_0 + tv$  in  $\mathbb{R}^n$  where  $x_0, v \in \mathbb{R}^n$ , we have that

$$\lim_{t \rightarrow 0} f(x_0 + tv) = f(x_0)$$

**Definition 1.2.9: Separate Continuity**

A function  $f : \mathbb{R}^2 \rightarrow \mathbb{R}$  is separately continuous at  $(x_0, y_0)$  if  $g(x) = f(x, y_0)$  and  $h(y) = f(x_0, y)$  are both continuous.

**Proposition 1.2.10**

Let  $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$  be continuous. If  $f$  is continuous, then  $f$  is linearly continuous. If  $f$  is linearly continuous, then  $f$  is separately continuous.

Beware that none of the inverse implications hold. But we can use the contrapositive to prove that functions are not continuous. Namely if we can prove that  $f$  is not separately continuous, then clearly  $f$  will not be continuous. This is precisely why we made this definitions in the first place.

**Proposition 1.2.11**

Let  $g : U \subset \mathbb{R} \rightarrow \mathbb{R}$ . Define  $U_i = \{x \in \mathbb{R}^n | x_i \in U\}$ . This  $U_i$  is the set of all  $\mathbb{R}^n$  such that the  $i$ th element is in  $U$ . Define  $f : U_i \subset \mathbb{R}^n \rightarrow \mathbb{R}$  by  $f(x) = g(x_i)$ . If  $g$  is continuous at  $a \in U$  then  $f$  is continuous at  $\{x \in \mathbb{R}^n | x_i = a\}$ .

**1.3 Topological Properties**

We will only develop sufficient topological properties so that we can apply it to our proofs. The complete development of topology is in another set of notes.

**Definition 1.3.1: Open Ball**

Let  $p \in \mathbb{R}^n$  and  $r > 0$ . The open ball at  $p$  with radius  $r$  is defined to be

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

**Definition 1.3.2: Open and Closed Sets**

We say that a set  $U \subset \mathbb{R}^n$  is

- open if for all  $p \in U$  there exists  $r > 0$  such that  $B_r(p) \subseteq U$
- closed if  $\mathbb{R}^n \setminus U$  is open

**Proposition 1.3.3**

Let  $U \subseteq \mathbb{R}^n$  is closed if and only if for every sequence  $(x_n)_{n \in \mathbb{N}} \subset U$  that converges to  $x \in \mathbb{R}^n$ ,  $x \in U$ .

*Proof.* Suppose that  $U$  is closed. Suppose that  $(x_n)_{n \in \mathbb{N}} \subset U$  is a sequence such that it converges to some  $x \in \mathbb{R}^n$ . Suppose for a contradiction that  $x \in \mathbb{R}^n \setminus U$ . Then  $\mathbb{R}^n \setminus U$  being open means that there exists some  $r > 0$  such that  $B_r(x) \subseteq \mathbb{R}^n \setminus U$ . By definition of convergent, there exists  $N \in \mathbb{N}$  such that  $x_n \in B_r(x)$  for all  $n > N$ . But this contradicts the fact that  $x_n \in U$ . Thus we must have  $x \in U$ .

Now suppose that every sequence  $(x_n)_{n \in \mathbb{N}} \subset U$  has limit  $x \in U$ . Suppose for a contradiction that  $U$  is not closed. Then  $\mathbb{R}^n \setminus U$  is not open and that there exists  $x \in \mathbb{R}^n \setminus U$  such that for all  $r > 0$ ,  $B_r(x)$  is not a subset of  $\mathbb{R}^n \setminus U$ . Let  $(y_n)_{n \in \mathbb{N}}$  be a sequence with the property that  $y_n \in B_{1/k}(x)$  but  $y_n \notin \mathbb{R}^n \setminus U$ . Then  $y_n \in U$  and  $y_n \rightarrow x \in \mathbb{R}^n \setminus U$ , a contradiction.  $\square$

**Definition 1.3.4: Bounded Sets**

We say that a set  $U \subset \mathbb{R}^n$  is bounded if there exists an open ball  $B_r(p)$  such that  $U \subset B_r(p)$ .

**Theorem 1.3.5**

Let  $f : \mathbb{R}^n \rightarrow \mathbb{R}^k$  be a function. Then the following are equivalent.

- $f$  is continuous
- $V \subseteq \mathbb{R}^k$  is open implies  $f^{-1}(V)$  is open
- $U \subseteq \mathbb{R}^k$  is closed implies  $f^{-1}(U)$  is closed

*Proof.* Suppose that  $f$  is continuous. Then  $V$  being open means that if  $f(x) \in V$ , then  $B_\epsilon(f(x)) \subseteq V$  for some  $\epsilon > 0$ . For this  $\epsilon$ , apply continuity. Then there exists  $\delta > 0$  such that  $f(B_\delta(x)) \subseteq B_\epsilon(f(x))$ . Then this means that

$$B_\delta(x) \subseteq f^{-1}(B_\epsilon(f(x))) \subseteq f^{-1}(V)$$

and we are done.

Now suppose that  $f$  has the second property. Let  $f(x) \in V \subseteq \mathbb{R}^k$  be open. Then there exists  $\epsilon > 0$  such that  $B_\epsilon(f(x)) \subseteq V$ . Thus  $f^{-1}(B_\epsilon(f(x)))$  is also open. This being open means that there exists  $\delta > 0$  such that  $B_\delta(x) \subseteq f^{-1}(B_\epsilon(f(x)))$  and

$$f(B_\delta(x)) \subseteq B_\epsilon(f(x))$$

and we are done.

For the closed property, simply take the complements of the entire proof.  $\square$

### Definition 1.3.6: Sequential Compactness

Let  $K \subset \mathbb{R}^n$  be a set. Then  $K$  is sequentially compact if every sequence  $(x_n)_{n \in \mathbb{N}} \subset K$  has a convergent subsequence  $(x_{n_j})_{j \in \mathbb{N}}$  that converges to some  $k \in K$ .

### Proposition 1.3.7: Heine-Borel Theorem

$K \subset \mathbb{R}^n$  is sequentially compact if and only if  $K$  is closed and bounded.

*Proof.* Let  $K$  be compact. Let  $(x_n)_{n \in \mathbb{N}} \subset K$  be a convergent sequence. By sequential compactness,  $(x_{n_k})_{k \in \mathbb{N}}$  is a subsequence that converges to  $x \in K$ . But  $(x_n)_{n \in \mathbb{N}}$  has the same limit as its subsequence thus  $x_n \rightarrow x$ . Now suppose for a contradiction that  $K$  is unbounded. Then there exists a sequence  $(x_n)_{n \in \mathbb{N}}$  such that  $|x_n| \geq n$  for all  $n \in \mathbb{N}$ . By sequential compactness, there is a subsequence  $(x_{n_k})_{k \in \mathbb{N}}$  such that its limit is  $x \in K$ . This means that  $(x_{n_k})_{k \in \mathbb{N}}$  is bounded, a contradiction.

Now suppose that  $K$  is closed and bounded. Let  $(x_n)_{n \in \mathbb{N}}$  be a sequence in  $K$ . Then  $K$  being bounded means that  $(x_n)_{n \in \mathbb{N}}$  is bounded. By the Bolzano-Weierstrass theorem, it has a convergent subsequence  $(x_{n_k})_{k \in \mathbb{N}}$  with limit in  $K$  since  $K$  is closed.  $\square$

### Theorem 1.3.8

Let  $f : K \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$  be continuous and  $K$  sequentially compact. Then  $f(K)$  is sequentially compact.

*Proof.* Let  $(y_n)_{n \in \mathbb{N}}$  be a sequence in  $f(K)$ . Then there exists a sequence  $(x_n)_{n \in \mathbb{N}}$  such that  $y_n = f(x_n)$  for all  $n \in \mathbb{N}$ . By sequential compactness, there exists a subsequence  $(x_{n_k})_{k \in \mathbb{N}}$  that is convergent to  $x \in K$ . By sequential continuity of  $f$  at  $x$ ,  $y_{n_k} \rightarrow f(x)$  and we are done.  $\square$

### Theorem 1.3.9: Extreme Value Theorem

Let  $f : K \subset \mathbb{R}^n \rightarrow \mathbb{R}$  be continuous and  $K$  sequentially compact. Then there exists  $a, b \in K$  such that

$$f(a) \leq f(x) \leq f(b)$$

for all  $x \in K$ .

*Proof.* We know that  $f(K)$  is closed and bounded. Thus it must have a supremum  $M$  and infimum  $m$  that are finite. Then we must have  $(a_n)_{n \in \mathbb{N}}$  and  $(b_n)_{n \in \mathbb{N}}$  such that  $a_n \rightarrow m$  and  $b_n \rightarrow M$  by definition of supremum and infimum. Since  $f(K)$  is closed,  $m, M \in f(K)$  and we are done.  $\square$

## 1.4 The Space of Matrices

### Definition 1.4.1: Frobenius Norm

Let  $A \in M_{k \times n}(\mathbb{R})$ . Define the Frobenius norm to be

$$\|A\|_F = \left( \sum_{j=1}^n \sum_{i=1}^k a_{ij}^2 \right)^{\frac{1}{2}}$$

### Definition 1.4.2: Operator Norm

Let  $T \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^k)$ . Define the operator norm to be

$$\|T\| = \sup_{x \in \mathbb{R}^n \setminus \{0\}} \frac{|T(x)|}{|x|} = \sup_{\substack{x \in \mathbb{R}^n \\ |x|=1}} |T(x)|$$

The definition of the operator norm given on  $S^{n-1}$  is particularly useful because  $S^{n-1}$  is sequentially compact. Since  $\|A\|$  is a supremum, there must exist a sequence  $(x_n)_{n \in \mathbb{N}}$  such that  $|Ax_n| \rightarrow \|A\|$ .

### Proposition 1.4.3

Let  $T, S \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^k)$ . The operator norm satisfies the following and thus is a norm on  $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^k)$

- $\|T\| = 0$  if and only if  $T$  is the zero mapping
- $\|\lambda T\| = |\lambda| \|T\|$  for any  $\lambda \in \mathbb{R}$
- $\|T + S\| \leq \|T\| + \|S\|$

We want to compare the two norms and show that they are equivalent. But they are in two different vector spaces. We need to associate the two vector space into the same one first.

### Proposition 1.4.4

Both  $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^k)$  and  $M_{k \times n}(\mathbb{R})$  is a vector space over  $\mathbb{R}$  where  $n, k \in \mathbb{N} \setminus \{0\}$ . Moreover, we have

$$\mathcal{L}(\mathbb{R}^n, \mathbb{R}^k) \cong M_{k \times n}(\mathbb{R}) \cong \mathbb{R}^{nk}$$

*Proof.* The fact that  $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^k)$  and  $M_{k \times n}(\mathbb{R})$  is a vector space is trivial. Since they have the same dimension and we know that all finite dimensional vector space is isomorphic to  $\mathbb{R}^p$  for some  $p \in \mathbb{N} \setminus \{0\}$ , we get the desired result.  $\square$

### Proposition 1.4.5

The Frobenius norm and the operator norm are equivalent.

*Proof.* Given that they actually act on the same vector space, we can associate every



linear transformation  $T$  with the matrix  $A$ . Now we have

$$\begin{aligned}
 \|A\|_F^2 &= \sum_{j=1}^n \sum_{i=1}^k a_{ij}^2 \\
 &= \sum_{j=1}^n |T(e_j)|^2 \\
 &= \sum_{j=1}^n |T(e_j)|^2 \\
 &\leq \|T\|^2 \sum_{j=1}^n |e_j|^2 && \left( \frac{|T(x)|}{|x|} \leq \|T\| \right) \\
 &= n\|T\|^2
 \end{aligned}$$

Thus we have  $\|A\|_F \leq \sqrt{n}\|T\|$ .

Now for any  $x \in \mathbb{R}^n$ , we also have

$$\begin{aligned}
 |T(x)|^2 &= |Ax|^2 \\
 &= \sum_{i=1}^k \left( \sum_{j=1}^n a_{ij} x_j \right)^2 \\
 &\leq \sum_{i=1}^k \left( \left( \sum_{j=1}^n a_{ij}^2 \right) \left( \sum_{j=1}^n x_j^2 \right) \right) && \text{(Cauchy-Schwarz Inequality)} \\
 &= \left( \sum_{i=1}^k \sum_{j=1}^n a_{ij}^2 \right) |x|^2 \\
 &= \|A\|_F^2 |x|^2
 \end{aligned}$$

If  $x \neq 0$ , we can write it as

$$\begin{aligned}
 \frac{|T(x)|^2}{|x|^2} &\leq \sup_{x \in \mathbb{R}^n \setminus \{0\}} \frac{|T(x)|^2}{|x|^2} \leq \|A\|_F^2 \\
 \|T\|^2 &\leq \|A\|_F^2
 \end{aligned}$$

Thus we now have

$$\|T\| \leq \|A\|_F \leq \sqrt{n}\|T\|$$

and we are done. □

Now we can make sense of using the operator norm for matrices, and the frobenius norm for linear maps.

#### Proposition 1.4.6

Let  $T \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^k)$  and  $S \in \mathcal{L}(\mathbb{R}^k, \mathbb{R}^m)$ . Then

$$\|S \circ T\| \leq \|S\| \|T\|$$

*Proof.* We have that

$$\begin{aligned}
 |S(T(x))| &\leq \|S\| |T(x)| \\
 &\leq \|S\| \|T\| |x|
 \end{aligned}$$

Thus we are done. □

#### Proposition 1.4.7

The function  $\det(\cdot) : \mathbb{R}^{n \times n} \rightarrow \mathbb{R}$ , which is the determinant, is continuous with respect to the elements of the matrix.

*Proof.* Note that  $\mathbb{R}^{n \times n} \cong \mathbb{R}^{n^2}$  thus the determinant really is just a linear form of  $\mathbb{R}^{n^2}$  since the determinant is defined by a polynomial. Polynomials are clearly continuous thus we are done. □

#### Proposition 1.4.8

Let  $n \in \mathbb{N} \setminus \{0\}$ . Then

$$GL(n, \mathbb{R}) \subset L(\mathbb{R}^n)$$

is open.

*Proof.* Consider the function  $\det(\cdot) : \mathbb{R}^{n \times n} \rightarrow \mathbb{R}$ . We have that  $\det(GL(n, \mathbb{R})) = \mathbb{R} \setminus \{0\}$  and image is clearly open in  $\mathbb{R}$ . Thus by continuity with open sets,  $GL(n, \mathbb{R})$  is open. □

#### Proposition 1.4.9

Let  $A \in GL(n, \mathbb{R})$ . If  $B \in M_{n \times n}(\mathbb{R})$  and  $\|B - A\| < \frac{1}{\|A^{-1}\|}$ , then  $B$  is invertible. This means that

$$B_{1/\|A^{-1}\|}(A) \subset GL(n, \mathbb{R})$$

is open in  $M_{n \times n}(\mathbb{R})$ . Furthermore, we must have

$$\|B^{-1}\| \leq \frac{1}{\frac{1}{\|A^{-1}\|} - \|B - A\|}$$

#### Proposition 1.4.10

Let  $f : \mathbb{R} \rightarrow M_{k \times n}(\mathbb{R})$  be a function. Then  $f$  is continuous if and only if every item in the matrix is a continuous function in the sense of a real function.

*Proof.* This is clear since  $M_{k \times n}(\mathbb{R}) \cong \mathbb{R}^{kn}$  and we have proved this in the above section. □

## 2 Differentiation

### 2.1 Frechet Derivative

#### Definition 2.1.1: Frechet Derivatives

Let  $f : U \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^m$  with  $U$  open and  $x \in U$ . We say that  $f$  is differentiable at  $x \in U$  if there exists a linear map  $T \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$  such that

$$\lim_{h \rightarrow 0} \frac{|f(x+h) - f(x) - T(h)|}{|h|} = 0$$

If  $T$  has a matrix representation  $A$  we write  $Df(x) = A$  as the derivative of  $f$  at  $x$ .

It is important that this definition uses a linear map as an approximation rather than a matrix so that in higher order derivatives, we can quit using matrices for the approximation and instead use bilinear forms and more.

#### Definition 2.1.2: Derivative Operator

Let  $f : U \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^m$  be differentiable for all  $x \in U$ . We define the operator  $D$  on  $f$  to be the function that takes  $x$  to  $Df(x)$ . This means that  $Df : U \subseteq \mathbb{R}^n \rightarrow M_{m \times n}(\mathbb{R})$ .

Note the notational differences.  $Df$  is a function that takes in  $x$  and spits out  $Df(x)$ .  $Df(x)$  itself is a linear map  $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ . It can also be a function  $Df(x) : \mathbb{R}^n \rightarrow \mathbb{R}^m$  that takes a direction  $v \in \mathbb{R}^n$  and spits out the directional derivative  $Df(x)v \in \mathbb{R}^m$ .

We then have the extension of theorems as in the one-dimensional case.

#### Theorem 2.1.3

The linear transformation representing the derivative is unique if it exists.

*Proof.* Suppose that  $A, B$  both represent the derivative of  $f : U \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^m$ . Then fix  $\epsilon > 0$ . There exists  $\delta_1, \delta_2$  such that  $|h| < \min\{\delta_1, \delta_2\}$  implies

$$\frac{|f(x+h) - f(x) - Ah|}{|h|} < \epsilon$$

and

$$\frac{|f(x+h) - f(x) - Bh|}{|h|} < \epsilon$$

Then we have

$$\begin{aligned} |(A - B)h| &= |f(x+h) - f(x) - Bh - f(x+h) - f(x) - Ah| \\ &\leq |f(x+h) - f(x) - Bh| + |f(x+h) - f(x) - Ah| \\ &\leq \epsilon|h| + \epsilon|h| \\ &= 2\epsilon|h| \end{aligned}$$

Thus  $A = B$  by definition of limit and we are done.  $\square$

#### Proposition 2.1.4

If  $f : U \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^m$  is differentiable at  $x \in U$ , then  $f$  is continuous at  $x$ .

*Proof.* Since  $f$  is continuous at  $x$ , there exists  $A \in \mathbb{R}^{m \times n}$  such that

$$\lim_{h \rightarrow 0} \frac{|f(x+h) - f(x) - Ah|}{|h|} = 0$$

Fix  $\epsilon > 0$ , there exists  $\delta_1 > 0$  such that  $|h| < \delta_1$

$$\begin{aligned} |f(x+h) - f(x) - Ah| &\leq \epsilon|h| \\ |f(x+h) - f(x)| &\leq |Ah| + \epsilon|h| \\ &< (\|A\| + \epsilon)|h| \end{aligned}$$

Set  $\delta_2 = \min \left\{ \delta_1, \frac{\epsilon}{\|A\| + \epsilon} \right\}$ . Then  $|h| < \delta_2$  implies

$$|f(x+h) - f(x)| < (\|A\| + \epsilon)\delta_2 < \epsilon$$

and we are done.  $\square$

Similar to continuity, differentiability in higher dimensions can be broken down by its individual components. This way we only need to show individual differentiability to save trouble.

#### Proposition 2.1.5: Componentwise Differentiability

Let  $f : U \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^k$ , where

$$f(x) = \begin{pmatrix} f_1(x) \\ \vdots \\ f_k(x) \end{pmatrix}$$

Then  $f$  is differentiable at  $x \in U$  if and only if for each  $i \in \{1, \dots, k\}$ ,  $f_i : U \rightarrow \mathbb{R}$  is differentiable at  $x$ .

The rest of this section are also extensions of one-dimensional case.

#### Proposition 2.1.6

If  $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$  is a constant function then  $Df(x) = 0$ .

*Proof.* Suppose that  $f(x) = k$  for some  $k \in \mathbb{R}^m$ . Then I claim that  $Df(x) = 0$ . Indeed since

$$\lim_{h \rightarrow 0} \frac{|0 + 0 - 0h|}{|h|} = 0$$

thus the definition of differentiability is satisfied.  $\square$

#### Proposition 2.1.7

If  $A \in \mathbb{R}^{m \times n}$  and  $f : U \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^m$  is defined by  $f(x) = Ax$ , then  $Df(x) = A$ .

*Proof.* Clearly  $Df(x) = A$  satisfies the definition of differentiability since

$$\lim_{h \rightarrow 0} \frac{|A(x+h) - Ax - Ah|}{|h|} = 0$$

$\square$

## 2.2 Jacobian Matrix and Directional Derivatives

### Definition 2.2.1: Directional Derivative

Let  $v \in \mathbb{R}^n$ . Let  $f : U \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^k$  be a function. Define the directional derivative along  $v$  passing through  $x$  to be

$$\partial_v f(x) = \lim_{t \rightarrow 0} \frac{f(x + tv) - f(x)}{t} = \left. \frac{d}{dt} f(x + tv) \right|_{t=0}$$

if the limit exists.

### Proposition 2.2.2: I

$f : U \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^k$  has a directional derivative along  $v \in \mathbb{R}^n$ , then  $f$  is linearly continuous along  $v$ .

### Proposition 2.2.3

Let  $f : U \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^k$  be a function. If  $Df$  exists then  $\partial_v f(x) = Df(x)v$ . Moreover,  $\partial_v f(x)$  is linear. Meaning

$$\partial_{av+bw} f(x) = a\partial_v f(x) + b\partial_w f(x)$$

for all  $a, b \in \mathbb{R}$  and for all  $v, w \in \mathbb{R}^n$ .

*Proof.* Suppose that  $Df$  exists. Then using the definition of differentiability, we have that

$$\begin{aligned} \lim_{t \rightarrow 0} \frac{f(x + tv) - f(x) - Df(x)(tv)}{t|v|} &= 0 \\ \lim_{t \rightarrow 0} \frac{f(x + tv) - f(x) - tDf(x)v}{t} &= 0 \\ \lim_{t \rightarrow 0} \frac{f(x + tv) - f(x)}{t} &= \lim_{t \rightarrow 0} \frac{tDf(x)v}{t} \\ \lim_{t \rightarrow 0} \frac{f(x + tv) - f(x)}{t} &= Df(x)v \\ \partial_v f(x) &= Df(x)v \end{aligned}$$

Linearity follows since  $Df(x)$  is matrix and matrix multiplication on a vector is linear.  $\square$

We should make absolutely clear distinction with the existence of frechet derivatives and the existence of directional derivatives. Even if the directional derivatives exists, as long as they are not linear, the matrix for the frechet derivative will not exist. In particular, it is possible to write out the Jacobian matrix as we see below, but this Jacobian matrix will not be equal to the matrix in the frechet derivative.

The above theorem is also useful for detecting non-differentiability. By verifying the directional derivatives are non-linear, one can use the contrapositive to prove that  $f$  would not be differentiable.

### Definition 2.2.4: Partial Derivatives

Let  $f : U \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^k$ . Define

$$\partial_j f(x) = \lim_{t \rightarrow 0} \frac{f(x + te_j) - f(x)}{t}$$

which is the directional derivatives with standard basis as the direction. In particular,

$$\partial_j f_i(x) = \lim_{t \rightarrow 0} \frac{f_i(x + t\mathbf{e}_j) - f_i(x)}{t}$$

is just the one dimensional differentiation in real analysis and

$$\partial_j f(x) = \begin{pmatrix} \partial_j f_1(x) \\ \vdots \\ \partial_j f_n(x) \end{pmatrix}$$

The partial derivatives are simply the special case of directional derivatives, when taken their direction to be unit vectors. Therefore we also have the following lemma.

**Proposition 2.2.5: I**

the partial derivatives of  $f : U \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^k$  exists, then  $f$  is separately continuous along  $v$ .

Now we have the six notions, namely continuity, linear continuity, separate continuity, differentiable, directional derivatives and partial derivatives interconnected with each other. Namely each type of differentiability implies its own continuity.

**Definition 2.2.6: Jacobian Matrix**

Let  $f : U \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^k$  be a function and let  $x \in U$ . Define the Jacobian matrix at  $x$  to be

$$\partial f(x) = \begin{pmatrix} \partial_1 f_1(x) & \cdots & \partial_n f_1(x) \\ \vdots & & \vdots \\ \partial_1 f_k(x) & \cdots & \partial_n f_k(x) \end{pmatrix}$$

Clearly the existence of the Jacobian matrix simply relies on the existence of partial derivatives. Note that in order for  $Df$  to exist, we require the directional derivatives to be linear.

**Theorem 2.2.7**

If  $f : U \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^k$  is differentiable at  $x \in U$  and  $v \in \mathbb{R}^n$ , then

$$Df(x)v = \partial f(x)v$$

*Proof.* Note that we have

$$\begin{aligned} Df(x)v &= Df(x) \sum_{k=1}^n v_k e_k \\ &= \sum_{k=1}^n v_k Df(x)e_k \\ &= \sum_{k=1}^n v_k \partial_k f(x) \\ &= \partial f(x)v \end{aligned}$$

□

The above theorem shows us that as long as  $f$  is differentiable, the Jacobian and the matrix given in the frechet derivative will be equal and interchangeable.

**Lemma 2.2.8**

If  $f : U \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^m$  is differentiable at  $x \in U$  and  $v \in \mathbb{R}^n$ , then

$$Df(x)v = \partial f(x)v = \partial_v f(x)$$

*Proof.* This immediately follows from previous theorems.  $\square$

This theorem also shows that if  $f$  is differentiable, then  $f$  must have directional derivatives.

**2.3 Properties of the Derivative****Theorem 2.3.1: Algebra of Differentiable functions**

If  $f, g : U \rightarrow \mathbb{R}^m$  are differentiable functions at  $x$  and  $\lambda, \mu \in \mathbb{R}$  are constants, then

$$D(\lambda f + \mu g)(x) = \lambda Df(x) + \mu Dg(x)$$

**Lemma 2.3.2**

Let  $f : U \subset \mathbb{R}^n \rightarrow \mathbb{R}^k$ ,  $x \in U$  and  $r > 0$  such that  $B_r(x) \subset U$  and  $A \in L(\mathbb{R}^n, \mathbb{R}^k)$ . Define  $\Delta_{x,A}f : B_r(0) \rightarrow \mathbb{R}^k$  by

$$\Delta_{x,A}f(h) = \begin{cases} \frac{f(x+h) - f(x) - Ah}{|h|} & \text{if } h \neq 0 \\ 0 & \text{if } h = 0 \end{cases}$$

Then  $f$  is differentiable at  $x$  with  $Df(x) = A$  if and only if  $\Delta_{x,A}$  is continuous at 0.

*Proof.* Let  $\Delta_{x,A}$  be continuous at 0. Then

$$\lim_{h \rightarrow 0} |\Delta_{x,A}f(h)| = |\Delta_{x,A}f(0)| = 0$$

Thus by definition of differentiability  $f$  is differentiable at  $x$  with  $Df(x) = A$ .

Now let  $f$  be differentiable at  $x$  and set  $A = Df(x)$ . Then by definition of differentiability we have that

$$\lim_{h \rightarrow 0} \left| \frac{f(x+h) - f(x) - Ah}{|h|} \right| = \lim_{h \rightarrow 0} |\Delta_{x,A}f(h)| = 0 = \Delta_{x,A}f(0)$$

Thus  $\Delta_{x,A}f(x)$  is continuous at 0.  $\square$

**Lemma 2.3.3**

Let  $\tau > 0$ . Let  $\xi : B_\tau(0) \rightarrow \mathbb{R}$  be bounded and  $\nu : B_\tau(0) \rightarrow \mathbb{R}^k$  be continuous at  $0 \in B_\tau(0)$  and  $\nu(0) = 0$ . Then

$$\delta(h) = \xi(h)\nu(h)$$

where  $0 < |h| < \tau$  and  $\delta(0) = 0$  is continuous at  $0 \in B_\tau(0)$ .

*Proof.* By continuity of  $\nu$  at 0, let  $\epsilon > 0$ . Then there exists  $\delta \in (0, \tau)$  such that  $|h| < \tau$  implies  $|\nu(h)| < \epsilon$ . By boundedness of  $\xi$ , there exists  $M > 0$  such that  $|\xi| < M$  for all  $h \in B_\tau(0) \setminus \{0\}$ . Thus  $0 < |h| < \delta$  implies  $|\delta(h)| < M\epsilon$ , meaning  $\lim_{h \rightarrow 0} \delta(h) = 0 = \delta(0)$  thus we are done.  $\square$

**Proposition 2.3.4: Chain Rule**

Let  $f : U \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$  and  $g : V \subset \mathbb{R}^m \rightarrow \mathbb{R}^k$  be two differentiable functions. Let  $x \in U$ . Then  $g \circ f$  is differentiable and

$$D(g \circ f)(x) = (Dg)(f(x)) \cdot Df(x)$$

*Proof.* Let

$$\Delta_x f(h) = \begin{cases} \frac{f(x+h) - f(x) - Df(x)h}{|h|} & \text{if } h \neq 0 \\ 0 & \text{if } h = 0 \end{cases}$$

and

$$\Delta_{f(x)} g(k) = \begin{cases} \frac{g(f(x)+k) - g(f(x)) - Dg(f(x))k}{|k|} & \text{if } k \neq 0 \\ 0 & \text{if } k = 0 \end{cases}$$

where both functions are continuous by lemma 2.3.2. Then we have that

$$f(x+h) = f(x) + Df(x)h + \Delta_x f(h)|h|$$

and

$$g(f(x)+k) = g(f(x)) + Dg(f(x))k + \Delta_{f(x)} g(k)|k|$$

Let  $k(h) = Df(x)h + \Delta_x f(h)|h|$ . Then by linearity of  $Dg(f(x))$ ,

$$g(f(x+h)) = g(f(x)) + Dg(f(x))Df(x)h + Dg(f(x))\Delta_x f(h)|h| + \Delta_{f(x)} g(k(h))|k(h)|$$

Let

$$\delta_1(h) = Dg(f(x))(\Delta_x f(h))$$

and

$$\delta_2(h) = \begin{cases} \frac{|k(h)|}{|h|} \Delta_{f(x)} g(k(h)) & \text{if } h \neq 0 \\ 0 & \text{if } h = 0 \end{cases}$$

Then we have

$$g(f(x+h)) - g(f(x)) - Dg(f(x)) \circ Df(x)h = |h|(\delta_1(h) + \delta_2(h))$$

We now show that  $\lim_{h \rightarrow 0} |\delta_1(h)| = 0$  and  $\lim_{h \rightarrow 0} |\delta_2(h)| = 0$ . Note that

$$|\delta_1(h)| \leq \|Dg(f(x))\| |\Delta_x f(h)|$$

Since  $\lim_{h \rightarrow 0} |\Delta_x f(h)| = 0$  by construction, we are done. Now let

$$\xi(h) = \frac{|k(h)|}{|h|} \leq \frac{|Df(x)h|}{|h|} + |\Delta_x f(h)| \leq \|Df(x)\| + |\Delta_x f(h)|$$

Since  $\Delta_x f$  is continuous,  $\xi$  is bounded on  $B_r(0) \setminus \{0\}$  for some  $r > 0$ . Now set  $\nu(h) = \Delta_{f(x)} g(k(h))$ . Since  $k$  is continuous at  $k(0) = 0$  and  $g$  is differentiable at  $f(x)$ , we have that  $\nu(h)$  is continuous and  $\nu(0) = 0$ . Thus we can apply lemma 2.3.3 and

$$\lim_{h \rightarrow 0} |\delta_2(h)| = 0$$

Thus we must have

$$g(f(x+h)) - g(f(x)) - Dg(f(x)) \circ Df(x)h = |h|(\delta_1(h) + \delta_2(h)) \rightarrow 0$$

and we are done.  $\square$



## 2.4 Mean Value Inequality

### Theorem 2.4.1: Mean Value Theorem

Let  $x, y \in \mathbb{R}^n$ . Let  $r : [a, b] \rightarrow \mathbb{R}^n$  be continuously differentiable. Let  $f : C^1(r([a, b]), \mathbb{R}^k)$  and there exists some  $M$  such that  $|\partial f(x)| \leq M$  for all  $x \in U$ . Then

$$|f(r(b)) - f(r(a))| \leq M \int_a^b |r'(t)| dt$$

*Proof.* We have that

$$\begin{aligned} |f(r(b)) - f(r(a))| &= \left| \int_a^b \frac{d}{dt} f(r(t)) dt \right| && \text{(By the FTC)} \\ &= \left| \int_a^b \partial f(r(t)) r'(t) dt \right| \\ &\leq \int_a^b |\partial f(r(t)) r'(t)| dt \\ &\leq \int_a^b |\partial f(r(t))| |r'(t)| dt \\ &\leq \int_a^b M |r'(t)| dt \end{aligned}$$

Thus we are done.  $\square$

Notice that  $\int_a^b |r'(t)| dt$  is exactly the length of the path  $r$  from  $a$  to  $b$ .

### Proposition 2.4.2

Suppose that  $U \subseteq \mathbb{R}^n$  is path-connected and every path is differentiable. Let  $f : U \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^m$  be differentiable and that  $\partial f(x) = 0$  for all  $x \in U$ . Then  $f$  is constant on  $U$ .

*Proof.* Let  $x, y \in U$ , then any path  $r : [a, b] \rightarrow U$  with  $r(a) = x$  and  $r(b) = y$  is differentiable. Then

$$\begin{aligned} f(y) - f(x) &= f(r(b)) - f(r(a)) \\ &= \int_a^b \frac{d}{dt} f(r(t)) dt && \text{(By the FTC)} \\ &= \int_a^b \partial f(r(t)) r'(t) dt \\ &= 0 \end{aligned}$$

Thus  $f(x) = f(y)$  for any  $x, y \in U$ .  $\square$

## 2.5 Conditions for Differentiability

### Theorem 2.5.1

Let  $f : U \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^k$  and suppose that  $B_r(x_0) \subset U$  for some  $r > 0$ . Suppose that the Jacobian  $\partial f(x)$  exists and is continuous for all  $x \in B_r(x_0)$ . Then  $f$  is differentiable.

*Proof.* We show this for the case that  $k = 1$  since  $f : \mathbb{R}^n \rightarrow \mathbb{R}^k$  is differentiable if and only if each component  $f_1, \dots, f_k$  is differentiable. Thus now  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ . Suppose that  $\partial f$  is continuous in  $B_r(x)$  for fixed  $x$ . This means that  $\partial_{x_j} f_i$  is continuous for any  $j \in \{1, \dots, n\}$

and  $i \in \{1, \dots, k\}$ . Take a point  $x + h \in B_r(x)$ . Define  $h = \begin{pmatrix} h_1 \\ \vdots \\ h_n \end{pmatrix}$  and

$$h'_k = \begin{pmatrix} h_1 \\ \vdots \\ h_k \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

Trivially define  $h'_0 = 0$  and  $h_0 = 0$  for convenience. Now  $f(x + h'_k) - f(x + h'_{k-1})$  is just a one variable function on the  $k$ th slot, thus we can apply the mean value theorem in real analysis and conclude that there exists  $\theta_k \in (0, 1)$  such that

$$f(x + h'_k) - f(x + h'_{k-1}) = \partial_{x_k} f(x + h'_{k-1}) + (\theta_k h_k) e_k h_k.$$

Summing all these up, we have that

$$\begin{aligned} \sum_{k=1}^n f(x + h'_k) - f(x + h'_{k-1}) &= \sum_{k=1}^n \partial_{x_k} f(x + h'_{k-1}) + (\theta_k h_k) e_k h_k \\ f(x + h) - f(x) &= \sum_{k=1}^n \partial_{x_k} f(x + h'_{k-1}) + (\theta_k h_k) e_k h_k \end{aligned}$$

We now use the continuity of the partial derivatives. Given  $\epsilon > 0$ , there  $\delta_k > 0$  such that  $|h| < \delta_k$  implies  $|\partial_{x_k} f(x + h) - \partial_{x_k} f(x)| < \epsilon$ . Choose  $\delta = \min\{\delta_1, \dots, \delta_n\}$ . Since  $\theta_k \in (0, 1)$  for all  $k \in \{1, \dots, n\}$ , we must have  $|h'_{k-1} + (\theta_k h_k) e_k| < |h| < \delta$ . Thus

$$\begin{aligned} \left| f(x + h) - f(x) - \sum_{k=1}^n \partial_{x_k} f(x) h_k \right| &= \left| \sum_{k=1}^n \partial_{x_k} f(x + h'_{k-1}) + (\theta_k h_k) e_k h_k - \sum_{k=1}^n \partial_{x_k} f(x) h_k \right| \\ &< \epsilon \sum_{k=1}^n |h_k| \\ &< \epsilon \sqrt{n} |h| \end{aligned}$$

if  $|h| < \delta$ . The second inequality is due to the fact that  $\sum_{k=1}^n h_k \leq \sqrt{n} |h|$ .

Now define  $A \in \mathcal{L}(\mathbb{R}^n, \mathbb{R})$  by  $A(h) = (\partial f)h$ . This makes sense in matrix multiplication since  $\partial f \in M_{1 \times n}(\mathbb{R})$ . Then  $0 < |h| < \delta$  implies

$$\begin{aligned} \frac{|f(x + h) - f(x) - Ah|}{|h|} &\leq \epsilon \sqrt{n} \frac{|h|}{|h|} \\ &= \epsilon \sqrt{n} \end{aligned}$$

since  $A(h) = \sum_{k=1}^n \partial_{x_k} f(x) h_k$ . Since this is true for all  $\epsilon$ , the condition for differentiability is satisfied and we are done.  $\square$

Notice that in the above proof, in order to show that  $f$  is differentiable at  $x$ , we need to know the partial derivatives of its neighbours.

The below theorem acts as a summary or recollection of the relationships between differentiability and

continuous Jacobian Matrices.

**Theorem 2.5.2**

Let  $U$  be an open subset of  $\mathbb{R}^n$ . Then  $f : U \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^k$  is continuously differentiable on  $U$  if and only if  $\partial f : U \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^{k \times n}$  is continuous on  $U$ .

*Proof.* Suppose that  $f$  is differentiable at  $x$  and its derivative  $Df(x)$  is continuous. Then  $Df(x) = \partial f(x)$  thus  $\partial f(x)$  is also continuous.

Now suppose that  $\partial f(x)$  is continuous in  $U$ . Then by the above theorem  $f$  is differentiable and thus we have  $Df(x) = \partial f(x)$ . Hence  $Df(x)$  is also continuous.  $\square$

### 3 More Properties of Differentiability

#### 3.1 Inverse Function Theorem

The inverse function theorem is a powerful multidimensional analog of finding the derivative of an inverse function. The function has a few conditions to satisfy in order for it to be applied. Since the proof is exceptionally long, I will split it to a number of theorems and prepositions.

##### Proposition 3.1.1

Suppose that  $\Psi : U \rightarrow V$  is a bijection which is differentiable at  $x \in U$ . Suppose that  $\Psi^{-1}$  is differentiable at  $y = \Psi(x) \in V$ . Then  $D\Psi(x)$  and  $D\Psi^{-1}(y)$  are both invertible and

$$(D\Psi^{-1})(y) = (D\Psi(\Psi^{-1}(y)))^{-1}$$

*Proof.* Differentiating the relation  $\Psi(\Psi^{-1}(y)) = y$  gives

$$D\Psi(\Psi^{-1}(y)) \circ D\Psi^{-1}(y) = I_n$$

which is the identity transformation thus we are done.  $\square$

##### Lemma 3.1.2

$T \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^k)$  is injective if and only if there exists  $a > 0$  such that  $|T(x)| \geq a|x|$  for all  $x \in \mathbb{R}^n$ .

*Proof.* Firstly suppose that there exists  $a > 0$  such that  $|T(x)| \geq a|x|$  for all  $x \in \mathbb{R}^n$ . Let  $T(x) = T(y)$ . Then

$$\begin{aligned} |T(x - y)| &\geq a|x - y| \\ |T(x) - T(y)| &\geq a|x - y| \\ 0 &\geq a|x - y| \end{aligned}$$

Since  $a > 0$  we must have  $x = y$ .

Now suppose that for all  $a > 0$ ,  $|T(x)| < a|x|$  for all  $x \in \mathbb{R}^n$ . This means that  $\frac{|T(x)|}{|x|} < a$  for all  $a > 0$ . This is precisely the definition of a limit. We can find a sequence  $(x_j)_{j \in \mathbb{N}}$  such that

$$\frac{|T(x_j)|}{|x_j|} \rightarrow 0$$

Now consider a new sequence defined by  $y_j = \frac{x_j}{|x_j|}$ . Clearly  $(y_j)_{j \in \mathbb{N}} \subset S^{n-1}$ . But  $S^{n-1}$  is compact by the Heine-Borel theorem. Thus there exists  $(y_{j_m})_{m \in \mathbb{N}}$  such that it has its limit in  $S^{n-1}$ . But then

$$T(y_j) = \frac{T(x_j)}{|x_j|}$$

thus  $|T(y_j)| \rightarrow 0$ . All the  $y_j$  are clearly nonzero, this means that  $T$  has a non-trivial kernel and thus  $T$  is not injective.  $\square$

We split the inverse function theorem into its injective and surjective part for better readability.

##### Proposition 3.1.3: Injective Part of Inverse Function Theorem

Let  $U$  be an open subset of  $\mathbb{R}^n$  and suppose that  $\Psi : \mathbb{R}^n \rightarrow \mathbb{R}^m$  such that it is differentiable and its derivative is continuous in  $U$ . Assume that  $D\Psi(p)$  is injective at a point  $p \in U$ . Then

there exists  $\delta > 0$  such that  $B_\delta(p) \subset U$  and such that  $f$  is injective on  $B_\delta(p)$ .

*Proof.* From the above lemma,  $D\Psi(p)$  being injective means that there exists  $a > 0$  such that

$$|Df(p)h| \geq \epsilon|h|$$

for all  $h \in \mathbb{R}^n$ . Since  $Df : U \rightarrow \mathcal{L}(\mathbb{R}^n, \mathbb{R}^k)$  is continuous, there exists  $\delta > 0$  such that  $x \in B_\delta(p) \subset U$  implies

$$\|Df(p) - Df(x)\| < \frac{1}{2}\epsilon$$

Define a new function  $F : U \rightarrow \mathbb{R}^k$  by  $F(x) = f(x) - Df(p)x$ . Then  $F$  is differentiable since  $Df(p)x$  is just a linear transformation and we have

$$DF(x) = Df(x) - Df(p)$$

Thus we now have

$$\|DF(x)\| = \|Df(x) - Df(p)\| < \frac{1}{2}\epsilon$$

We can now apply the mean value inequality to get

$$|F(z) - F(x)| \leq \frac{1}{2}\epsilon|z - x|$$

for all  $x, z \in B_\delta(p)$ . Finally we have

$$\begin{aligned} |f(x) - f(z)| &= |Df(p)(x - z) - (F(z) - F(x))| \\ &\geq \epsilon|x - z| - \frac{1}{2}\epsilon|x - z| \\ &= \frac{1}{2}\epsilon|x - z| \end{aligned}$$

This means that  $x \neq z$  implies  $f(x) \neq f(z)$  and we are done.  $\square$

### Proposition 3.1.4: Surjective Part of Inverse Function Theorem

Let  $U$  be an open subset of  $\mathbb{R}^n$  and suppose that  $\Psi : \mathbb{R}^n \rightarrow \mathbb{R}^n$  such that it is differentiable and its derivative is continuous in  $U$ . Assume that  $D\Psi(p)$  is surjective at a point  $p \in U$ . Then there exists  $\rho > 0$  such that  $B_\rho(\Psi(p)) \subset \Psi(U)$ .

*Proof.* By the rank nullity theorem,  $D\Psi(p)$  is injective (surjectivity implies bijectivity in linear maps). By the above proposition, there exists  $\epsilon > 0$  such that

$$|D\Psi(p)h| \geq \epsilon|h|$$

for all  $h \in \mathbb{R}^n$ . Again define  $F : U \rightarrow \mathbb{R}^n$  by  $F(x) = \Psi(x) - Df(p)x$ . Then exactly the same as the above proof, we must have  $|F(x) - F(z)| \leq \frac{1}{2}\epsilon|x - z|$  and  $|\Psi(x) - \Psi(z)| \geq \frac{1}{2}\epsilon|x - z|$ . Set

$$K = \overline{B_{\frac{1}{2}\delta}(p)} = \{x \in \mathbb{R}^n \mid |x - p| \leq \frac{1}{2}\delta\}$$

and  $\partial K = \{x \in \mathbb{R}^n \mid |x - p| = \frac{1}{2}\delta\}$ . Now we have

$$\begin{aligned} |\Psi(x) - \Psi(p)| &\geq \frac{1}{2}\epsilon|x - p| \\ &= \frac{1}{4}\epsilon\delta \end{aligned}$$

For all  $x \in \partial K$ . Set  $\rho = \frac{1}{8}\epsilon\delta$  and fix  $y \in B_\rho(\Psi(p))$ . We will show that

$$y \in \Psi(B_{\frac{1}{2}\delta}(p))$$

to finish the proof. Define a new function  $\phi : K \rightarrow \mathbb{R}$  by  $\phi(x) = |\Psi(x) - y|$ . Then  $\phi$  is continuous by composition of continuous functions. By the extreme value theorem, there exists  $c \in K$  such that  $\phi(c) \leq \phi(x)$  for all  $x \in K$ .

Now I will show that  $c \in B_{\frac{1}{2}\delta}(p)$  by showing that  $c \notin \partial K$  and using the fact that  $B_{\frac{1}{2}\delta}(p) = K \setminus \partial K$ .  $c \notin \partial K$  can be proved by  $\phi(x) > \phi(p)$  for all  $x \in \partial K$ . Now if  $x \in \partial K$ , we have

$$\begin{aligned}\phi(x) &= |\Psi(x) - y| \\ &\geq |\Psi(x) - \Psi(p)| - |y - \Psi(p)| \\ &\geq \frac{1}{4}\epsilon\delta - \frac{1}{8}\epsilon\delta \\ &= \rho \\ &> |y - \Psi(p)| \\ &= \phi(p)\end{aligned}$$

Thus we are done with this part.

Now since  $D\Psi(p)$  is surjective, there exists  $h \in \mathbb{R}^n$  such that  $D\Psi(p)h = y - \Psi(p)$  (by nuisances). Since  $c \in B_{\frac{1}{2}\delta}(p)$ , there exists  $\nu > 0$  such that  $|t| < \nu$  implies  $|c + th - p| < \frac{1}{2}\delta$ . Thus we have

$$\begin{aligned}\Psi(c + th) - y &= \Psi(c + th) - \Psi(c) - tD\Psi(p)h + (\Psi(c) - y + tD\Psi(p)h) \\ &= F(c + th) - F(c) + (1 - t)(\Psi(c) - y)\end{aligned}$$

for  $|t| < \nu$ . Now, we have that

$$\begin{aligned}\phi(c) &\leq \phi(c + th) \\ &= |\Psi(c + th) - y| \\ &\leq \frac{1}{2}|tD\Psi(p)h| + (1 - t)|\Psi(c) - y| \\ &= \left(1 - \frac{1}{2}t\right)(\Psi(c) - y)\end{aligned}$$

Since this holds for all  $|t| < \nu$ , we must have that  $\phi(c) = 0$ , which implies that  $y = \Psi(c)$ . Thus we have shown that

$$B_\rho(\Psi(p)) \subset \Psi(B_{\frac{1}{2}\delta}(p)) \subset \Psi(U)$$

□

### Corollary 3.1.5

Let  $U \subset \mathbb{R}^n$  be open. Let  $\Psi \in \mathcal{C}^1(U, \mathbb{R}^n)$  and suppose that  $D\Psi(p)$  is invertible for all points  $p \in U$ , then  $\Psi$  maps open subsets of  $U$  to open subsets of  $\mathbb{R}^n$ .

*Proof.* Let  $V \subset U$  be open. By the above proposition applied to  $\Psi|_V$ , we have for all  $p \in V$ , there exists  $\rho > 0$  such that

$$B_\rho(\Psi(p)) \subset \Psi(V)$$

, thus  $\Psi(V)$  is open.

□

### Theorem 3.1.6: Inverse Function Theorem

Let  $U$  be an open subset of  $\mathbb{R}^n$  and suppose that  $\Psi \in \mathcal{C}^1(U, \mathbb{R}^n)$ . Assume that  $D\Psi(p)$  is invertible at a point  $p \in U$ . Let  $q = \Psi(p)$ . Then there exists a neighbourhood  $N_p$  and  $N_q$  such

that  $\Psi : N_p \rightarrow N_q$  is a bijection and  $\Psi^{-1} : N_q \rightarrow N_p$  is continuously differentiable and

$$(D\Psi^{-1})(y) = (D\Psi(\Psi^{-1}(y)))^{-1}$$

for all  $y \in N_q$ .

*Proof.* Clearly  $\Psi$  satisfies the above two propositions. Thus this means that there exists  $\epsilon > 0$  such that  $|Df(p)h| \geq \epsilon|h|$  for all  $h \in \mathbb{R}^n$  and that there exists  $\delta > 0$  such that  $x \in B_\delta(p) \subset U$  implies

$$\|Df(p) - Df(x)\| < \frac{1}{2}\epsilon$$

Using these two facts and the rank nullity theorem we deduce that  $D\Psi(x)$  is invertible for all  $x \in B_\delta(p)$ . The above corollary implies that  $\Psi(B_\delta(p))$  is open. Thus  $\Psi|_{B_\delta(p)} : B_\delta(p) \rightarrow \Psi(B_\delta(p))$  is a bijection.

Now we show that  $\Psi^{-1}$  is continuously differentiable. □

Note that there could be multiple  $p$  such that  $\Psi$  maps  $p$  to  $q$ . The inverse function theorem guarantees that as long as  $D\Psi(p)$  is invertible, then they will have bijective neighbourhoods. This is why we can have multiple branches for the inverse. A good example would be  $y = x^2$ . It has two inverses, namely  $\sqrt{x}$  and  $-\sqrt{x}$  precisely because of this reason. In this case the two neighbourhoods would be  $\mathbb{R}^+$  and  $\mathbb{R}^-$  respectively.

Furthermore, even if there are multiple  $p$  mapping to  $q$ , not every  $p$  could have a neighbourhood such that they are bijective because we must require the fact that  $D\Psi(p) \neq 0$ .

### 3.2 Implicit Function Theorem

#### Theorem 3.2.1: Implicit Function Theorem

Let  $U$  be an open subset of  $\mathbb{R}^{n+l}$  and  $c \in \mathbb{R}^l$ . Suppose that  $F \in \mathcal{C}^1(U, \mathbb{R}^l)$  and that the equation  $F(x, y) = c$  has a solution  $(x_0, y_0) \in U$  such that  $\det(\partial_y F(x_0, y_0)) \neq 0$ . Then there exists an open set  $x_0 \in N_{x_0} \subset \mathbb{R}^n$  and  $g \in \mathcal{C}^1(N_{x_0}, \mathbb{R}^l)$  such that

- $g(x_0) = y_0$ ,  $\{(x, g(x)) : x \in N_{x_0}\} \subset U$  and  $F(x, g(x)) = c$  for all  $x \in N_{x_0}$
- $\partial_y F(x, g(x))$  is invertible for all  $x \in N_{x_0}$  and

$$\partial g(x) = -(\partial_y F(x, g(x)))^{-1} \cdot \partial_x F(x, g(x))$$

for all  $x \in N_{x_0}$

Beware that more often we seen that the invertible matrix is not necessarily on the right hand side of  $\partial F$ . It could consist of multiple columns of  $\partial F$  simply because of the ordering of the variables which is completely by the writer's choice. Moreover, there may be more than one  $g$  to convert the implicit function into an explicit one if you consider different variables for the domain and the codomain.

### 3.3 Higher Order Derivatives

From here onwards, our function  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  will be a scalar function else the Hessian Matrix cannot be defined. Do note that general second order differential operators for  $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$  do exists. It is just that we will not discuss it here.

**Definition 3.3.1: Second Order Partial Derivatives**

Suppose that  $f : U \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$  is differentiable with partial derivative operator  $\partial_j f : U \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$  for  $1 \leq j \leq n$ . Define the second order partial derivative at  $x_0 \in U$  to be

$$\partial_{ij} f(x_0) = \frac{\partial}{\partial x_i} \partial_j f(x) \Big|_{x=x_0} = \frac{\partial^2}{\partial x_i \partial x_j} f(x) \Big|_{x=x_0}$$

for  $1 \leq i \leq n$  if it exists. This is done by treating  $\partial_j f$  as a function from  $\mathbb{R}^n$  to  $\mathbb{R}$  and taking the partial derivative of it.

**Definition 3.3.2: Hessian Matrix**

Suppose that all second order partial derivatives of  $f : U \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$  exists. Define the Hessian Matrix  $f$  to be

$$H_f(x) = \partial^2 f(x) = \begin{pmatrix} \partial_{11} f(x) & \cdots & \partial_{1n} f(x) \\ \vdots & \ddots & \vdots \\ \partial_{n1} f(x) & \cdots & \partial_{nn} f(x) \end{pmatrix}$$

With  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ , to say that  $Df$  is differentiable means that there exists a linear operator  $T$  such that

$$\lim_{h \rightarrow 0} \frac{|Df(x+h) - Df(x) - T(h)|}{|h|} = 0$$

as defined by the definition of differentiability. Notice that  $Df(x) \in \mathcal{L}(\mathbb{R}^n, \mathbb{R})$ . This means that  $T$  must take the form  $T : \mathbb{R}^n \rightarrow \mathcal{L}(\mathbb{R}^n, \mathbb{R})$  so that  $T(h) \in \mathcal{L}(\mathbb{R}^n, \mathbb{R})$  makes sense. Thus  $T \in \mathcal{L}(\mathbb{R}^n, \mathcal{L}(\mathbb{R}^n, \mathbb{R}))$ . This leads to the following identification:

**Proposition 3.3.3**

Let  $n \in \mathbb{N} \setminus \{0\}$ . Then

$$\mathcal{L}(\mathbb{R}^n, \mathcal{L}(\mathbb{R}^n, \mathbb{R})) \cong \mathcal{L}^2(\mathbb{R}^n, \mathbb{R})$$

where  $\mathcal{L}^2(\mathbb{R}^n, \mathbb{R})$  is the space of all bilinear forms  $\mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$ .

*Proof.* Let  $T \in \mathcal{L}(\mathbb{R}^n, \mathcal{L}(\mathbb{R}^n, \mathbb{R}))$ . Then for every  $x \in \mathbb{R}^n$ ,  $T(x) : \mathbb{R}^n \rightarrow \mathbb{R}$ . Thus  $T(x)(y) \in \mathbb{R}$  with  $x, y \in \mathbb{R}^n$ . Obviously we can write it into  $T(x)(y) = B(x, y)$ . Conversely any bilinear form can be written into the above form. In particular they both are vector spaces of dimension  $n^2$ .

From linear algebra we know that  $B(x, y) = x^T A y$  for some matrix  $A$ . If  $T(x) = Cx$  with  $C$  representing  $T$ , then

$$B(x, y) = (T(x)) \cdot y = (Cx)^T y = x^T C^T y$$

□

**Proposition 3.3.4**

Let  $f : U \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$  be differentiable for all  $x \in U$  and  $Df : U \subseteq \mathbb{R}^n \rightarrow \mathcal{L}(\mathbb{R}^n, \mathbb{R})$  be differentiable. Then  $D(Df)(x) = H_f(x)$ . Moreover,  $H_f(x)$  is symmetric and all second order partial derivatives commute.



**Lemma 3.3.5**

If all second order partial derivatives of  $f : U \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$  at  $x$  is continuous for  $x \in U$ , then

$$\frac{\partial^2}{\partial x_i \partial x_j} f(x) = \frac{\partial^2}{\partial x_j \partial x_i} f(x)$$

**Definition 3.3.6: Twice differentiable and Continuous derivatives**

Let  $f : U \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^m$  where  $f(x) = \begin{pmatrix} f_1(x) \\ \vdots \\ f_m(x) \end{pmatrix}$  with  $f_k(x) : \mathbb{R}^n \rightarrow \mathbb{R}$  for  $k \in \{1, \dots, m\}$ . Define

$$\mathcal{C}^2(U, \mathbb{R}^m) = \{f : U \rightarrow \mathbb{R}^m \mid \partial^2 f_k(x) : U \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^n \text{ is continuous for } k \in \{1, \dots, m\}\}$$

**Theorem 3.3.7**

[Second Order Taylor Expansion] Let  $U \subset \mathbb{R}^n$  be convex and  $x, x+h \in U$ . If  $f \in \mathcal{C}^2(U)$  then

$$f(x+h) = f(x) + \sum_{k=1}^n h_k \frac{\partial f(x)}{\partial x_k} + \frac{1}{2} \sum_{i,j=1}^n h_i h_j \frac{\partial^2 f(x)}{\partial x_i \partial x_j} + R(h)$$

where

$$\lim_{h \rightarrow 0} \frac{|R(h)|}{|h|^2} = 0$$

**3.4 Second Order Derivative Test****Definition 3.4.1: Critical Point**

We say that  $p \in U$  is a critical point of  $f \in \mathcal{C}^1(U)$  if  $\nabla f(p) = 0$ .

**Proposition 3.4.2**

If  $f$  has a local minimum or maximum at  $p$  then  $\nabla f(p) = 0$ .

**Theorem 3.4.3: Second Order Derivative Test**

Suppose that  $f : U \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$  such that  $H_f(x)$  exists and is continuous ( $f \in \mathcal{C}^2(U)$ ). Suppose that  $\nabla f(p) = 0$  for some  $p \in U$  (All first order derivatives are 0 at  $p$  or  $p$  is a critical point).

- If  $x^T H_f(p) x > 0$  for all  $x \in \mathbb{R}^n \setminus \{0\}$  then  $f$  has a strict local minimum at  $p$
- If  $x^T H_f(p) x < 0$  for all  $x \in \mathbb{R}^n \setminus \{0\}$  then  $f$  has a strict local maximum at  $p$
- If there exists  $x, y \in \mathbb{R}^n \setminus \{0\}$  such that  $x^T H_f(p) x > 0$  and  $y^T H_f(p) y < 0$  then  $p$  is a saddle point
- The test is inconclusive otherwise.

## 4 Vector Calculus

### 4.1 Regions

Before we define integration of functions, we need to make sense of the domain of our integral.

#### Definition 4.1.1: Regions

A region in  $\mathbb{R}^n$  is a bounded open subset  $\Omega$  of  $\mathbb{R}^n$  such that there exists a function  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  with the property that

- all partial derivatives of  $f$  are continuous
- $\Omega = \{x \in \mathbb{R}^n | f(x) < 0\}$
- $\nabla f(p) \neq 0$  for all  $p \in \mathbb{R}^n$  such that  $f(p) = 0$ .

Also define the boundary of a region,  $\partial\Omega$  to be  $\{x \in \mathbb{R}^n | f(x) = 0\}$ .

#### Proposition 4.1.2

Let  $\Omega \subset \mathbb{R}^n$  be a region with boundary  $f(x) = 0$ . Then there exists  $0 < l < n$  and  $r : \mathbb{R}^l \rightarrow \mathbb{R}^n$  such that  $f(r(x)) = 0$ . We say that  $r$  is the parametrization of  $\partial\Omega$ .

*Proof.* Since  $\nabla f(p) \neq 0$  for all  $p \in \mathbb{R}^n$ , we can apply the implicit function theorem to find function  $g : \mathbb{R}^l \rightarrow \mathbb{R}^{n-l}$  such that  $f(x, g(x)) = 0$  for any  $p \in f^{-1}(0)$ . Now define  $r : \mathbb{R}^l \rightarrow \mathbb{R}^n$  by  $r(x) = (x, g(x))$ . Then clearly  $f(r(x)) = 0$ .  $\square$

In particular, if  $l = 1$ , then we call  $r$  a curve. If  $l = 2$ , then  $r$  would be a surface. This is precisely the parametrization of a boundary, be it a curve or a surface.

#### Definition 4.1.3: Tangent Space

Let  $\Omega \subset \mathbb{R}^n$  be a region with boundary  $f(x) = 0$  and parametrization  $r : \mathbb{R}^l \rightarrow \mathbb{R}^n$ . Define the tangent space at  $p \in \partial\Omega$  to be

$$T_p(\partial\Omega) = \text{span}\{\partial_{x_1}r, \dots, \partial_{x_l}r\}$$

Tangent spaces will be given a more rigorous definition in differential geometry. But the notion does coincide.

#### Proposition 4.1.4

Let  $\Omega \subset \mathbb{R}^n$  be a region with boundary  $f(x) = 0$  and parametrization  $r : \mathbb{R}^l \rightarrow \mathbb{R}^n$ . Let  $p \in \partial\Omega$ . Then

$$T_p(\partial\Omega) = \ker(\partial f)$$

*Proof.* Notice that from the equation  $f(r(x)) = 0$ , we have that by the chain rule,

$$(\partial f)(\partial r) = 0$$

But  $\partial r$  is precisely  $(\partial_{x_1}r, \dots, \partial_{x_l}r)$ . Thus we are done.  $\square$

### 4.2 Differentiation and Integration of Curves

This section is dedicated to vector functions, which are functions with domain  $\mathbb{R}$  and codomain  $\mathbb{R}^n$ .

**Definition 4.2.1: Curves**

Let  $r : I \subset \mathbb{R} \rightarrow \mathbb{R}^n$ . Then

$$C = \{\mathbf{r}(t) | t \in I\}$$

is called a curve in  $\mathbb{R}^n$ .  $r$  is a parametrization of  $C$ , and is called a vector function.

**Definition 4.2.2: Classifying Curves**

Let  $r : I \subseteq \mathbb{R} \rightarrow \mathbb{R}^3$  induce a curve  $C$ .

- We say  $r \in \mathcal{C}^k$  if it can be differentiated  $k$  times. It is smooth if  $r \in \mathcal{C}^\infty$
- $C$  is simple if it does not intersect itself
- $C$  is closed if  $I = [a, b]$  and  $r(a) = \mathbf{r}(b)$
- $C$  is regular if  $r'(t) \neq 0$  for all  $t \in I$

From regions, curves arise. On the other hand, we can construct regions from sufficiently nice curves.

**Proposition 4.2.3**

A curve  $r : I \rightarrow \mathbb{R}^n$  gives rise to a region  $\Omega$  if  $r$  is in  $\mathcal{C}^1$ ,  $r$  is simple, closed and regular.

**Definition 4.2.4: Arc Length Function and Parametrization**

Let  $r : I \rightarrow \mathbb{R}^n$  be a vector function. Define the Arc Length Function by

$$s(t) = \int_{t_0}^t |\mathbf{r}'(u)| du$$

Define the Arc Length Parametrization by  $\mathbf{r}(s)$ .

**Lemma 4.2.5**

$\mathbf{r}(s)$  has unit speed.

*Proof.* We show that  $|\mathbf{r}'(s)| = 1$ . By the chain rule,

$$\begin{aligned} \left| \frac{d}{ds} \mathbf{r}(s(t)) \right| &= \left| \frac{d\mathbf{r}}{dt} \cdot \frac{dt}{ds} \right| \\ &= \frac{|\mathbf{r}'(t)|}{|\mathbf{r}'(t)|} && \text{(By the FTC)} \\ &= 1 \end{aligned}$$

□

**Definition 4.2.6: Line Integrals on Scalar Functions**

Let  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ . Let  $r : [a, b] \rightarrow \mathbb{R}^n$  a smooth curve with image  $C$ . Define the line integral along a function  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  to be

$$\int_C f dr = \int_a^b f(r(t)) |r'(t)| dt$$

If  $r$  is a closed curve then we write the integral as  $\oint_C f$ .

$dr$  in the integral is justified by noting that if  $r$  is a function of  $t$ , then  $dr = r'(t)dt$ . But since  $dr$  is a

positive line segment, we take the absolute value, giving us

$$dr = |dr| = |r'(t)||dt| = dt$$

There is no proper infinite sum that we can justify for  $\int_C f$  because it is simply a notation short hand for the integral on the right hand side. Notice that the parameterization  $r$  does not appear on this notation because there can be mutiple parametrizations with the same image.

### 4.3 Vector Fields

#### Definition 4.3.1: Vector Fields

We say that a function  $v : U \subset \mathbb{R}^n \rightarrow \mathbb{R}^n$  is a vector field. For every  $x \in \mathbb{R}^n$  it assigns a vector  $v(x) \in \mathbb{R}^n$  to it. If each  $v_k \in \mathcal{C}^1$  then we say that  $v \in \mathcal{C}^1$ .

#### Definition 4.3.2: Conservative Vector Fields

We say that a vector field  $v : U \subset \mathbb{R}^n \rightarrow \mathbb{R}^n$  is conservative if there exsits some scalar function  $\phi : \mathbb{R}^n \rightarrow \mathbb{R}$  such that

$$v = \nabla \phi$$

#### Definition 4.3.3: Divergence

Let  $f : U \subset \mathbb{R}^n \rightarrow \mathbb{R}^n$  be a vector field. Define the divergence of  $f$  to be

$$\nabla \cdot f = \left( \frac{\partial}{\partial x_1} \quad \cdots \quad \frac{\partial}{\partial x_n} \right) \begin{pmatrix} f_1 \\ \vdots \\ f_n \end{pmatrix} = \sum_{k=1}^n \frac{\partial f_k}{\partial x_k}$$

#### Definition 4.3.4: Line Integrals on Vector Fields

Let  $v : U \subset \mathbb{R}^n \rightarrow \mathbb{R}^n$  be a vector field. Let  $r : [a, b] \rightarrow \mathbb{R}^n$  be a curve with image  $C \in U$ . Define the integral of  $v$  along the curve  $C$  to be

$$\int_C v \cdot dr = \int_a^b v(r(t)) \cdot r'(t) dt$$

If  $r$  is a closed curve then we write the integral as  $\oint_C v \cdot dr$ .

#### Proposition 4.3.5

Let  $v : U \subset \mathbb{R}^n \rightarrow \mathbb{R}^n$  be a vector field. Let  $r : [a, b] \rightarrow \mathbb{R}^n$  be a curve with image  $C \in U$ . Then

$$\int_C v \cdot dr = \sum_{k=1}^n \int_C v_k dx_k$$

#### Proposition 4.3.6

If  $v : U \subset \mathbb{R}^n \rightarrow \mathbb{R}^n$  is conservative with  $v = \nabla f$  and  $r : [a, b] \rightarrow \mathbb{R}^n$  is a curve with image  $C$  then

$$\int_C v \cdot dr = f(r(b)) - f(r(a))$$

**Theorem 4.3.7: Equivalent Characterization of Conservative Vector Fields**

Let  $\gamma : [a, b] \rightarrow \mathbb{R}^n$  be a path. Let  $F : \mathbb{R}^n \rightarrow \mathbb{R}^n$  be a vector field. Then the following are equivalent.

- $F$  is conservative
- $\int_{\gamma} F(r) \cdot dr$  is independent of the choice of  $\gamma$
- $\int_{\gamma} F(r) \cdot dr = 0$  if  $\gamma$  is closed

## 5 2 Dimensional Calculus

### 5.1 Double Integrals

#### Definition 5.1.1: Positively Oriented Tangents and Outward Normals

Let  $\Omega$  be a region in  $\mathbb{R}^2$  with boundary  $f(x) = 0$  and parametrization  $r : \mathbb{R} \rightarrow \mathbb{R}^2$ . We say that  $r$  is positively oriented if as the domain of  $r$  increases, the tangent vector  $r'(t)$  moves anticlockwise. In this case, we say that  $\begin{pmatrix} -r_2(p) \\ r_1(p) \end{pmatrix}$  for  $p \in \partial\Omega$  is the outward normal vector which is perpendicular to the tangent space  $T_p(\partial\Omega)$ .

#### Definition 5.1.2: Double Integrals

Let  $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ . Define the double integral over  $f$  on a region  $\Omega$  to be

$$\iint_{\Omega} f(x, y) \, dx \, dy$$

#### Proposition 5.1.3

Let  $f, g : \mathbb{R}^2 \rightarrow \mathbb{R}$  be integrable on a region  $\Omega$  and  $a, b \in \mathbb{R}$ . Then the following are true for double integrals.

- $\iint_{\Omega} (af(x, y) + bg(x, y)) \, dA = a \iint_{\Omega} f(x, y) \, dA + b \iint_{\Omega} g(x, y) \, dA$
- If  $f(x, y) \geq g(x, y)$  for all  $x, y \in \Omega$ , then  $\iint_{\Omega} f(x, y) \, dA \geq \iint_{\Omega} g(x, y) \, dA$
- If  $\Omega = D_1 \cup D_2$  and  $D_1 \cap D_2 = \emptyset$ , then  $\iint_{\Omega} f(x, y) \, dA = \iint_{D_1} f(x, y) \, dA + \iint_{D_2} f(x, y) \, dA$

#### Proposition 5.1.4: Fubini's Theorem

If  $f : \mathbb{R}^2 \rightarrow \mathbb{R}$  is continuous and bounded on the rectangle  $R = \{(x, y) \in \mathbb{R}^2 \mid a \leq x \leq b, c \leq y \leq d\}$ . Then

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

#### Proposition 5.1.5

If  $f : \mathbb{R}^2 \rightarrow \mathbb{R}$  is a function such that  $f(x, y) = g(x)h(y)$  where  $g, h : \mathbb{R} \rightarrow \mathbb{R}$  and  $R = \{(x, y) \in \mathbb{R}^2 \mid a \leq x \leq b, c \leq y \leq d\}$ , then

$$\iint_R f(x, y) \, dA = \int_a^b g(x) \, dx \int_c^d h(y) \, dy$$

#### Theorem 5.1.6

Let  $R$  be a region. Let  $f : \mathbb{R}^2 \rightarrow \mathbb{R}^2$  defined by  $f(x, y) = \begin{pmatrix} u(x, y) \\ v(x, y) \end{pmatrix}$  such that  $f|_R$  is a bijection and  $\partial f \in C^1(R)$ . Let  $g : \mathbb{R}^2 \rightarrow \mathbb{R}$ . Then

$$\iint_S g(u, v) \, du \, dv = \iint_R g(x, y) |\partial f| \, dx \, dy$$

The conditions for  $f$  being a local bijection is given by the inverse function theorem.

**Theorem 5.1.7: Moments**

Let  $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ . Let  $\rho(x, y)$  give the moment at  $(x, y)$ . The moment about the  $x$ -axis on a region  $D \subset \mathbb{R}^2$  is given by

$$M_x = \iint_D y \rho(x, y) dA$$

and the moment about the  $y$ -axis is given by

$$M_y = \iint_D x \rho(x, y) dA$$

**Theorem 5.1.8: Mass**

Let  $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ . Let  $\rho(x, y)$  give the moment at  $(x, y)$ . Let  $D \subset \mathbb{R}^2$  be a region. Then the mass is given by

$$m = \iint_D \rho(x, y) dA$$

**Theorem 5.1.9: Center of Mass**

Let  $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ . Let  $\rho(x, y)$  give the moment at  $(x, y)$ . Let  $D \subset \mathbb{R}^2$  be a region. The center of mass is given by

$$(\bar{x}, \bar{y}) = \left( \frac{M_y}{m}, \frac{M_x}{m} \right)$$

**5.2 Green's Theorem for a Planar Region****Definition 5.2.1: Curl**

Let  $v : U \subseteq \mathbb{R}^2 \rightarrow \mathbb{R}^2$  be a continuously differentiable planar vector field given by  $v(x, y) = (a(x, y), b(x, y))$ . Define the curl of  $v$  to be

$$\text{curl}(v) = \frac{\partial b}{\partial x} - \frac{\partial a}{\partial y}$$

**Theorem 5.2.2: Green's Theorem for a Rectangular Region**

Let  $v : U \subseteq \mathbb{R}^2 \rightarrow \mathbb{R}^2$  be a continuously differentiable planar vector field where  $v(x, y) = \begin{pmatrix} v_1(x, y) \\ v_2(x, y) \end{pmatrix}$ . Let  $\Omega = \{(x, y) \in \mathbb{R}^2 \mid a \leq x \leq b, c \leq y \leq d\}$  such that  $\partial\Omega \subset U$ . Then

$$\iint_{\Omega} \text{curl}(v) dA = \int_{\partial\Omega} v \cdot dr$$

where  $r$  is positively oriented.

*Proof.* Clearly, we have

$$\iint_{\Omega} \text{curl}(v) dA = \int_a^b v_1(x, c) dx + \int_b^c v_2(d, y) dy + \int_c^a v_1(x, d) dx + \int_d^b v_2(a, y) dy$$

□

**Theorem 5.2.3: Green's Theorem for a Planar Region**

Let  $v : U \subseteq \mathbb{R}^2 \rightarrow \mathbb{R}^2$  be a continuously differentiable planar vector field. Let  $\Omega$  be a region such that  $\Omega \cup \partial\Omega \subset U$  and  $r$  a parametrization of  $\partial\Omega$ . Then

$$\iint_{\Omega} \text{curl}(v) \, dA = \oint_{\partial\Omega} v \cdot dr$$

where  $r$  is positively oriented.

**Theorem 5.2.4: Conservative Implies Zero Curl**

If  $v : U \subseteq \mathbb{R}^2 \rightarrow \mathbb{R}^2$  is a conservative vector field with continuous partial derivatives on  $U$ , then

$$\text{curl}(v) = 0$$

**5.3 Divergence Theorem for a Planar Region****Definition 5.3.1: Flux across a Curve**

Let  $v : \mathbb{R}^2 \rightarrow \mathbb{R}^2$  be a planar vector field. Let  $\Omega \subset \mathbb{R}^2$  be a region with  $\partial\Omega$  parametrized by  $r$ . Define the flux of  $v$  across  $\Omega$  to be

$$\text{flux of } v = \int_{\partial\Omega} v \cdot n \, dr$$

where  $r$  is positively oriented such that  $n$  is the outwards unit normal.

**Theorem 5.3.2: Divergence Theorem for a Planar Region**

Let  $v : \mathbb{R}^2 \rightarrow \mathbb{R}^2$  be a continuously differentiable planar vector field. Let  $\Omega$  be a region such that  $\Omega \cup \partial\Omega \subset U$  and  $r$  a parametrization of  $\partial\Omega$ . Then

$$\iint_{\Omega} \nabla \cdot v \, dA = \int_{\partial\Omega} v \cdot n \, dr$$

where  $r$  is positively oriented such that  $n$  is the outwards unit normal.



## 6 3 Dimensional Calculus

### 6.1 Triple Integrals

#### Definition 6.1.1: Triple Integral

Let  $f : \mathbb{R}^3 \rightarrow \mathbb{R}$  be a function. Define the triple integral of  $f$  in a region  $\Omega$  to be

$$\iiint_{\Omega} f(x, y, z) \, dx \, dy \, dz$$

#### Theorem 6.1.2

Let  $R \subset \mathbb{R}^3$  be a region. Let  $f : \mathbb{R}^3 \rightarrow \mathbb{R}^3$  defined by  $f(x, y, z) = \begin{pmatrix} u(x, y, z) \\ v(x, y, z) \\ w(x, y, z) \end{pmatrix}$  such that  $f|_R$  is a bijection and  $\partial f \in \mathcal{C}^1(R)$ . Let  $g : \mathbb{R}^3 \rightarrow \mathbb{R}$ . Then

$$\iiint_S g(u, v, w) \, du \, dv \, dw = \iiint_R g(x, y, z) |\partial f| \, dx \, dy \, dz$$

#### Theorem 6.1.3: Cylindrical Coordinates

Let  $f : \mathbb{R}^3 \rightarrow \mathbb{R}$ . Let  $E \subset \mathbb{R}^3$  be a region. Then

$$\iiint_R f(x, y, z) \, dV = \iiint_R f(r \cos \theta, r \sin \theta, z) r \, dr \, d\theta \, dz$$

with  $x = r \cos(\theta)$  and  $y = r \sin(\theta)$  and  $z = z$

#### Theorem 6.1.4: Spherical Coordinates

Let  $f : \mathbb{R}^3 \rightarrow \mathbb{R}$ . Let  $E \subset \mathbb{R}^3$  be a region. Then

$$\iiint_R f(x, y, z) \, dV = \iiint_R f(r \cos \theta \sin \phi, r \sin \theta \sin \phi, r \cos \phi) r^2 \sin(\phi) \, dr \, d\theta \, d\phi$$

with  $x = r \sin(\phi) \cos(\theta)$  and  $y = r \sin(\phi) \sin(\theta)$  and  $z = r \cos(\phi)$

#### Theorem 6.1.5: Mass

Consider a solid occupying a region  $\Omega \subset \mathbb{R}^3$  with density function  $\rho(x, y, z)$ . Then its total mass is

$$M = \iiint_{\Omega} \rho(x, y, z) \, dV$$

#### Theorem 6.1.6: Center of Mass

Consider a solid occupying a region  $\Omega \subset \mathbb{R}^3$  with density function  $\rho(x, y, z)$  and total mass  $M$ . Then

$$(\bar{x}, \bar{y}, \bar{z}) = \left( \frac{1}{M} \iiint_{\Omega} \rho x \, dV, \frac{1}{M} \iiint_{\Omega} \rho y \, dV, \frac{1}{M} \iiint_{\Omega} \rho z \, dV \right)$$

## 6.2 Surface Integrals

### Theorem 6.2.1

Let  $S$  be a surface parametrized by  $\mathbf{r}(u, v)$ . The surface area is given by

$$A = \iint_{\Omega} |\mathbf{r}_u \times \mathbf{r}_v| \, du dv$$

where  $\Omega$  is the parameter domain.

### Lemma 6.2.2

The formula of the area of a surface given by  $z = f(x, y)$  where  $(x, y) \in R \subset \mathbb{R}^2$  is given by

$$A = \iint_R \sqrt{1 + \left(\frac{\partial f}{\partial x}\right)^2 + \left(\frac{\partial f}{\partial y}\right)^2} \, dx dy$$

## 6.3 Divergence Theorem

### Definition 6.3.1: Curl

Let  $f : U \subset \mathbb{R}^3 \rightarrow \mathbb{R}^3$  be a vector field. Then the curl of  $f$  is defined as

$$\nabla \times f = \begin{vmatrix} i & j & k \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ f_1 & f_2 & f_3 \end{vmatrix}$$

### Proposition 6.3.2

If  $v : U \subseteq \mathbb{R}^3 \rightarrow \mathbb{R}^3$  is a conservative vector field with continuous partial derivatives on  $U$ , then

$$\text{curl}(v) = 0$$

### Definition 6.3.3: Flux of a Surface

Let  $v : \mathbb{R}^3 \rightarrow \mathbb{R}^3$  be a vector field. Let  $\Omega \subset \mathbb{R}^3$  be a volume. Define the flux of  $v$  across  $\Omega$  to be

$$\text{flux of } v = \iint_{\partial\Omega} v \cdot n \, dA$$

where  $n$  is the outward pointing normal.

### Lemma 6.3.4: Practical Definition of Flux

Let  $\mathbf{v} : \mathbb{R}^3 \rightarrow \mathbb{R}^3$ . Let  $\Omega \subset \mathbb{R}^3$ . Let  $r : \mathbb{R}^2 \rightarrow \mathbb{R}^3$  parameterize the surface  $\partial\Omega$ . Then

$$\iint_{\partial\Omega} v \cdot n \, dA = \iint_{\partial\Omega} \mathbf{v}(r(u, v)) \cdot \left( \frac{\partial r}{\partial u} \times \frac{\partial r}{\partial v} \right) \, du dv$$

### Theorem 6.3.5: The Divergence Theorem

Let  $v : U \subset \mathbb{R}^3 \rightarrow \mathbb{R}^3$  be a  $C^1$  vector field. Let  $\Omega$  be a region in  $\mathbb{R}^3$  with  $\partial\Omega$  parametrized by  $r$  such that  $\Omega \cup \partial\Omega \subset U$ . Then

$$\iiint_{\Omega} \nabla \cdot v \, dV = \iint_{\partial\Omega} v \cdot n \, dA$$

where  $r$  is positively oriented such that  $n$  is the outwards unit normal.

## 6.4 Stoke's Theorem

### Definition 6.4.1: Pointwise Circulation

Let  $\mathbf{F}(x, y, z)$  give the velocity of a fluid. The pointwise circulation of  $\mathbf{F}$  is given by

$$\nabla \times \mathbf{F} \cdot \hat{n}$$

### Theorem 6.4.2: Net Circulation

The net circulation over a surface  $S$  is given by

$$\iint_S \nabla \times \mathbf{F} \cdot \hat{n} \, dS$$

### Theorem 6.4.3: Stoke's Theorem

Let  $\mathbf{F}(x, y, z)$  be a vector field. Let  $S$  be a surface with unit normal  $\hat{n}$  and boundary curve  $C$ , oriented according to the right hand rule. Then

$$\iint_S \nabla \times \mathbf{F} \cdot \hat{n} \, dS = \int_C \mathbf{F} \cdot d\mathbf{r}$$