ANALYSIS OF SEVERAL VARIABLES

Jaydeb Sarkar, notes by Ramdas Singh

Third Semester

List of Symbols

Placeholder

Contents

Chapter 1

INTRODUCTION TO $\mathbb{R}^n$

1.1 Translation into Higher Dimensions

July 21st.

We begin with a definition.

Definition 1.1. The space $\mathbb{R}^n$ is defined as $\mathbb{R} \times \mathbb{R} \times \cdots \times \mathbb{R}$ (n times) $= \{(x_1, x_2, \ldots, x_n) : x_i \in \mathbb{R}\}$.

In the context of analysis, we will talk about open sets, closed sets, sequences, compact sets, and connected sets. In contrast, algebra considers $\mathbb{R}^n$ as a vector space with the operators $+$ and $\cdot$. Combining both these aspects results in the study of analysis of several variables. In this course, we will mainly focus on dealing with functions of the form $f : \mathbb{R}^n \to \mathbb{R}^m$, and talk about their properties such as continuity, differentiability, and integrability.

1.1.1 Algebraic and Analytic Structure

We note that $\mathbb{R}^n$ is also an inner product space with the following properties:

- $\langle x, y\rangle = \sum_{i=1}^n x_i y_i$ for all $x, y \in \mathbb{R}^n$.
- The set $\{e_i\}_{i=1}^n$ consisting of the unit vectors is an orthonormal basis for $\mathbb{R}^n$.
- The simplest maps from $\mathbb{R}^n$ to $\mathbb{R}^m$ are linear maps that send lines to lines.

Example 1.2. Suppose the function f is a linear map from $\mathbb{R}$ to $\mathbb{R}$. This implies that $f(x) = xf(1)$ for all $x \in \mathbb{R}$. Thus, $f(x) = cx$ for all $x \in \mathbb{R}$, where c is a constant. Conversely, if $c \in \mathbb{R}$, then $x \mapsto cx$ is a linear map. Therefore, we conclude that $\{f : \mathbb{R} \to \mathbb{R}, \text{linear}\} \leftrightarrow \mathbb{R}$, with a possible bijection given by $f \mapsto f(1)$.

In the above example, we note that 1 is not special; we could simply fix any $\alpha \in \mathbb{R} \setminus \{0\}$, and notice that $f(x) = \frac{x}{\alpha} f(\alpha)$ for all $x \in \mathbb{R}$. Here, replacing $f(1)$ by $f(\alpha)$ is a kind of 'change of variable'.

Remark 1.3. Let $L : \mathbb{R}^n \to \mathbb{R}^m$ be a linear map. Then, $Le_j = \sum_{i=1}^m a_{ij} e_i$ for all $j = 1, 2, \ldots, n$. We may write L as $(a_{ij})_{m\times n} \in M_{m\times n}(\mathbb{R})$, the set of all $m \times n$ matrices with real entries.

Coming to the analysis side, there is a need for defining a distance between points in $\mathbb{R}^n$. Previously, we have seen that the *norm* may be given as $\|x\| = \sqrt{\sum_{i=1}^n x_i^2}$ for all $x \in \mathbb{R}^n$. We can use this norm to define our required distance function.

Definition 1.4. The *distance* function between two points $d : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}_{\geq 0}$ is defined as $d(x, y) = \|x - y\| = \left(\sum_{i=1}^n (x_i - y_i)^2\right)^{1/2}$ for all $x, y \in \mathbb{R}^n$.

For $n = 1$, we note that $d(x, y) = |x - y|$, from the previous analysis courses. $\mathbb{R}^n$ equipped with the function d is called a *metric space*. Coming to the properties of the inner product, we have

- $\|x\| = \langle x, x\rangle^{1/2}$ for all $x \in \mathbb{R}^n$.
- $\langle x, y\rangle = \langle y, x\rangle$ for all $x, y \in \mathbb{R}^n$.
- The function $\langle , \rangle$ is linear with respect to the first and second arguments.

We also have the important Cauchy-Schwarz inequality.

Theorem 1.5 (*Cauchy-Schwarz inequality*)**.** *For all $x, y \in \mathbb{R}^n$, we have $|\langle x, y\rangle| \le \|x\| \|y\|$.*

Proof. Note that

$$0 \le \sum_{i=1}^{n}\sum_{j=1}^{n}(x_i y_j - x_j y_i)^2 = 2\left(\sum_{i,j} x_i^2 y_j^2 - \sum_{i,j} x_i x_j y_i y_j\right) = 2\left(\|x\|^2 \|y\|^2 - \langle x, y\rangle^2\right) \tag{1.1}$$

$$\implies |\langle x, y\rangle| \le \|x\| \|y\| . \tag{1.2}$$

■

We note that equality occurs if and only if the first quantity in the above equation is zero, *i.e.*, if and only if $x_i y_j = x_j y_i$ for all i, j, or $\frac{x_i}{y_i} = \frac{x_j}{y_j}$ for all i, j showing that x and y are linearly dependent.

Corollary 1.6 (*Triangle inequality*)**.** *For all $x, y \in \mathbb{R}^n$, we have $\|x + y\| \le \|x\| + \|y\|$.*

Proof. We have

$$\|x + y\|^2 = \langle x + y, x + y\rangle = \|x\|^2 + 2\langle x, y\rangle + \|y\|^2 \le \|x\|^2 + 2\|x\| \|y\| + \|y\|^2 = (\|x\| + \|y\|)^2 \tag{1.3}$$

where the inequality follows from Cauchy-Schwarz. ■

The following will prove to be an important result.

Theorem 1.7. *Let $L : \mathbb{R}^n \to \mathbb{R}^m$ be a linear map. Then, there exists a $M > 0$ such that $\|Lx\| \le M \|x\|$ for all $x \in \mathbb{R}^n$.*

Proof. Rewriting x as $x = \sum_{i=1}^{n} x_i e_i$, we have

$$Lx = \sum_{i=1}^{n} x_i L e_i$$

$$\implies \|Lx\| = \left\|\sum_{i=1}^{n} x_i L e_i\right\| \le \sum_{i=1}^{n} |x_i| \|Le_i\| \le \|x\| \sum_{i=1}^{n} \|Le_i\| = \|x\| M. \tag{1.4}$$

The first inequality follows from the triangle inequality, and the second from Cauchy-Schwarz. In the last step, M is set to be $\sum_{i=1}^{n} \|Le_i\|$, which is a constant. ■

We also term $(\mathbb{R}^n, d)$ as a Euclidean metric space. There is now a need to define open sets in $\mathbb{R}^n$ to talk more about the analysis of several variables.

Definition 1.8. For $a \in \mathbb{R}^n$ and $r > 0$, the *open ball* centred at a of radius r is $B_r(a) := \{x \in \mathbb{R}^n : d(x, a) < r\}$, the set of all points in $\mathbb{R}^n$ that are at a distance less than r from a.

From the notion of open balls, we can define open sets.

Definition 1.9. A set $S \subseteq \mathbb{R}^n$ is said to be an *open set* if for all $x \in S$, there exists an $r > 0$ such that $B_r(x) \subseteq S$.

We now bring the notion of convergence of sequences.

Definition 1.10. Let $\{x_m\} \subseteq \mathbb{R}^n$ be a sequence and $x \in \mathbb{R}^n$. We say that $\{x_m\}$ *converges* to x if for every $\varepsilon > 0$, there exists a natural N such that $\|x_m - x\| < \varepsilon$ for all $m \geq N$.

July 23rd.

Definition 1.11. Let $S \subseteq \mathbb{R}^n$ and $a \in \mathbb{R}^n$. We say that a is a *limit point* of S if $S \cap (B_r(a) \setminus \{a\})$ is non-empty for all $r > 0$.

We introduce more notation; for all $i = 1, 2, \ldots, n$, the mapping $\Pi_i : \mathbb{R}^n \to \mathbb{R}$ is called the i^{th} *projection* where $\Pi_i(x) = x_i$. Note that $x = (x_1, x_2, \ldots, x_n) = (\Pi_1(x), \Pi_2(x), \ldots, \Pi_n(x))$. This notation allows us to formulate the following useful fact a little more neatly.

Theorem 1.12. *Let $\{x_m\} \subseteq \mathbb{R}^n$ be a sequence and $x \in \mathbb{R}^n$. Then $x_m \to x$ if and only if $\Pi_i(x_m) \to \Pi_i(x)$ for all $i = 1, 2, \ldots, n$.*

Proof. Suppose $x_m \to x$. Then for all $\varepsilon > 0$, there exists a natural N such that $\|x_m - x\| < \varepsilon$ for all $N \geq N$. Restating, we have

$$\sum_{i=1}^{n} (\Pi_i(x_m) - \Pi_i(x))^2 < \varepsilon^2 \text{ for all } n \geq N \tag{1.5}$$

$$\implies \text{For all } i,\ |\Pi_i(x_m) - \Pi_i(x)| < \varepsilon^2 \text{ for all } n \geq N. \tag{1.6}$$

For the converse, we simply work backwards with $\varepsilon/\sqrt{n}$ as our choice of epsilon. ∎

For example, the sequence $\{(\frac{1}{n}, \frac{1}{2n+3})\}_{n=1}^{\infty}$ converges to $(0, 0)$. However, the sequence $\{(\frac{1}{n}, n^2)\}_{n=1}^{\infty}$ does not.

Definition 1.13. Let $S \subseteq \mathbb{R}^n$. $a \in S$ is termed an *interior point* of S if for some $r > 0$, $B_r(a) \subseteq S$ holds. Thus, a set S is open if a is an interior point for all $a \in S$. The *interior* of set S is defined as $\operatorname{int} S := \{a \in S \mid a \text{ is an interior point.}\}$.
If $a \in \operatorname{int}(S^c)$, then a is termed an *exterior point* of S. a is termed a *boundary point* if $B_r(a)$ meets both S and S^c for all $r > 0$. The set of *boundary points* of S is denoted as ∂S.

We also term a set $S \subseteq \mathbb{R}^n$ as a *closed set* if $\mathbb{R}^n \setminus S$ is open. The following facts will only be stated and will be left as an exercise to the reader:

- A set $C \subseteq \mathbb{R}^n$ is closed if and only if for all sequences $\{x_m\}_{m=1}^{\infty} \subseteq C$ that converge to x implies $x \in C$.
- The open ball $B_r(a)$ is an open set.
- The intersection of an arbitrary collection of closed sets is closed; likewise, the union of an arbitrary collection of open sets is open.
- The set $S \subseteq \mathbb{R}^n$ is open if and only if $S = \operatorname{int} S$.

Fix $O \subseteq \mathbb{R}^n$.

- O is open if and only if $O \cap \partial O = \emptyset$.
- O is closed if and only if $\partial O \subseteq O$.

For $S \subseteq \mathbb{R}^n$, we define the *closure* of set S as $\overline{S} = \operatorname{int} S \cup \partial S$.

- $S \subseteq \mathbb{R}^n$ is closed if and only if $\overline{S} = S$.
- Let $C_i \subseteq R$ be closed sets and $O_i \subseteq \mathbb{R}$ be open sets, for $i = 1, 2, \ldots, n$. Then $C_1 \times C_2 \times \cdots \times C_n \subseteq \mathbb{R}^n$ is a closed set, and $O_1 \times O_2 \times \cdots \times O_n \subseteq \mathbb{R}^n$ is an open set.
- The n *dimensional unit sphere* $S^{n-1} := \{x \in \mathbb{R}^n : \|x\| = 1\}$ is closed in $\mathbb{R}^n$.

Definition 1.14. For $S \subseteq \mathbb{R}^n$ and $a \in \mathbb{R}^n$, a is termed an *isolated point* if a is *not* a limit points; that there exists an $r > 0$ such that $S \cap (B_r(a) \setminus \{a\}) = \emptyset$.

With the pesky definitions and translation of one dimensional concept into being defined over several variables, we come to limits and continuity.

1.2 Limits and Continuity

Recall that given $f : (a, b) \setminus \{c\} \to \mathbb{R}$, we say that $\lim_{x \to c} f(x) = b$ if for every $\varepsilon 0$, there exists a $\delta > 0$ such that $|f(x) - b| < \varepsilon$ for all x satisfying $0 < |x - c| < \delta$. Note that in this definition of the limit, we have $f(x) \in B_\varepsilon(b)$ and $x \in B_\delta(c) \setminus \{c\}$; this can easily be rewritten as $f(B_\delta(c) \setminus \{c\}) \subseteq B_\varepsilon(b)$. However, for our definition we would not require f to be defined on an open set. We defind it over any arbitrary set.

Definition 1.15. Let $a \in S \subseteq \mathbb{R}^n$ be a limit point of S and let $f : S \setminus \{a\} \to \mathbb{R}^m$ be a function and $b \in \mathbb{R}^m$. We say $\lim_{x \to a} f(x) = b$ if for every $\varepsilon > 0$, there exists a $\delta > 0$ such that $f((B_\delta(a) \setminus \{a\}) \cap S) \subseteq B_\varepsilon(b)$. Again, this is equivalent to saying that $\|f(x) - b\| < \varepsilon$ for all $x \in S \setminus \{a\}$ satisfying $\|x - a\| < \delta$.

It is important to get accustomed to the definition that works with open balls.

Remark 1.16. In the above definition, if we instead write $x - a = h$, then $\lim_{x \to a} f(x) = b$ is equivalent to saying that for every $\varepsilon > 0$, there exists a $\delta > 0$ such that $\|f(a + h) - b\| < \varepsilon$ for all $\|h\| < \delta$. We can further rewrite to get the usual notation of

$$\lim_{\|h\| \to 0} \|f(a + h) - b\| = 0. \tag{1.7}$$

Note that the above limit is in the real numbers, making it eaiser to deal with.

A notion of continuity also comes in handy.

Definition 1.17. For $S \subseteq \mathbb{R}^n$, let $f : S \to R^m$ with $a \in S$. We say f is *continuous* at a if $\lim_{x \to a} f(x) = f(a)$. In other words, for every $\varepsilon > 0$, there exists a $\delta > 0$ such that

$$\|f(x) - f(a)\| < \varepsilon \text{ for all } x \in S \text{ satisfying } \|x - a\| < \delta \tag{1.8}$$

or

$$f(B_\delta(a) \cap S) \subseteq B_\varepsilon(f(a)). \tag{1.9}$$

Note that if a is an isolated point of S, then any $f : S \to \mathbb{R}^m$ is continuous at a since $f(\{a\}) \subseteq B_\varepsilon(f(a))$ holds true, trivially.

Remark 1.18. Similar to the previous remark, f is continuous at a if and only if

$$\lim_{\|h\| \to 0} \|f(a + h) - f(a)\| = 0. \tag{1.10}$$

Functions defined on $S \subseteq \mathbb{R}^n$ can be broken down into components; given $f : S \to R^m$, define $f_j := \Pi_j \circ f$ for all $j = 1, 2, \ldots, m$. Thus, f can be rewritten as $(f_1, f_2, \ldots, f_m)$. We can conclude that f is continuous at $a \in S$ if and only if $f_j : S \to \mathbb{R}$ is continuous at a for all $j = 1, 2, \ldots, m$. The proof of this observation is left as an exercise to the reader.

Theorem 1.19. *Let $a \in \mathbb{R}^n$ be a limit point of a set $S \subseteq \mathbb{R}^n$, with $b \in \mathbb{R}^m$ and $f : S \to R^m$ a function. Then, the following are equivalent—*

1. $\lim_{x \to a} f(x) = b$.
2. *If* $\{x_p\} \subseteq S \setminus \{a\}$ *and* $x_p \to a$, *then* $f(x_p) \to b$.
3. $\lim_{x \to a} \|f(x) - b\| = 0$.

The proof of this theorem is left as an exercise to the reader.

Definition 1.20. For a set $S \subseteq \mathbb{R}^n$, a function $f : S \to \mathbb{R}^m$ is termed a continuous function if f is continuous at a for all $a \in S$.

Theorem 1.21. *Let $f : S \to \mathbb{R}^m$ be a function, where $S \subseteq \mathbb{R}^n$. The following are, then, equivalent—*

1. f *is continuous.*
2. *For all* $a \in S$ *and* $\{x_n\} \subseteq S$ *with* $x_n \to a$, *we have* $f(x_n) \to f(a)$.
3. *For all open sets* $O \subseteq \mathbb{R}^m$, *the set* $f^{-1}(O) \subseteq S$ *is also open.*
4. *For all closed sets* $C \subseteq \mathbb{R}^m$, *the set* $f^{-1}(C) \subseteq S$ *is also closed.*

Proof. For 1. implies 3. , let $O \subseteq \mathbb{R}^m$ be open. Pick some $a \in f^{-1}(O)$. Then, since $f(a) \in O$, there exists $r > 0$ such that $B_r(f(a)) \subseteq O$. Also, f is continuous at a; for $\frac{r}{2} > 0$, there exists $\delta > 0$ such that

$$f(B_\delta(a)) \subseteq B_{\frac{r}{2}}(f(a)) \subseteq B_r(f(a)) \implies a \in B_\delta(a) \subseteq f^{-1}(B_r(f(a))) \subseteq f^{-1}(O). \tag{1.11}$$

Thus, $f^{-1}(O)$ is open. For 3. implies 1. , let $a \in S$. Fix $\varepsilon > 0$. Then the set $f^{-1}(B_\varepsilon(f(a)))$ is open; there exists a $\delta > 0$ such that $B_\delta(a) \subseteq f^{-1}(B_\varepsilon(f(a)))$. ∎

July 28th.

We look at a few examples.

Example 1.22. Let $f : \mathbb{R}^2 \setminus \{(0,0)\} \to \mathbb{R}$ be defind as $f(x,y) = \frac{2xy}{x^2+y^2}$. We find the limit $\lim_{(x,y)\to(0,0)} f(x,y)$. Let us approach from different directions, starting with the line L_1 defined as $y = 0$ with $x > 0$. Then $\lim_{(x,y)\to(0,0);(x,y)\in L_1} f(x,y) = \lim_{(x,y)\to(0,0)} 0 = 0$. However, along the line L_2 defned as $\{(x,y) \mid x = y\ x,y > 0\}$, we have $\lim_{(x,y)\to(0,0);(x,y)\in L_2} f(x,y) = 1$. We conclude that this limit cannot exist. Going along the line $y = mx$ gives several possible values for the limit.

The above method is good only for showing that the limit does not exist; if the limit does exist, we need to use theory.

Example 1.23. We compute the limit $\lim_{(x,y)\to(0,0)} \frac{x^3}{x^2+y^2}$. Here, we can prove that the limit exists as follows:

$$\left|\frac{x^3}{x^2+y^2}\right| \leq \left|\frac{x^3}{x^2}\right| = |x| \to 0 \implies \lim_{(x,y)\to(0,0)} \frac{x^3}{x^2+y^2} = 0. \tag{1.12}$$

Example 1.24. We solve the limit $\lim_{(x,y)\to(0,0)} \frac{\sin(x^2+y^2)}{x^2+y^2}$. Simply rewriting $z = x^2 + y^2$ gives us $z \to 0$ as $(x,y) \to (0,0)$, so $\lim_{z\to 0} \frac{\sin z}{z} = 1$.

Before the next example, we write down a few properties. Let $S \subseteq \mathbb{R}^n$, let $f, g : S \to \mathbb{R}$ be functions, and let $a \in \mathbb{R}^n$ be a limit point of S. Suppose $\lim_{x\to a} f(x) = \alpha$ and $\lim_{x\to b} g(x) = \beta$. Then,

1. $\lim_{x\to a}(cf(x) + g(x)) = c\alpha + \beta$ for all $c \in \mathbb{R}$,

2. $\lim_{x \to a} f(x)g(x) = \alpha\beta$,
3. $\lim_{x \to a} \frac{f(x)}{g(x)} = \frac{\alpha}{\beta}$, provided that $\beta \neq 0$,
4. if $f(x) \leq h(x) \leq g(x)$ for all $x \in S$ and if $\alpha = \beta$, then $\lim_{x \to a} h(x)$ exists and equals α.

A similar set of corresponding statements also hold true for continuous functions. Note that the function $\Pi_i : \mathbb{R}^n \to \mathbb{R}$ is also continuous.

Example 1.25. The function $f(x, y) = \frac{\sin(x^2+y^2)}{x^2+y^2}$ for $(x, y) \neq (0, 0)$ and $f(x, y) = 1$ otherwise is a continuous function since it has been assigned its limit at $(x, y) = (0, 0)$. However, the function $f(x, y) = \frac{2xy}{x^2+y^2}$ for $(x, y) \neq (0, 0)$ and $f(x, y) = \alpha$ otherwise is continuous only at $\mathbb{R}^2 \setminus \{(0, 0)\}$ for all $\alpha \in \mathbb{R}$.

Example 1.26. We look at the continuity of the function

$$f(x, y) = \begin{cases} \frac{xy}{\sqrt{x^2+y^2}} & \text{if } (x, y) \neq (0, 0), \\ 0 & \text{if } (x, y) = (0, 0). \end{cases} \tag{1.13}$$

Then, we have

$$\left| \frac{xy}{\sqrt{x^2 + y^2}} \right| \leq \frac{1}{2} \cdot \frac{x^2 + y^2}{\sqrt{x^2 + y^2}} = \frac{1}{2} \|(x, y)\| \tag{1.14}$$

which shows that $\lim_{(x,y) \to (0,0)} f(x, y) = 0 = f(0, 0)$. Thus, f is continuous on $\mathbb{R}^2$.

Example 1.27. Set $\mathcal{D} = \{(x, y) \in \mathbb{R}^2 \mid y \neq 0\}$. Since $\mathcal{D} = (\Pi_2^{-1}(\{0\}))^c$, $\mathcal{D}$ is an open set. Define $f : \mathcal{D} \to \mathbb{R}$ by $f(x, y) = x \sin \frac{1}{y}$. Here, we simply work as

$$|f(x, y)| = \left| x \sin \frac{1}{y} \right| \leq |x| \text{ on } \mathcal{D}. \tag{1.15}$$

Thus, the limit becomes $f(x, y)$.

Hereforth, O_n denotes an open subset of $\mathbb{R}^n$.

Remark 1.28. Let $(a, b) \in \mathbb{R}^2$ be a limit point of O_2. Suppose $\lim_{(x,y) \to (a,b)} f(x, y)$ exists and equals $\alpha \in \mathbb{R}$. It is natural to ask whether

$$\alpha = \lim_{y \to b} \lim_{x \to a} f(x, y) = \lim_{x \to a} \lim_{y \to b} f(x, y). \tag{1.16}$$

For someone looking at multivariable limits for the first time, it is tempting to believe this holds true always. We leave this question unanswered for now, and come back to it later.

A notion of uniform continuity may also be explored.

Definition 1.29. A function $f : S \to \mathbb{R}$, for $S \subseteq \mathbb{R}^n$, is said to be a *uniformly continuous* function if for every $\varepsilon > 0$ there exists $\delta > 0$ such that

$$\|f(x) - f(y)\| < \varepsilon \text{ for all } \|x - y\| < \delta \text{ in } S. \tag{1.17}$$

We urge the reader to compute examples for uniformly continuous functions. The exercise of uniform contiuity implying continuity but not the other way around is left as an exercise to the reader.

1.3 Differentiability

As a little convention, for any $f : O_n \to \mathbb{R}^m$, we prefer to rewrite it as $f = (f_1, \ldots, f_n)$ where $f_j = \Pi_j f$. We now ask the question of derivatives; what does it mean for the derivative of a function $f : O_n \to \mathbb{R}^m$? What about $f'(a)$ for some $a \in O_n$?

For the case of $n = m = 1$, we recall that f is termed differentiable at a if and only if $\lim_{h\to 0} \frac{f(a+h)-f(a)}{h}$ exists. If the limit is λ, then this limit exists if and only if $\lim_{h\to 0} \frac{f(a+h)-f(a)-h\lambda}{h}$, which is really a function of h. Thus, the function $h \mapsto \lambda h$ matters the most, that is, $L : \mathbb{R} \to \mathbb{R}$ where $Lh = \lambda h$. So, we can twist our words a little and say that f is differentiable at a if and only if there exists a linear map $L : \mathbb{R} \to \mathbb{R}$ such that

$$\lim_{h\to 0} \frac{f(a+h) - f(a) - Lh}{h} = 0. \tag{1.18}$$

In this case, $f'(a) = L1 = \lambda$. We translate this exact idea into higher dimensions.

Definition 1.30. Let $f : O_n \to \mathbb{R}^m$. We say that f is *differentiable* at $a \in O_n$ if there exists a linear map $L : \mathbb{R}^n \to \mathbb{R}^m$, that depends on a, such that

$$\lim_{h\to 0} \frac{1}{\|h\|} (f(a+h) - f(a) - Lh) = 0. \tag{1.19}$$

In this case, we write $Df(a) = L$ and call it the total derivative of f at a. We say f is differentiable on O_n if f is differentiable at a for all $a \in O_n$.

Observe that the above limit is equivalent to saying that $\lim_{h\to 0} \frac{1}{\|h\|} \|f(a+h) - f(a) - Lh\| = 0$. Note that $Df(a) = L$ is unique. To show this, suppose there exists another linear map $\tilde{L} : \mathbb{R}^n \to \mathbb{R}^m$ such that $\lim_{h\to 0} \frac{1}{\|h\|} \left\|f(a+h) - f(a) - \tilde{L}h\right\| = 0$. Let there exist $h_0 \in \mathbb{R}^n$ such that $Lh_0 \neq \tilde{L}h_0$ and $\|h_0\| = 1$. Define $h : \mathbb{R} \to \mathbb{R}^n$ by $ht = th_0$. Then as $t \to 0$, $ht \to 0$. Therefore,

$$\left\|L(h(t)) - \tilde{L}(h(t))\right\| \leq \|f(a+h) - f(a) - Lh(t)\| + \left\|f(a+h) - f(a) - \tilde{L}h(t)\right\| \tag{1.20}$$

$$\implies \lim_{t\to 0} \frac{\left\|Lh(t) - \tilde{L}h(t)\right\|}{\|h(t)\|} = 0 \implies \lim_{t\to 0} \frac{1}{|t|} |t| \cdot \left\|Lh_0 - \tilde{L}h_0\right\| = 0 \implies Lh_0 = \tilde{L}h_0 \tag{1.21}$$

which is a contradiction.

Example 1.31. Let $f : \mathbb{R}^n \to \mathbb{R}^m$ be a linear map. In this case, $\frac{f(a+h)-f(a)-f(h)}{\|h\|} \to 0$ as $\|h\| \to 0$. Thus, f is differentiable at a and $Df(a) = f$ for all $a \in \mathbb{R}^n$.

Example 1.32. Suppose $f : \mathbb{R}^n \to \mathbb{R}^m$ be defined as $f(x) = c$ for all $x \in \mathbb{R}^n$. Then, we simply have $Df(a) = 0$, the null linear mapping.

We now truly ask how to compute $Df(a)$. Observe that $Df(a) : \mathbb{R}^n \to \mathbb{R}^m$ is a linear map. Then, one can represent it as a matrix $Df(a) \in M_{m\times n}(R)$.

July 30th.

Theorem 1.33. *Let $f : O_n \to \mathbb{R}^m$ be a function. Then f is differentiable at $a \in O_n$ if and only if $f_i : O_n \to \mathbb{R}$ is differentiable at a for all $i = 1, 2, \ldots, n$. Moreover, in this case,*

$$[Df(a)]_{m\times n} = \begin{bmatrix} [Df_1(a)]_{1\times n} \\ \vdots \\ [Df_m(a)]_{1\times n} \end{bmatrix}_{m\times n}. \tag{1.22}$$

From the above theorem it is clear that $D\Pi_i f = \Pi_i Df$. We now provide a proof.

Proof. For the forward implication, let f be differentiable at $a \in O_n$. Set $L := Df(a)$ and $L_i := \Pi_i Df(a)$. Note that $L_i : \mathbb{R}^n \to \mathbb{R}$ since $Df(a) : \mathbb{R}^n \to \mathbb{R}^m$ and $\Pi_i : \mathbb{R}^m \to \mathbb{R}$. Observe

$$f(a+h) - f(a) - Df(a)h = (\tilde{f}_1(h), \tilde{f}_2(h), \ldots, \tilde{f}_m(h)) \tag{1.23}$$

where $\tilde{f}_i(h) = f_i(a+h) - f_i(a) - L_i h$ for $i = 1, 2, \ldots, m$. Thus, for all $i = 1, 2, \ldots, m$,

$$\left|\tilde{f}_i(h)\right| \leq \left(\sum_{j=1}^{m} \left|\tilde{f}_j(h)\right|^2\right)^{1/2} = \|f(a+h) - f(a) - Lh\|. \tag{1.24}$$

Dividing by $\|h\|$ and taking $h \to 0$, we have

$$\lim_{h \to 0} \frac{\left|\tilde{f}_i(h)\right|}{\|h\|} = 0 \tag{1.25}$$

which shows that f_i is differentiable with $Df_i(a) = L_i$ $(= \Pi_i Df(a))$.

For the converse, let f_i be differentiable at a for all $i = 1, 2, \ldots, m$ and set $L_i = Df_i(a) : \mathbb{R}^n \to \mathbb{R}$. Set $L = \begin{bmatrix} L_1 \\ \vdots \\ L_m \end{bmatrix} : \mathbb{R}^n \to \mathbb{R}^m$, a linear map. Therefore,

$$\frac{1}{\|h\|} \|f(a+h) - f(a) - Lh\| = \frac{1}{\|h\|} \left(\sum_{j=1}^{m} \left|\tilde{f}_i(h)\right|^2\right)^{1/2} \to 0. \tag{1.26}$$

where $\tilde{f}_i(h) = f_i(a+h) - f_i(a) - L_i h$ for all i. ∎

Corollary 1.34. *Let $f : O_1 \to \mathbb{R}^m$ be a function. f is, then, differentiable at $a in O_1$ if and only if f_i is differentiable at a for all $1 \leq i \leq m$. Moreover, in this case,*

$$Df(a) = \begin{bmatrix} f_1'(a) \\ \cdots \\ f_m'(a) \end{bmatrix}. \tag{1.27}$$

This is just a special case when $n = 1$.

Remark 1.35. Let $f : O_n \to \mathbb{R}^m$ be differentiable at a. Then f is continuous at a.

Proof. We have

$$\begin{aligned} \|f(x) - f(a)\| &\leq \|f(x) - f(a) - (Df(a))(x-a)\| + \|(Df(a))(x-a)\| \\ &\leq \frac{1}{\|x-a\|} \|f(x) - f(a) - (Df(a))(x-a)\| \cdot \|x-a\| + M\|x-a\| \to 0 \end{aligned} \tag{1.28}$$

as $x \to a$. Note that such an $M > 0$ exists because $Df(a)$ is a linear map. ∎

1.3.1 Chain Rule

To simplify our study of derivatives in highder dimensions, we look at the so called chain rule. This will prove to be a very important tool to study the differentiability of any function of several variables.

Theorem 1.36 (The *chain rule*)**.** *Let $f : O_n \to O_m$ be differentiable at $a \in O_n$, and $g : O_m \to \mathbb{R}^p$ be differentiable at $b = f(a) \in O_m$. Then $g \circ f : O_n \to \mathbb{R}^p$ is differentiable at $a \in O_n$, and*

$$(Dg \circ f)(a) = Dg(f(a)) \circ Df(a). \tag{1.29}$$

For the proof, we will denote $A := Df(a)$ and $B := Dg(f(a))$, and $b = f(a)$. Moreover, we will write $r_f(x) := f(x) - f(a) + Df(a)(x-a)$.

Proof. There exists r_f in the neighbourhood of a and r_g in the neighbourhood of b such that $r_f(x) = f(x) - f(a) - A(x-a)$ and $r_g(y) = g(y) - g(b) - B(y-b)$. Now set

$$r(x) = g(f(x)) - g(b) - BA(x-a). \tag{1.30}$$

We claim that $\lim_{x\to a} \frac{r(x)}{\|x-a\|} = 0$. We know that $\lim_{x\to a} \frac{\|r_f(x)\|}{\|x-a\|} = 0 = \lim_{y\to b} \frac{\|r_g(y)\|}{\|y-b\|}$. Now,

$$\begin{aligned} r(x) &= g(f(x)) - g(b) - B(A(x-a)) = g(f(x)) - g(b) + B(r_f(x) - f(x) + f(a)) \\ &= (g(f(x)) - g(f(a)) - B(f(x) - f(a))) + Br_f(x) = r_g(f(x)) + Br_f(x). \end{aligned}$$

We show that both terms on the right hand side, when divided by $\|x-a\|$, tend to zero. Now,

$$\frac{\|Br_f(x)\|}{\|x-a\|} \leq M\frac{\|r_f(x)\|}{\|x-a\|} \to 0. \tag{1.31}$$

For the remaining term, we work as follows: for $\varepsilon > 0$ fixed, there exists a $\delta > 0$ such that $\|r_g(y)\| < \varepsilon\,\|y-b\|$ for all $0 < \|y-b\| < \delta$. By continuity of f at a, there exists a $\tilde{\delta} > 0$ such that $\|f(x) - f(a)\| < \delta$ for all $\|x-a\| < \tilde{\delta}$. Thus, for all $0 < \|x-a\| < \tilde{\delta}$, we have

$$\|r_g(f(x))\| < \varepsilon\,\|f(x) - f(a)\| < \varepsilon\,\|r_f(x) + A(x-a)\| \tag{1.32}$$

$$\implies \|r_g(f(x))\| < \varepsilon\,(\|r_f(x)\| + \|A(x-a)\|) < \varepsilon\left(\|r_f(x)\| + \tilde{M}\,\|x-a\|\right)$$

$$\implies \frac{\|r_g(f(x))\|}{\|x-a\|} \to 0 \text{ as } x \to a. \tag{1.33}$$

■

August 4th.

The derivative also satisfies nice properties in higher dimensions.

Proposition 1.37. *Let $f, g : O_n \to \mathbb{R}^m$ be differentiable at $a \in O_n$. Then*

1. *$D(\alpha f + g)(a) = \alpha Df(a) + Dg(a)$ for all $\alpha \in \mathbb{R}$.*
2. *If $m = 1$, then $(f \times g)'(a) = f(a)g'(a) + f'(a)g(a)$.*
3. *If $m = 1$, then $\left(\frac{f}{g}\right)'(a) = \frac{1}{(g(a))^2}(f'(a)g(a) - f(a)g'(a))$.*

Proof. The proof is left as an exercise to the reader. ■

1.4 Partial Derivatives

Given $a \in O_n$ and $f : O_n \to \mathbb{R}$, define $\eta_i : (-\varepsilon, \varepsilon) \to \mathbb{R}$ for $i = 1, 2, \ldots, n$ by

$$\eta_i(t) = f(a + te_i). \tag{1.34}$$

We define $\frac{\partial f}{\partial x_i}(a) = \frac{d\eta_i}{dt}(0)$, if the latter term is defined. This is called the *partial derivative* of f in the direction, or with respect to, x_i at a. Therefore,

$$\frac{\partial f}{\partial x_i}(a) = \lim_{t\to 0} \frac{f(a+te_i) - f(a)}{t}. \tag{1.35}$$

Remark 1.38.

1. Note that $\frac{\partial f}{\partial x_i}$ is easy to compute since we are essentially holding all other x_j's with $j \neq i$ as constant and just differentiating f with respect to x_i.
2. $\frac{\partial f}{\partial x_i}$ is the total derivative of f with the limit taken in the x_i-direction.

We want $Df(a)$ for a function $f : O_n \to \mathbb{R}$. We show further that the idea of partial derivatives solves this issues of computing $Df(a)$.

Definition 1.39. $f : O_n \to \mathbb{R}$ is termed a function in $C^1(O_n)$ if $\frac{\partial f}{\partial x_i}$ for all $i = 1, 2, \ldots, n$ exists and $x \mapsto \frac{\partial f}{\partial x_i}(x)$ is continuous on O_n.

We discuss some examples.

Example 1.40. Let $f(x, y) = x^3 + y^4 + \sin(xy)$ on $\mathbb{R}^2$. Taking y as a constant, $x \mapsto f(x, y)$ is differentiable. Thus,

$$\frac{\partial f}{\partial x} = 3x^2 + y\cos(xy). \tag{1.36}$$

Similarly,

$$\frac{\partial f}{\partial y} = 4y^3 + x\cos(xy). \tag{1.37}$$

The above two partial derivatives are continuous functions. Thus, $f \in C^1(\mathbb{R}^2)$.

Example 1.41. Recall the function

$$f(x, y) = \begin{cases} \frac{xy}{x^2+y^2} & \text{if } (x, y) \neq (0, 0), \\ 0 & \text{if } (x, y) = (0, 0) \end{cases} \tag{1.38}$$

which was discontinuous at the origin. At the origin, we have

$$\frac{\partial f}{\partial x}(0, 0) = \lim_{t \to 0} \frac{f(t, 0) - f(0, 0)}{t} = 0. \tag{1.39}$$

Similarly,

$$\frac{\partial f}{\partial y}(0, 0) = 0. \tag{1.40}$$

Thus we conclude that the partial derivatives exist at $(0, 0)$. But f is not continuous at $(0, 0)$. Thus, we conclude that the existence of partial derivatives does *not* imply the existence of the total derivative.

1.4.1 Higher Order Partials

Hereforth, we will denote $f_{x_i} := \frac{\partial f}{\partial x_i}$. Let us assume that the partials f_{x_i} exist for all $i = 1, 2, \ldots, n$. Therefore, $f_{x_i} : O_n \to \mathbb{R}$ are functions. Define

$$\frac{\partial^2 f}{\partial x_j \partial x_i} := \frac{\partial}{\partial x_j}\left(\frac{\partial f}{\partial x_i}\right) = \frac{\partial}{\partial x_j}\left(f_{x_i}\right) \text{ for all } j = 1, 2, \ldots, n. \tag{1.41}$$

This is known as the second order partial derivative. We may write this as $f_{x_i x_j} = (f_{x_i})_{x_j}$. Similarly, $\frac{\partial^3 f}{\partial x_i \partial x_j \partial x_k}$ may be defined as $\frac{\partial}{\partial x_i}(f_{x_j x_k})$. This idea can be extended even further into higher dimensions.

Example 1.42. Define $f(x, y) = \sin x + e^y + xy$. Then $f_x = \cos x + y$ and $f_y = e^y + x$. From here, we further have $f_{xy} = 1$ and $f_{yx} = 1$. Coincidentally, we have $f_{xy} = f_{yx}$. Thus, we question whether the order of the variables even matters.

The following example shows that the order of the variables does matter.

Example 1.43. Define

$$f(x, y) = \begin{cases} \frac{xy(x^2-y^2)}{x^2+y^2} & \text{if } (x, y) \neq (0, 0), \\ 0 & \text{if } (x, y) = (0, 0). \end{cases} \tag{1.42}$$

In this case, we get $f_{xy}(0,0) = 1 \neq f_{yx}(0,0) = -1$. Partial differentiation is not commutative.

The following result shows that with a little more constraints, the commutativity does hold.

Theorem 1.44 (*Clairaut's theorem*)**.** *Let $(a,b) \in O_2$, $f : O_2 \to \mathbb{R}$, and assume that f_{xy} and f_{yx} exist on O_2. Also suppose that f_{xy} is continuous at (a,b). Then $f_{xy}(a,b) = f_{yx}(a,b)$.*

The result does hold in higher dimensions too, but we only show for two dimensions. The rest of the theorem and proof are left as exercises for the reader.

Proof. Without the loss of generality, let $(a,b) = (0,0)$ and $O_2 = B_1(0,0)$. Choose $h,k > 0$ such that $[0,h] \times [0,k] \subseteq B_1(0,0)$. Then,

$$\begin{aligned}
\frac{\partial^2 f}{\partial y \partial x} &= \lim_{k\to 0} \frac{1}{k}\left(\frac{\partial f}{\partial x}(x, y+k) - \frac{\partial f}{\partial x}(x,y)\right) && (1.43)\\
&= \lim_{k\to 0} \frac{1}{k} \lim_{h\to 0} \frac{1}{h}(f(x+h,y+k) - f(x,y+k) - f(x+h,y) + f(x,y))\\
&= \lim_{k\to 0}\lim_{h\to 0} \frac{1}{hk}(f(x+h,y+k) - f(x,y+k) - f(x+h,y) + f(x,y)) && (1.44)\\
\implies \frac{\partial^2 f}{\partial y \partial x}(0,0) &= \lim_{k\to 0}\lim_{h\to 0} \frac{1}{hk}(f(h,k) - f(0,k) - f(h,0) + f(0,0)) := \lim_{k\to 0}\lim_{h\to 0}\frac{1}{hk}F(h,k). && (1.45)
\end{aligned}$$

Similarly,

$$\frac{\partial^2 f}{\partial x \partial y}(0,0) = \lim_{h\to 0}\lim_{k\to 0}\frac{1}{hk}F(h,k). \tag{1.46}$$

Fix k and h for a moment. Set $f_1(x) = f(x,k) - f(x,0)$ for all $x \in [0,h]$. Then f_1 is continuous on $[0,h]$, since f_x exists, and is differentiable on $(0,h)$. By the mean value theorem, there exists $c_1 \in (0,h)$ such that $f_1(h) - f_1(0) = f_1'(c_1) \times h_1$

$$\implies \frac{1}{h}F(h,k) = \left(\frac{\partial f}{\partial x}(c_1,k) - \frac{\partial f}{\partial x}(c_1,0)\right). \tag{1.47}$$

Next, consider $f_2(y) = \frac{\partial f}{\partial x}(c_1,y)$ for all $y \in [0,k]$ which is continuous on $[0,k]$ and differentiable on $(0,k)$. Again, by the mean value theorem, there exists $c_2 \in (0,k)$ such that $f_2(k) - f_2(0) = f_2'(c_2) \times k$

$$\implies \frac{1}{hk}F(h,k) = \frac{\partial^2 f}{\partial y \partial x}(c_1,c_2) \tag{1.48}$$

with $0 < c_1 < h$ and $0 < c_2 < k$. Similarly, if we had redefined f_1 and f_2, we would have received

$$\frac{1}{hk}F(h,k) = \frac{\partial^2 f}{\partial x \partial y}(\tilde{c}_1,\tilde{c}_2) \tag{1.49}$$

with $0 < \tilde{c}_1 < h$ and $0 < \tilde{c}_2 < k$. Thus,

$$\frac{\partial^2 f}{\partial y \partial x}(c_1,c_2) = \frac{\partial^2 f}{\partial x \partial y}(\tilde{c}_1,\tilde{c}_2). \tag{1.50}$$

As $(h,k) \to (0,0)$, $f_{xy}(0,0) = f_{yx}(0,0)$. ∎

August 6th.

Theorem 1.45 (*Schwarz theorem*)**.** *Let $f : O_2 \to \mathbb{R}$ be a function such that $(0,0) \in O_2$. Also suppose that f_x, f_y, f_{xy} exist on O_2 and f_{xy} is continuous on O_2. Then $f_{yx}(0,0)$ exists and $f_{yx}(0,0) = f_{xy}(0,0)$.*

Proof. As f_{xy} is continuous at $(0,0)$, for $\varepsilon > 0$, there exists $\delta > 0$ such that

$$|f_{xy}(s,t) - f_{xy}(0,0)| < \varepsilon \text{ for all } \sqrt{s^2+t^2} < \delta. \tag{1.51}$$

We already know $F(h,k) = f_{xy}(c_1,c_2)$ with $0 < c_1 < h$ and $0 < c_2 < k$. Choose h,k small enough such that $\sqrt{h^2+k^2} < \delta$. Therefore, for $\sqrt{c_1^2+c_2^2} < \delta$,

$$|f_{xy}(c_1,c_2) - f_{xy}(0,0)| < \varepsilon \implies |F(h,k) - f_{xy}(0,0)| < \varepsilon \text{ for } \sqrt{h^2+k^2} < \varepsilon. \tag{1.52}$$

From the above, we infer $-\varepsilon + f_{xy}(0,0) < F(h,k) < \varepsilon + f_{xy}(0,0)$. Rewriting the middle term,

$$F(h,k) = \frac{1}{h}\left(\frac{f(h,k)-f(h,0)}{k} - \frac{f(0,k)-f(0,0)}{k}\right) \xrightarrow{k\to 0} \frac{1}{h}(f_y(h,0) - f_y(0,0)) \tag{1.53}$$

Rebounding gives us

$$\left|\frac{1}{h}(f_y(h,0) - f_y(0,)) - f_{xy}(0,0)\right| \le \varepsilon \implies \lim_{h\to 0}\frac{1}{h}(f_y(h,0) - f_y(0,0)) = f_{xy}(0,0). \tag{1.54}$$

■

Note that in the above theorem, $(0,0)$ was chosen for the sake of simplifying the proof. Any $(\alpha,\beta) \in O_2$ would have worked. We move our focus back to the total derivative.

Theorem 1.46. *Let $f : O_n \to \mathbb{R}^m$ be differentiable at $a \in O_n$. Then $\frac{\partial f_i}{\partial x_j}$ exists at a for all $i = 1,2,\ldots,m$ and $j = 1,2,\ldots,n$. Moreover,*

$$\big[Df(a)\big]_{m\times n} = \left(\frac{\partial f_i}{\partial x_j}(a)\right)_{m\times n}. \tag{1.55}$$

Proof. Let $m = 1$, and fix $j \in \{1,2,\ldots,n\}$. Suppose $a = (a_1,\ldots,a_j,\ldots,a_n)$. Consider the mapping $\eta_j : (a_j-\varepsilon, a_j+\varepsilon) \to \mathbb{R}^n$ defined as $x \mapsto (a_1,\ldots,a_{j-1},x,a_{j+1},\ldots,a_n)$. Since this image is in the neighbourhood of a, we can apply f on it to get an image in $\mathbb{R}$; the mapping maps x to $f(a_1,\ldots,a_{j-1},x,a_{j+1},\ldots,a_n)$.

As $x \to a_1$, $x \to a_2$, ..., $x \to x$, $x \to a_{j+1}$, ... differentiable on $(a_j-\varepsilon, a_j+\varepsilon)$, η_j is differentiable on a_j and $\eta_j'(a_j) = e_j$. Thus, $f \circ n_j$ is differentiable at a_j by the chain rule.

$$(Df \circ n_j)(a_j) = \frac{d}{dx}(f\circ n_j)(a_j) = \lim_{h\to 0}\frac{f(\eta_j(a_j+h)) - f(a)}{h} = \lim_{h\to 0}\frac{f(a_1,\ldots,a_j,\ldots,a_n) - f(a)}{h} = \frac{\partial f}{\partial x_j}(a). \tag{1.56}$$

The chain rule implies that $(Df\circ n_j)(a_j) = Df(\eta_j(a_j))D\eta_j(a_j) \implies \frac{\partial f}{\partial x_j}(a) = Df(a)e_j$. Thus, we must have

$$\big(Df(a)\big) = \begin{pmatrix} f_{x_1(a)} & \cdots & f_{x_n}(a)\end{pmatrix}. \tag{1.57}$$

We now work the case for a general m; write f as $(f_1,\ldots,f_m)$. f is differentiable at a implies that

$$\big[Df(a)\big] = \begin{bmatrix} Df_1(a) \\ \vdots \\ Df_m(a)\end{bmatrix} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1}(a) & \cdots & \frac{\partial f_1}{\partial x_n}(a) \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1}(a) & \cdots & \frac{\partial f_m}{\partial x_n}(a)\end{bmatrix}. \tag{1.58}$$

■

Definition 1.47. Let $f : O_n \to \mathbb{R}^m$ be differentiable at a. The matrix representation $\left(\frac{\partial f_i}{\partial x_j}(a)\right)_{m\times n}$ of the total derivative $Df(a)$ is termed the *Jacobian* of f at a. We prefer to write it as $J_f(a)$.

Since we have a matrix to deal with now, it is only natural to ask questions regarding its nature. For instance, what does the rank of the Jacobian tell us? What about its determinant?

Theorem 1.48. *Let $f : O_n \to \mathbb{R}^m$ be a function with $a \in O_n$. Suppose f is a C^1 function in the neighbourhood of a. Then f is differentiable at a.*

There is a *gap* between this theorem and the previous one; we have the extra requirement of continuity of the partial derivatives here.

August 11th.

Example 1.49. Consider the function

$$f(x,y) = \begin{cases} (x^2+y^2)\sin\frac{1}{\sqrt{x^2+y^2}} & \text{if } (x,y) \neq (0,0), \\ 0 & \text{if } (x,y) = (0,0). \end{cases} \tag{1.59}$$

Then f is continuous at $(0,0)$. Moreover, one can show that f is also differentiable at $(0,0)$ with $Df(0,0) = \begin{bmatrix} 0 & 0 \end{bmatrix}$. However, both f_x and f_y are not continuous at $(0,0)$.

We now provide the proof of the above theorem, setting $a = 0$ without the loss of generality.

Proof. Let $m = 1$; the general case will be handled later. We claim that

$$\lim_{h\to 0} \frac{1}{\|h\|} \left| f(h) - f(0) - \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(0) h_i \right| = 0. \tag{1.60}$$

For $h \in \mathbb{R}^n$, with $\|h\|$ sufficiently small, we write $\hat{h}_i = (h_1, h_2, \ldots, h_i, 0, \ldots, 0)$ with $(n-i)$ zeroes at the end, for all $i = 1, 2, \ldots, n$, and $\hat{h}_0 = (0, \ldots, 0)$. Therefore,

$$\begin{aligned} f(h) - f(0) = f(\hat{h}_n) - f(\hat{h}_0) &= (f(\hat{h}_1) - f(\hat{h}_0)) + (f(\hat{h}_2) - f(\hat{h}_1)) + \cdots + (f(\hat{h}_n) - f(\hat{h}_{n-1})) \\ &= \sum_{i=1}^{n} (f(\hat{h}_i) - f(\hat{h}_{i-1})). \end{aligned} \tag{1.61}$$

Define $\eta_i(t) = f(h_1, \ldots, h_{i-1}, t, 0, \ldots, 0)$ for $t \in [0, h_i]$. Thus, each η_i is a single variable function and the chain rule tells us $\eta_i : [0, h_i] \to \mathbb{R}$ is a C^1-function. The mean value theorem then tell us that there exists $c_i \in (0, h_i)$ such that

$$\eta_i(h_i) - \eta_i(0) = h_i \eta_i'(c_i) \tag{1.62}$$

$$\implies f(\hat{h}_i) - f(\hat{h}_{i-1}) = h_i \frac{\partial f}{\partial x_i}(h_1, \ldots, h_{i-1}, c_i, 0, \ldots, 0). \tag{1.63}$$

Therefore,

$$\begin{aligned} \frac{1}{\|h\|} \left| f(h) - f(0) - \sum_{i=1}^{n} f_{x_i}(0) h_i \right| &= \frac{1}{\|h\|} \left| \sum_{i=1}^{n} h_i \left(\frac{\partial f}{\partial x_i}(h_1, \ldots, h_{i-1}, c_i, 0, \ldots, 0) - f_{x_i}(0) \right) \right| \\ &\leq \sum_{i=1}^{n} \frac{|h_i|}{\|h\|} \left| f_{x_i}(h_1, \ldots, h_i, c_i, 0, \ldots, 0) - f_{x_i}(0) \right| \xrightarrow{h\to 0} 0. \end{aligned} \tag{1.64}$$

Hence, the function f is differentiable at 0. ■

Example 1.50. We compute the total derivative of $f(x,y,z) = (x+2y+3z, xyz, \cos x, \sin x)$. Clearly, f is a C^1 function. Thus,

$$J_f(x,y,z) = \begin{bmatrix} \frac{\partial f_1}{\partial x} & \frac{\partial f_1}{\partial y} & \frac{\partial f_1}{\partial z} \\ \frac{\partial f_2}{\partial x} & \frac{\partial f_2}{\partial y} & \frac{\partial f_2}{\partial z} \\ \frac{\partial f_3}{\partial x} & \frac{\partial f_3}{\partial y} & \frac{\partial f_3}{\partial z} \\ \frac{\partial f_4}{\partial x} & \frac{\partial f_4}{\partial y} & \frac{\partial f_4}{\partial z} \end{bmatrix} = \begin{bmatrix} 1 & 2 & 3 \\ yz & xz & xy \\ -\sin x & 0 & 0 \\ \cos x & 0 & 0 \end{bmatrix}. \tag{1.65}$$

1.5 Gradient and The Chain Rule

We first extend the notion of the derivative along the coordinate directions to the derivative along *any* direction.

Definition 1.51. Let $u \in \mathbb{R}^n$ be a unit vector, that is, $\|u\| = 1$. Also let $f : O_n \to \mathbb{R}$ be a function. The *directional derivative* of f at $x \in O_n$ in the direction of u is defined as

$$D_u f(x) = \lim_{t\to 0} \frac{f(x+tu) - f(x)}{t}, \text{ if exists.} \tag{1.66}$$

The reader may verify that $D_{e_i} f(x) = f_{x_i}(x)$. If we denote $t \mapsto f(x+tu)$ as $\eta(t)$, then we directional derivative is simply $D_u f(x) = \eta'(0)$. If we additionally assume that f is differentiable at x, then the chain rule gives us

$$D_u f(x) = \eta'(0) = Df(x) \circ u. \tag{1.67}$$

Notice that u can be thought of as a linear map from $\mathbb{R}$ to $\mathbb{R}^n$ and $Df(x)$ is a linear map from $\mathbb{R}^n$ to $\mathbb{R}$.

Theorem 1.52. *Let $f : O_n \to \mathbb{R}$ be differentiable at $x \in O_n$ and let u be a unit vector in $\mathbb{R}^n$. Then the directional derivative $D_u f(x)$ exists and is given by*

$$(D_u f)(x) = Df(x)u. \tag{1.68}$$

The idea of the gradient is introduced.

Definition 1.53. Let $f : O_n \to \mathbb{R}$ be a function with $a \in O_n$. Also suppose $f_{x_i}(a)$ exists for all $i = 1, 2, \ldots, n$. Then

$$(\nabla f)(a) = (f_{x_1}(a), \ldots, f_{x_n}(a)) \tag{1.69}$$

is called the *gradient* of f at a. Therefore,

$$\nabla : \{f : \mathbb{R}^n \to \mathbb{R} \mid f_{x_i} \text{ exists}\} \to \mathbb{R}^n. \tag{1.70}$$

Corollary 1.54. *Let $f : O_n \to \mathbb{R}$ be differentiable at $x \in O_n$ and let u be a unit vector. Then, $(D_u f)(x) = (\nabla f)(x) \cdot u$.*

Remark 1.55. Suppose $f : O_n \to \mathbb{R}$ is a function whose partial derivatives exist at $x \in O_n$. Then

$$(\nabla f)(x) \cdot u = \|(\nabla f)(x)\| \cdot \|u\| \cos\theta = \|(\nabla f)(x)\| \cos\theta. \tag{1.71}$$

As $|\cos\theta| \leq 1$, the right hand side is maximum if $\theta = 0$.

August 13th.

What follows is an important result.

Theorem 1.56. *Let $f : O_n \to \mathbb{R}$ be differentiable at $x \in O_n$ and suppose $(\nabla f)(x) \neq 0$. Then the vector $(\nabla f)(x)$ points in the direction of the steepest ascent of the function f at x and $\|(\nabla f)(x)\|$ is the greatest possible rate of change.*

In other words, the maximum possible directional derivative of f at a occurs at $\nabla f(a)$.

Example 1.57. Let $f(x, y, z) = x^2yz$. We find the directional derivatives of f at $(1, 1, 0)$ in the direction of $\langle 1, 1, -1\rangle$. Note that the unit vector in this case is $u = \frac{1}{\sqrt{3}}\langle 1, 1, -1\rangle$, and that f is differentiable since f is a polynomial. Thus, $f_x = 2xyz$, $f_y = x^2z$, and $f_z = x^2y$. Therefore,

$$\nabla f = \langle 2xyz, x^2z, x^2y\rangle \implies \nabla f(1,1,0) = \langle 0,0,1\rangle. \tag{1.72}$$

Finally, we get

$$D_u f(1,1,0) = \langle 0,0,1\rangle \cdot \frac{1}{\sqrt{3}}\langle 1,1,-1\rangle = -\frac{1}{\sqrt{3}}. \tag{1.73}$$

Also, the maximum possible derivative of f at $(1,1,0)$ occurs at $\langle 0,0,1\rangle$ and it is $\langle 0,0,1\rangle \cdot \langle 0,0,1\rangle = 1$.

Example 1.58. We look at

$$f(x,y) = \begin{cases} \frac{x^2y}{x^2+y^2} & \text{if } (x,y) \neq (0,0), \\ 0 & \text{if } (x,y) = (0,0). \end{cases} \tag{1.74}$$

Clearly, $|f(x,y) - f(0,0)| = \frac{x^2|y|}{x^2+y^2} \leq |y|$, so f is continuous at $(0,0)$. For a given direction $u = \langle u_1, u_2\rangle$, we have

$$D_u f(0,0) = \lim_{t\to 0} \frac{f(tu_1,tu_2) - f(0,0)}{t} = \lim_{t\to 0} \frac{t^3u_1^2u_2}{t^2(u_1^2+u_2^2)} \cdot \frac{1}{t} = u_1^2u_2 < \infty. \tag{1.75}$$

Therefore, $D_u f(0,0) = u_1^2u_2$. Also, $f_x(0,0) = 0 = f_y(0,0)$ implying that $\nabla f(0,0) = \langle 0,0\rangle$. This gives us

$$\nabla f(0,0) \cdot u = 0 \neq u_1^2u_2 = D_u f(0,0) \tag{1.76}$$

making f *not* differentiable at $(0,0)$.

Hereforth, given $a, b \in \mathbb{R}^n$, we denote $L_{a,b}$ to be the line segment joining a to b. Essentially, $L_{a,b} = \{(1-t)a + tb \mid 0 \leq t \leq 1\}$.

Theorem 1.59 (The *mean value theorem*)**.** *Let $f : O_n \to \mathbb{R}$ be differentiable and $L_{a,b} \subseteq O_n$. Then there exists $c \in L_{a,b}$ such that*

$$f(b) - f(a) = \nabla f(c) \cdot (b-a). \tag{1.77}$$

Proof. Define $\eta : [0,1] \to L_{a,b}$ as $\eta(t) = (1-t)a + tb$. Then $f \circ \eta : [0,1] \to \mathbb{R}$ is differentiable and $\eta'(t) = b - a$, a column vector. We apply the one-dimensional mean value theorem on $f \circ \eta$ to get

$$(f\circ\eta)(1) - (f\circ\eta)(0) = (f\circ\eta)'(t_0) \tag{1.78}$$

for some $t_0 \in (0,1)$. This implies that

$$f(b) - f(a) = f'(\eta(t_0)) \cdot \eta'(t_0) = \nabla f(\eta(t_0)) \cdot (b-a) = \nabla f(c) \cdot (b-a) \tag{1.79}$$

where $c = \eta(t_0) \in L_{a,b}$. ■

1.5.1 More Partials

Suppose we have $f : O_n \to O_m$ and $g : O_m \to \mathbb{R}^p$ where $a \in O_n$ and $b = f(a) \in O_m$. Assume that f and g are differentiable at a and b, respectively. Then

$$J_{g\circ f}(a) = J_g(b)J_f(a). \tag{1.80}$$

Comparing the $(i,j)^{\text{th}}$ entry of both sides (for $1 \leq i \leq p$ and $1 \leq j \leq n$), we get

$$\frac{\partial (g\circ f)_i}{\partial x_j}(a) = \sum_{k=1}^{m} \frac{\partial g_i}{\partial y_k}(b)\frac{\partial f_k}{\partial x_j}(a). \tag{1.81}$$

In a more natural way, set $y_k = f_k(x_1, \ldots, x_n)$ for $k = 1, \ldots, m$ and set $z_i = g_i(y_1, \ldots, y_m)$ for $i = 1, \ldots, p$. Then we can write

$$\frac{\partial z_i}{\partial x_j} = \frac{\partial z_i}{\partial y_1} \cdot \frac{\partial y_1}{\partial x_j} + \cdots + \frac{\partial z_i}{\partial y_m} \cdot \frac{\partial y_m}{\partial x_j} = \sum_{t=1}^{m} \frac{\partial z_i}{\partial y_t} \cdot \frac{\partial y_t}{\partial x_j}. \tag{1.82}$$

This is termed the *chain rule for partials*.

Example 1.60. Suppose on $O_1 \to O_m \to \mathbb{R}$, we have the mapping(s) $t \mapsto (x_1(t), \ldots, x_m(t)) \mapsto f(x_1(t), \ldots, x_m(t))$. If we call the first mapping $\eta(t)$, and second mapping $(f(\eta(t)))$, then

$$\frac{df}{dt} = \frac{\partial f}{\partial x_1}\frac{dx_1}{dt} + \cdots + \frac{\partial f}{\partial x_m}\frac{dx_m}{dt} = \sum_{i=1}^{m} \frac{\partial f}{\partial x_i}\frac{dx_i}{dt}. \tag{1.83}$$

Example 1.61. Suppose $f(x, y, z) = xy^2z$ with $x = t$, $y = e^t$, and $z = 1 + t$. Then

$$\frac{df}{dt} = \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt} + \frac{\partial f}{\partial z}\frac{dz}{dt} = y^2z + 2xyze^t + xy = e^{2t}((1 + 2t) + 2t(1 + t)). \tag{1.84}$$

Of course, in this example, it would have been preferable to substitute back in the variables in terms of t and then derivating. This may not always be the case, especially in abstract computations.

Example 1.62. Suppose $g(z, w) = f(x(z, w), y(z, w))$. Making all the necessary assumptions, we have

$$\frac{\partial g}{\partial z} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial z} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial z}. \tag{1.85}$$

Example 1.63. We introduce the idea of *polar coordinates*. $g(y, x)$ can be written as $f(r, \theta)$ where $r = \sqrt{x^2 + y^2}$ and $\theta = \tan^{-1}(\frac{y}{x})$. The converse may also be done. Here,

$$\frac{\partial g}{\partial x} = \frac{\partial g}{\partial r}\frac{\partial r}{\partial x} + \frac{\partial g}{\partial \theta}\frac{\partial \theta}{\partial x}. \tag{1.86}$$

Example 1.64. Suppose $z = z(u, v)$ with $u = x^2y$ and $v = 3x + 2y$. Then

$$\frac{\partial z}{\partial y} = \frac{\partial z}{\partial u}\frac{\partial u}{\partial y} + \frac{\partial z}{\partial v}\frac{\partial v}{\partial y} = \frac{\partial z}{\partial u}x^2 + \frac{\partial z}{\partial v} \cdot 2. \tag{1.87}$$

One can reuse the chain rule and think of double partial differentiation too as

$$\frac{\partial^2 y}{\partial y^2} = x^2 \frac{\partial}{\partial y}\left(\frac{\partial z}{\partial u}\right) + 2\frac{\partial}{\partial y}\left(\frac{\partial z}{\partial v}\right). \tag{1.88}$$

August 18th.

The Laplacian is introduced.

Definition 1.65. For $u \in C^2(O_n)$, the *Laplacian* of u is defined as

$$\Delta u := \nabla \circ \nabla u = \langle \frac{\partial}{\partial x_1}, \ldots, \frac{\partial}{\partial x_n} \rangle \cdot \langle u_{x_1}, \ldots, u_{x_n} \rangle = \sum_{j=1}^{n} \frac{\partial^2 u}{\partial x_j^2}. \tag{1.89}$$

Example 1.66. Let $x = r\cos\theta$ and $y = r\sin\theta$ and let $u := u(x, y)$. The Laplacian of u is $\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2}$.

Now,

$$\frac{\partial u}{\partial r} = \frac{\partial u}{\partial x}\frac{\partial x}{\partial r} + \frac{\partial u}{\partial y}\frac{\partial y}{\partial r} = \frac{\partial u}{\partial x}\cos\theta + \frac{\partial u}{\partial y}\sin\theta. \tag{1.90}$$

In a case where u is considered in its polar coordinates, the Laplacian of u is also written in its polar coordinate form.

1.6 Extremum and Critical Points

Of course, one can extend the idea of the maximum (minimum) value of a function into higher dimensions.

Definition 1.67. Let $S_n \subseteq \mathbb{R}^n$ and let $a \in S_n$ be an interior point. We say $f : S_n \to \mathbb{R}$ attains a *local maximum* at a if there exists $r > 0$ such that $f(x) \leq f(a)$ for all $x \in B_r(a) \subseteq S_n$. Similarly, it attains a *local minimum* at a if there exists $r > 0$ such that $f(x) \geq f(a)$ for all $x \in B_r(a) \subseteq S_n$.
a is, instead, termed a *saddle point* of f if for all $r > 0$ satisfying $B_r(a) \subseteq S_n$, there exist $h_1, h_2 \in B_r(a)$ such that $f(h_1) > f(a)$ and $f(h_2) < f(a)$.

Similarly, one has critical points.

Definition 1.68. $a \in S_n$ is called a *critical point* of $f : S_n \to \mathbb{R}$ if $\nabla f(a) = 0$. In other words, $f_{x_i}(a) = 0$ for all $i = 1, 2, \ldots, n$.

Theorem 1.69. *Suppose $f : O_n \to \mathbb{R}$ is differentiable at $a \in O_n$. If a is a local extremum, then $\nabla f(a) = 0$.*

Proof. We claim that $f_{x_i}(a) = 0$ for each i. Fix i and define $\varphi_i(t) = f(a_1, \ldots, a_{i-1}, t, a_{i+1}, \ldots, a_n)$ for $t \in (a_i - \varepsilon, a_i + \varepsilon)$. This implies that a_i is a local extremum of φ_i giving $\frac{d\varphi_i}{dt}(a_i) = 0$. Thus, $f_{x_i}(a) = 0$. ■

To find these extrema, one would find the critical points and apply the second derivative test in the one variable case. In higher dimensions, we do the same; however, one needs to formulate the idea of a 'second derivative test' here.

Definition 1.70. Given $f \in C^2(O_n)$ with $a \in O_n$. The *Hessian matrix*, or Hessian, of f at a is defined as

$$H_f(a) = \begin{bmatrix} \dfrac{\partial^2 f(a)}{\partial x_i \partial x_j} \end{bmatrix}_{n\times n} \tag{1.91}$$

Note that H_f is a symmetric matrix.

Example 1.71. Consider $f : \mathbb{R}^2 \to \mathbb{R}$ defined as $f(x, y) = \sin^2 x + x^2 y + y^2$. Then $f_x = \sin 2x + 2xy$, $f_y = x^2 + 2y$, and $f_{xy} = 2x$. The Jacobian in this case is

$$J_f = \begin{bmatrix} \sin 2x + 2xy & x^2 + 2y \end{bmatrix} \tag{1.92}$$

and the Hessian is

$$H_f = \begin{bmatrix} 2\cos 2x + 2y & 2x \\ 2x & 2 \end{bmatrix}. \tag{1.93}$$

Definition 1.72. Given $A = (a_{ij})_{n\times n} \in M_n(\mathbb{R})$, we define $Q_A(x) = x^t A x = \langle Ax, x\rangle = \sum_{i,j=1}^n a_{ij} x_i x_j$ for all $x \in \mathbb{R}^n$. A *quadratic form* is Q_A when A is symmetric.

Thus, $x^t H_f(a)x$ is a quadratic form which is also a homogenous polynomial of degree 2. One calls a symmetric matrix $A \in M_n(\mathbb{R})$ a *positive definite matrix* if $h^t Ah > 0$ for all $h \in \mathbb{R}^n \setminus \{0\}$ and, likewise, a *negative definite matrix*. It is termed a *positive semidefinite matrix* if $h^t Ah \geq 0$ for all $h \in \mathbb{R}^n$ and, likewise, a *negative semidefinite matrix*. It is called a *indefinite matrix* if it satisfies none of the above conditions.

Theorem 1.73. *Consider a symmetric matrix* $A = \begin{bmatrix} a & b \\ b & c \end{bmatrix} \in M_2(\mathbb{R})$. *Then*

1. *A is positive definite if and only if* $a > 0$ *and* $ac - b^2 > 0$,
2. *A is negative definite if and only if* $a < 0$ *and* $ac - b^2 > 0$,
3. *A is indefinite if and only if* $ac - b^2 < 0$.

Lemma 1.74. *Let* $a \in O_n$, *and suppose* $A(x) = \begin{bmatrix} a_{11}(x) & a_{12}(x) \\ a_{21}(x) & a_{22}(x) \end{bmatrix} \in M_2(\mathbb{R})$ *is a symmetric matrix for all* $x \in O_n$. *Suppose A is continuous at a, that is, the functions* a_{ij} *are continuous at a for all pairs* i, j. *If* $A(a)$ *is positive definite, then A is positive definite in a neighbourhood of a.*

Proof. $A(a)$ is positive definite implies that $a_{11}(a) > 0$ and $a_{11}(a)a_{22}(a) - a_{12}(a)^2 > 0$. As $x \mapsto a_{ij}(x)$ is continuous at a, there exists a neighbourhood of a such that $a_{11}(x) > 0$ and $a_{11}(x)a_{22}(x) - a_{12}(x)^2 > 0$ for all x in that neighbourhood. Therefore, $A(x)$ is positive definite in a neighbourhood of a. ■

Recall Taylor's polynomial and approximation; given a function $f \in C^k(I)$ with $a \in I \subseteq \mathbb{R}$, we would have

$$p_{a,k}(a+h) = \sum_{m=0}^{k} \frac{f^{(m)}(a)}{m!} h^m \tag{1.94}$$

where $a + h \in I$. In terms of x, it was

$$p_{a,k}(x) = \sum_{m=0}^{k} \frac{f^{(m)}(a)}{m!} (x-a)^m \tag{1.95}$$

One would also have Taylor's theorem, where if the above f were a $C^{k+1}(I)$-function, then for all $x \in I$ there exists a $\zeta \equiv \zeta(x, a)$ such that

$$f(x) = p_{a,k}(x) + \frac{f^{(k+1)}(\zeta)}{(k+1)!}(x-a)^{k+1}. \tag{1.96}$$

This theorem extends to higher dimensions.

Theorem 1.75 (*Taylor's theorem*)**.** *Let* $a \in O_n$ *where* O_n *is a convex set. Also suppose* $f : O_n \to \mathbb{R}$ *is a* C^{k+1}*-function. If* $h \in O_n$ *and* $a + h \in O_n$, *then*

$$f(a+h) = \sum_{|\alpha| \leq k} \frac{\partial^\alpha f(a)}{\alpha!} h^\alpha + \gamma_{a,k}(h) \tag{1.97}$$

where

$$\gamma_{a,k}(h) = \sum_{|\alpha| = k+1} \frac{\partial^\alpha f(a+ch)}{\alpha!} h^\alpha \tag{1.98}$$

for some $c \in (0, 1)$.

We clear up some notation. Here, $\alpha = (\alpha_1, \ldots, \alpha_n) \in \mathbb{N}^n$ is termed a multi-index, with $|\alpha| = \sum \alpha_i$. The notation $\partial^\alpha f(a)$ is shorthand for $\frac{\partial^{|\alpha|} f(a)}{\partial x_1^{\alpha_1} \cdots \partial x_n^{\alpha_n}}$. Also, $\alpha! = \alpha_1! \cdots \alpha_n!$, and finally, $h^\alpha = h_1^{\alpha_1} \cdots h_n^{\alpha_n}$.

Proof. Fix $a, a+h \in O_n$. Consider the mappings $t \mapsto a+th \mapsto f(a+th)$, with $\eta : [0,1] \to \mathbb{R}$ defined as $\eta(t) = f(a+th)$. Note that η is also a C^{k+1}-function, and $\eta'(t) = \nabla f(a+th) \circ h = \sum_{i=1}^n h_i f_{x_i}(a+th)$. The double derivatives is

$$\eta''(t) = \sum_{i=1}^n h_i f_{x_i}(a+th) = \nabla f_{x_i}(a+th) \circ h = \sum_{i,j=1}^n h_i h_j f_{x_i x_j}(a+th). \tag{1.99}$$

Call $\nabla \circ h = \sum_{i=1}^n h_i \frac{\partial}{\partial x_i}$. Therefore, $\eta'(t) = (\nabla \circ h) f(a+th)$ and $\eta''(t) = (\nabla \circ h)^2 f(a+th)$; we have

$$\eta^{(m)}(t) = (\nabla \circ h)^m f(a+th) \tag{1.100}$$

for all $0 \le m \le k+1$. Note that

$$(\nabla \circ h)^m = \sum_{|\alpha|=m} \frac{\partial^\alpha}{\alpha!} h^\alpha. \tag{1.101}$$

We apply the one-dimensional Taylor's theorem to get

$$\eta(1) = \eta(0) + \eta'(0) + \frac{1}{2!}\eta''(0) + \cdots + \frac{1}{k!}\eta^{(k)}(0) + \frac{1}{(k+1)!}\eta^{(k+1)}(c) \tag{1.102}$$

for some $c \in (0,1)$. Expanding in terms of f gives the desired result. ■

Index